

Eckhard Platen
Nicola Bruti-Liberati

STOCHASTIC MODELLING
AND APPLIED PROBABILITY

64

Numerical Solution of Stochastic Differential Equations with Jumps in Finance



Springer

*Stochastic Mechanics
Random Media
Signal Processing
and Image Synthesis*
*Mathematical Economics and Finance
Stochastic Optimization
Stochastic Control
Stochastic Models in Life Sciences*

**Stochastic Modelling
and Applied Probability**
(Formerly:
Applications of Mathematics)

64

Edited by B. Rozovskii
 G. Grimmett

Advisory Board M. Hairer
 I. Karatzas
 F.P. Kelly
 A. Kyprianou
 Y. Le Jan
 B. Øksendal
 G. Papanicolaou
 E. Pardoux
 E. Perkins

For other titles in this series, go to
<http://www.springer.com/series/602>

Eckhard Platen • Nicola Bruti-Liberati

Numerical Solution of Stochastic Differential Equations with Jumps in Finance



Eckhard Platen
Nicola Bruti-Liberati (1975–2007)
School of Finance and Economics
Department of Mathematical Sciences
University of Technology, Sydney
PO Box 123
Broadway NSW 2007
Australia
eckhard.platen@uts.edu.au

Managing Editors

Boris Rozovskii
Division of Applied Mathematics
Brown University
182 George St
Providence, RI 02912
USA
rozovsky@dam.brown.edu

Geoffrey Grimmett
Centre for Mathematical Sciences
University of Cambridge
Wilberforce Road
Cambridge CB3 0WB
UK
g.r.grimmett@statslab.cam.ac.uk

ISSN 0172-4568

ISBN 978-3-642-12057-2

e-ISBN 978-3-642-13694-8

DOI 10.1007/978-3-642-13694-8

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010931518

Mathematics Subject Classification (2010): 60H10, 65C05, 62P05

© Springer-Verlag Berlin Heidelberg 2010

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: deblik

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This research monograph concerns the design and analysis of discrete-time approximations for stochastic differential equations (SDEs) driven by Wiener processes and Poisson processes or Poisson jump measures. In financial and actuarial modeling and other areas of application, such jump diffusions are often used to describe the dynamics of various state variables. In finance these may represent, for instance, asset prices, credit ratings, stock indices, interest rates, exchange rates or commodity prices. The jump component can capture event-driven uncertainties, such as corporate defaults, operational failures or insured events. The book focuses on efficient and numerically stable strong and weak discrete-time approximations of solutions of SDEs. Strong approximations provide efficient tools for simulation problems such as those arising in filtering, scenario analysis and hedge simulation. Weak approximations, on the other hand, are useful for handling problems via Monte Carlo simulation such as the evaluation of moments, derivative pricing, and the computation of risk measures and expected utilities. The discrete-time approximations considered are divided into regular and jump-adapted schemes. Regular schemes employ time discretizations that do not include the jump times of the Poisson jump measure. Jump-adapted time discretizations, on the other hand, include these jump times.

The first part of the book provides a theoretical basis for working with SDEs and stochastic processes with jumps motivated by applications in finance. This part also introduces stochastic expansions for jump diffusions. It further proves powerful results on moment estimates of multiple stochastic integrals. The second part presents strong discrete-time approximations of SDEs with given strong order of convergence, including derivative-free and predictor-corrector schemes. The strong convergence of higher order schemes for pure jump SDEs is established under conditions weaker than those required for jump diffusions. Estimation and filtering methods are discussed. The third part of the book introduces a range of weak approximations with jumps. These weak approximations include derivative-free, predictor-corrector, and

simplified schemes. The final part of the research monograph raises questions on numerical stability and discusses powerful martingale representations and variance reduction techniques in the context of derivative pricing.

The book does not claim to be a complete account of the state of the art of the subject. Rather it attempts to provide a systematic framework for an understanding of the basic concepts and tools needed when implementing simulation methods for the numerical solution of SDEs. In doing so the book aims to follow up on the presentation of the topic in [Kloeden & Platen \(1999\)](#) where no jumps were considered and no particular field of application motivated the numerical methods. The book goes significantly beyond [Kloeden & Platen \(1999\)](#). It is covering many new results for the approximation of continuous solutions of SDEs. The discrete time approximation of SDEs with jumps represents the focus of the monograph. The reader learns about powerful numerical methods for the solution of SDEs with jumps. These need to be implemented with care. It is directed at readers from different fields and backgrounds.

The area of finance has been chosen to motivate the methods. It has been also a focus of research by the first author for many years that culminated in the development of the benchmark approach, see [Platen & Heath \(2006\)](#), which provides a general framework for modeling risk in finance, insurance and other areas and may be new to most readers. The book is written at a level that is appropriate for a reader with an engineer's or similar undergraduate training in mathematical methods. It is readily accessible to many who only require numerical recipes.

Together with Nicola Bruti-Liberati we had for several years planned a book to follow on the book with Peter Kloeden on the “Numerical Solution of Stochastic Differential Equations”, which first appeared in 1992 at Springer Verlag and helped to develop the theory and practice of this field. Nicola's PhD thesis was written to provide proofs for parts of such a book. It is very sad that Nicola died tragically in a traffic accident on 28 August 2007. This was an enormous loss for his family and friends, his colleagues and the area of quantitative methods in finance.

The writing of such a book was not yet started at the time of Nicola's tragic death. I wish to express my deep gratitude to Katrin Platen, who then agreed to typeset an even more comprehensive book than was originally envisaged. She carefully and patiently wrote and revised several versions of the manuscript under difficult circumstances. The book now contains not only results that we obtained with Nicola on the numerical solution of SDEs with jumps, but also presents methods for exact simulation, parameter estimation, filtering and efficient variance reduction, as well as the simulation of hedge ratios and the construction of martingale representations.

I would like to thank several colleagues for their collaboration in related research and valuable suggestions on the manuscript, including Kevin Burrage, Leunglung Chan, Kristoffer Glover, David Heath, Des Higham, Hardy Hulley, Constantinos Kardaras, Peter Kloeden, Uwe Küchler, Herman Lukito,

Remigius Mikulevicius, Renata Rendek, Wolfgang Runggaldier, Lei Shi and Anthony Tooman. Particular thanks go to Rob Lynch, the former Dean of the Faculty of Business at the University of Technology Sydney, who made the writing of the book possible through his direct support. Finally, I like to thank the Editor, Catriona Byrne, at Springer for her excellent work and her encouragement to write this book as a sequel of the previous book with Peter Kloeden.

It is greatly appreciated if readers could forward any errors, misprints or suggested improvements to: eckhard.platen@uts.edu.au. The interested reader is likely to find updated information about the numerical solution of stochastic differential equations on the webpage of the first author under “Numerical_Methods”:

[http://www.business.uts.edu.au/
finance/staff/Eckhard/Numerical_Methods.html](http://www.business.uts.edu.au/finance/staff/Eckhard/Numerical_Methods.html)

Sydney, January 2010

Eckhard Platen

Contents

Preface	V
Suggestions for the Reader	XV
Basic Notation	XIX
Motivation and Brief Survey.....	XXIII
1 Stochastic Differential Equations with Jumps	1
1.1 Stochastic Processes	1
1.2 Supermartingales and Martingales	16
1.3 Quadratic Variation and Covariation	23
1.4 Itô Integral	26
1.5 Itô Formula.....	34
1.6 Stochastic Differential Equations	38
1.7 Linear SDEs	45
1.8 SDEs with Jumps	53
1.9 Existence and Uniqueness of Solutions of SDEs	57
1.10 Exercises	59
2 Exact Simulation of Solutions of SDEs	61
2.1 Motivation of Exact Simulation	61
2.2 Sampling from Transition Distributions	63
2.3 Exact Solutions of Multi-dimensional SDEs	78
2.4 Functions of Exact Solutions	99
2.5 Almost Exact Solutions by Conditioning.....	105
2.6 Almost Exact Simulation by Time Change.....	113
2.7 Functionals of Solutions of SDEs	123
2.8 Exercises	136

3	Benchmark Approach to Finance and Insurance	139
3.1	Market Model	139
3.2	Best Performing Portfolio	142
3.3	Supermartingale Property and Pricing	145
3.4	Diversification	149
3.5	Real World Pricing Under Some Models	158
3.6	Real World Pricing Under the MMM	168
3.7	Binomial Option Pricing	176
3.8	Exercises	185
4	Stochastic Expansions	187
4.1	Introduction to Wagner-Platen Expansions	187
4.2	Multiple Stochastic Integrals	195
4.3	Coefficient Functions	202
4.4	Wagner-Platen Expansions	206
4.5	Moments of Multiple Stochastic Integrals	211
4.6	Exercises	230
5	Introduction to Scenario Simulation	233
5.1	Approximating Solutions of ODEs	233
5.2	Scenario Simulation	245
5.3	Strong Taylor Schemes	252
5.4	Derivative-Free Strong Schemes	266
5.5	Exercises	271
6	Regular Strong Taylor Approximations with Jumps	273
6.1	Discrete-Time Approximation	273
6.2	Strong Order 1.0 Taylor Scheme	278
6.3	Commutativity Conditions	286
6.4	Convergence Results	289
6.5	Lemma on Multiple Itô Integrals	292
6.6	Proof of the Convergence Theorem	302
6.7	Exercises	307
7	Regular Strong Itô Approximations	309
7.1	Explicit Regular Strong Schemes	309
7.2	Drift-Implicit Schemes	316
7.3	Balanced Implicit Methods	321
7.4	Predictor-Corrector Schemes	326
7.5	Convergence Results	331
7.6	Exercises	346

8	Jump-Adapted Strong Approximations	347
8.1	Introduction to Jump-Adapted Approximations	347
8.2	Jump-Adapted Strong Taylor Schemes	350
8.3	Jump-Adapted Derivative-Free Strong Schemes	355
8.4	Jump-Adapted Drift-Implicit Schemes	356
8.5	Predictor-Corrector Strong Schemes	359
8.6	Jump-Adapted Exact Simulation	361
8.7	Convergence Results	362
8.8	Numerical Results on Strong Schemes	368
8.9	Approximation of Pure Jump Processes	375
8.10	Exercises	388
9	Estimating Discretely Observed Diffusions	389
9.1	Maximum Likelihood Estimation	389
9.2	Discretization of Estimators	393
9.3	Transform Functions for Diffusions	397
9.4	Estimation of Affine Diffusions	404
9.5	Asymptotics of Estimating Functions	409
9.6	Estimating Jump Diffusions	413
9.7	Exercises	417
10	Filtering	419
10.1	Kalman-Bucy Filter	419
10.2	Hidden Markov Chain Filters	424
10.3	Filtering a Mean Reverting Process	433
10.4	Balanced Method in Filtering	447
10.5	A Benchmark Approach to Filtering in Finance	456
10.6	Exercises	475
11	Monte Carlo Simulation of SDEs	477
11.1	Introduction to Monte Carlo Simulation	477
11.2	Weak Taylor Schemes	481
11.3	Derivative-Free Weak Approximations	491
11.4	Extrapolation Methods	495
11.5	Implicit and Predictor-Corrector Methods	497
11.6	Exercises	504
12	Regular Weak Taylor Approximations	507
12.1	Weak Taylor Schemes	507
12.2	Commutativity Conditions	514
12.3	Convergence Results	517
12.4	Exercises	522

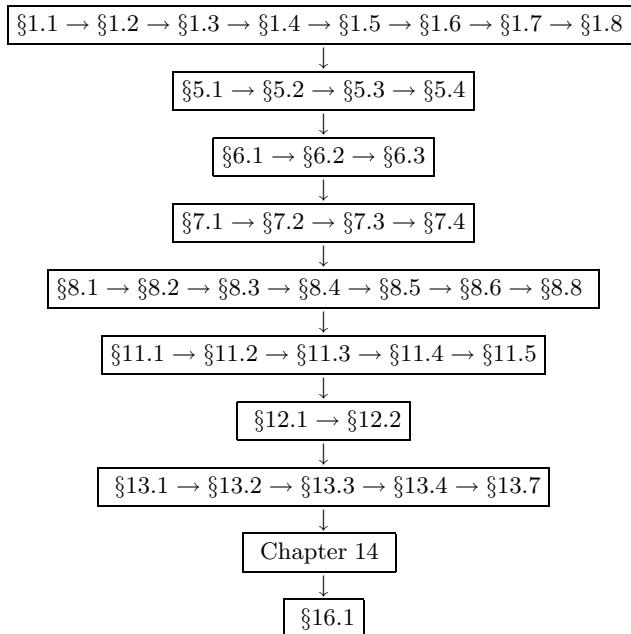
13	Jump-Adapted Weak Approximations	523
13.1	Jump-Adapted Weak Schemes	523
13.2	Derivative-Free Schemes	529
13.3	Predictor-Corrector Schemes	530
13.4	Some Jump-Adapted Exact Weak Schemes	533
13.5	Convergence of Jump-Adapted Weak Taylor Schemes	534
13.6	Convergence of Jump-Adapted Weak Schemes	543
13.7	Numerical Results on Weak Schemes	548
13.8	Exercises	569
14	Numerical Stability	571
14.1	Asymptotic p -Stability	571
14.2	Stability of Predictor-Corrector Methods	576
14.3	Stability of Some Implicit Methods	583
14.4	Stability of Simplified Schemes	586
14.5	Exercises	590
15	Martingale Representations and Hedge Ratios	591
15.1	General Contingent Claim Pricing	591
15.2	Hedge Ratios for One-dimensional Processes	595
15.3	Explicit Hedge Ratios	601
15.4	Martingale Representation for Non-Smooth Payoffs	606
15.5	Absolutely Continuous Payoff Functions	616
15.6	Maximum of Several Assets	621
15.7	Hedge Ratios for Lookback Options	627
15.8	Exercises	635
16	Variance Reduction Techniques	637
16.1	Various Variance Reduction Methods	637
16.2	Measure Transformation Techniques	645
16.3	Discrete-Time Variance Reduced Estimators	658
16.4	Control Variates	669
16.5	HP Variance Reduction	677
16.6	Exercises	694
17	Trees and Markov Chain Approximations	697
17.1	Numerical Effects of Tree Methods	697
17.2	Efficiency of Simplified Schemes	712
17.3	Higher Order Markov Chain Approximations	720
17.4	Finite Difference Methods	734
17.5	Convergence Theorem for Markov Chains	744
17.6	Exercises	753
18	Solutions for Exercises	755
	Acknowledgements	781

Bibliographical Notes	783
References	793
Author Index	835
Index	847

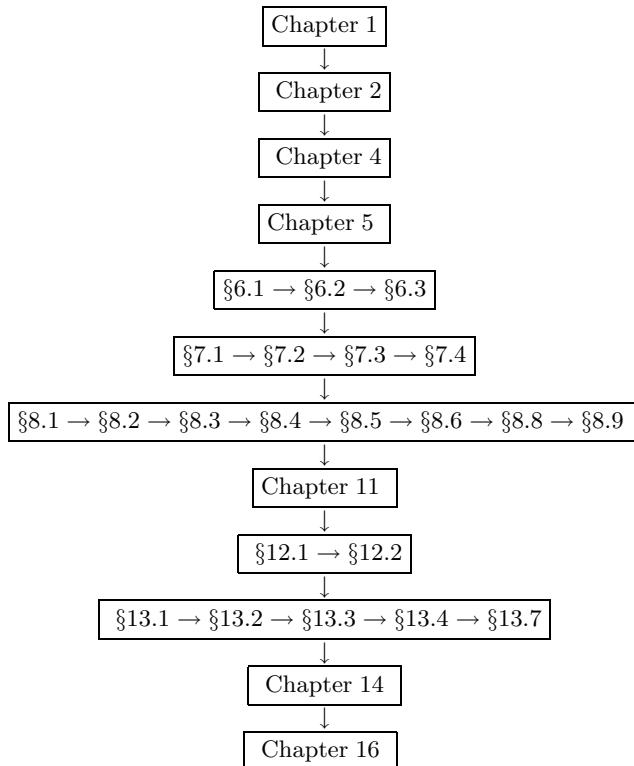
Suggestions for the Reader

It has been mentioned in the Preface that the material of this book has been arranged in a way that should make it accessible to as wide a readership as possible. Prospective readers will have different backgrounds and objectives. The following four groups are suggestions to help use the book more efficiently.

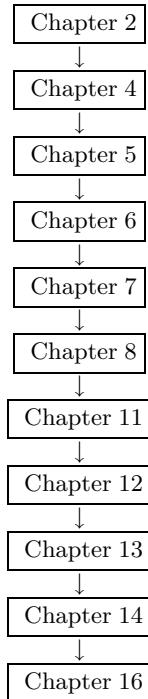
- (i) Let us begin with those readers who aim for a sufficient understanding, to be able to apply stochastic differential equations with jumps and appropriate simulation methods in their field of application, which may not be finance. Deeper mathematical issues are avoided in the following suggested sequence of reading, which provides a guide to the book for those without a strong mathematical background:



- (ii) Engineers, quantitative analysts and others with a more technical background in mathematical and quantitative methods who are interested in applying stochastic differential equations with jumps, and in implementing efficient simulation methods or developing new schemes could use the book according to the following suggested flowchart. Without too much emphasis on proofs the selected material provides the underlying mathematics.

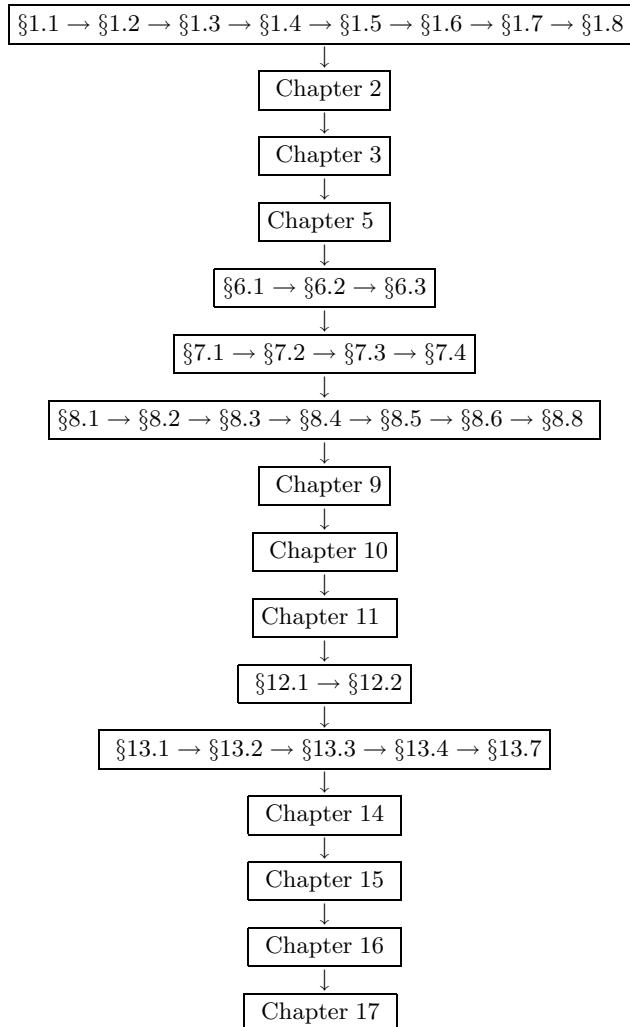


- (iii) Readers with strong mathematical background and mathematicians may omit the introductory Chap. 1. The following flowchart focuses on the theoretical aspects of the numerical approximation of solutions of stochastic differential equations with jumps while avoiding well-known or applied topics.



XVIII Suggestions for the Reader

- (iv) Financial engineers, quantitative analysts, risk managers, fund managers, insurance professionals and others who have no strong mathematical background and are interested in finance, insurance and other areas of risk management will find the following flowchart helpful. It suggests the reading for an introduction into quantitative methods in finance and related areas.



Basic Notation

μ_X	mean of X
σ_X^2 , $\text{Var}(X)$	variance of X
$\text{Cov}(X, Y)$	covariance of X and Y
$\inf\{\cdot\}$	greatest lower bound
$\sup\{\cdot\}$	smallest upper bound
$\max(a, b) = a \vee b$	maximum of a and b
$\min(a, b) = a \wedge b$	minimum of a and b
$(a)^+ = \max(a, 0)$	maximum of a and 0
\boldsymbol{x}^\top	transpose of a vector or matrix \boldsymbol{x}
$\boldsymbol{x} = (x^1, x^2, \dots, x^d)^\top$	column vector $\boldsymbol{x} \in \Re^d$ with i th component x^i
$ \boldsymbol{x} $	absolute value of \boldsymbol{x} or Euclidean norm
$\boldsymbol{A} = [a^{i,j}]_{i,j=1}^{k,d}$	$(k \times d)$ -matrix \boldsymbol{A} with ij th component $a^{i,j}$
$\det(\boldsymbol{A})$	determinant of a matrix \boldsymbol{A}
\boldsymbol{A}^{-1}	inverse of a matrix \boldsymbol{A}
$(\boldsymbol{x}, \boldsymbol{y})$	inner product of vectors \boldsymbol{x} and \boldsymbol{y}
$\mathcal{N} = \{1, 2, \dots\}$	set of natural numbers

∞	infinity
(a, b)	open interval $a < x < b$ in \mathbb{R}
$[a, b]$	closed interval $a \leq x \leq b$ in \mathbb{R}
$\mathbb{R} = (-\infty, \infty)$	set of real numbers
$\mathbb{R}^+ = [0, \infty)$	set of nonnegative real numbers
\mathbb{R}^d	d -dimensional Euclidean space
Ω	sample space
\emptyset	empty set
$A \cup B$	the union of sets A and B
$A \cap B$	the intersection of sets A and B
$A \setminus B$	the set A without the elements of B
$\mathcal{E} = \mathbb{R} \setminus \{0\}$	\mathbb{R} without origin
$[X, Y]_t$	covariation of processes X and Y at time t
$[X]_t$	quadratic variation of process X at time t
$n! = 1 \cdot 2 \cdot \dots \cdot n$	factorial of n
$[a]$	largest integer not exceeding $a \in \mathbb{R}$
i.i.d.	independent identically distributed
a.s.	almost surely
f'	first derivative of $f : \mathbb{R} \rightarrow \mathbb{R}$
f''	second derivative of $f : \mathbb{R} \rightarrow \mathbb{R}$
$f : Q_1 \rightarrow Q_2$	function f from Q_1 into Q_2
$\frac{\partial u}{\partial x^i}$	i th partial derivative of $u : \mathbb{R}^d \rightarrow \mathbb{R}$
$\left(\frac{\partial}{\partial x^i}\right)^k u$	k th order partial derivative of u with respect to x^i
\exists	there exists
$F_X(\cdot)$	distribution function of X
$f_X(\cdot)$	probability density function of X
$\phi_X(\cdot)$	characteristic function of X
$\mathbf{1}_A$	indicator function for event A to be true

$N(\cdot)$	Gaussian distribution function
$\Gamma(\cdot)$	gamma function
$\Gamma(\cdot; \cdot)$	incomplete gamma function
$(\text{mod } c)$	modulo c
\mathcal{A}	collection of events, sigma-algebra
$\underline{\mathcal{A}}$	filtration
$E(X)$	expectation of X
$E(X \mathcal{A})$	conditional expectation of X under \mathcal{A}
$P(A)$	probability of A
$P(A B)$	probability of A conditioned on B
\in	element of
\notin	not element of
\neq	not equal
\approx	approximately equal
$a \ll b$	a is significantly smaller than b
$\lim_{N \rightarrow \infty}$	limit as N tends to infinity
$\liminf_{N \rightarrow \infty}$	lower limit as N tends to infinity
$\limsup_{N \rightarrow \infty}$	upper limit as N tends to infinity
i	square root of -1 , imaginary unit
$\delta(\cdot)$	Dirac delta function at zero
\mathbf{I}	unit matrix
$\text{sgn}(x)$	sign of $x \in \Re$
\mathcal{L}_T^2	space of square integrable, progressively measurable functions on $[0, T] \times \Omega$
$\mathcal{B}(U)$	smallest sigma-algebra on U
$\ln(a)$	natural logarithm of a
MM	Merton model
MMM	minimal market model

GIG	generalized inverse Gaussian
GH	generalized hyperbolic
VG	variance gamma
GOP	growth optimal portfolio
EWI	equi-value weighted index
ODE	ordinary differential equation
SDE	stochastic differential equation
PDE	partial differential equation
PIDE	partial integro differential equation
$I_\nu(\cdot)$	modified Bessel function of the first kind with index ν
$K_\lambda(\cdot)$	modified Bessel function of the third kind with index λ
Δ	time step size of a time discretization
$\binom{i}{l} = \frac{i!}{l!(i-l)!}$	combinatorial coefficient
$\mathcal{C}^k(\mathcal{R}^d, \mathcal{R})$	set of k times continuously differentiable functions
$\mathcal{C}_P^k(\mathcal{R}^d, \mathcal{R})$	set of k times continuously differentiable functions which, together with their partial derivatives of order up to k , have at most polynomial growth

Letters such as $K, K_1, \dots, \tilde{K}, C, C_1, \dots, \tilde{C}, \dots$ represent finite positive real constants that can vary from line to line. All these constants are assumed to be independent of the time step size Δ .

Motivation and Brief Survey

Key features of advanced models in many areas of application with uncertainties are often event-driven. In finance and insurance one has to deal with events such as corporate defaults, operational failures or insured accidents. By analyzing time series of historical data, such as prices and other financial quantities, many authors have argued in the area of finance for the presence of jumps, see [Jorion \(1988\)](#) and [Ait-Sahalia \(2004\)](#) for foreign exchange and stock markets, and [Johannes \(2004\)](#) for short-term interest rates. Jumps are also used to generate the short-term smile effect observed in implied volatilities of option prices, see [Cont & Tankov \(2004\)](#). Furthermore, jumps are needed to properly model credit events like defaults and credit rating changes, see for instance [Jarrow, Lando & Turnbull \(1997\)](#). The short rate, typically set by a central bank, jumps up or down, usually by some quarters of a percent, see [Babbs & Webber \(1995\)](#). Models for the dynamics of financial quantities specified by stochastic differential equations (SDEs) with jumps have become increasingly popular. Models of this kind can be found, for instance, in [Merton \(1976\)](#), [Björk, Kabanov & Rungaldier \(1997\)](#), [Duffie, Pan & Singleton \(2000\)](#), [Kou \(2002\)](#), [Schönbucher \(2003\)](#), [Glasserman & Kou \(2003\)](#), [Cont & Tankov \(2004\)](#) and [Geman & Roncoroni \(2006\)](#). The areas of application of SDEs with jumps go far beyond finance. Other areas of application include economics, insurance, population dynamics, epidemiology, structural mechanics, physics, chemistry and biotechnology. In chemistry, for instance, the reactions of single molecules or coupled reactions yield stochastic models with jumps, see, for instance, [Turner, Schnell & Burrage \(2004\)](#), to indicate just one such application.

Since only a small class of jump diffusion SDEs admits explicit solutions, it is important to construct discrete-time approximations. The focus of this monograph is the numerical solution of SDEs with jumps via simulation. We consider pathwise scenario simulation, for which strong schemes are used, and Monte Carlo simulation, for which weak schemes are employed. Of course, there exist various alternative methods to Monte Carlo simulation that we only consider peripherally in this book when it is related to the idea of discrete-time

numerical approximations. These methods include Markov chain approximations, tree-based, and finite difference methods. The class of SDEs considered here are those driven by Wiener processes and Poisson random measures. Some authors consider the smaller class of SDEs driven by Wiener processes and homogeneous Poisson processes, while other authors analyze the larger class of SDEs driven by fairly general semimartingales. The class of SDEs driven by Wiener processes and Poisson jump measures with finite intensity appears to be large enough for realistic modeling of the dynamics of quantities in finance. Here continuous trading noise and a few single events model the typical sources of uncertainty. Furthermore, stochastic jump sizes and stochastic intensities, can be conveniently covered by using a Poisson jump measure. As we will explain, there are some numerical and theoretical advantages when modeling jumps with predescribed size. The simulation of some Lévy process driven dynamics will also be discussed. The development of a rich theory on simulation methods for SDEs with jumps, similar to that established for pure diffusion SDEs in Kloeden & Platen (1992), is still under way. This book aims to contribute to this theory motivated by applications in the area of finance. However, challenging problems in insurance, biology, chemistry, physics and many other areas can readily apply the presented numerical methods.

We consider discrete-time approximations of solutions of SDEs constructed on time discretizations $(t)_\Delta$, with maximum step size $\Delta \in (0, \Delta_0)$, with $\Delta_0 \in (0, 1)$. We call a time discretization *regular* if the jump times, generated by the Poisson measure, are not discretization times. On the other hand, if the jump times are included in the time discretization, then a *jump-adapted* time discretization is obtained. Accordingly, discrete-time approximations constructed on regular time discretizations are called *regular schemes*, while approximations constructed on jump-adapted time discretizations are called *jump-adapted schemes*.

Discrete-time approximations can be divided into two major classes: *strong* approximations and *weak* approximations, see Kloeden & Platen (1999). We say that a discrete-time approximation \mathbf{Y}^Δ , constructed on a time discretization $(t)_\Delta$, with maximum step size $\Delta > 0$, converges with *strong* order γ at time T to the solution \mathbf{X} of a given SDE, if there exists a positive constant C , independent of Δ , and a finite number $\Delta_0 \in (0, 1)$, such that

$$E(|\mathbf{X}_T - \mathbf{Y}_T^\Delta|) \leq C\Delta^\gamma, \quad (0.0.1)$$

for all $\Delta \in (0, \Delta_0)$. From the definition of the strong error on the left hand side of (0.0.1) one notices that strong schemes provide pathwise approximations of the original solution \mathbf{X} of the given SDE. These methods are therefore suitable for problems such as filtering, scenario simulation and hedge simulation, as well as the testing of statistical and other quantitative methods. In insurance, the area of dynamic financial analysis is well suited for applications of strong approximations.

On the other hand, we say that a discrete-time approximation \mathbf{Y}^Δ converges *weakly* with order β to \mathbf{X} at time T , if for each $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ there exists a positive constant C , independent of Δ , and a finite number, $\Delta_0 \in (0, 1)$, such that

$$|E(g(\mathbf{X}_T)) - E(g(\mathbf{Y}_T^\Delta))| \leq C\Delta^\beta, \quad (0.0.2)$$

for each $\Delta \in (0, \Delta_0)$. Here $\mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ denotes the set of $2(\beta + 1)$ continuously differentiable functions which, together with their partial derivatives of order up to $2(\beta + 1)$, have polynomial growth. This means that for $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ there exist constants $K > 0$ and $r \in \mathcal{N}$, possibly depending on g , such that

$$|\partial_y^j g(\mathbf{y})| \leq K(1 + |\mathbf{y}|^{2r}), \quad (0.0.3)$$

for all $\mathbf{y} \in \mathbb{R}^d$ and any partial derivative $\partial_{y^i}^j g(\mathbf{y})$ of order $j \leq 2(\beta + 1)$. Weak schemes provide approximations of the probability measure generated by the solution of a given SDE. These schemes are appropriate for problems such as derivative pricing, the evaluation of moments and the computation of risk measures and expected utilities.

Let us briefly discuss some relationships between strong and weak approximations. Let \mathbf{Y}^Δ be a discrete-time approximation, constructed on a time discretization $(t)_\Delta$, with strong order of convergence γ , see (0.0.1). Consider a function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfying the Lipschitz condition

$$|g(\mathbf{x}) - g(\mathbf{y})| \leq K|\mathbf{x} - \mathbf{y}|, \quad (0.0.4)$$

for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, where K is a positive constant. Then there exists a positive constant C , independent of Δ , and a finite number, $\Delta_0 \in (0, 1)$, such that by the Lipschitz condition (0.0.4) and the strong order γ we have

$$\left| E(g(\mathbf{X}_T)) - E\left(g\left(\mathbf{Y}_T^\Delta\right)\right) \right| \leq K E\left(\left|\mathbf{X}_T - \mathbf{Y}_T^\Delta\right|\right) \leq \Delta^\gamma \quad (0.0.5)$$

for each $\Delta \in (0, \Delta_0)$.

Since the set of Lipschitz continuous functions includes the set $\mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$, the above result implies that if a discrete-time approximation \mathbf{Y}^Δ achieves an order γ of strong convergence, then it also achieves *at least* an order $\beta = \gamma$ of weak convergence. We emphasize that the weak order obtained above is usually not sharp and, thus, the order of weak convergence could actually be higher than that of strong convergence. For instance, it is well-known and will later be shown that the Euler scheme typically achieves only strong order $\gamma = 0.5$ but weak order $\beta = 1.0$.

In the light of the estimate (0.0.5), one could think that the design of strong approximations is sufficient for any type of application, since these approximations can be also applied to weak problems. This is in principle true, but the resulting schemes might remain far from being optimal in terms of

computational efficiency. Let us consider as an example the strong Milstein scheme for pure diffusion SDEs, see [Milstein \(1974\)](#). By adding the double Wiener integrals to the Euler scheme one obtains the Milstein scheme, thus enhancing the order of strong convergence from $\gamma = 0.5$ to $\gamma = 1.0$. Nonetheless, the order of weak convergence of the Milstein scheme equals $\beta = 1.0$, which is not an improvement over the order of weak convergence of the Euler scheme. Therefore, to price a European call option, for example, the Euler scheme is often computationally more efficient than the Milstein scheme, since it has fewer terms and the same order of weak convergence. Furthermore, the numerical stability of the Milstein scheme can be worse than that of the Euler scheme. This simple example indicates that to construct efficient higher order weak approximations, one should not take the naive approach of just using higher order strong approximations. Furthermore, as will be discussed, when designing weak schemes one has the freedom of using simple multi-point distributed random variables to approximate the underlying multiple stochastic integrals. These multi-point distributed random variables lead to highly efficient implementations of weak schemes.

For the approximation of the expected value of a function g of the solution \mathbf{X}_T at a final time T , there exist alternative numerical methods. Under suitable conditions, the pricing function $u(\mathbf{x}, t) = E(g(\mathbf{X}_T) | \mathbf{X}_t = \mathbf{x})$ can be expressed as a solution of a *partial integro differential equation* (PIDE). Therefore, an approximation of the pricing function $u(\mathbf{x}, t)$ can be obtained by solving the corresponding PIDE via finite difference or finite element methods, see, for instance, [D'Halluin, Forsyth & Vetzal \(2005\)](#) and [Cont & Voltchkova \(2005\)](#). These methods are computationally efficient when we have a low dimensional underlying factor process \mathbf{X} . Moreover, it is easy to incorporate early exercise features, as those arising in the pricing of Bermudan and American options. However, when the underlying stochastic process \mathbf{X} has dimension higher than two or three, finite difference and finite element methods become difficult to be implemented and turn out to be computationally, prohibitively expensive.

Monte Carlo simulation is well suited to tackle high dimensional problems. It has the great advantage that its computational complexity increases, in principle, polynomially with the dimension of the problem. Consequently, the curse of dimensionality applies in a milder fashion to Monte Carlo simulation than it does to most other numerical methods. Additionally, Monte Carlo simulation is well suited to parallel hardware devices and seems to provide solutions where no alternative is known.

The focus of this book is on the numerical solution of stochastic differential equations (SDEs) with jumps via simulation methods, motivated by problems in finance. The monograph is divided into three parts. The first part, covering Chaps. 1 up to 4, introduces SDEs with jumps, presents exact simulation methods, describes the benchmark approach as a general financial modeling framework and introduces Wagner-Platen expansions. The second part, comprising Chaps. 5 up to 10, considers strong approximations of jump diffusion

and pure jump SDEs. It also includes some discussions on parameter estimation and filtering as well as their relation to strong approximation methods. Finally, the third part, which is composed of Chaps. 11 up to 17, introduces weak approximations for Monte Carlo simulation and discusses efficient implementations of weak schemes and numerical stability. Here the simulation of hedge ratios, efficient variance reduction techniques, Markov chain approximations and finite difference methods are discussed in the context of weak approximation.

The monograph Kloeden & Platen (1992) and its printings in (1995) and (1999) aimed to give a reasonable overview on the literature on the numerical solution of SDEs via simulation methods. Over the last two decades the field has grown so rapidly that it is no longer possible to provide a reasonably fair presentation of the area. This book is, therefore, simply presenting results that the authors were in some form involved with. There are several other lines of research that may be of considerable value to those who have an interest in this field. We apologize to those who would have expected other interesting topics to be covered by the book. Unfortunately, this was not possible due to limitations of space.

For further reading also in areas that are related but could not be covered we may refer the reader to various well-written books, including Ikeda & Watanabe (1989), Niederreiter (1992), Elliott, Aggoun & Moore (1995), Milstein (1995a), Embrechts, Klüppelberg & Mikosch (1997), Björk (1998), Karatzas & Shreve (1998), Mikosch (1998), Kloeden & Platen (1999), Shiryaev (1999), Bielecki & Rutkowski (2002), Borodin & Salminen (2002), Jäckel (2002), Joshi (2003), Schönbucher (2003), Shreve (2003a, 2003b), Cont & Tankov (2004), Glasserman (2004), Higham (2004), Milstein & Tretjakov (2004), Achdou & Pironneau (2005), Brigo & Mercurio (2005), Elliott & Kopp (2005), Klebaner (2005), McLeish (2005), McNeil, Frey & Embrechts (2005), Musiela & Rutkowski (2005), Øksendal & Sulem (2005), Protter (2005), Chan & Wong (2006), Delbaen & Schachermayer (2006), Elliott & van der Hoek (2006), Malliavin & Thalmaier (2006), Platen & Heath (2006), Seydel (2006), Asmussen & Glynn (2007), Lamberton & Lapeyre (2007) and Jeanblanc, Yor & Chesney (2009).

The book has been used as reference for the Masters in Quantitative Finance and the PhD program at the University of Technology in Sydney, as well as for courses and workshops that the first author has presented in various places.

The formulas in the book are numbered according to the chapter and section where they appear. Assumptions, theorems, lemmas, definitions and corollaries are numbered sequentially in each section. The most common notations are listed at the beginning and an *Index of Keywords* is given at the end of the book. Some readers may find the *Author Index* at the end of the book useful. Each chapter finishes with some *Exercises* with *Solutions* given in Chap. 18. These are aimed to support the study of the material.

XXVIII Motivation and Brief Survey

We conclude this brief survey with the remark that the practical application and theoretical understanding of numerical methods for stochastic differential equations with jumps are still under development. This book shall stimulate interest and further work on such methods. The *Bibliographical Notes* at the end of this research monograph may be of assistance.

Stochastic Differential Equations with Jumps

Stochastic differential equations (SDEs) with jumps provide the most flexible, numerically accessible, mathematical framework that allows us to model the evolution of financial and other random quantities over time. In particular, feedback effects can be easily modeled and jumps enable us to model events. This chapter introduces SDEs driven by Wiener processes, Poisson processes and Poisson random measures. We also discuss the Itô formula, the Feynman-Kac formula and the existence and uniqueness of solutions of SDEs. These tools and results provide the basis for the application and numerical solution of stochastic differential equations with jumps.

1.1 Stochastic Processes

Stochastic Process

If not otherwise stated, throughout the book we shall assume that there exists a common underlying probability space (Ω, \mathcal{A}, P) consisting of the sample space Ω , the sigma-algebra or collection of events \mathcal{A} , and the probability measure P , see for instance [Shiryayev \(1984\)](#). One typically observes a collection of random variables X_{t_0}, X_{t_1}, \dots , which describe the evolution of financial quantities, for instance, daily closing prices of an index at the observation times $t_0 < t_1 < \dots$. The collection of random variables is indexed by the time t , and we call \mathcal{T} the *time set*. The *state space* of X is here the d -dimensional Euclidean space \mathbb{R}^d , $d \in \mathcal{N} = \{1, 2, \dots\}$, or a subset of it.

Definition 1.1.1. *We call a family $X = \{X_t, t \in \mathcal{T}\}$ of random variables $X_t \in \mathbb{R}^d$ a d -dimensional stochastic process, where the totality of its finite-dimensional distribution functions*

$$F_{X_{t_{i_1}}, \dots, X_{t_{i_j}}}(x_{i_1}, \dots, x_{i_j}) = P(X_{t_{i_1}} \leq x_{i_1}, \dots, X_{t_{i_j}} \leq x_{i_j}) \quad (1.1.1)$$

for $i_j \in \{0, 1, \dots\}$, $j \in \mathcal{N}$, $x_{i_j} \in \mathbb{R}^d$ and $t_{i_j} \in \mathcal{T}$ determines its probability law.

We set the time set to the interval $\mathcal{T} = [0, \infty)$ if not otherwise stated. On some occasions the time set may become the bounded interval $[0, T]$ for $T \in (0, \infty)$ or a set of discrete time points $\{t_0, t_1, t_2, \dots\}$, where $t_0 < t_1 < t_2 < \dots$.

One can distinguish between various classes of stochastic processes according to specific properties. First, we aim to identify stochastic processes that are suitable as basic building blocks for realistic models. From a quantitative point of view it is essential to select stochastic processes that allow explicit formulas or at least fast and efficient numerical methods for calculating moments, probabilities, option prices or other quantities. In the following we set usually the dimension d equal to one. However, most concepts and results we will introduce generalize to the multi-dimensional case.

Stationary Process

Important information about a stochastic process is provided by the *mean*

$$\mu(t) = E(X_t) \quad (1.1.2)$$

and the *variance*

$$v(t) = \text{Var}(X_t) = E((X_t - \mu(t))^2) \quad (1.1.3)$$

for $t \geq 0$, as well as the *covariance*

$$C(s, t) = \text{Cov}(X_s, X_t) = E((X_s - \mu(s))(X_t - \mu(t))) \quad (1.1.4)$$

for $s, t \in \mathcal{T}$.

The concept of a market equilibrium has been very important in economics and finance. The corresponding class of stochastic processes for modeling equilibria is that of *stationary processes* since they allow us to express probabilistically a form of equilibrium. For instance, interest rates, dividend rates, inflation rates, volatilities, consumption rates, hazard rates and credit spreads are likely to be modeled by stationary processes since they typically exhibit in reality some type of equilibrium.

Definition 1.1.2. *We say that a stochastic process $X = \{X_t, t \geq 0\}$ is stationary if its joint distributions are all invariant under time displacements, that is if*

$$F_{X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h}} = F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}} \quad (1.1.5)$$

for all $h > 0$, $t_i \geq 0$, $i \in \{1, 2, \dots, n\}$ and $n \in \mathcal{N}$.

The random values X_t of a stationary process X have the same distribution for all $t \in \mathcal{T}$. Therefore, means, variances and covariances satisfy the equations

$$\mu(t) = \mu(0), \quad v(t) = v(0) \quad \text{and} \quad C(s, t) = c(t - s) \quad (1.1.6)$$

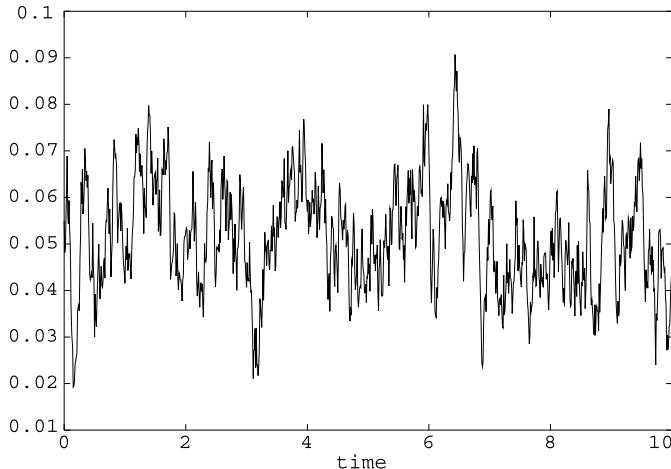


Fig. 1.1.1. Sample path for the Vasicek interest rate model, $T = 10$

for all $s, t \geq 0$, where $c : \mathbb{R} \rightarrow \mathbb{R}$ is a function. However, to ensure this property its initial value needs to be modeled as an approximate random variable. Therefore, in practice we may typically model equilibrium type dynamics by employing a stationary process conditioned on an observed initial value. The concept of stationarity extends in a straightforward manner to multi-dimensional stochastic processes.

For illustration, in Fig. 1.1.1 we display a trajectory of a stationary continuous Gaussian process with mean $\mu(t) = 0.05$, variance $v(t) = 0.1$ and initial value $X_0 = 0.05$. We may interpret it as the sample path of some interest rate obtained under the, so called, Vasicek interest rate model, see [Vasicek \(1977\)](#). The process fluctuates around a reference level and appears to revert back towards its mean. As a stationary process we may continue to observe its movements over longer time horizons with its mean, variance and covariances not changing.

Filtration as Information Structure

As we will see later, in finance the notion of a stochastic process for which its last observed value provides the best forecast for its future values, plays a fundamental role. Forecasting is primarily based on current available information. Financial markets are strongly influenced by information. Essentially, it is information that drives the dynamics of markets. Therefore, in financial modeling a precise definition of the information structure available is important.

As already indicated previously, our modeling is based on a given probability space (Ω, \mathcal{A}, P) , which consists of the sample space Ω , the sigma-algebra \mathcal{A} that is generated by all events and the given probability measure P . On

such a probability space we consider dynamics, typically of a financial market model, that is based on the observation of a continuous time stochastic vector process $\mathbf{X} = \{\mathbf{X}_t \in \Re^d, t \geq 0\}$, $d \in \mathcal{N}$. We denote by $\hat{\mathcal{A}}_t$ the time t *information set*, which is the sigma-algebra generated by the events that are known at time $t \geq 0$, see [Shiryayev \(1984\)](#). Our interpretation of $\hat{\mathcal{A}}_t$ is that it represents the information available at time t , which is obtained from the observed values of the vector process \mathbf{X} up to time t . More precisely, it is the sigma-algebra

$$\hat{\mathcal{A}}_t = \sigma\{\mathbf{X}_s : s \in [0, t]\}$$

generated from all observations of \mathbf{X} up to time t . Since information is not lost, the increasing family

$$\underline{\mathcal{A}} = \{\hat{\mathcal{A}}_t, t \geq 0\}$$

of information sets $\hat{\mathcal{A}}_t$ satisfies, for any sequence $0 \leq t_1 < t_2 < \dots < \infty$ of observation times, the relation $\hat{\mathcal{A}}_{t_1} \subseteq \hat{\mathcal{A}}_{t_2} \subseteq \dots \subseteq \hat{\mathcal{A}}_\infty = \cup_{t \geq 0} \hat{\mathcal{A}}_t$.

For technical reasons one introduces the information set \mathcal{A}_t as the *augmented* sigma-algebra of $\hat{\mathcal{A}}_t$ for each $t \geq 0$. It is augmented by every null set in $\hat{\mathcal{A}}_\infty$ such that it belongs to \mathcal{A}_0 , and also to each $\hat{\mathcal{A}}_t$. One also says that \mathcal{A}_t is *complete*. Define $\mathcal{A}_{t+} = \cap_{\varepsilon > 0} \mathcal{A}_{t+\varepsilon}$ as the sigma-algebra of events immediately after $t \in [0, \infty)$. The family $\underline{\mathcal{A}} = \{\mathcal{A}_t, t \geq 0\}$ is called *right continuous* if $\mathcal{A}_t = \mathcal{A}_{t+}$ holds for every $t \geq 0$. Such a right-continuous family $\underline{\mathcal{A}} = \{\mathcal{A}_t, t \geq 0\}$ of information sets one calls a *filtration*. Such a filtration can model the evolution of information as it becomes available over time. We define for our purposes \mathcal{A} as the smallest sigma-algebra that contains $\mathcal{A}_\infty = \cup_{t \geq 0} \mathcal{A}_t$. From now on, if not stated otherwise, we always assume in this book a *filtered probability space* $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ to be given.

Any right-continuous stochastic process $Y = \{Y_t, t \geq 0\}$ generates its *natural filtration* $\underline{\mathcal{A}}^Y = \{\mathcal{A}_t^Y, t \geq 0\}$, which is the sigma-algebra generated by Y up to time t . For a given model with a vector process \mathbf{X} we typically set $\underline{\mathcal{A}} = \underline{\mathcal{A}}^{\mathbf{X}}$ with $\mathcal{A}_t = \mathcal{A}_t^{\mathbf{X}}$.

If for a process $Z = \{Z_t, t \geq 0\}$ and each time $t \geq 0$ the random variable Z_t is $\mathcal{A}_t^{\mathbf{X}}$ -measurable, then Z is called *adapted* to $\underline{\mathcal{A}}^{\mathbf{X}} = \{\mathcal{A}_t^{\mathbf{X}}, t \geq 0\}$. The history of the process Z until time t is then covered by the information set $\mathcal{A}_t^{\mathbf{X}}$. The completeness of the information set $\mathcal{A}_t^{\mathbf{X}}$, which means that it includes all null events, allows us to conclude that for two random variables Z_1 and Z_2 , where $Z_1 = Z_2$ almost surely (a.s.) and Z_1 is $\mathcal{A}_t^{\mathbf{X}}$ -measurable, Z_2 is also $\mathcal{A}_t^{\mathbf{X}}$ -measurable.

Conditional Expectations

The notion of conditional expectation is central to many of the concepts that arise in applications of stochastic processes and also in finance. The mean value or expectation $E(X)$ is the coarsest estimate that we have for an *integrable random variable* X , that is, for which $E(|X|) < \infty$, see [Shiryayev \(1984\)](#). If we know that some event A has occurred we may be able to improve on this

estimate. For instance, suppose that the event $A = \{\omega \in \Omega : X(\omega) \in [a, b]\}$ has occurred. Then in evaluating our estimate of the value of X we need only to consider corresponding values of X in $[a, b]$ and weight them according to their likelihood of occurrence, which thus becomes the conditional probability given this event, see [Shiryayev \(1984\)](#).

The resulting estimate is called the *conditional expectation* of X given the event A and is denoted by $E(X|A)$. For a continuous random variable X with a density function f_X the corresponding *conditional density* is

$$f_X(x|A) = \begin{cases} 0 & \text{for } x < a \text{ or } b < x \\ \frac{f_X(x)}{\int_a^b f_X(s) ds} & \text{for } x \in [a, b] \end{cases},$$

with the *conditional expectation*

$$E(X|A) = \int_{-\infty}^{\infty} x f_X(x|A) dx = \frac{\int_a^b x f_X(x) dx}{\int_a^b f_X(x) dx}, \quad (1.1.7)$$

which is conditioned on the event A and is, thus, a number.

More generally let (Ω, \mathcal{A}, P) be a given probability space with an integrable random variable X . We denote by \mathcal{S} a sub-sigma-algebra of \mathcal{A} , thus representing a coarser type of information than is given by $\mathcal{S} \subset \mathcal{A}$. We then define the *conditional expectation* of X with respect to \mathcal{S} , which we denote by $E(X|\mathcal{S})$, as an \mathcal{S} -measurable function satisfying the equation

$$\int_Q E(X|\mathcal{S})(\omega) dP(\omega) = \int_Q X(\omega) dP(\omega), \quad (1.1.8)$$

for all $Q \in \mathcal{S}$. The Radon-Nikodym theorem, see [Shiryayev \(1984\)](#), guarantees the existence and uniqueness of the random variable $E(X|\mathcal{S})$. Note that $E(X|\mathcal{S})$ is a random variable defined on the coarser probability space (Ω, \mathcal{S}, P) and thus on (Ω, \mathcal{A}, P) . However, X is usually not a random variable on (Ω, \mathcal{S}, P) , but when it is we have

$$E(X|\mathcal{S}) = X, \quad (1.1.9)$$

which is the case when X is \mathcal{S} -measurable.

The following results are important when handling the evolution of stochastic processes in conjunction with an evolving information structure, as is often the case in finance.

For nested sigma-algebras $\mathcal{S} \subset \mathcal{T} \subset \mathcal{A}$ and an integrable random variable X we have the *iterated conditional expectations*

$$E(E(X|\mathcal{T})|\mathcal{S}) = E(X|\mathcal{S}) \quad (1.1.10)$$

almost surely. Since most equations and relations we will formulate in this book hold almost surely, we will typically suppress these words from now on. When X is independent of the events in \mathcal{S} we have

$$E(X | \mathcal{S}) = E(X). \quad (1.1.11)$$

By setting $\mathcal{S} = \{\emptyset, \Omega\}$ it can be seen that

$$E(E(X | \mathcal{S})) = E(X). \quad (1.1.12)$$

Conditional expectations have similar properties to those of ordinary integrals such as the linearity property

$$E(\alpha X + \beta Y | \mathcal{S}) = \alpha E(X | \mathcal{S}) + \beta E(Y | \mathcal{S}), \quad (1.1.13)$$

where X and Y are integrable random variables and $\alpha, \beta \in \mathbb{R}$ are deterministic constants. Furthermore, if X is \mathcal{S} -measurable, then

$$E(XY | \mathcal{S}) = X E(Y | \mathcal{S}). \quad (1.1.14)$$

Finally, one has the order preserving property

$$E(X | \mathcal{S}) \leq E(Y | \mathcal{S}) \quad (1.1.15)$$

if $X \leq Y$ a.s.

The conditional expectation $E(X | \mathcal{S})$ is in some sense obtained by smoothing X over the events in \mathcal{S} . Thus, the finer the information set \mathcal{S} , the more $E(X | \mathcal{S})$ resembles the random variable X .

Wiener Process

In contrast to stationary processes we consider now *stochastic processes with stationary independent increments*. These basic processes have mathematical properties that make them suitable as fundamental building blocks in stochastic modeling and, thus, in financial modeling. The random increments $X_{t_{j+1}} - X_{t_j}$, $j \in \{0, 1, \dots, n-1\}$, of these processes are independent for any sequence of time instants $t_0 < t_1 < \dots < t_n$ in $[0, \infty)$ for all $n \in \mathbb{N}$. If $t_0 = 0$ is the smallest time instant, then the initial value X_0 and the random increment $X_{t_j} - X_0$ for any other $t_j \in [0, \infty)$ are also required to be independent. Additionally, the increments $X_{t+h} - X_t$ are assumed to be stationary, that is $X_{t+h} - X_t$ has the same distribution as $X_h - X_0$ for all $h > 0$ and $t \geq 0$.

The most important continuous process with stationary independent increments is the *Wiener process*. Bachelier was the first who employed, already in 1900, such a mathematical object in his modeling of asset prices at the Paris Bourse, see [Bachelier \(1900\)](#) or [Davis & Etheridge \(2006\)](#). He did this even before Einstein used an equivalent mathematical construct that we call now the Wiener process or *Brownian motion*, see [Einstein \(1905\)](#).

Definition 1.1.3. We define the standard Wiener process $W = \{W_t, t \geq 0\}$ as an \mathcal{A} -adapted process with Gaussian stationary independent increments and continuous sample paths for which

$$W_0 = 0, \quad \mu(t) = E(W_t) = 0, \quad \text{Var}(W_t - W_s) = t - s \quad (1.1.16)$$

for all $t \geq 0$ and $s \in [0, t]$.

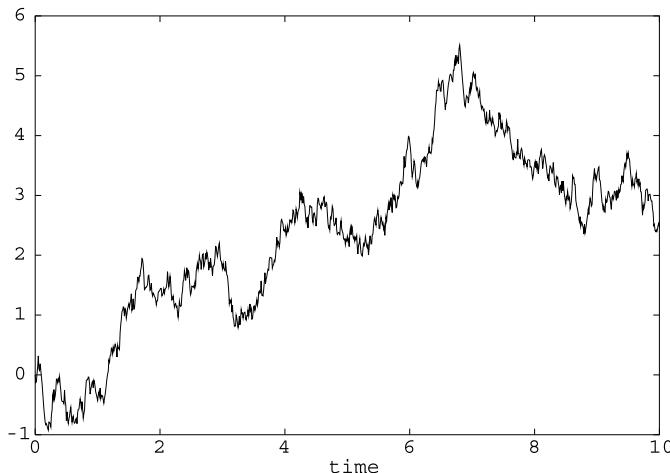


Fig. 1.1.2. Sample path of a standard Wiener process, $T = 10$

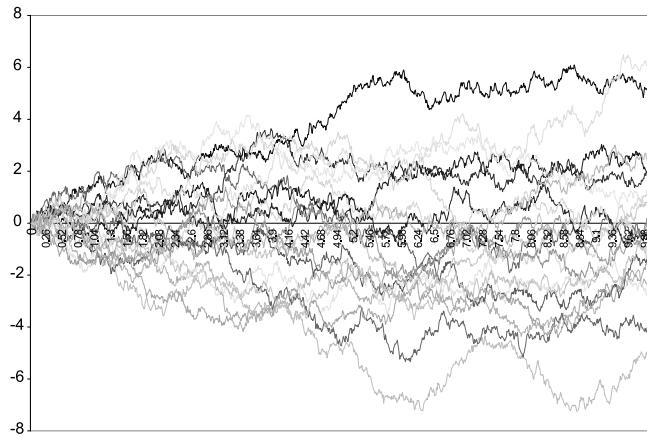


Fig. 1.1.3. Sample paths of a Wiener process, $T = 10$

The Wiener process is a continuous time stochastic process with independent Gaussian distributed increments generating continuous sample paths. This process is also known as Brownian motion, since it can model the motion of a pollen grain under the microscope, as observed by Robert Brown in the early 19th century.

In Fig. 1.1.2 we plot a sample path of a standard Wiener process. To visualize more of its probabilistic properties we display 20 such trajectories in Fig. 1.1.3

The Wiener process has fundamental mathematical properties and is used as a basic building block in many applications, in particular in finance. The Wiener process is the most basic stochastic process that allows us to model continuous uncertainty, as it arises for instance as trading noise. Note that

the Wiener process is *not* a stationary process that is conditioned on its known initial value, as can be seen from its increasing variance, see (1.1.16). However, its independent increments have the same distribution when taken over periods of the same time length and are therefore stationary in this sense.

There exists also a multi-dimensional version of the above Wiener process. We call the vector process $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, W_t^2, \dots, W_t^m)^\top, t \geq 0\}$ an *m-dimensional standard Wiener process* if each of its components $W^j = \{W_t^j, t \geq 0\}$, $j \in \{1, 2, \dots, m\}$ is a scalar \mathcal{A} -adapted standard Wiener process and the Wiener processes W^k and W^j are independent for $k \neq j$, $k, j \in \{1, 2, \dots, m\}$.

This means that according to Definition 1.1.3, each random variable W_t^j is Gaussian and \mathcal{A}_t -measurable with

$$E(W_t^j | \mathcal{A}_0) = 0 \quad (1.1.17)$$

and we have independent increments $W_t^j - W_s^j$ such that

$$E(W_t^j - W_s^j | \mathcal{A}_s) = 0 \quad (1.1.18)$$

for $t \geq 0$, $s \in [0, t]$ and $j \in \{1, 2, \dots, m\}$. Moreover, one has the additional property that

$$E((W_t^j - W_s^j)(W_t^k - W_s^k) | \mathcal{A}_s) = \begin{cases} (t-s) & \text{for } k=j \\ 0 & \text{otherwise} \end{cases} \quad (1.1.19)$$

for $t \geq 0$, $s \in [0, t]$ and $j, k \in \{1, 2, \dots, m\}$.

Poisson Process

In finance and insurance one also needs to be able to model various random events, such as defaults, catastrophes or operational failures. There exists a piecewise constant stochastic process with independent increments which counts events. This is the *Poisson process* which can be conveniently used to model event driven uncertainty.

Definition 1.1.4. A Poisson process $N = \{N_t, t \geq 0\}$ with intensity $\lambda > 0$ is a piecewise constant process with stationary independent increments with initial value $N_0 = 0$ such that $N_t - N_s$ is Poisson distributed with intensity λ_{t-s} , that is, with probability

$$P(N_t - N_s = k) = \frac{e^{-\lambda(t-s)} (\lambda(t-s))^k}{k!} \quad (1.1.20)$$

for $k \in \{0, 1, \dots\}$, $t \geq 0$ and $s \in [0, t]$.

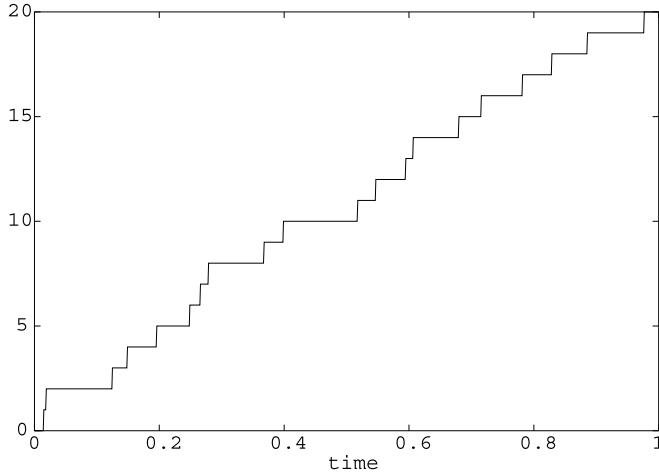


Fig. 1.1.4. Trajectory of a standard Poisson process with intensity $\lambda = 20$

For the Poisson process N with intensity λ we have the mean

$$\mu(t) = E(N_t) = \lambda t, \quad (1.1.21)$$

and the variance

$$v(t) = \text{Var}(N_t) = E((N_t - \mu(t))^2) = \lambda t \quad (1.1.22)$$

for $t \geq 0$.

In Fig. 1.1.4 we plot a graph for a Poisson process with intensity $\lambda = 20$. According to (1.1.21), we should expect on average 20 events to happen during the time period $[0, 1]$, which is about the case for this trajectory. We emphasize that also the Poisson process is *not* a stationary process conditioned on its known initial value. However, its increments have the same Poisson distribution over time intervals of the same length and are in this sense stationary.

A Poisson process N counts events and generates an increasing sequence of *jump times* τ_1, τ_2, \dots related to each event that it counts. Thus, N_t equals the number of events that occurred up until time $t \geq 0$. For $t \geq 0$ let the time τ_{N_t} denote the last time that N_t made a jump, that is,

$$\tau_k = \inf\{t \geq 0 : N_t \geq k\} \quad (1.1.23)$$

for $k \in \mathcal{N}$, $t \geq 0$.

In some applications, as in the modeling of defaults, the intensity λ_t that a certain type of event occurs may depend on the time $t \geq 0$. This leads to a *time transformed Poisson process* $N = \{N_t, t \geq 0\}$, where

$$P(N_t - N_s = k) = \frac{\exp\left\{-\int_s^t \lambda_z dz\right\} \left(\int_s^t \lambda_z dz\right)^k}{k!} \quad (1.1.24)$$

for $k \in \{0, 1, \dots\}$, $t \geq 0$ and $s \in [0, t]$.

To obtain more flexibility in modeling we introduce the *mark* ξ_k of the k th event which, for instance, could be the *recovery rate* of the k th default.

Let us assume for the moment that the k th mark ξ_k is deterministic, $k \in \{1, 2, \dots\}$. We can then consider the process $Y = \{Y_t, t \geq 0\}$ with

$$Y_t = \sum_{k=1}^{N_t} \xi_k \quad (1.1.25)$$

for $t \geq 0$. If the intensity process $\lambda = \{\lambda_t, t \geq 0\}$ is deterministic and the k th mark ξ_k is deterministic for each $k \in \{1, 2, \dots\}$, then it follows from (1.1.24) that the mean $\mu(t)$ of Y_t is given by the expression

$$\mu(t) = E(Y_t) = \sum_{k=1}^{\infty} \xi_k P(N_t = k), \quad (1.1.26)$$

where the above probabilities are expressed in (1.1.24). For instance, if one chooses $\xi_1 = 1$ and $\xi_k = 0$ for $k \in \{2, 3, \dots\}$, then this simple process allows us to model the credit worthiness C_t of an obligor at time t by setting

$$C_t = 1 - Y_t.$$

The credit worthiness may start at time zero with $C_0 = 1$. It declines to zero at the time when the first default arises. The expected value $E(C_t)$ of the above credit worthiness at time t equals by (1.1.26) and (1.1.24) the expression

$$E(C_t) = P(N_t = 0) = \exp \left\{ - \int_0^t \lambda_z dz \right\}$$

for $t \geq 0$.

Compound Poisson Process

One often needs to differentiate between certain types of events, for instance, those with different recovery rates at a default. A corresponding intensity for the occurrence of each type of possible event is then required.

To construct a process that models sequences of different types of events, we consider a Poisson process N with intensity $\lambda > 0$, together with a sequence of independent identically distributed (i.i.d.) random variables ξ_1, ξ_2, \dots that are independent of N . Here

$$F_{\xi_1}(z) = P(\xi_1 \leq z) \quad (1.1.27)$$

denotes for $z \in \mathbb{R}$ the corresponding value of the distribution function of ξ_1 .

This allows us to construct the *compound Poisson process* $Y = \{Y_t, t \geq 0\}$, where $Y_0 = 0$ and

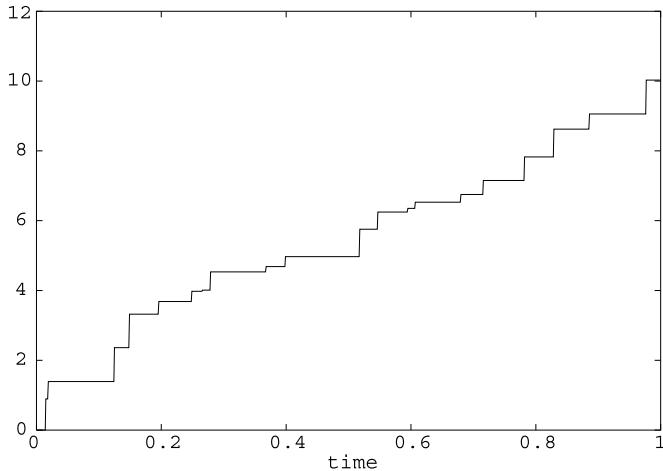


Fig. 1.1.5. Compound Poisson process

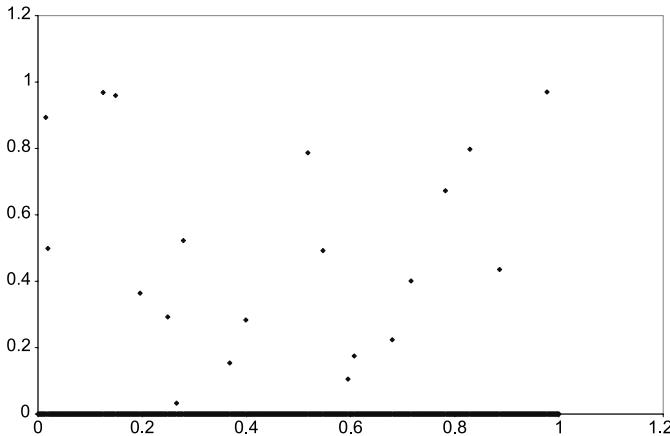


Fig. 1.1.6. Pairs of jump times and marks

$$Y_t = \sum_{k=1}^{N_t} \xi_k \quad (1.1.28)$$

for $t \geq 0$. A compound Poisson process generates a sequence of pairs $(\tau_k, \xi_k)_{k \in \mathcal{N}}$ of *jump times* τ_k and *marks* ξ_k . In Fig. 1.1.5 we show the trajectory of a compound Poisson process Y where the i.i.d. random variables are uniformly distributed on the interval $[0, 1]$, and N is as in Fig. 1.1.4. In Fig. 1.1.6 we plot the points generated by the pairs of jump times and marks of the path of the compound Poisson process Y shown in Fig. 1.1.5.

A compound Poisson process is again a process with stationary independent increments. Its trajectory is fully characterized by the sequence of pairs $(\tau_k, \xi_k)_{k \in \mathcal{N}}$.

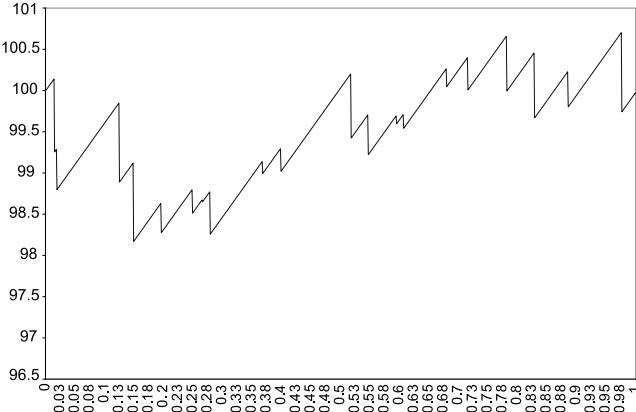


Fig. 1.1.7. Risk reserve process

The following simple, but important insurance model involves the above compound Poisson process Y . The *Cramér-Lundberg model* for the *risk reserve* X_t at time t of an insurance company can be described in the form

$$X_t = X_0 + c t - Y_t, \quad (1.1.29)$$

where the claim process $Y = \{Y_t, t \geq 0\}$ is modeled as a compound Poisson process with i.i.d. claim sizes $\xi_k > 0$, $k \in \mathcal{N}$. Here c denotes the premium rate that describes the premium payment per unit of time that the insurance company collects. We show in Fig. 1.1.7 the trajectory of a risk reserve process X when choosing $c = 10$ and $X_0 = 100$.

Poisson Measure

For the modeling of events that arise with high intensity and which have different marks, the notion of a *Poisson measure* is useful. A simple example for a Poisson measure is given by a compound Poisson process with marks that are uniformly distributed on $[0, 1]$, as we will see below.

It is a fundamental feature of the Poisson process that due to the independence of its increments the location of the set of points in the time interval $[0, 1]$, see Fig. 1.1.6, can be intuitively interpreted as if one has generated N_1 independent uniformly $U(0, 1)$ distributed random variables for obtaining the jump times. On the other hand, the independent marks ξ_k of the compound Poisson process Y , displayed in Fig. 1.1.6, are also independent uniformly distributed on $[0, 1]$. Consequently, the pairs (τ_k, ξ_k) , for $k \in \mathcal{N}$, shown in Fig. 1.1.6, are independent uniformly distributed in the square $[0, 1] \times [0, 1]$. Such a graph can be interpreted as the realization of some Poisson measure.

The general notion of a Poisson measure is rather technical. However, it represents simply a compound Poisson process as long as the total intensity

remains finite. It is also important that a Poisson measure can accommodate infinite total intensity.

We now introduce the *mark set*

$$\mathcal{E} = \mathbb{R} \setminus \{0\}. \quad (1.1.30)$$

Here the element $\{0\}$ is excluded which allows us to avoid conveniently in our modeling jumps of size zero. Let $\mathcal{B}(\Gamma)$ denote the smallest sigma-algebra containing all open sets of a set Γ . Now, we consider on $\mathcal{E} \times [0, \infty)$ a given *intensity measure* of the form

$$\nu_\varphi(dv \times dt) = \varphi(dv) dt, \quad (1.1.31)$$

where $\varphi(\cdot)$ is a measure on $\mathcal{B}(\mathcal{E})$ with

$$\int_{\mathcal{E}} \min(1, v^2) \varphi(dv) < \infty. \quad (1.1.32)$$

The corresponding *Poisson measure* $p_\varphi(\cdot)$ on $\mathcal{E} \times [0, \infty)$, see [Protter \(2005\)](#), is assumed to be such that for $T \in (0, \infty)$ and each set A from the product-sigma-algebra of $\mathcal{B}(\mathcal{E})$ and $\mathcal{B}([0, T])$ the random variable $p_\varphi(A)$, which counts the number of points in $A \subseteq \mathcal{E} \times [0, \infty)$, is Poisson distributed with intensity

$$\nu_\varphi(A) = \int_0^T \int_{\mathcal{E}} \mathbf{1}_{\{(v,t) \in A\}} \varphi(dv) dt. \quad (1.1.33)$$

This means, one has

$$P(p_\varphi(A) = \ell) = \frac{\nu_\varphi(A)^\ell}{\ell!} e^{-\nu_\varphi(A)} \quad (1.1.34)$$

for $\ell \in \{0, 1, \dots\}$. For disjoint sets $A_1, \dots, A_r \subseteq \mathcal{E} \times [0, T]$, $r \in \mathcal{N}$, the random variables $p_\varphi(A_1), \dots, p_\varphi(A_r)$ are assumed to be independent.

For example, the points displayed in Fig. 1.1.6 can be interpreted as a realization of a Poisson measure on $[0, 1] \times [0, T]$ with $\varphi(dv) = \lambda dv$, where $\frac{\varphi(dv)}{dv} = \lambda = 20$ and $T = 1$. Whenever it is convenient, we shall use the equivalent notations $p_\varphi(dv, dt) = p_\varphi(dv \times dt)$ for a Poisson measure and $\nu_\varphi(dv, dt) = \nu_\varphi(dv \times dt)$ for the corresponding intensity measure.

Lévy Processes

The Wiener process, the Poisson process and the compound Poisson process are processes with stationary independent increments. We emphasize that they are not stationary processes for any initial values. Clearly, they do not reflect equilibrium dynamics. However, they are basic objects in stochastic modeling which we will exploit later for applications in finance and insurance. The following class of *Lévy processes* generalizes the just mentioned processes

and has been used in recent years by a number of researchers in financial and insurance modeling, see, for instance, Barndorff-Nielsen & Shephard (2001), Geman, Madan & Yor (2001), Eberlein (2002) and Kou (2002). It includes as special cases the Poisson process and the Wiener process. These processes have stationary independent increments and enjoy a number of convenient mathematical properties.

Definition 1.1.5. A stochastic process $X = \{X_t, t \geq 0\}$ with $X_0 = 0$ a.s. is called a Lévy process if

1. X is a process with independent increments, where $X_t - X_s$ for $0 \leq s < t < \infty$ is independent of X_r with $0 \leq r \leq s$;
2. X has stationary increments, which means that $X_t - X_s$ has for $0 \leq s < t < \infty$ the same distribution as X_{t-s} ;
3. X is continuous in probability, that is $X_s \xrightarrow{P} \lim_{t \rightarrow s} X_t$ for $s \in [0, \infty)$.

A popular asset price model is the *exponential Lévy model* $S = \{S_t, t \in [0, \infty)\}$, where

$$S_t = S_0 \exp\{X_t\} \quad (1.1.35)$$

denotes the asset price at time $t \geq 0$ with $S_0 > 0$ and X is a given Lévy process. Some exponential Lévy models have been already studied, for instance, in Samuelson (1955), Black & Scholes (1973) and Merton (1973b, 1976). Note that, since the asset price here is an exponential, it is guaranteed to stay positive, which is a desirable property of an asset price.

If X is a Lévy process, then it can be shown that it can be decomposed into the form

$$X_t = \alpha t + \beta W_t + \int_0^t \int_{|v|<1} v (p_\varphi(dv, ds) - \varphi(dv) ds) + \int_0^t \int_{|v|\geq 1} v p_\varphi(dv, ds) \quad (1.1.36)$$

for $t \geq 0$, $X_0 \in \mathfrak{R}$, see, for instance, Protter (2005). Here $W = \{W_t, t \geq 0\}$ is a standard Wiener process. Furthermore, p_φ is a *Poisson measure* as defined previously. For any set $A \in \mathcal{E}$ the process $p_\varphi(A) = \{p_\varphi(A, [0, t]), t \geq 0\}$ is a Poisson process independent of W with intensity $\varphi(A)$, where $\varphi(dv)$ is called a *Lévy measure*. This measure characterizes the jumps of a Lévy process and is defined on \mathcal{E} assuming that (1.1.32) holds. Note that for disjoint sets $A, B \in \mathcal{E}$ the resulting Poisson processes $p_\varphi(A)$ and $p_\varphi(B)$ are independent. The decomposition (1.1.36) shows that a Lévy process is a superposition of a constant trend, a scaled Wiener process, and some jump process with stationary independent increments.

A Lévy process is fully characterized by the two parameters $\alpha, \beta \in \mathfrak{R}$ together with the very flexible Lévy measure φ . In the case $\beta = 0$ and $\varphi(A) = 0$ for all $A \in \mathcal{E}$, the resulting Lévy process X is deterministic and has at time t the value $X_t = \alpha t$. If $\varphi(A) = 0$ for all $A \in \mathcal{E}$, then the Lévy process is a transformed Wiener process of the form $X_t = \alpha t + \beta W_t$. This is the only possible form that a continuous Lévy process can have. As an

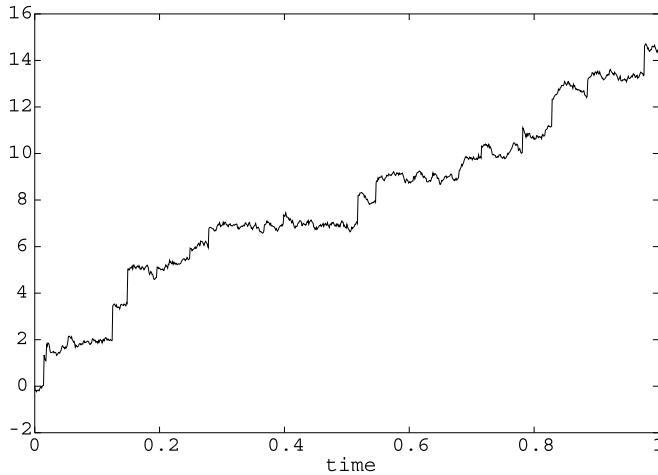


Fig. 1.1.8. A linearly interpolated sample path of a Lévy process

example, for $\alpha = \beta = 0$ and $\varphi(\mathcal{E}) = \varphi(\{1\}) \in (0, \infty)$ the Lévy process $X = p_\varphi(\{1\}) = \{p_\varphi(\{1\}, [0, t]), t \geq 0\}$ is a Poisson process with intensity $\varphi(\{1\})$. For illustration, in Fig. 1.1.8 the path of a Lévy process is shown for the case $\alpha = 0$, $\beta = 1$ and

$$\varphi(dv) = \begin{cases} \lambda dv & \text{for } v \in (0, 1) \\ 0 & \text{otherwise,} \end{cases}$$

where $\lambda = 20$. This is simply a combination of a Poisson and a Wiener process. Note that we used in this example the jump times and marks of the compound Poisson process exhibited in Figs. 1.1.5 and 1.1.6.

If not directly required by the modeling task at hand, it is wise to focus only on the important major events that arise with finite intensity. The continuous noise of a Wiener process can conveniently aggregate many small very frequent events. This way of modeling reduces significantly the computational complexity and keeps the quantitative analysis tractable. In principle, the infinite dimensional Lévy measure $\varphi(\cdot)$ provides almost too much freedom in modeling. It may also be extremely difficult to estimate $\varphi(\cdot)$ in practice. Furthermore, the numerical burden of working with a general Lévy process instead of a much simpler compound Poisson process can be substantial. Empirically it is almost impossible to distinguish a Wiener process from the high intensity part of a Lévy process with small jumps.

We will typically use a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ in our financial modeling, where the sources of continuous uncertainty are modeled by independent standard Wiener processes W^1, W^2, \dots, W^m and the sources of event driven uncertainty are generated via independent Poisson processes N^1, N^2, \dots, N^{d-m} , $d \in \mathcal{N} = \{1, 2, \dots\}$, $m \in \{1, 2, \dots, d\}$. In some cases we may involve a Poisson measure $p_\varphi(\cdot)$. Throughout the book we always assume

that these Wiener processes, Poisson processes and Poisson measures are $\underline{\mathcal{A}}$ -adapted and that their increments over a period $(s, t]$ are independent of \mathcal{A}_s for $t \geq 0$, $s \in [0, t]$.

1.2 Supermartingales and Martingales

Martingales

As we will see later in the context of asset pricing and investing under the benchmark approach, investors have reasons to determine, in some sense, the best estimate for the actual value of a future payoff when expressed in units of their best performing portfolio, the benchmark. If they use different information sets, then they usually generate different estimates. Obviously, an investor has to employ a probability measure for forming her or his expectations. This probability measure will be given by an underlying model.

To identify the best estimate for the value of a future payoff we define the quantity F_s for $s \in [0, \infty)$ as the least-squares estimate of the future value X_t at the future time $t \in [s, \infty)$ under the information given by \mathcal{A}_s at time s . This estimate is \mathcal{A}_s -measurable and minimizes the error

$$\varepsilon_s = E((X_t - F_s)^2)$$

over all possible \mathcal{A}_s -measurable estimates. The random variable F_s is simply the least-squares projection of X_t given the information at time $s \in [0, t]$. It is obtained by the conditional expectation

$$F_s = E(X_t | \mathcal{A}_s), \quad (1.2.1)$$

for all $s \in [0, t]$. This leads to the following definition:

Definition 1.2.1. *A continuous time stochastic process $X = \{X_t, t \geq 0\}$ is called an $(\underline{\mathcal{A}}, P)$ -martingale or simply a martingale, if it satisfies the equation*

$$X_s = E(X_t | \mathcal{A}_s) \quad (1.2.2)$$

for all $s \in [0, t]$ and the integrability condition

$$E(|X_t|) < \infty \quad (1.2.3)$$

for all $t \geq 0$.

An example of an $(\underline{\mathcal{A}}, P)$ -martingale is a Wiener process $W = \{W_t, t \geq 0\}$ on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$.

There are many other continuous time stochastic processes that form martingales. For example, using again the standard Wiener process W it can be shown that the process

$$X = \{X_t = W_t^2 - t, t \geq 0\} \quad (1.2.4)$$

is an $(\underline{\mathcal{A}}, P)$ -martingale. The process

$$Y = \left\{ Y_t = \exp \left\{ \sigma W_t - \frac{1}{2} \sigma^2 t \right\}, t \geq 0 \right\},$$

which is the exponential of a transformed Wiener process, is also an $(\underline{\mathcal{A}}, P)$ -martingale. From a practical perspective one can say that a martingale is simply a trendless process.

Super- and Submartingales

Systematically upward or downward trending stochastic processes can be captured by the following notions:

Definition 1.2.2. An $\underline{\mathcal{A}}$ -adapted process $X = \{X_t, t \geq 0\}$ is an $(\underline{\mathcal{A}}, P)$ -supermartingale (submartingale) if

$$E(|X_t|) < \infty \quad (1.2.5)$$

and

$$X_s \stackrel{(\leq)}{\geq} E(X_t | \mathcal{A}_s) \quad (1.2.6)$$

for all $s \in [0, \infty)$ and $t \in [s, \infty)$.

This means, a supermartingale is trending systematically downward or has no trend. We call a supermartingale a *strict supermartingale* (strict submartingale) if the inequality in (1.2.6) is always a strict inequality.

As an example for a supermartingale we show in Fig. 1.2.1 a path of the exponential function

$$X_t = \exp \{ -\sigma^2 t + \sigma W_t \} \quad (1.2.7)$$

of the value of a drifted, scaled Wiener process $W = \{W_t, t \geq 0\}$ with volatility parameter $\sigma = 0.2$.

We will see later that supermartingales play a fundamental role in finance and insurance modeling. They appear naturally when nonnegative securities are expressed in units of the best performing portfolio, the benchmark, of a given market.

Assume for a Poisson process N that it is $\underline{\mathcal{A}}$ -adapted. We can then show for $0 \leq s < t < \infty$ by (1.1.21) that

$$\begin{aligned} E(N_t | \mathcal{A}_s) &= E(N_t - N_s | \mathcal{A}_s) + N_s \\ &= \lambda(t-s) + N_s \geq N_s, \end{aligned} \quad (1.2.8)$$

which proves that the Poisson process is a submartingale.

On the other hand, the *compensated Poisson process* $q = \{q_t, t \geq 0\}$ with

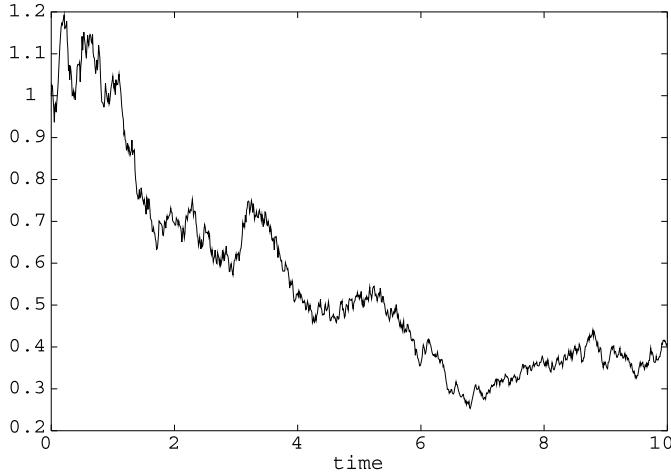


Fig. 1.2.1. Path of a supermartingale

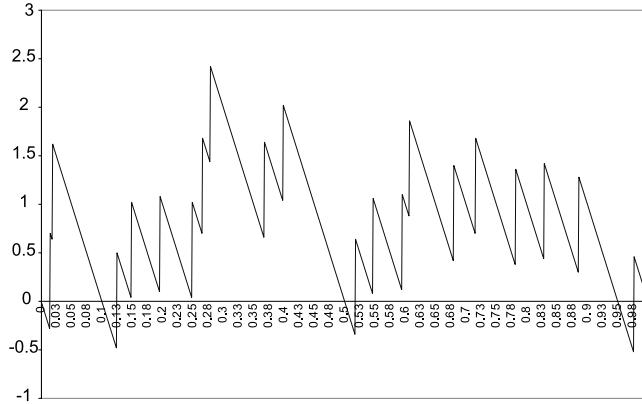


Fig. 1.2.2. Path of a compensated Poisson process, $\lambda = 20$

$$q_t = N_t - \lambda t \quad (1.2.9)$$

is a martingale since we have by similar arguments as in (1.2.8)

$$\begin{aligned} E(q_t | \mathcal{A}_s) &= E(q_t - q_s | \mathcal{A}_s) + q_s \\ &= E(N_t - N_s) - \lambda(t-s) + q_s = q_s \end{aligned} \quad (1.2.10)$$

for $0 \leq s \leq t < \infty$. In Fig. 1.2.2 we plot the path of a compensated Poisson process q with intensity $\lambda = 20$, see Fig. 1.1.4 for the corresponding trajectory of the Poisson process N .

Stopping Times

Observable random times arise in many ways in financial modeling. For instance, times of default, the first time of reaching a critical threshold or the early exercise time of an American option are such random times. Let us make their definition precise, and consider a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ as introduced above.

Definition 1.2.3. A random variable $\tau : \Omega \rightarrow [0, \infty)$ is called a stopping time with respect to the filtration $\underline{\mathcal{A}}$ if for all $t \geq 0$ one has

$$\{\tau \leq t\} \in \mathcal{A}_t. \quad (1.2.11)$$

The relation (1.2.11) expresses the fact that the event $\{\tau \leq t\}$ is \mathcal{A}_t -measurable and thus observable at time t . The information set associated with a stopping time τ is defined as

$$\mathcal{A}_\tau = \sigma \{A \in \mathcal{A} : A \cap \{\tau \leq t\} \in \mathcal{A}_t \text{ for } t \geq 0\}. \quad (1.2.12)$$

It represents the information available before and at the stopping time τ . For example, the first time

$$\tau(a) = \inf\{t \geq 0 : W_t = a\} \quad (1.2.13)$$

when a Wiener process W reaches a level $a \in \mathbb{R}$ is a stopping time.

Trading strategies can react almost immediately to jumps in securities. However, in practice they always lag slightly behind. In this sense they are predictable even though the triggering jump times may be not predictable. One calls a sigma-algebra *predictable* when it is generated by left-continuous $\underline{\mathcal{A}}$ -adapted processes with right hand limits. We exclude in a predictable sigma-algebra, in principle, all information about the time instant when a sudden non predictable event, like a default, occurs. Immediately after the event a predictable sigma-algebra already contains this information.

A stochastic process $X = \{X_t, t \geq 0\}$, where X_τ is for each stopping time τ measurable with respect to a predictable sigma-algebra, is called *predictable*. For example, all continuous stochastic processes are predictable. From a right-continuous process with left hand limits $X = \{X_t, t \geq 0\}$ we obtain its predictable version $\tilde{X} = \{\tilde{X}_t, t \geq 0\}$ by taking at each time point the left hand limit, that is

$$\tilde{X}_t = X_{t-} = \lim_{\varepsilon \downarrow 0} X_{t-\varepsilon} \quad (1.2.14)$$

for all $t \geq 0$.

A stopping time is called *predictable*, if \mathcal{A}_τ is predictable. This means, \mathcal{A}_τ is generated by left-continuous stochastic processes with right hand limits. A stopping time that is not predictable is called *inaccessible*. For example, the jump times of a Poisson process are inaccessible. However, the first hitting time $\tau(a)$, given in (1.2.13), of the level a by the Wiener process W is predictable.

For $a, b \in \mathbb{R}$ we employ the notation $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. Let us summarize the following useful properties of stopping times τ and τ' , see, for instance, Karatzas & Shreve (1991) and Elliott (1982):

- (i) τ is \mathcal{A}_τ -measurable.
- (ii) For a continuous $\underline{\mathcal{A}}$ -adapted process $X = \{X_t, t \geq 0\}$ the random variable X_τ is \mathcal{A}_τ -measurable.
- (iii) If $P(\tau \leq \tau') = 1$, then $\mathcal{A}_\tau \subseteq \mathcal{A}_{\tau'}$.
- (iv) The random variables $\tau \wedge \tau'$, $\tau \vee \tau'$ and $(\tau + \tau')$ are stopping times.
- (v) If for a real valued random variable Y we have $E(|Y|) < \infty$ and $P(\tau \leq \tau') = 1$, then

$$E(Y | \mathcal{A}_\tau) = E(Y | \mathcal{A}_{\tau \wedge \tau'}) \quad (1.2.15)$$

and

$$E(E(Y | \mathcal{A}_\tau) | \mathcal{A}_{\tau'}) = E(Y | \mathcal{A}_\tau). \quad (1.2.16)$$

If $X = \{X_t, t \geq 0\}$ is a right continuous $(\underline{\mathcal{A}}, P)$ -supermartingale, then the supermartingale property (1.2.2) still holds if the times s and t in (1.2.2) are bounded stopping times. More precisely, Doob's *Optional Sampling Theorem* states the following result, see Doob (1953):

Theorem 1.2.4. (Doob) *If $X = \{X_t, t \geq 0\}$ is a right continuous $(\underline{\mathcal{A}}, P)$ -supermartingale on $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$, then it holds for two bounded stopping times τ and τ' with $\tau \leq \tau'$ almost surely that*

$$E(X_{\tau'} | \mathcal{A}_\tau) \leq X_\tau. \quad (1.2.17)$$

Furthermore, if X is additionally an $(\underline{\mathcal{A}}, P)$ -martingale, then equality holds in (1.2.17).

For example, this theorem will become important when one has to apply a pricing formula at a stopping time or when a claim matures at a stopping time.

Martingale Inequalities

In the proofs of convergence results for discrete-time approximations and at many other occasions one has to heavily rely on powerful inequalities. For this reason we list below several important inequalities related to martingales and supermartingales.

A continuous martingale $X = \{X_t, t \geq 0\}$ with finite p th moment satisfies the *maximal martingale inequality*

$$P \left(\sup_{s \in [0, t]} |X_s| > a \right) \leq \frac{1}{a^p} E(|X_t|^p) \quad (1.2.18)$$

and the *Doob inequality*

$$E \left(\sup_{s \in [0,t]} |X_s|^p \right) \leq \left(\frac{p}{p-1} \right)^p E(|X_t|^p) \quad (1.2.19)$$

for $a > 0$, $p > 1$ and $t \geq 0$. If X is a continuous martingale, then the maximal martingale inequality provides an estimate for the probability that a level a will be exceeded by the maximum of X . In particular the Doob inequality provides for $p = 2$ for the maximum of the square the estimate

$$E \left(\sup_{s \in [0,t]} |X_s|^2 \right) \leq 4 E(|X_t|^2)$$

for $t \geq 0$.

If $X = \{X_t, t \geq 0\}$ is a right continuous supermartingale, then it can be shown, see [Doob \(1953\)](#), that for any $\lambda > 0$ it holds

$$\lambda P \left(\sup_{t \geq 0} X_t \geq \lambda \mid \mathcal{A}_0 \right) \leq E(X_0 \mid \mathcal{A}_0) + E(\max(0, -X_0) \mid \mathcal{A}_0). \quad (1.2.20)$$

Standard Inequalities

We list here some standard inequalities, see [Ash \(1972\)](#) and [Ikeda & Watanabe \(1989\)](#), that we will use later repeatedly. Consider a sequence of pairs of real numbers $\{(a_i, b_i), i \in \{1, \dots, n\}\}$, with $n \in \mathcal{N}$. Then one can formulate the *Cauchy-Schwarz inequality*

$$\left(\sum_{i=1}^n a_i b_i \right)^2 \leq \sum_{i=1}^n a_i^2 \sum_{i=1}^n b_i^2 \quad (1.2.21)$$

and the inequality

$$\left(\sum_{i=1}^n a_i \right)^2 \leq n \sum_{i=1}^n a_i^2. \quad (1.2.22)$$

Let $1 < p < \infty$, $1 < q < \infty$ and $(1/p) + (1/q) = 1$, then we have the *Hölder inequality*

$$\left| \sum_{i=1}^n a_i b_i \right| \leq \left(\sum_{i=1}^n |a_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n |b_i|^q \right)^{\frac{1}{q}} \quad (1.2.23)$$

and

$$\left(\sum_{i=1}^n a_i \right)^p \leq n^{p-1} \sum_{i=1}^n a_i^p. \quad (1.2.24)$$

Let $(\Omega, \mathcal{A}, \mu)$ be a measurable space and consider two functions $f, g : \Omega \rightarrow \mathbb{R}$. If f and $g \in L^2(\mu)$, then $fg \in L^1(\mu)$ and one has the following version of the *Cauchy-Schwarz inequality*

$$\left| \int_{\Omega} f g d\mu \right| \leq \left(\int_{\Omega} f^2 d\mu \right)^{\frac{1}{2}} \left(\int_{\Omega} g^2 d\mu \right)^{\frac{1}{2}} \quad (1.2.25)$$

and the inequality

$$\left(\left| \int_{\Omega} f d\mu \right| \right)^2 \leq \mu(\Omega) \int_{\Omega} f^2 d\mu. \quad (1.2.26)$$

Here $L^p(\mu)$ denotes the space of p -integrable functions. Let $1 < p < \infty$, $1 < q < \infty$ and $(1/p) + (1/q) = 1$, if $f \in L^p(\mu)$ and $g \in L^q(\mu)$, then $fg \in L^1(\mu)$ and the following *Hölder inequality* holds

$$\left| \int_{\Omega} f g d\mu \right| \leq \left(\int_{\Omega} |f|^p d\mu \right)^{\frac{1}{p}} \left(\int_{\Omega} |g|^q d\mu \right)^{\frac{1}{q}}. \quad (1.2.27)$$

Let $1 < p < \infty$ and $f \in L^p(\mu)$, then

$$\left(\left| \int_{\Omega} f d\mu \right| \right)^p \leq \mu(\Omega)^{p-1} \int_{\Omega} |f|^p d\mu. \quad (1.2.28)$$

Let $\alpha, \beta : [t_0, T] \rightarrow \mathbb{R}$ be integrable with

$$0 \leq \alpha(t) \leq \beta(t) + C \int_{t_0}^t \alpha(s) ds \quad (1.2.29)$$

for $t \in [t_0, T]$ and $C > 0$. Then the following version of the *Gronwall inequality* holds

$$\alpha(t) \leq \beta(t) + C \int_{t_0}^t e^{C(t-s)} \beta(s) ds. \quad (1.2.30)$$

Let $f : [t_0, T] \rightarrow \mathbb{R}$ be a nonnegative integrable function and $\alpha, \beta : [t_0, T] \rightarrow \mathbb{R}$ be continuous functions on $[t_0, T]$, if

$$\alpha(t) \leq \beta(t) + \int_{t_0}^t f(s) \alpha(s) ds \quad (1.2.31)$$

for $t \in [t_0, T]$, then the following alternative Gronwall inequality can be used

$$\alpha(t) \leq \beta(t) + \int_{t_0}^t f(s) \beta(s) \exp \left(\int_s^t f(u) du \right) ds. \quad (1.2.32)$$

Moreover, if β is non-decreasing, then

$$\alpha(t) \leq \beta(t) \exp \left(\int_{t_0}^t f(u) du \right). \quad (1.2.33)$$

Let X be a random variable with finite first moment and $g : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function, then we have *Jensen's inequality*

$$g(E(X)) \leq E(g(X)). \quad (1.2.34)$$

If $E(|X|^s)$ is finite for some $s > 0$, then for all $r \in (0, s]$ and $a \in \mathbb{R}$ we have the *Lyapunov inequality*

$$(E(|X - a|^r))^{\frac{1}{r}} \leq (E(|X - a|^s))^{\frac{1}{s}}. \quad (1.2.35)$$

For an integrable random variable X we have the *Markov inequality*

$$P(X \geq a) \leq \frac{1}{a} E(|X|). \quad (1.2.36)$$

This yields for $a > 0$ the *Chebyshev inequality*

$$P(|X - E(X)| > a) \leq \frac{1}{a^2} \text{Var}(X). \quad (1.2.37)$$

1.3 Quadratic Variation and Covariation

Quadratic Variation

For simplicity, let us consider an *equidistant time discretization*

$$\{t_k = k h : k \in \{0, 1, \dots\}\}, \quad (1.3.1)$$

with small time steps of length $h > 0$, such that $0 = t_0 < t_1 < t_2 < \dots$. Thus, we have the discretization times $t_k = k h$ for $k \in \{0, 1, \dots\}$. The specific structure of the time discretization is not essential for the definition below, as long as the maximum time step size vanishes. There is no need to have the time discretization equidistant. However, it makes our presentation simpler.

For a given stochastic process X the *quadratic variation process* $[X] = \{[X]_t, t \geq 0\}$ is defined as the limit in probability as $h \rightarrow 0$ of the sums of squared increments of the process X , provided this limit exists and is unique, see [Jacod & Shiryaev \(2003\)](#) and [Protter \(2005\)](#). More precisely, we have at time t the *quadratic variation*

$$[X]_t = \lim_{h \rightarrow 0} [X]_{h,t}, \quad (1.3.2)$$

where the *approximate quadratic variation* $[X]_{h,t}$ is given by the sum

$$[X]_{h,t} = \sum_{k=1}^{n_t} (X_{t_k} - X_{t_{k-1}})^2. \quad (1.3.3)$$

Here n_t denotes the integer

$$n_t = \max\{k \in \mathcal{N} : t_k \leq t\}, \quad (1.3.4)$$

which is the index of the last discretization point before or including $t \geq 0$.

The value of the quadratic variation process $[W] = \{[W]_t, t \geq 0\}$ at time t for a standard Wiener process W is given by the relation

$$[W]_t = t \quad (1.3.5)$$

for $t \geq 0$.

For a continuous, square integrable (\mathcal{A}, P) -martingale X , a new continuous (\mathcal{A}, P) -martingale $Y = \{Y_t, t \geq 0\}$ is obtained by setting

$$Y_t = (X_t)^2 - [X]_t \quad (1.3.6)$$

for $t \geq 0$, see [Protter \(2005\)](#). In the case of a standard Wiener process W , we obtain the martingale $Y = \{Y_t = (W_t)^2 - t, t \geq 0\}$, see [\(1.2.4\)](#).

Covariation

For the definition of *covariation* the same equidistant time discretization, as given in [\(1.3.1\)](#), is now used. For two continuous stochastic processes Z_1 and Z_2 the *covariation process* $[Z_1, Z_2] = \{[Z_1, Z_2]_t, t \geq 0\}$ is defined as the limit in probability as $h \rightarrow 0$ of the values of the *approximate covariation process* $[Z_1, Z_2]_{h,\cdot}$, with

$$[Z_1, Z_2]_{h,t} = \sum_{k=1}^{n_t} (Z_1(t_k) - Z_1(t_{k-1}))(Z_2(t_k) - Z_2(t_{k-1})) \quad (1.3.7)$$

for $t \geq 0$ and $h > 0$, see [\(1.3.4\)](#). More precisely, at time $t \geq 0$ we obtain the *covariation*

$$[Z_1, Z_2]_t \stackrel{P}{=} \lim_{h \rightarrow 0} [Z_1, Z_2]_{h,t}, \quad (1.3.8)$$

where $[Z_1, Z_2]_{h,t}$ is the above approximate covariation.

For a right-continuous stochastic process $\xi = \{\xi(t), t \geq 0\}$ we denote by

$$\xi(t-) \stackrel{\text{a.s.}}{=} \lim_{h \rightarrow 0+} \xi(t-h) \quad (1.3.9)$$

the almost sure *left hand limit* of $\xi(t)$ at time $t \in (0, \infty)$. The *jump size* $\Delta\xi(t)$ at time t is then obtained as

$$\Delta\xi(t) = \xi(t) - \xi(t-) \quad (1.3.10)$$

for $t \in (0, \infty)$.

In the case of a pure jump process $p = \{p_t, t \geq 0\}$ the corresponding quadratic variation is obtained as

$$[p]_t = \sum_{0 \leq s \leq t} (\Delta p_s)^2 \quad (1.3.11)$$

for $t \geq 0$, where $\Delta p_s = p_s - p_{s-}$. Interestingly, in the case when p is a Poisson process, its quadratic variation equals the process itself, that is, $[N]_t = N_t$ for all $t \geq 0$.

Denote by Z_1 and Z_2 two stochastic processes with continuous part

$$Z_i^c(t) = Z_i(t) - Z_i(0) - \sum_{0 < s \leq t} \Delta Z_i(s) \quad (1.3.12)$$

for $t \geq 0$ and $i \in \{1, 2\}$. We assume that the sum in (1.3.12) is almost surely finite. The covariation $[Z_1, Z_2]_t$ of Z_1 and Z_2 at time t is then defined as

$$[Z_1, Z_2]_t = [Z_1^c, Z_2^c]_t + \sum_{0 < s \leq t} (\Delta Z_1(s)) (\Delta Z_2(s)) \quad (1.3.13)$$

for $t \geq 0$, as long as the quantities involved remain almost surely finite. This also means that the quadratic variation of a process Z_1 equals the quadratic variation $[Z_1^c]_t$ of its continuous part plus the sum of the squares of its jumps, that is

$$[Z_1]_t = [Z_1^c]_t + \sum_{0 < s \leq t} (\Delta Z_1(s))^2 \quad (1.3.14)$$

for $t \geq 0$.

Local Martingales

As we will see, in continuous time finance certain stochastic processes become martingales when properly stopped but are not true martingales.

Definition 1.3.1. A stochastic process $X = \{X_t, t \geq 0\}$ is an $(\underline{\mathcal{A}}, P)$ -local martingale if there exists an increasing sequence $(\tau_n)_{n \in \mathbb{N}}$ of stopping times, that may depend on X , such that $\lim_{n \rightarrow \infty} \tau_n \stackrel{a.s.}{=} \infty$ and each stopped process

$$X^{\tau_n} = \{X_t^{\tau_n} = X_{t \wedge \tau_n}, t \geq 0\} \quad (1.3.15)$$

is an $(\underline{\mathcal{A}}, P)$ -martingale, where $t \wedge \tau_n = \min(t, \tau_n)$.

If X is a local martingale, then the value X_s does, in general, not equal the conditional expectation $E(X_t | \mathcal{A}_s)$ for $s \in [0, \infty)$ and $t \in [s, \infty)$. A local martingale that is not a martingale is called a *strict local martingale*.

The following holds, see [Protter \(2005\)](#) and [Shiryayev \(1984\)](#):

Lemma 1.3.2

- (i) An almost surely nonnegative (negative) $(\underline{\mathcal{A}}, P)$ -local martingale is an $(\underline{\mathcal{A}}, P)$ -supermartingale (submartingale).
- (ii) An almost surely uniformly bounded $(\underline{\mathcal{A}}, P)$ -local martingale is an $(\underline{\mathcal{A}}, P)$ -martingale.
- (iii) A square integrable $(\underline{\mathcal{A}}, P)$ -local martingale X is a square integrable $(\underline{\mathcal{A}}, P)$ -martingale if and only if

$$E([X]_T) < \infty \quad (1.3.16)$$

for all $T \in [0, \infty)$.

- (iv) A nonnegative $(\underline{\mathcal{A}}, P)$ -local martingale $X = \{X_t, t \geq 0\}$ with $E(X_t | \mathcal{A}_s) < \infty$ for all $0 \leq s \leq t < \infty$ is an $(\underline{\mathcal{A}}, P)$ -supermartingale.

Identification of Martingales as Wiener Processes

The Wiener process plays a central role in stochastic calculus and is the basic building block for modeling continuous uncertainty in many applications as in finance and insurance. We saw that the Wiener process is a martingale and from (1.3.5) it follows that its quadratic variation $[W]_t$ equals the time t . Note that the converse can also be shown. *Lévy's Theorem* provides this important result. Its derivation can be found, for instance in [Platen & Heath \(2006\)](#).

Theorem 1.3.3. (Lévy) *For $m \in \mathcal{N}$ let A be a given m -dimensional vector process $\mathbf{A} = \{\mathbf{A}_t = (A_t^1, A_t^2, \dots, A_t^m)^\top, t \geq 0\}$ on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. If each of the processes $A^i = \{A_t^i, t \geq 0\}$ is a continuous, square integrable $(\underline{\mathcal{A}}, P)$ -martingale that starts at 0 at time $t = 0$ and their covariations are of the form*

$$[A^i, A^k]_t = \begin{cases} t & \text{for } i = k \\ 0 & \text{for } i \neq k \end{cases} \quad (1.3.17)$$

for all $i, k \in \{1, 2, \dots, m\}$ and $t \geq 0$, then the vector process \mathbf{A} is an m -dimensional standard Wiener process on $[0, \infty)$. This means that each process A^i is a one-dimensional Wiener process that is independent of the other Wiener processes A^k for $k \neq i$.

This result implies that a continuous process $X = \{X_t, t \geq 0\}$ is a one-dimensional Wiener process if and only if both the process X and the process $Y = \{Y_t = X_t^2 - t, t \geq 0\}$ are martingales. Furthermore, if one is able to construct for an observed vector process a transformation such that the transformed processes are square integrable continuous martingales with covariations of the form (1.3.17), then one has found a vector of independent Wiener processes. In this case one can take the inverse of that transformation to arrive at a realistic model.

1.4 Itô Integral

Itô Integral as Gains from Trade

The notion of a stochastic integral is highly relevant and also very natural in finance because it can be interpreted as gains from trade. Let an investor choose his or her trading strategy as a piecewise constant buy and hold allocation process $\xi = \{\xi(t), t \in [0, T]\}$ with $\xi(t) = \xi(t_k)$ units of shares held at time $t \in [t_k, t_{k+1})$, $k \in \{0, 1, \dots\}$ and $t_k = kh$ for $h > 0$. The value of a share is assumed to evolve according to the asset price process $X = \{X_t, t \geq 0\}$. The gains from trade over the period $[0, t]$ can then be expressed in the form

$$\int_0^t \xi(s) dX_s = \sum_{k=1}^{n_t} \xi(t_{k-1}) \{X_{t_k} - X_{t_{k-1}}\} + \xi(t_{n_t}) \{X_t - X_{t_{n_t}}\}, \quad (1.4.1)$$

where

$$n_t = \max\{k \in \mathcal{N} : t_k \leq t\} \quad (1.4.2)$$

is the integer index of the latest discretization time before and including t .

For a left continuous, predictable stochastic process $\xi = \{\xi(t), t \geq 0\}$ as integrand with

$$\int_0^T \xi(s)^2 d[X]_t < \infty \quad (1.4.3)$$

for all $T \in [0, \infty)$ almost surely, the *Itô integral* with respect to X is defined as the left continuous limit in probability

$$\int_0^t \xi(s) dX_s \stackrel{P}{=} \lim_{h \rightarrow 0} \sum_{k=1}^{n_t} \xi(t_{k-1}) \{X_{t_k} - X_{t_{k-1}}\} \quad (1.4.4)$$

of a sequence of corresponding approximating sums for progressively finer time discretizations for $t \geq 0$.

For details on the definition of Itô integrals we refer to [Karatzas & Shreve \(1991\)](#), [Kloeden & Platen \(1999\)](#), [Protter \(2005\)](#) or [Platen & Heath \(2006\)](#). The most important characteristic of the Itô integral is that the evaluation point t_{k-1} for the integrand ξ is always taken at the left hand side of the discretization interval $[t_{k-1}, t_k]$. This feature is essential for investment applications because an investor needs to decide at the beginning of a holding period how many units of a security he or she wants to hold.

A simple example of how the Itô integral differs from most classical integrals is the following double Wiener integral

$$\begin{aligned} \int_0^t W_s dW_s &\stackrel{P}{=} \lim_{h \rightarrow 0} \sum_{k=1}^{n_t} W_{t_{k-1}} (W_{t_k} - W_{t_{k-1}}) \\ &\stackrel{P}{=} \lim_{h \rightarrow 0} \frac{1}{2} \sum_{k=1}^{n_t} \left\{ \left(W_{t_k}^2 - W_{t_{k-1}}^2 \right) - (W_{t_k} - W_{t_{k-1}})^2 \right\} \\ &\stackrel{P}{=} \frac{1}{2} W_t^2 - \frac{1}{2} W_0^2 - \lim_{h \rightarrow 0} \frac{1}{2} \sum_{k=1}^{n_t} (W_{t_k} - W_{t_{k-1}})^2. \end{aligned}$$

For the quadratic variation of the standard Wiener processes in [\(1.3.3\)](#) we have $[W]_t = t$, see [\(1.3.5\)](#), and in addition $W_0 = 0$, see [\(1.1.16\)](#). Consequently, the value of the above double Wiener integral is

$$\int_0^t W_s dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} [W]_t = \frac{1}{2} W_t^2 - \frac{1}{2} t, \quad (1.4.5)$$

which is clearly less than $\frac{1}{2} W_t^2$ which would not be expected under classical integration rules. This already indicates that ad hoc conclusions from analogies with deterministic calculus may fail in a stochastic setting. There exist significant differences between the rules for Itô integration and those for classical

integration. These differences turn out to be crucial for rigorous modeling in finance and insurance which are based on the gains from trade of wealth processes. They also have significant impact on the stochastic numerical methods that we will present when compared to classical numerical methods.

To model more general dynamics than those of the Wiener process let $e = \{e_t, t \geq 0\}$ and $f = \{f_t, t \geq 0\}$ be predictable stochastic processes. Consider a stochastic process $Y = \{Y_t, t \geq 0\}$, where

$$Y_t = y_0 + \int_0^t e_s ds + \int_0^t f_s dW_s \quad (1.4.6)$$

for $t \geq 0$ and initial value $Y_0 = y_0$. Here $W = \{W_t, t \geq 0\}$ is a standard Wiener process. The first integral is a random ordinary Riemann-Stieltjes integral, assuming

$$\int_0^t |e_s| ds < \infty \quad (1.4.7)$$

for all $t \geq 0$ a.s. The second integral is an Itô integral with respect to the Wiener process W , see (1.4.4), where we assume that

$$\int_0^t |f_s|^2 ds < \infty \quad (1.4.8)$$

almost surely for all $t \geq 0$. It is common to express the integral equation (1.4.6) in an equivalent short hand notation, i.e. Itô differential equation in the form

$$dY_t = e_t dt + f_t dW_t \quad (1.4.9)$$

for $t \geq 0$ with $Y_0 = y_0$. Equation (1.4.9) is simply another way of writing (1.4.6). The processes e and f are called *drift* and *diffusion coefficients*, respectively. The concept of an Itô differential allows the modeling of rather general dynamics.

The above definitions of an Itô integral and Itô differential extend to the case of multi-dimensional integrands ξ and integration with respect to several independent standard Wiener processes. Furthermore, one can generalize the notion of an Itô integral also to hold for more general processes as integrators including those with jumps, as we will see later.

Some Properties of Itô Integrals

Consider two \mathcal{A} -adapted independent Wiener processes W^1 and W^2 . Recall that $(W_t^i - W_s^i)$ is independent of \mathcal{A}_s for $t \geq 0$, $s \in [0, t]$ and $i \in \{1, 2\}$. It is useful to specify for $T \geq 0$ the class \mathcal{L}_T^2 of predictable, square integrable integrands $f = \{f_t, t \in [0, T]\}$ in the form that

$$\int_0^T E(f_t^2) dt < \infty. \quad (1.4.10)$$

The Itô integral exhibits the following important properties which we will exploit frequently throughout the book:

1. *Linearity*: For $T \geq 0$, $t \in [0, T]$, $s \in [0, t]$, $Z_1, Z_2 \in \mathcal{L}_T^2$, and \mathcal{A}_s -measurable square integrable random variables A and B it holds

$$\int_s^t (A Z_1(u) + B Z_2(u)) dW_u^1 = A \int_s^t Z_1(u) dW_u^1 + B \int_s^t Z_2(u) dW_u^1. \quad (1.4.11)$$

2. *Local martingale property*: For ξ predictable with

$$\int_0^T \xi(u)^2 du < \infty \quad (1.4.12)$$

a.s. for all $T \in [0, \infty)$ the Itô integral $I_{\xi, W} = \{I_{\xi, W}(t) = \int_0^t \xi(s) dW_s, t \geq 0\}$ forms an (\mathcal{A}, P) -local martingale.

3. *Martingale property*: Assume that $I_{\xi, W}(t)$ forms a square integrable process, then it is a square integrable (\mathcal{A}, P) -martingale if and only if

$$E \left(\int_0^T \xi(u)^2 du \right) < \infty \quad (1.4.13)$$

for all $T \geq 0$.

4. *Correlation property*: For $T \geq 0$, $t \in [0, T]$, independent Wiener processes W^1 and W^2 and $Z_1, Z_2 \in \mathcal{L}_T^2$ it holds that

$$\begin{aligned} E \left(\int_0^t Z_1(u) dW_u^i \int_0^t Z_2(u) dW_u^j \mid \mathcal{A}_s \right) \\ = \begin{cases} \int_0^t E(Z_1(u) Z_2(u) \mid \mathcal{A}_s) du & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (1.4.14)$$

for $i, j \in \{1, 2\}$.

5. *Covariation property*: For $t \geq 0$, independent Wiener processes W^1 and W^2 and predictable integrands Z_1 and Z_2 with $\int_0^t |Z_1(u) Z_2(u)| du < \infty$ almost surely it holds

$$\left[\int_0^t Z_1(u) dW_u^i, \int_0^t Z_2(u) dW_u^j \right]_t = \begin{cases} \int_0^t Z_1(u) Z_2(u) du & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \quad (1.4.15)$$

for $i, j \in \{1, 2\}$.

6. *Finite variation property*: For $t \geq 0$ and predictable Z_1 and Z_2 one has

$$\left[\int_0^t Z_1(u) dW_u^1, \int_0^t Z_2(u) du \right]_t = 0. \quad (1.4.16)$$

Note that some of the above employed conditions can be weakened, see Protter (2005).

Semimartingales

The most general class of stochastic processes that we will consider in this book is that of *semimartingales*, see Protter (2005). It is a very rich class of processes which appears to be closed under various operations, and turns out to be sufficient for the modeling of most quantitative problems that appear in finance and insurance.

In a filtered probability space $(\Omega, \mathcal{A}, \mathcal{A}, P)$ a *semimartingale* is an \mathcal{A} -adapted, right-continuous stochastic process $X = \{X_t, t \geq 0\}$ with left hand limits, where X_t can be expressed as a sum of the form

$$X_t = X_0 + A_t + M_t \quad (1.4.17)$$

for all $t \geq 0$. Here $A = \{A_t, t \geq 0\}$ is a process of finite total variation, that is $\lim_{h \rightarrow 0} \sum_{k=1}^{n_t} |A_{t_k} - A_{t_{k-1}}| < \infty$, with n_t given in (1.4.2), and $M = \{M_t, t \geq 0\}$ is an (\mathcal{A}, P) -local martingale.

If A is predictable, then X is called a *special semimartingale* and the decomposition (1.4.17) is unique. For an \mathcal{A} -adapted, right-continuous stochastic process $X = \{X_t, t \geq 0\}$ with left hand limits let

$$X_{t-} \stackrel{\text{a.s.}}{=} \lim_{\varepsilon \downarrow 0} X_{t-\varepsilon}$$

denote the almost sure left hand limit of X at time $t \geq 0$. Such a process may jump, and we denote by

$$\Delta X_t = X_t - X_{t-} \quad (1.4.18)$$

its jump size at time $t \geq 0$. If in (1.4.17) the discontinuous part

$$A_t^d = \sum_{0 \leq s \leq t} \Delta A_s$$

of A and the discontinuous part

$$M_t^d = \sum_{0 \leq s \leq t} \Delta M_s$$

of M are almost surely finite, then each of the processes A and M can be decomposed into a continuous and discontinuous part, being

$$A_t = A_t^c + A_t^d \quad (1.4.19)$$

and

$$M_t = M_t^c + M_t^d \quad (1.4.20)$$

for $t \geq 0$, respectively.

For example, the Wiener process $W = \{W_t, t \geq 0\}$, given in Definition 1.1.3, is a semimartingale. Here the decomposition (1.4.17) is simply such that $X_0 = 0$, $A_t = 0$ and $M_t = M_t^c = W_t$. Also a Poisson process N with

intensity λ , as given by Definition 1.1.4, is a semimartingale, as well as the compensated Poisson process q , defined in (1.2.9). For N we obtain in (1.4.17) $N_0 = X_0 = 0$, where $M_t = M_t^d = N_t - \lambda t = q_t$ is the local martingale, which is here a martingale. Furthermore, $A_t = A_t^c = \lambda t$ characterizes the predictable process A which is of finite total variation.

In the case of a Lévy process X , with the notation given in (1.1.36) and almost surely finite discontinuous martingale part

$$M_t^d = \int_0^t \int_{\mathcal{E}} v(p_\varphi(dv, ds) - \varphi(ds)) ds, \quad (1.4.21)$$

the initial value is $X_0 = 0$, the continuous local martingale part of X is

$$M_t^c = \beta W_t \quad (1.4.22)$$

and the predictable finite total variation part is continuous and equals

$$A_t = A_t^c = \alpha t + \int_0^t \int_{|v| \geq 1} v \varphi(dv) ds \quad (1.4.23)$$

for $t \geq 0$, always assuming A_t to be finite.

Itô Integral for Semimartingales

We define for a semimartingale $X = \{X_t, t \geq 0\}$ and a predictable semimartingale $\xi = \{\xi(s), s \geq 0\}$ the corresponding *Itô integral* as the limit in probability

$$I_{\xi, X}(t) = \int_0^t \xi(s) dX_s \stackrel{P}{=} \lim_{h \rightarrow 0} \sum_{k=1}^{n_t} \xi(t_{k-1}) (X_{t_k} - X_{t_{k-1}}), \quad (1.4.24)$$

using, for simplicity, an equidistant time discretization with step size h . It turns out that the limit does not depend on the particular time discretization, see Protter (2005). It is important to note that the integrand is effectively a predictable stochastic process, since we always take the left hand value in each discretization interval when refining the time discretization. We will sometimes make the integrand ξ act as a predictable process by writing $\xi(s-)$ instead of $\xi(s)$. Then the Itô integral is again a semimartingale. If the integrator X is an (\mathcal{A}, P) -local martingale with appropriate integrands, for example continuous or locally bounded integrands, then the Itô integral is also an (\mathcal{A}, P) -local martingale, see Protter (2005). In the case when X is of zero quadratic variation, the Itô integral coincides with the random ordinary Riemann-Stieltjes integral.

Let us consider a semimartingale X which has jumps, that is, the difference

$$\Delta X_t = X_t - X_{t-} \quad (1.4.25)$$

does not vanish for all $t \geq 0$. In this case the jumps of the Itô integral $I_{\xi,X}(t)$ are obtained as

$$\Delta I_{\xi,X}(t) = I_{\xi,X}(t) - I_{\xi,X}(t-) = \xi(t-) \Delta X_t \quad (1.4.26)$$

for $t \geq 0$. For example, if N is a Poisson process, then at the k th jump time τ_k we have $\Delta N_{\tau_k} = N_{\tau_k} - N_{\tau_{k-1}} = 1$ for $k \in \{1, 2, \dots\}$. The Itô integral for an integrand $\xi = \{\xi(t), t \geq 0\}$ takes then the form

$$I_{\xi,N}(t) = \int_0^t \xi(s-) dN_s = \sum_{k=1}^{N_t} \xi(\tau_k-) \Delta N_{\tau_k} = \sum_{k=1}^{N_t} \xi(\tau_k-) \quad (1.4.27)$$

for $t \geq 0$. Another example is that of a finite *pure jump process* $X = \{X_t = \sum_{0 \leq s \leq t} \Delta X_s, t \geq 0\}$, which jumps at the jump times τ_1, τ_2, \dots of a counting process $p = \{p_t, t \geq 0\}$ with corresponding jump size $\Delta X_{\tau_k} = c(k, \tau_k-)$ at the k th jump time τ_k . We then obtain for an integrand $\xi = \{\xi(t), t \geq 0\}$ the Itô integral

$$I_{\xi,X}(t) = \int_0^t \xi(s-) dX_s = \sum_{k=1}^{p_t} \xi(\tau_k-) \Delta X_{\tau_k} = \sum_{k=1}^{p_t} \xi(\tau_k-) c(k, \tau_k-) \quad (1.4.28)$$

for $t \geq 0$. Here we assume that the terms involved are almost surely finite, such that the sums $\sum_{k=1}^{p_t} \xi(\tau_k-) \Delta X_{\tau_k}$ and $\sum_{k=1}^{p_t} \Delta X_{\tau_k}$ almost surely converge to a finite value for all $t \geq 0$.

Itô Integral for Jump Measures

In the case when jump sizes are continuously distributed and total jump intensities may not be finite, one may use the Itô integral with respect to a Poisson measure as introduced in Sect. 1.1. Assume that $p_\varphi(dv, dt)$ is the Poisson measure on $\mathcal{E} \times [0, \infty)$ with intensity measure $q_\varphi(dv, dt) = \varphi(dv) dt$, satisfying condition (1.1.32).

For a family $(\xi(v))_{v \in \mathcal{E}}$ of almost surely finite adapted processes $\xi(v) = \{\xi(v, t), t \geq 0\}$ with $v \in \mathcal{E}$ the Itô integral

$$I_{\xi,p_\varphi}(t) = \int_0^t \int_{\mathcal{E}} \xi(v, s-) p_\varphi(dv, ds) \quad (1.4.29)$$

with respect to p_φ is defined, such that

$$\begin{aligned} \Delta I_{\xi,p_\varphi}(t) &= \int_0^t \int_{\mathcal{E}} \xi(v, s-) p_\varphi(dv, ds) - \int_0^{t-} \int_{\mathcal{E}} \xi(v, s-) p_\varphi(dv, ds) \\ &= \int_{\mathcal{E}} \xi(v, t-) p_\varphi(dv, \{t\}) \end{aligned} \quad (1.4.30)$$

for all $t \geq 0$. This means that if at a jump time τ the Poisson measure p_φ generates an event with mark v , then the change of the value of the corresponding Itô integral is given by the value $\xi(v, \tau-)$, the integrand ξ for the mark v just before the jump time.

For the case when $\mathcal{E} = (0, \lambda)$ is an interval with $\lambda \in (0, \infty)$, where

$$\varphi(v) = \begin{cases} 1 & \text{for } v \in \mathcal{E} \\ 0 & \text{otherwise,} \end{cases}$$

we obtain for the special case $\xi(v, t) = 1$ the Itô integral

$$I_{1,p_\varphi}(t) = \int_0^t \int_0^\lambda p_\varphi(dv, ds) = N_t \quad (1.4.31)$$

for $t \geq 0$. Here

$$N = \{N_t = p_\varphi((0, \lambda) \times [0, t]), t \geq 0\} \quad (1.4.32)$$

is a Poisson process with intensity λ .

Let us consider a jump measure $\tilde{p}_{\tilde{\varphi}}(dv, ds)$ on $\mathcal{E} \times \mathcal{T}$, where $\mathcal{T} = \{\tau_1, \tau_2, \dots\}$ is a set of predictable stopping times $0 \leq \tau_1 < \tau_2 < \dots$. At these times the measure $\tilde{p}_{\tilde{\varphi}}(dv, ds)$ may jump and create an independent mark v with some given probability $\tilde{\varphi}(dv)$. Such a jump measure allows us to model jumps at predictable times, for instance, at given dates. The Itô integral is then again given by equation (1.4.24).

Finally, let us mention that there exist other types of stochastic integrals. Important is, for instance, the symmetric or *Stratonovich integral* $J_{\xi, X}(t)$, which follows the classical rules of calculus. It can be obtained from the Itô integral by the equation

$$J_{\xi, X}(t) = I_{\xi, X}(t) + \frac{1}{2} [\xi, X^c]_t \quad (1.4.33)$$

for $t \geq 0$ and ξ and X as in (1.4.24), see Protter (2005). Here X^c denotes the continuous part of X and $[\cdot, \cdot]_t$ the covariation, see (1.3.8). For a Stratonovich or symmetric integral of ξ with respect to X we write

$$J_{\xi, X}(t) = \int_0^t \xi(s) \circ dX_s \quad (1.4.34)$$

and have

$$J_{\xi, X}(t) \stackrel{P}{=} \lim_{h \rightarrow 0} \sum_{k=1}^{n_t} \frac{(\xi(t_k) + \xi(t_{k-1}))}{2} (X_{t_k} - X_{t_{k-1}}) \quad (1.4.35)$$

as the limit in probability.

1.5 Itô Formula

One-dimensional Continuous Itô Formula

To be able to handle functions of solutions to stochastic processes, a chain rule, as it is known in classical calculus, is needed. Let $X = \{X_t, t \geq 0\}$ be a continuous stochastic process characterized by the Itô differential

$$dX_t = e_t dt + f_t dW_t \quad (1.5.1)$$

for $t \geq 0$ with initial value $X_0 = x_0$, see (1.4.9). Here $e = \{e_t, t \geq 0\}$ and $f = \{f_t, t \geq 0\}$ are two predictable stochastic processes.

Consider a function $u : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$ that is differentiable with respect to time t and twice continuously differentiable with respect to the spatial component x , that is, the partial derivatives $\frac{\partial u}{\partial t}$, $\frac{\partial u}{\partial x}$ and $\frac{\partial^2 u}{\partial x^2}$ exist and are continuous. To quantify the changes in $u(t, X_t)$ caused by changes in X_t one has the *Itô formula*

$$\begin{aligned} du(t, X_t) &= \left(\frac{\partial u(t, X_t)}{\partial t} + e_t \frac{\partial u(t, X_t)}{\partial x} + \frac{1}{2} (f_t)^2 \frac{\partial^2 u(t, X_t)}{\partial x^2} \right) dt \\ &\quad + f_t \frac{\partial u(t, X_t)}{\partial x} dW_t \end{aligned} \quad (1.5.2)$$

for $t \geq 0$. To derive this important formula one can use a Taylor expansion for increments over small time steps, where a second order term remains due to the nonvanishing quadratic variation of the Wiener process, see, for instance, Protter (2005) or Platen & Heath (2006).

When using the Stratonovich or symmetric integral we characterize, the dynamics (1.5.1) in the form

$$dX_t = \underline{e}_t dt + f_t \circ dW_t,$$

the classical chain rule emerges for $u(t, X_t)$, that is

$$du(t, X_t) = \left(\frac{\partial u(t, X_t)}{\partial t} + \underline{e}_t \frac{\partial u(t, X_t)}{\partial x} \right) dt + f_t \frac{\partial u(t, X_t)}{\partial x} \circ dW_t \quad (1.5.3)$$

for $t \geq 0$. Due to (1.4.33) it holds $\underline{e}_t = e_t + \frac{1}{2}[f, W]_t$.

By using the notion of quadratic variation, see (1.3.2), we can rewrite the Itô formula (1.5.2) conveniently in the form

$$du(t, X_t) = \frac{\partial u(t, X_t)}{\partial t} dt + \frac{\partial u(t, X_t)}{\partial x} dX_t + \frac{1}{2} \frac{\partial^2 u(t, X_t)}{\partial x^2} d[X]_t \quad (1.5.4)$$

for $t \geq 0$, which holds for continuous processes X .

Multi-dimensional Continuous Itô Formula

For multi-factor financial models the following more general version of the Itô formula is needed. Consider a d -dimensional vector process $\mathbf{e} = \{\mathbf{e}_t = (e_t^1, \dots, e_t^d)^T, t \geq 0\}$ with predictable components e^k , $k \in \{1, 2, \dots, d\}$. Assume that

$$\int_0^T |e_z^k| dz < \infty \quad (1.5.5)$$

almost surely for all $k \in \{1, 2, \dots, d\}$. The $d \times m$ -matrix valued process $\mathbf{F} = \{\mathbf{F}_t = [F_t^{i,j}]_{i,j=1}^{d,m}, t \geq 0\}$ is assumed to have predictable elements $F^{i,j}$ with

$$\int_0^T (F_z^{i,j})^2 dz < \infty \quad (1.5.6)$$

almost surely for $i \in \{1, 2, \dots, d\}$, $j \in \{1, 2, \dots, m\}$ and all $T \in (0, \infty)$, see [Protter \(2005\)](#). This allows us to introduce a d -dimensional continuous stochastic vector process $\mathbf{X} = \{\mathbf{X}_t = (X_t^1, X_t^2, \dots, X_t^d)^\top, t \geq 0\}$, where the k th component X^k is defined via the Itô differential

$$dX_t^k = e_t^k dt + \sum_{j=1}^m F_t^{k,j} dW_t^j \quad (1.5.7)$$

for $t \geq 0$ and given \mathcal{A}_0 -measurable initial value $\mathbf{X}_0 = (X_0^1, \dots, X_0^d)^\top \in \mathbb{R}^d$.

Consider now a function $u : [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}$ that has continuous partial derivatives $\frac{\partial u}{\partial t}$, $\frac{\partial u}{\partial x^k}$ and $\frac{\partial^2 u}{\partial x^k \partial x^i}$ for all $k, i \in \{1, 2, \dots, d\}$, $t \geq 0$ and $\mathbf{x} = (x^1, x^2, \dots, x^d)^\top$. The scalar stochastic process $u = \{u(t, X_t^1, X_t^2, \dots, X_t^d), t \geq 0\}$ satisfies then the Itô differential characterized by the *Itô formula*

$$\begin{aligned} du(t, X_t^1, X_t^2, \dots, X_t^d) \\ = \left\{ \frac{\partial u}{\partial t} + \sum_{k=1}^d e_t^k \frac{\partial u}{\partial x^k} + \frac{1}{2} \sum_{j=1}^m \sum_{i,k=1}^d F_t^{i,j} F_t^{k,j} \frac{\partial^2 u}{\partial x^i \partial x^k} \right\} dt \\ + \sum_{j=1}^m \sum_{i=1}^d F_t^{i,j} \frac{\partial u}{\partial x^i} dW_t^j, \end{aligned} \quad (1.5.8)$$

for $t \geq 0$ with initial value $u(0, X_0^1, X_0^2, \dots, X_0^d)$, where the partial derivatives of the function u are evaluated at $(t, X_t^1, X_t^2, \dots, X_t^d)$, which we suppressed in our notation.

We can rewrite the multi-dimensional Itô formula (1.5.8) also in the form

$$du(t, X_t^1, X_t^2, \dots, X_t^d) = \frac{\partial u}{\partial t} dt + \sum_{i=1}^m \frac{\partial u}{\partial x^i} dX_t^i + \frac{1}{2} \sum_{i,k=1}^m \frac{\partial^2 u}{\partial x^i \partial x^k} d[X^i, X^k]_t \quad (1.5.9)$$

for all $t \geq 0$ using covariations, see (1.3.8), which holds for continuous vector processes \mathbf{X} .

The Stratonovich version of this chain rule has the form

$$du(t, X_t^1, X_t^2, \dots, X_t^\ell) = \frac{\partial u}{\partial t} dt + \sum_{i=1}^m \frac{\partial u}{\partial x^i} \circ dX_t^i \quad (1.5.10)$$

for all $t \geq 0$, which is less complex than (1.5.9).

Itô Formula for Jump Processes

For a pure jump process \mathbf{X} with $\mathbf{X}_t = \sum_{s \in (0,t]} \Delta \mathbf{X}_s$ and a measurable function $u : \mathbb{R}^d \rightarrow \mathbb{R}$ we have by our notation the trivial Itô formula

$$du(\mathbf{X}_t) = u(\mathbf{X}_t) - u(\mathbf{X}_{t-}) = \Delta u(\mathbf{X}_t) \quad (1.5.11)$$

for all $t \geq 0$, where $\Delta u(\mathbf{X}_t) = u(\mathbf{X}_t) - u(\mathbf{X}_{t-})$. In (1.5.11) the jumps of \mathbf{X} are simply transferred through the function u as they arise.

Assume that the vector process $\mathbf{X} = \{\mathbf{X}_t = (X_t^1, \dots, X_t^\ell)^\top, t \geq 0\}$ has as its i th component the semimartingale X_t^i with decomposition

$$X_t^i = X_0^i + X_t^{i,c} + X_t^{i,d} \quad (1.5.12)$$

for $t \geq 0$ and $i \in \{1, 2, \dots, \ell\}$. Here

$$X_t^{i,c} = A_t^{i,c} + M_t^{i,c} \quad (1.5.13)$$

denotes the i th component of the continuous part of X_t^i and $X_t^{i,d}$ of the pure jump part. All jumps of X_t^i are expressed by

$$X_t^{i,d} = \sum_{s \in [0,t]} \Delta X_s^i \quad (1.5.14)$$

for all $t \geq 0$ and $i \in \{1, 2, \dots, \ell\}$. In (1.5.13) $M^{i,c}$ denotes a continuous (\underline{A}, P) -local martingale and $A^{i,d}$ a continuous process of finite total variation.

For a twice continuously differentiable function $u : [0, \infty) \times \mathbb{R}^\ell \rightarrow \mathbb{R}$, with continuous first partial derivative with respect to time and continuous second partial derivatives with respect to the spatial variables, the following Itô formula holds

$$\begin{aligned} du(t, X_t^1, \dots, X_t^\ell) &= \frac{\partial}{\partial t} u(t, X_t^1, \dots, X_t^\ell) dt + \sum_{i=1}^{\ell} \frac{\partial}{\partial x^i} u(t, X_t^1, \dots, X_t^\ell) dX_t^{i,c} \\ &\quad + \frac{1}{2} \sum_{i,k=1}^{\ell} \frac{\partial^2}{\partial x^i \partial x^k} u(t, X_t^1, \dots, X_t^\ell) d[M^{i,c}, M^{k,c}]_t \\ &\quad + \Delta u(t, X_t^1, \dots, X_t^\ell) \end{aligned} \quad (1.5.15)$$

for $t \geq 0$, where as before

$$\Delta u(t, X_t^1, \dots, X_t^\ell) = u(s, X_t^1, \dots, X_t^\ell) - u(t-, X_{t-}^1, \dots, X_{t-}^\ell), \quad (1.5.16)$$

see [Protter \(2005\)](#).

To provide an example for the above Itô formula (1.5.15) let us add to the dynamics of a compensated Poisson process $q = \{q_t = N_t - \lambda t, t \geq 0\}$ a Wiener process $W = \{W_t, t \geq 0\}$ and a trend. We interpret now the process $X = \{X_t, t \geq 0\}$ as the logarithm of an asset price, when X_t has the Itô differential

$$dX_t = g dt + \sigma dW_t + c(dN_t - \lambda dt) \quad (1.5.17)$$

for $t \geq 0$. Here we use as parameters the growth rate g , the Poisson intensity λ , the volatility σ and the jump size c . The Itô formula (1.5.15) yields then for the asset price

$$u(X_t) = \exp\{X_t\}$$

the Itô differential

$$d(\exp\{X_t\}) = \exp\{X_{t-}\} \left(\left(g + \frac{1}{2} \sigma^2 + \lambda(e^c - 2) \right) dt + \sigma dW_t + (e^c - 1) dq_t \right)$$

for $t \geq 0$. The above jump diffusion dynamics in (1.5.17) represent a special case of the well-known Merton jump-diffusion model, see [Merton \(1976\)](#), and generalizes the Black-Scholes model, see [Black & Scholes \(1973\)](#), when there are no jumps.

Finally, let us consider a particular case of the Itô formula (1.5.15) when only Wiener processes and Poisson jump measures drive the dynamics, as will be the case in several parts of the book. Let us assume that $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \geq 0\}$ is an m -dimensional standard Wiener process and $p_{\varphi_r}^r(dv, dt)$ denotes the r th Poisson measure on $\mathcal{E} \times [0, \infty)$ with the intensity measure

$$\nu_{\varphi_r}^r(dv, dt) = \varphi_r(dv) dt, \quad (1.5.18)$$

$r \in \{m+1, m+2, \dots, \bar{\ell}\}$, $\bar{\ell} \in \{m+1, m+2, \dots\}$. Suppose that the i th component X_t^i at time t of the underlying d -dimensional process \mathbf{X} has the Itô differential

$$dX_t^i = a_t^i dt + \sum_{k=1}^m b_t^{i,k} dW_t^k + \sum_{r=m+1}^{\bar{\ell}} \int_{\mathcal{E}} c^{i,r}(v, t-) p_{\varphi_r}^r(dv, dt) \quad (1.5.19)$$

for $t \geq 0$ and $i \in \{1, 2, \dots, d\}$, where a^i , $b^{i,k}$ and $c^{i,r}$ are predictable processes and the mark space is given as $\mathcal{E} = \mathfrak{R} \setminus \{0\}$. Then for a function $u : [0, \infty) \times \mathfrak{R}^d \rightarrow \mathfrak{R}$, which is assumed to be differentiable with respect to t and twice differentiable with respect to x , the Itô formula has the form

$$\begin{aligned}
du(t, X_t^1, \dots, X_t^d) = & \left(\frac{\partial u(t, X_t^1, \dots, X_t^d)}{\partial t} + \sum_{i=1}^d a_t^i \frac{\partial}{\partial x^i} u(t, X_t^1, \dots, X_t^d) \right. \\
& + \frac{1}{2} \sum_{i,j=1}^d \sum_{k=1}^m b_t^{i,k} b_t^{j,k} \frac{\partial^2 u(t, X_t^1, \dots, X_t^d)}{\partial x^i \partial x^j} \Big) dt \\
& + \sum_{k=1}^m \sum_{i=1}^d b_t^{i,k} \frac{\partial u(t, X_t^1, \dots, X_t^d)}{\partial x^i} dW_t^k \\
& + \sum_{r=m+1}^{\bar{\ell}} \int_{\mathcal{E}} (u(t, X_t^1, \dots, X_t^d) \\
& - u(t-, X_{t-}^1, \dots, X_{t-}^d)) p_{\varphi_r}^r(dv, dt) \tag{1.5.20}
\end{aligned}$$

for $t \geq 0$. This version of the Itô formula allows one to handle problems which may include Lévy processes with infinite jump intensity as underlying factors.

It is obvious that in the case with jumps the jump part of the Stratonovich chain rule is analogous to the jump part of the Itô formula above.

1.6 Stochastic Differential Equations

Feedback in Stochastic Dynamics

In the Itô differentials that we have considered we left the specification of the drift coefficient, diffusion coefficient and jump coefficient open. For instance, these coefficients can be made state dependent. In financial markets the modeling of quantities which have some feedback in their dynamics need to be mastered, which leads to *stochastic differential equations* (SDEs). An SDE contains an unknown, which is its solution process. To be useful in modeling, such a solution needs to *exist* in an appropriate sense. Furthermore, the *uniqueness* of the solution of an SDE has to be guaranteed to make sure that one achieves the modeling goal without any ambiguity.

Since in reality time is evolving continuously, a financial market model should originally be formulated in continuous time. However numerically, we typically simulate discrete-time approximations of solutions to SDEs because we have to implement these on a digital computer, which can only handle discrete objects. An SDE, as a modeling device, is rather robust when its solution is observed for sufficiently small time step sizes. The elegant framework of SDEs permits continuous time transformations and other manipulations via stochastic calculus without the need to deal with error terms that typically result when manipulating discrete-time models, see [Merton \(1992\)](#). In continuous time finance, the dynamics can be efficiently modeled by drift and

diffusion coefficient functions that determine an SDE. If jumps are involved, then a jump coefficient and corresponding jump intensity are also needed in an SDE.

Solution of Continuous SDEs

We consider now a continuous stochastic process $Y = \{Y_t, t \geq 0\}$ which is a *solution* of a given SDE

$$dY_t = a(t, Y_t) dt + b(t, Y_t) dW_t \quad (1.6.1)$$

for $t \geq 0$, with initial value $Y_0 = y_0$. Here W denotes the driving standard Wiener process. If the process Y has for all $t \geq 0$ an Itô differential of the form (1.6.1), then $Y = \{Y_t, t \geq 0\}$ is called a solution of the SDE (1.6.1). More precisely, a solution of the SDE (1.6.1) is a pair (Y, W) of adapted stochastic processes, defined on a given filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$, where the continuous process Y satisfies for each $t \geq 0$ the Itô integral equation

$$Y_t = y_0 + \int_0^t a(s, Y_s) ds + \int_0^t b(s, Y_s) dW_s. \quad (1.6.2)$$

Of course, we need to assume that both integrals on the right hand side of (1.6.2) exist. It is sufficient to request that

$$\int_0^t |a(s, Y_s)| ds + \int_0^t |b(s, Y_s)|^2 ds < \infty$$

for all $t \geq 0$. Note that the SDE (1.6.1) is only a shorthand notation for the integral equation (1.6.2). The existence and uniqueness of a solution of an SDE is not trivially given, and needs to be ensured, as will be discussed later. In addition to the initial value, not only the drift coefficient, the diffusion coefficient and the jump coefficient need to be given for certain SDEs, additionally also the behavior of its solution at certain boundaries may have to be defined when establishing the uniqueness of the solution of the SDE. For instance, absorption or reflection has to be declared for certain SDEs at the level zero. However, in many cases this is not necessary since there is no ambiguity. In finance, asset prices are typically absorbed at zero which excludes obvious arbitrage opportunities.

Examples of One-Factor Continuous Asset Price Models

Let us list, together with the already mentioned examples, a few one-dimensional diffusion processes that have been applied in the literature of asset price modeling:

The *linearly transformed Wiener process*, which shifts the trajectories shown in Fig. 1.1.3 to a positive initial level and scales time, is an example of

a diffusion process with drift $a(s, x) = 0$ and diffusion coefficient $b(s, x) = 1$. As previously mentioned in Sect. 1.1, it was used in [Bachelier \(1900\)](#) for stock price modeling. One of the disadvantages of the *Bachelier model* is given by the fact that it can generate negative asset prices. In a simplistic way it is sometimes argued that if one freezes the trajectory of the Bachelier model when it first hits zero, then one obtains a basic asset price model. This model has still several deficiencies due to its simplicity. In particular, the asset price will hit the level zero with positive probability, which is usually not intended when modeling asset prices or indices.

The *Black-Scholes* (BS) *model* has drift coefficient

$$a(s, x) = x \left(g + \frac{1}{2} b^2 \right) \quad (1.6.3)$$

and diffusion coefficient

$$b(s, x) = x b. \quad (1.6.4)$$

The BS model was suggested in [Samuelson \(1955, 1965a\)](#) and used in Black & Scholes ([\(1973\)](#)). Despite the fact that this model became the standard financial market model, it has some shortcomings. For instance, it does not generate a random, fluctuating volatility, which is usually observed in practice.

The *geometric Ornstein-Uhlenbeck* (GOU) *model* can be shown to have drift coefficient

$$a(s, x) = x (1 - \ln(x)) \quad (1.6.5)$$

and diffusion coefficient

$$b(s, x) = \sqrt{2} x. \quad (1.6.6)$$

This asset price model permits equilibrium type dynamics for its logarithm around a straight line. It was used, for instance, in [Föllmer & Schweizer \(1993\)](#), [Platen & Rebolledo \(1996\)](#) and [Fleming & Sheu \(1999\)](#). A disadvantage of this model is again that it does not accommodate a fluctuating volatility.

The *constant elasticity of variance* (CEV) *model* introduced in [Cox \(1975\)](#), see also [Schroder \(1989\)](#), has drift

$$a(s, x) = x r \quad (1.6.7)$$

and diffusion coefficient

$$b(s, x) = \sigma x^\alpha \quad (1.6.8)$$

with constants r , σ and $\alpha \in (0, 1)$. It does generate a fluctuating volatility. The, so called, elasticity of the changes of the variance of log-returns can be shown to be constant due to the power structure of the diffusion coefficient. However, as shown in [Delbaen & Shirakawa \(2002\)](#), the model has a deficiency since the asset price will hit zero with positive probability in finite time, which is not what one usually intends to model, say, for a market index.

The *minimal market model* (MMM) was introduced in [Platen \(2001, 2002\)](#) and has in its stylized version the deterministic drift

$$a(s, x) = \alpha \exp\{\eta s\} \quad (1.6.9)$$

and the diffusion coefficient

$$b(s, x) = \sqrt{\alpha \exp\{\eta s\} x} \quad (1.6.10)$$

with initial scaling parameter $\alpha > 0$ and net growth rate $\eta > 0$. This model generates a rather realistic, fluctuating volatility and never hits zero. In particular, its volatility dynamics reflect those of observed equity index volatilities which increase when the index falls and vice versa, known as the leverage effect, see [Black \(1976\)](#).

Examples of One-Factor Interest Rate Models

A large variety of interest rate models has been developed that are formed by solutions of SDEs. In the following we mention several one-factor interest rate models by only specifying their drift and diffusion coefficients.

One of the simplest stochastic interest rate models arises if the Wiener process is linearly transformed by a deterministic, time dependent drift coefficient $a(s, x) = a_s$ and a deterministic time dependent diffusion coefficient $b(s, x) = b_s$. This leads to the *Merton model*, see [Merton \(1973a\)](#). It can also be interpreted as some specification of the continuous time version of the *Ho-Lee model*, see [Ho & Lee \(1986\)](#). Under this model the interest rate may not remain positive.

A widely used interest rate model is the *Vasicek model*, see [Vasicek \(1977\)](#), or more generally the *extended Vasicek model*, with linear drift coefficient $a(s, x) = \gamma_s (\bar{x}_s - x)$ and deterministic diffusion coefficient $b(s, x) = b_s$. Here γ_s , \bar{x}_s and b_s are deterministic functions of time. Also this model yields a Gaussian interest rate and, thus, allows negative rates.

In [Black \(1995\)](#) it was suggested that one considers the nonnegative value of an interest rate, similar to an option value, which only takes the positive part of an underlying quantity. The *Black model* results, when using a Vasicek model $u = \{u_t, t \in [0, T]\}$ to model an underlying shadow interest rate u_t . The interest rate is then obtained by its truncation $x_s = (u_s)^+ = \max(0, u_s)$. Such type of interest rate model, which allows the consideration of low interest rate regimes, has been studied, for instance, in [Gorovoi & Linetsky \(2004\)](#) and [Miller & Platen \(2005\)](#).

[Cox, Ingersoll & Ross \(1985\)](#) suggested the well-known *CIR model*, which is sometimes called square root (SR) process and uses a process with linear drift coefficient $a(s, x) = \gamma_s (\bar{x}_s - x)$ and square root diffusion coefficient of the form $b(s, x) = b_s \sqrt{x}$. Here γ_s , \bar{x}_s and b_s are deterministic functions of time with $\gamma_s \bar{x}_s > 2 b_s^2$. This model has the desirable feature that it excludes negative interest rates. Furthermore, it yields equilibrium dynamics. Unfortunately, when calibrated to market data, it shows a number of deficiencies which mainly concern the possible shapes of the, so-called, *forward rate* or *yield curves*.

An extended model of the CIR type is the *Pearson-Sun model*, see [Pearson & Sun \(1989\)](#), which assumes $a(s, x) = \gamma(\bar{x}_s - x)$ and $b(s, x) = \sqrt{b_1 + b_2 x}$. Here the parameters are usually assumed to be constants which fulfill certain conditions, such as $\gamma(\bar{x} + \frac{b_1}{b_2}) > 0$. These ensure that the solution is contained in a certain region. [Duffie & Kan \(1994\)](#) generalized this model, which belongs to the affine class of processes, because the drift a and squared diffusion coefficient b^2 are affine. This model is therefore sometimes called *affine model*.

[Marsh & Rosenfeld \(1983\)](#) and also [Dothan \(1978\)](#) considered an interest rate model with $a(s, x) = a_s x$ and $b(s, x) = b_s x$. This specification is known as the *lognormal model*. Here the interest rate remains positive, however, it does not admit a stationary regime.

A generalized lognormal model, also called the *Black-Karasinski model*, see [Black & Karasinski \(1991\)](#), is obtained by setting $a(s, x) = x(a_s + g_s \ln(x))$ and $b(s, x) = b_s x$. This generates a geometric Ornstein-Uhlenbeck process, see (1.6.5)–(1.6.6). If $g_s = -\frac{b'_s}{b_s}$, then the above model can also be called the continuous-time version of the *Black-Derman-Toy model*, see [Black, Derman & Toy \(1990\)](#). This type of model keeps interest rates positive and allows them to have equilibrium dynamics.

Another model arises if one sets $a(s, x) = \gamma_s (\bar{x}_s - x)$ and $b(s, x) = b_s x$. In the case of constant parameters this formulation is known as the *Courtadon model*, see [Courtadon \(1982\)](#). The *Longstaff model*, see [Longstaff \(1989\)](#) is obtained by setting $a(s, x) = \gamma_s (\sqrt{\bar{x}_s} - \sqrt{x})$ and $b(s, x) = b_s \sqrt{x}$.

A rather general model is the *Hull-White model*, see [Hull & White \(1990\)](#). It has linear mean-reverting drift $a(s, x) = \gamma_s (\bar{x}_s - x)$ and power diffusion coefficient $b(s, x) = b_s x^q$ for some choice of exponent $q \geq 0$. Obviously, this structure includes several of the above models. For instance, in the case $q = 0$ the Hull-White model reduces to the extended Vasicek model.

The *Sandmann-Sondermann model*, see [Sandmann & Sondermann \(1994\)](#), was motivated by the aim to consider annual, continuously compounded interest rates. It has drift $a(s, x) = (1 - e^{-x})(a_s - \frac{1}{2}(1 - e^{-x}) b_s^2)$ and diffusion coefficient $b(s, x) = (1 - e^{-x})c_s$.

An alternative interest rate model was proposed in [Platen \(1999b\)](#), which suggests a drift $a(s, x) = \gamma(x - a_s)(c_s - x)$ and a $\frac{3}{2}$ power diffusion coefficient of the type $b(s, x) = b_s |x - c_s|^{\frac{3}{2}}$. This model provides a reasonable reflection of the interest rate drift and diffusion coefficient as estimated from market data in [Ait-Sahalia \(1996\)](#).

As can be seen from these examples one can, in principle, choose quite general functions for the drift and diffusion coefficients in SDEs to form meaningful models of asset prices, interest rates and other financial quantities. These coefficient functions then characterize, together with the initial conditions, the dynamics of the process in an elegant and compact way. Note that in some cases it is necessary to determine also the behavior of the solution of an SDE at its boundaries. For instance, reflection or absorption has to be declared

which may not be automatically clear from the drift and diffusion coefficients of an SDE alone.

Continuous Vector SDEs

For the modeling of financial markets, multi-dimensional solutions of SDEs are needed. We recall that $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, W_t^2, \dots, W_t^m)^\top, t \geq 0\}$ is an m -dimensional standard Wiener process with components W^1, W^2, \dots, W^m .

Given a d -dimensional vector function $\mathbf{a} : [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and a $d \times m$ -matrix function $\mathbf{b} : [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$, then we can form the d -dimensional *vector stochastic differential equation*

$$d\mathbf{X}_t = \mathbf{a}(t, X_t) dt + \mathbf{b}(t, X_t) d\mathbf{W}_t \quad (1.6.11)$$

for $t \geq 0$ with initial value $\mathbf{X}_0 \in \mathbb{R}^d$. The vector stochastic differential (1.6.11) should be interpreted as an Itô integral equation of the form

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t \mathbf{a}(s, X_s) ds + \int_0^t \mathbf{b}(s, X_s) d\mathbf{W}_s, \quad (1.6.12)$$

for $t \geq 0$, where the integrals are defined componentwise. Thus, the i th component of (1.6.12) is then given by the SDE

$$X_t^i = X_0^i + \int_0^t a^i(s, X_s) ds + \sum_{\ell=1}^m \int_0^t b^{i,\ell}(s, X_s) dW_s^\ell, \quad (1.6.13)$$

for $t \geq 0$ and $i \in \{1, 2, \dots, d\}$. Note that the drift and diffusion coefficients of each component can depend on all other components, which allows us to model a wide range of possible feedback mechanisms.

Before applying the above type of model, questions concerning the existence and uniqueness of solutions of vector SDEs must be answered. For this we refer to existence and uniqueness results, to be formulated later in Sect. 1.9, see also Krylov (1980) and Protter (2005).

SDEs with Poissonian Jumps

It is important to be able to incorporate into a model event driven uncertainty expressed by jumps. This arises, for instance, when modeling credit risk, insurance risk or operational risk. Again, an SDE will allow the modeling of feedback effects in the presence of jumps and in dependence on the level of the state variable itself.

Let $W = \{W_t, t \geq 0\}$ denote a standard Wiener process and $N = \{N_t, t \geq 0\}$ a Poisson process with intensity $\lambda > 0$. A scalar SDE for an asset price X_t could then take the form

$$dX_t = a(t, X_t) dt + b(t, X_t) dW_t + c(t-, X_{t-}) dN_t \quad (1.6.14)$$

for $t \geq 0$, with initial value $X_0 > 0$. Here the jump coefficient $c(\cdot, \cdot)$ determines the jump size at an event. One often calls the resulting solution to process X of the SDE (1.6.14) an *Itô process with jumps* or a *jump diffusion*.

In a financial market several interacting quantities need to play a role. Let us denote m independent standard Wiener processes by W^1, \dots, W^m . Furthermore, let N^1, \dots, N^n denote n independent Poisson processes with corresponding intensities $\lambda^1(t, \mathbf{X}_t), \dots, \lambda^n(t, \mathbf{X}_t)$ at time t . As indicated, these intensities may depend on time and also on the state vector $\mathbf{X}_t = (X_t^1, X_t^2, \dots, X_t^d)^\top$. The model for the components of the state vector is described by the following system of SDEs

$$dX_t^i = a^i(t, \mathbf{X}_t) dt + \sum_{k=1}^m b^{i,k}(t, \mathbf{X}_t) dW_t^k + \sum_{\ell=1}^n c^{i,\ell}(t-, \mathbf{X}_{t-}) dN_t^\ell \quad (1.6.15)$$

for $t \geq 0$ with $X_0^i \in \mathfrak{R}$ and $i \in \{1, 2, \dots, d\}$. To guarantee the existence and uniqueness of a solution of the system of SDEs (1.6.15) the drift coefficients $a^i(\cdot, \cdot)$, diffusion coefficients $b^{i,k}(\cdot, \cdot)$ and jump coefficients $c^{i,\ell}(\cdot, \cdot)$ need to satisfy appropriate conditions. Such conditions will be outlined in Sect. 1.9, see also Ikeda & Watanabe (1989) and Protter (2005).

SDEs Driven by Jump Measures

When events with randomly distributed jump sizes are modeled, it is appropriate to use SDEs driven by Poisson measures. Instead of Poisson processes, Poisson measures $p_{\varphi_\ell}^\ell(\cdot, \cdot)$, $\ell \in \{1, 2, \dots, n\}$, will now be employed, see Sect. 1.1. Leaving the other terms in the SDE (1.6.15) unchanged, we consider an SDE for the i th component of \mathbf{X}_t in the form

$$dX_t^i = a^i(t, \mathbf{X}_t) dt + \sum_{k=1}^m b^{i,k}(t, \mathbf{X}_t) dW_t^k + \sum_{\ell=1}^n \int_{\mathcal{E}} c^{i,\ell}(v, t-, \mathbf{X}_{t-}) p_{\varphi_\ell}^\ell(dv, dt) \quad (1.6.16)$$

for $t \geq 0$ with $X_0^i \in \mathfrak{R}$, $i \in \{1, 2, \dots, d\}$. Here the (i, ℓ) th jump coefficient may not only be a function of time and other components, but may also depend on the mark $v \in \mathcal{E}$, see (1.1.30). It is important to note that φ_ℓ satisfies condition (1.1.32) for $\ell \in \{1, 2, \dots, n\}$. One can model by the SDE (1.6.16) dynamics involving infinitely many small jumps per unit of time, for instance, those that arise for some Lévy processes. The intensity for those jumps that impact the i th component is controllable via the jump coefficients which depend on the mark v . Even though the Poisson jump measure $p_{\varphi_\ell}^\ell$ is modeled in a standard way without depending on state variables, its impact can depend, in a very flexible manner, on the actual values of all components of the SDE. To recover the SDE (1.6.15) for jump diffusions we may choose

$$\varphi_\ell(dv) = \mathbf{1}_{\{v \in [0, \lambda]\}} dv,$$

and may leave the jump coefficients independent of the marks. SDEs that are driven by jump measures have been used in financial modeling, for instance, in [Björk, Kabanov & Rungaldier \(1997\)](#) in a more classical framework or [Christensen & Platen \(2005\)](#) in a rather general setup.

1.7 Linear SDEs

We call a solution of an SDE an explicit solution, if it has a representation where the terms on its right hand side do not use the solution itself. The best example for SDEs with explicit solutions are linear SDEs which have linear drift and diffusion coefficients. As we will show below, they form an interesting class of SDEs that is very convenient in some applications in finance.

Continuous Linear SDE with Additive Noise

Let us consider the linear SDE with *additive noise*

$$dX_t = (a_1(t) X_t + a_2(t)) dt + b_2(t) dW_t \quad (1.7.1)$$

for $t \geq 0$, $X_0 \in \Re$, where the coefficients a_1 , a_2 and b_2 are deterministic functions of time t . If $a_1(t) \neq 0$, then the drift coefficient can model a feedback effect. W is here, again a standard Wiener process. With the *fundamental solution*

$$\Phi_t = \exp \left\{ \int_0^t a_1(s) ds \right\}$$

for $t \geq 0$ one can write the *explicit solution* for (1.7.1) in the form

$$X_t = \Phi_t \left(X_0 + \int_0^t a_2(s) \Phi_s^{-1} ds + \int_0^t b_2(s) \Phi_s^{-1} dW_s \right) \quad (1.7.2)$$

for $t \geq 0$. Note that this explicit solution is Gaussian.

The corresponding expectation equals

$$E(X_t) = \Phi_t \left(X_0 + \int_0^t a_2(s) \Phi_s^{-1} ds \right) \quad (1.7.3)$$

and the variance is given as

$$E \left((X_t - E(X_t))^2 \right) = \int_0^t (b_2(s) \Phi_t \Phi_s^{-1})^2 ds \quad (1.7.4)$$

for $t \geq 0$.

Ornstein-Uhlenbeck Process

An important example for a solution of the above linear SDE is the *Ornstein-Uhlenbeck process*, which is given by (1.7.2) when the drift $a_1(t)X_t + a_2(t) = \gamma(\bar{x} - X_t)$ is linear and mean reverting, with given constant speed of adjustment γ and reference level \bar{x} . The diffusion coefficient $b_2(t) = \beta$ is, in this case, constant. For the linear SDE

$$dX_t = \gamma(\bar{x} - X_t) dt + \beta dW_t \quad (1.7.5)$$

for $t \geq 0$ with initial value $X_0 \in \Re$ one has by (1.7.2) the explicit solution

$$X_t = X_0 \exp\{-\gamma t\} + \bar{x}(1 - \exp\{-\gamma t\}) + \beta \int_0^t \exp\{-\gamma(t-s)\} dW_s \quad (1.7.6)$$

for $t \geq 0$.

The solution (1.7.6) can be interpreted as that of the [Vasicek \(1977\)](#) interest rate model. The reference level \bar{x} is the long-term average value of the interest rate. The parameter β characterizes the magnitude of the fluctuations of the interest rate. The speed of adjustment parameter γ determines how fast a shock to the interest rate loses its impact. From (1.7.3) and (1.7.4) it follows that the mean and the variance converge over time to \bar{x} and $\frac{\beta^2}{2\gamma}$, respectively.

Continuous Linear SDE

We now consider a *linear SDE*

$$dX_t = (a_1(t)X_t + a_2(t)) dt + (b_1(t)X_t + b_2(t)) dW_t \quad (1.7.7)$$

for $t \geq 0$, $X_0 \in \Re$, which also involves multiplicative noise. Here a_1 , a_2 , b_1 and b_2 are deterministic functions of time and W is a standard Wiener process. The SDE (1.7.7) has an explicit solution of the form

$$X_t = \Psi_t \left(X_0 + \int_0^t (a_2(s) - b_1(s)b_2(s)) \Psi_s^{-1} ds + \int_0^t b_2(s) \Psi_s^{-1} dW_s \right), \quad (1.7.8)$$

where

$$\Psi_t = \exp \left\{ \int_0^t \left(a_1(s) - \frac{1}{2} b_1^2(s) \right) ds + \int_0^t b_1(s) dW_s \right\} \quad (1.7.9)$$

for $t \geq 0$. Obviously, for $b_1(t) = 0$ the expression (1.7.8) recovers (1.7.2).

The mean has the form

$$\mu(t) = E(X_t) = E(X_0) \exp \left\{ \int_0^t a_1(s) ds \right\} + \int_0^t a_2(s) \exp \left\{ \int_s^t a_1(u) du \right\} ds \quad (1.7.10)$$

for $t \geq 0$. By application of the Itô formula it can be shown that the second moment

$$P(t) = E(X_t^2) \quad (1.7.11)$$

satisfies the ordinary differential equation (ODE)

$$dP(t) = \left((2a_1(t) + b_1^2(t)) P(t) + 2\mu(t) (a_2(t) + b_1(t) b_2(t)) + b_2^2(t) \right) dt \quad (1.7.12)$$

for $t \geq 0$ and $P(0) = E(X_0^2) < \infty$.

Geometric Brownian Motion

The standard example of an asset price model is that of a *geometric Brownian motion* given by the linear SDE

$$dX_t = r_t X_t dt + \sigma_t X_t dW_t, \quad (1.7.13)$$

for $t \geq 0$ with deterministic initial value $X_0 > 0$. The solution of this SDE is also known as the *Black-Scholes model*, see [Black & Scholes \(1973\)](#) and [Merton \(1973b\)](#). It was originally mentioned in [Osborne \(1959\)](#) and [Samuelson \(1965b\)](#) as potential asset price model. Its key advantages are that it generates strictly positive dynamics, and it has an explicit solution

$$X_t = X_0 \exp \left\{ \int_0^t r_s ds - \frac{1}{2} \int_0^t \sigma_s^2 ds + \int_0^t \sigma_s dW_s \right\} \quad (1.7.14)$$

for $t \geq 0$ and permits easy manipulations. Obviously, its logarithm $\ln(X_t)$ is Gaussian with mean

$$E(\ln(X_t)) = \ln(X_0) + \int_0^t \left(r_s - \frac{\sigma_s^2}{2} \right) ds \quad (1.7.15)$$

and variance

$$E \left((\ln(X_t) - E(\ln(X_t)))^2 \right) = \int_0^t \sigma_s^2 ds. \quad (1.7.16)$$

If in (1.7.14) $\sigma_t = \sigma$ and $r_t = r$ are chosen as constants, then the explicit solution X_t , represents simply a function of time and Wiener process. More precisely, $\ln(X_t)$ represents a linearly transformed Wiener process.

Using the moment equation (1.7.10) we have for X_t with constant drift $r_t = r$ and constant volatility $\sigma_t = \sigma$ the mean

$$\mu(t) = E(X_t) = X_0 \exp\{r t\} \quad (1.7.17)$$

and variance

$$v(t) = P(t) - (\mu(t))^2 = (X_0)^2 \exp\{2 r t\} (\exp\{\sigma^2 t\} - 1) \quad (1.7.18)$$

for $t \geq 0$.

Furthermore, for $q \in \Re$ we can obtain the q th moment of X_t as

$$E((X_t)^q) = (X_0)^q \exp \left\{ q \left(r - \frac{\sigma^2}{2} (1-q) \right) t \right\} \quad (1.7.19)$$

for $t \geq 0$.

Continuous Multi-dimensional Linear SDEs

Consider a d -dimensional *linear SDE* of the form

$$d\mathbf{X}_t = (\mathbf{A}_t \mathbf{X}_t + \boldsymbol{\alpha}_t) dt + \sum_{\ell=1}^m \left(\mathbf{B}_t^\ell \mathbf{X}_t + \boldsymbol{\beta}_t^\ell \right) dW_t^\ell, \quad (1.7.20)$$

where $\mathbf{A}, \mathbf{B}^1, \mathbf{B}^2, \dots, \mathbf{B}^m$ are $d \times d$ -matrix valued and $\boldsymbol{\alpha}, \boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \dots, \boldsymbol{\beta}^m$ d -dimensional vector valued deterministic functions of time. We have an explicit solution of (1.7.20) in the form

$$\mathbf{X}_t = \Psi_t \left(\mathbf{X}_0 + \int_0^t \Psi_s^{-1} \left(\boldsymbol{\alpha}_s - \sum_{\ell=1}^m \mathbf{B}_s^\ell \boldsymbol{\beta}_s^\ell \right) ds + \sum_{\ell=1}^m \int_0^t \Psi_s^{-1} \boldsymbol{\beta}_s^\ell dW_s^\ell \right) \quad (1.7.21)$$

for $t \geq 0$. Here Ψ_t is the $d \times d$ *fundamental matrix* at time t with $\Psi_0 = \mathbf{I}$, where \mathbf{I} denotes the identity matrix. The fundamental matrix process satisfies the *matrix SDE*

$$d\Psi_t = \mathbf{A}_t \Psi_t dt + \sum_{\ell=1}^m \mathbf{B}_t^\ell \Psi_t dW_t^\ell. \quad (1.7.22)$$

Recall that for d -dimensional vectors \mathbf{x} and \mathbf{y} the product $\mathbf{x}\mathbf{y}^\top$ is a $d \times d$ matrix with (i,j) th component $x^i y^j$. One can then show that the *vector mean* $\boldsymbol{\mu}(t) = E(\mathbf{X}_t)$ satisfies the vector ODE

$$d\boldsymbol{\mu}(t) = (\mathbf{A}_t \boldsymbol{\mu}(t) + \boldsymbol{\alpha}_t) dt. \quad (1.7.23)$$

Furthermore, the $d \times d$ *matrix of second moments* $\mathbf{P}(t) = E(\mathbf{X}_t \mathbf{X}_t^\top)$ satisfies the matrix ODE

$$\begin{aligned} d\mathbf{P}(t) &= \left(\mathbf{A}_t \mathbf{P}(t) + \mathbf{P}(t) \mathbf{A}_t^\top + \sum_{\ell=1}^m \mathbf{B}_t^\ell \mathbf{P}(t) \left(\mathbf{B}_t^\ell \right)^\top + \boldsymbol{\alpha}_t \boldsymbol{\mu}(t)^\top + \boldsymbol{\mu}(t) \boldsymbol{\alpha}_t^\top \right. \\ &\quad \left. + \sum_{\ell=1}^m \left(\mathbf{B}_t^\ell \boldsymbol{\mu}(t) \left(\boldsymbol{\beta}_t^\ell \right)^\top + \boldsymbol{\beta}_t^\ell \boldsymbol{\mu}(t)^\top \mathbf{B}_t^\ell + \boldsymbol{\beta}_t^\ell \left(\boldsymbol{\beta}_t^\ell \right)^\top \right) \right) dt, \end{aligned} \quad (1.7.24)$$

with initial conditions $\boldsymbol{\mu}(0) = E(\mathbf{X}_0)$ and $\mathbf{P}(0) = E(\mathbf{X}_0 \mathbf{X}_0^\top)$, assuming that these are finite.

If the matrices $\mathbf{A}, \mathbf{B}^1, \mathbf{B}^2, \dots, \mathbf{B}^m$ commute, that is, if

$$\mathbf{A}_t \mathbf{B}_t^\ell = \mathbf{B}_t^\ell \mathbf{A}_t \quad \text{and} \quad \mathbf{B}_t^\ell \mathbf{B}_t^k = \mathbf{B}_t^k \mathbf{B}_t^\ell \quad (1.7.25)$$

for all $k, \ell \in \{1, 2, \dots, m\}$ and $t \geq 0$, then an explicit solution of the fundamental matrix SDE (1.7.22) can be obtained in the form

$$\Psi_t = \exp \left\{ \int_0^t \left(\mathbf{A}_s - \frac{1}{2} \sum_{\ell=1}^m \left(\mathbf{B}_s^\ell \right)^2 \right) ds + \sum_{\ell=1}^m \int_0^t \mathbf{B}_s^\ell dW_s^\ell \right\}, \quad (1.7.26)$$

where the exponential is a matrix exponential. Of particular practical interest is the case of additive noise, where the matrices \mathbf{B}_t^ℓ , $\ell \in \{1, 2, \dots, m\}$, all vanish. In this situation we obtain vectors of correlated Ornstein-Uhlenbeck processes. Their joint probability is Gaussian, where mean and second moments follow from (1.7.23) and (1.7.24), respectively. In the case when α and β^1, \dots, β^m vanish we obtain the following multi-dimensional Black-Scholes (BS) model:

Multi-Asset Black-Scholes Model

The SDE for the j th Black-Scholes asset price S_t^j can be defined in the form

$$dS_t^j = S_t^j \left(a_t^j dt + \sum_{k=1}^d b_t^{j,k} dW_t^k \right) \quad (1.7.27)$$

for $t \geq 0$ and $j \in \{1, 2, \dots, d\}$, $S_0^j > 0$. To fit this SDE into the framework given by the fundamental matrix SDE (1.7.22) we use the diagonal matrices $\mathbf{A}_t = [A_t^{i,j}]_{i,j=1}^d$ with

$$A_t^{i,j} = \begin{cases} a_t^j & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \quad (1.7.28)$$

and $\mathbf{B}_t^k = [B_t^{k,i,j}]_{i,j=1}^d$ with

$$B_t^{k,i,j} = \begin{cases} b_t^{j,k} & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \quad (1.7.29)$$

for $k, i, j \in \{1, 2, \dots, d\}$ and $t \geq 0$. Note that these matrices commute due to their diagonal structure. We denote by $\mathbf{S}_t = \Psi_t$ a diagonal matrix with j th diagonal element S_t^j , $j \in \{1, 2, \dots, d\}$, representing the j th stock price at time $t \geq 0$. This allows us to write the SDE (1.7.27) as matrix SDE

$$d\mathbf{S}_t = \mathbf{A}_t \mathbf{S}_t dt + \sum_{k=1}^d \mathbf{B}_t^k \mathbf{S}_t dW_t^k \quad (1.7.30)$$

for $t \geq 0$. Consequently, by (1.7.26) we obtain the explicit solution

$$S_t^j = S_0^j \exp \left\{ \int_0^t \left(a_s^j - \frac{1}{2} \sum_{k=1}^d (b_s^{j,k})^2 \right) ds + \sum_{k=1}^d b_s^{j,k} dW_s^k \right\} \quad (1.7.31)$$

for $t \geq 0$ and $j \in \{1, 2, \dots, d\}$ for the multi-dimensional BS model.

ARCH Diffusion Model

A class of discrete-time models generating stochastic volatility has been developed by Engle (1982) and a number of subsequent authors known as ARCH,

GARCH, NGARCH type models. In Nelson (1990) it was pointed out that these models, when appropriately scaled lead to a continuous time diffusion limit that satisfies a pair of SDEs. The resulting *ARCH diffusion model* follows the dynamics

$$dS_t = S_t (a dt + \sigma_t dW_t) \quad (1.7.32)$$

and the squared volatility satisfies the SDE

$$d\sigma_t^2 = k (\bar{\sigma}^2 - \sigma_t^2) dt + \gamma \sigma_t^2 d\tilde{W}_t \quad (1.7.33)$$

for $t \geq 0$, with $S_0 > 0$ and $\sigma_0^2 > 0$. Here W and \tilde{W} are independent Wiener processes. We note that the SDE (1.7.33) is linear.

It is appropriate to generalize the ARCH diffusion model to the multi-dimensional case with correlated driving noise. We have then the j th ARCH diffusion asset price S_t^j satisfying the SDE

$$dS_t^j = S_t^j \left(a_t^j dt + \sum_{k=1}^d \sigma_t^{j,k} dW_t^k \right) \quad (1.7.34)$$

and their (j, k) th squared volatilities can be modeled with the dynamics

$$d(\sigma_t^{j,k})^2 = \kappa_t^{j,k} \left((\bar{\sigma}_t^{j,k})^2 - (\sigma_t^{j,k})^2 \right) dt + \sum_{\ell=1}^m \gamma_t^{j,k,\ell} (\sigma_t^{j,k})^2 dW_t^\ell \quad (1.7.35)$$

for $t \geq 0$, $S_0^j > 0$, $\sigma_0^{j,k} \geq 0$, $j, k \in \{1, 2, \dots, d\}$, $m \in \{1, 2, \dots\}$. Here a_t^j , $\kappa_t^{j,k}$, $(\bar{\sigma}_t^{j,k})^2$ and $\gamma_t^{j,k,\ell}$ are deterministic functions of time, and W^1, W^2, \dots are independent standard Wiener processes. The above model for the volatility is the solution of a multi-dimensional linear SDE which we will discuss in further detail in the next chapter.

Merton's Jump Diffusion Model

Let us now discuss some SDEs with jumps that have explicit solutions. Merton (1976) modeled the dynamics of a stock price S_t by an SDE of the type

$$dS_t = S_{t-} (a dt + \sigma dW_t + dY_t) \quad (1.7.36)$$

for $t \geq 0$ and $S_0 > 0$. In the above *Merton model* the process $Y = \{Y_t, t \geq 0\}$ denotes a compound Poisson process, see (1.1.28), where

$$Y_t = \sum_{k=1}^{N_t} \xi_k \quad (1.7.37)$$

and thus

$$dY_t = \xi_{N_{t-}+1} dN_t \quad (1.7.38)$$

for $t \geq 0$. Here the process $N = \{N_t, t \geq 0\}$ is a Poisson process, see (1.1.20), with intensity $\lambda > 0$, and ξ_1, ξ_2, \dots are i.i.d random variables independent of W and N with given distribution and mean

$$\hat{\xi} = E(\xi_i) < \infty.$$

The constant a is the instantaneous expected rate of return of the stock if there were no jumps, whereas σ denotes the constant volatility parameter. Recall that the left hand limit S_{t-} of S_t at time t describes the value just before a potential jump at time t . We denote the k th jump time of the Poisson process N by τ_k . Then we have for Y the jump size

$$\Delta Y_{\tau_k} = Y_{\tau_k} - Y_{\tau_k-} = \xi_k \quad (1.7.39)$$

at time τ_k , $k \in \{1, 2, \dots\}$. The dynamics specified in the SDE (1.7.36) leads, for S , to the k th jump increment

$$\Delta S_{\tau_k} = S_{\tau_k} - S_{\tau_k-} = S_{\tau_k-} \Delta Y_{\tau_k} = S_{\tau_k-} \xi_k$$

at time τ_k . This means that we have

$$S_{\tau_k} = S_{\tau_k-} (\xi_k + 1). \quad (1.7.40)$$

Obviously,

$$\frac{S_{\tau_k}}{S_{\tau_k-}} = \xi_k + 1 \quad (1.7.41)$$

is the *jump ratio* of S at τ_k . To avoid S_t becoming negative we assume that

$$\xi_k \geq -1. \quad (1.7.42)$$

Equivalently, we may request that the jump ratio is nonnegative, that is, by (1.7.41) and (1.7.40)

$$\frac{S_{\tau_k}}{S_{\tau_k-}} \geq 0 \quad (1.7.43)$$

for all $k \in \{1, 2, \dots\}$. The quantity $\lambda \hat{\xi} t$ compensates the jumps of Y that arise until time t so that the compensated compound Poisson process

$$\tilde{Y} = \left\{ \tilde{Y}_t = Y_t - \lambda \hat{\xi} t, t \geq 0 \right\} \quad (1.7.44)$$

forms an (\mathcal{A}, P) -martingale. We can rewrite the SDE (1.7.36) in the form

$$dS_t = S_t (a + \lambda \hat{\xi}) dt + S_t \sigma dW_t + S_{t-} d\tilde{Y}_t \quad (1.7.45)$$

for $t \geq 0$. The last term on the right hand side of the SDE (1.7.45) forms a martingale. The expected return of S over a period of length Δ equals $(a + \lambda \hat{\xi})\Delta$.

The *explicit solution* of the SDE (1.7.36) has the form

$$S_t = S_0 \exp \left\{ \left(a - \frac{1}{2} \sigma^2 \right) t + \sigma W_t \right\} \prod_{k=1}^{N_t} (\xi_k + 1) \quad (1.7.46)$$

for $t \geq 0$, which generalizes the case of geometric Brownian motion arising under the Black-Scholes model, see (1.7.14). In addition to the Black-Scholes case, here the product of jump ratios, see (1.7.41), appears. Merton derived a formula for the value of an option on S , see [Merton \(1976\)](#), where it was assumed that the k th jump ratio $\frac{S_{\tau_k}}{S_{\tau_{k-}}} = \xi_k + 1$ is lognormal. Other distributions for the jump ratio are possible. For instance, a log-Laplace distribution was assumed in [Kou \(2002\)](#). The above *Merton model* can be interpreted as an extension of the Cramér-Lundberg model, which has been designed to reflect the surplus process of an insurance company, see (1.1.29).

Explicit Solution of an SDE with Jumps

Now, we consider a stochastic process $Z = \{Z_t, t \geq 0\}$ which exhibits jumps and that we characterize by the stochastic differential

$$dZ_t = g_t dt + \sigma_t dW_t + c_t dN_t \quad (1.7.47)$$

for $t \geq 0$ with $Z_0 = z_0 \in \mathfrak{R}$. Here $W = \{W_t, t \geq 0\}$ denotes a standard Wiener process and $N = \{N_t, t \geq 0\}$ a Poisson process with intensity $\lambda > 0$. The quantities g_t , σ_t and c_t are assumed to be deterministic functions of time. By application of the Itô formula (1.5.20) it then follows for the exponential

$$X_t = \exp\{Z_t\} = \exp\{Z_0\} \exp \left\{ \int_0^t g_s ds + \int_0^t \sigma_s dW_s \right\} \prod_{k=1}^{N_t} (\xi_k + 1) \quad (1.7.48)$$

with $\xi_k = \exp\{c_{\tau_k-}\} - 1$ that it solves the corresponding linear SDE

$$dX_t = X_{t-} \left[\left(g_t + \frac{1}{2} \sigma_t^2 \right) dt + \sigma_t dW_t + (\exp\{c_{t-}\} - 1) dN_t \right] \quad (1.7.49)$$

for $t \geq 0$. Note that W is a martingale and the compensated Poisson process $q = \{q_t, t \geq 0\}$ with

$$q_t = N_t - \lambda t$$

is also a martingale. Therefore, the mean $\mu(t) = E(X_t)$ satisfies the ordinary differential equation (ODE)

$$d\mu(t) = \mu(t) \left(g_t + \frac{1}{2} \sigma_t^2 + \lambda(\exp\{c_t\} - 1) \right) dt \quad (1.7.50)$$

for $t \geq 0$ with $\mu(0) = \exp\{Z_0\}$. By the exponential expression (1.7.48) one obtains

$$X_t^2 = \exp\{2 Z_t\}. \quad (1.7.51)$$

Hence, by similar arguments as above one obtains for the second moment

$$P(t) = E(X_t^2)$$

the ODE

$$dP(t) = P(t) \left(2g_t + \sigma_t^2 + \lambda (\exp\{2c_t\} - 1) \right) dt \quad (1.7.52)$$

for $t \geq 0$ with $P(0) = \exp\{2Z_0\}$. One obtains then directly the variance of X_t using (1.7.52) and (1.7.50).

Exponential Lévy Models

An important generalization of Merton's jump diffusion model (1.7.46) is given by an exponential of the form

$$S_t = S_0 \exp\{X_t\}. \quad (1.7.53)$$

Here $X = \{X_t, t \geq 0\}$ is a Lévy process, as defined in (1.1.36), and we have

$$X_t = \alpha t + \beta W_t + \int_0^t \int_{|v|<1} v(p_\varphi(dv, ds) - \varphi(dv) ds) + \int_0^t \int_{|v|\geq 1} v p_\varphi(dv, ds) \quad (1.7.54)$$

for $t \geq 0$, where W is a standard Wiener process and p_φ is a Poisson measure with Lévy measure $\varphi(dv)$ so that (1.1.32) is satisfied. By application of the Itô formula (1.5.15) one obtains for the resulting *exponential Lévy model* the SDE

$$\begin{aligned} dS_t &= S_{t-} \left[\left(\alpha + \frac{1}{2} \beta^2 - \int_{|v|<1} v \varphi(dv) \right) dt + \beta dW_t \right. \\ &\quad \left. + \int_{-\infty}^{\infty} (\exp\{v\} - 1) p_\varphi(dv, dt) \right] \end{aligned} \quad (1.7.55)$$

for $t \geq 0$. Assume that all expressions on the right hand sides of (1.7.54) and (1.7.55) exist. Then inserting formula (1.7.54) into (1.7.53) provides an explicit solution for the SDE (1.7.55).

Models of this type have been proposed and studied by various authors, for instance, by Madan & Seneta (1990), Madan & Milne (1991), Eberlein & Keller ((1995)), Barndorff-Nielsen & Shephard (2001), Miyahara & Novikov (2002) and Carr, Geman, Madan & Yor (2003).

1.8 SDEs with Jumps

Multi-dimensional SDE

Let there be given a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$, with $\underline{\mathcal{A}} = (\underline{\mathcal{A}}_t)_{t \geq 0}$, satisfying the usual conditions, and with a set of marks $\mathcal{E} = \mathfrak{R} \setminus \{0\}$. Note that

one could also use a multi-dimensional mark space without any complication. We define on $\mathcal{E} \times [0, T]$ an \mathcal{A} -adapted Poisson measure $p_\varphi(dv, dt)$, with intensity measure $\nu_\varphi(dv, dt) = \varphi(dv)dt$, $T \in [0, \infty)$. To keep our notation transparent we assume at this stage that the total intensity

$$\lambda = \varphi(\mathcal{E}) < \infty \quad (1.8.1)$$

of p_φ is finite. This is also the case that one can numerically handle if one focuses on single events. Thus, $p_\varphi = \{p_\varphi(t) = p_\varphi(\mathcal{E}, [0, t]), t \in [0, T]\}$ is a stochastic process that counts the number of jumps occurring in the time interval $[0, T]$. The Poisson random measure $p_\varphi(dv, dt)$ generates a sequence of pairs $\{(\tau_i, \xi_i), i \in \{1, 2, \dots, p_\varphi(T)\}\}$, where $\{\tau_i \in [0, T], i \in \{1, 2, \dots, p_\varphi(T)\}\}$ is a sequence of increasing nonnegative random variables representing the jump times of a standard Poisson process with intensity λ , and $\{\xi_i \in \mathcal{E}, i \in \{1, 2, \dots, p_\varphi(T)\}\}$ is a sequence of independent, identically distributed random variables. Here ξ_i is distributed according to $\frac{\varphi(dv)}{\lambda} = F(dv)$. Therefore, we call $F(\cdot)$ the distribution function of the marks. We can interpret τ_i as the time of the i th event and the mark ξ_i as its amplitude. For a more general presentation of random measures we refer to [Elliott \(1982\)](#).

We consider the d -dimensional SDE with jumps

$$d\mathbf{X}_t = \mathbf{a}(t, \mathbf{X}_t) dt + \mathbf{b}(t, \mathbf{X}_t) d\mathbf{W}_t + \int_{\mathcal{E}} \mathbf{c}(t, \mathbf{X}_{t-}, v) p_\varphi(dv, dt), \quad (1.8.2)$$

for $t \in [0, T]$, with initial value $\mathbf{X}_0 \in \Re^d$, an \mathcal{A} -adapted m -dimensional Wiener process $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m), t \in [0, T]\}$ and the previously introduced Poisson random measure p_φ . Note that in (1.8.2) we denote by \mathbf{X}_{t-} the almost sure left-hand limit of \mathbf{X} at time t . A solution of an SDE of the type (1.8.2) is called a jump diffusion or an Itô process with jumps.

The coefficients $\mathbf{a}(t, x)$ and $\mathbf{c}(t, x, v)$ are d -dimensional vectors of Borel measurable real valued functions on $[0, T] \times \Re^d$ and on $[0, T] \times \Re^d \times \mathcal{E}$, respectively. Additionally, $\mathbf{b}(t, x)$ is a $d \times m$ -matrix of Borel measurable real valued functions on $[0, T] \times \Re^d$. As before, for a given vector \mathbf{a} we adopt the notation a^i to denote its i th component. Similarly with $b^{i,j}$ we will denote the component of the i th row and j th column of a given matrix \mathbf{b} .

The SDE (1.8.2) is only a short hand notation for its integral form

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t \mathbf{a}(s, \mathbf{X}_s) ds + \int_0^t \mathbf{b}(s, \mathbf{X}_s) d\mathbf{W}_s + \sum_{i=1}^{p_\varphi(t)} \mathbf{c}(\tau_i, \mathbf{X}_{\tau_i-}, \xi_i), \quad (1.8.3)$$

where $\{(\tau_i, \xi_i), i \in \{1, 2, \dots, p_\varphi(t)\}\}$ is the above described sequence of pairs of jump times and corresponding marks generated by the Poisson random measure p_φ . In this book we adopt the convention that the summation $\sum_{i=j_1}^{j_2} c_i$ yields zero for all possible summands c_i if $j_1 > j_2$. Therefore, if there are no jumps up to time t , which means that $p_\varphi(t) = 0$, the last term in (1.8.3) vanishes.

In the SDE (1.8.2) we have defined the impact of a jump via an Itô stochastic integral with respect to a Poisson random measure as

$$\int_0^t \int_{\mathcal{E}} c(s, \mathbf{X}_{s-}, v) p_\varphi(dv, ds). \quad (1.8.4)$$

This stochastic integral allows us to model rather general jump behavior. The only restriction we impose on the jump component is the finiteness of the total intensity as requested in condition (1.8.1). By using a Poisson measure there is no real need for relying on this condition other than being able to deal with finite sums as in (1.8.3).

A Special Case

Often we will consider in our discussions on numerical schemes the special case with coefficient functions $a(t, x) = \mu x$, $b(t, x) = \sigma x$ and $c(t, x, v) = x v$. Then the SDE (1.8.2) reduces to

$$dX_t = X_{t-} \left(\mu dt + \sigma dW_t + \int_{\mathcal{E}} v p_\varphi(dv, dt) \right), \quad (1.8.5)$$

for $t \in [0, T]$ and admits, according to (1.7.48), the explicit solution

$$X_t = X_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t} \prod_{i=1}^{p_\varphi(t)} (\xi_i + 1), \quad (1.8.6)$$

where the marks $\xi_i \geq -1$ are distributed according to a given probability measure $F(dv) = \frac{\varphi(dv)}{\lambda}$. When we choose a lognormal probability measure, which means that $\zeta_i = \ln(\xi_i + 1)$ is an independent Gaussian random variable, $\zeta_i \sim N(\varrho, \varsigma)$, with mean ϱ and variance ς , then equation (1.8.6) represents a specification of the *Merton model*, see (1.7.46).

Specifications of Jump Integrals

The flexibility provided in the definition of the jump integral (1.8.4) is illustrated in the following examples:

One can construct several independent driving Poisson processes by splitting the mark space \mathcal{E} into corresponding disjoint subsets $\mathcal{E}_1, \dots, \mathcal{E}_r$, for $r \in \mathcal{N}$, and obtaining with $P^j = \{P_t^j = p_\varphi(\mathcal{E}_j, [0, t]), t \in [0, T]\}$ the j th Poisson process having intensity $\lambda^j = \varphi(\mathcal{E}_j)$.

It is also possible to specify a jump component with time-dependent intensity by choosing

$$c(t, x, v) = \mathbf{1}_{\{v \in [\eta_1(t), \eta_2(t)]\}} f(t, x, v), \quad (1.8.7)$$

where η_1 and η_2 are deterministic functions of time and $\mathbf{1}_{\{\cdot\}}$ is the indicator function defined by

$$\mathbf{1}_{\{v \in A\}} = \begin{cases} 1 & \text{for } v \in A \\ 0 & \text{for } v \notin A \end{cases} \quad (1.8.8)$$

for a given set $A \subseteq \mathcal{E}$. Then we obtain a jump integral of the form

$$\int_0^t \int_{\eta_1(s)}^{\eta_2(s)} f(s, X_{s-}, v) p_\varphi(dv, ds).$$

Moreover, if we allow the functions η_1 and η_2 to be time and state-dependent, then via the SDE (1.8.2) we obtain a jump integral of the form

$$\int_0^t \int_{\eta_1(s, X_{s-})}^{\eta_2(s, X_{s-})} f(s, X_{s-}, v) p_\varphi(dv, ds),$$

which allows the modeling of the effect of a stochastic intensity. Note that in this case the coefficient function c is, in general, discontinuous in the state variable and, therefore, could not be Lipschitz. The above specifications can also be used for advanced credit risk models with multiple obligors and correlated intensities, as discussed in [Schönbucher \(2003\)](#). In the fast growing energy markets, quantities such as electricity prices are often described by rather complex jump diffusion SDEs of the type (1.8.2); see, for instance, [Geman & Roncoroni \(2006\)](#) for a class of jump diffusion models that capture the typical mean reversion feature of electricity prices.

Other important examples of jump diffusions of the form (1.8.2) arise in the pricing and hedging of complex interest rate derivatives in a jump diffusion term structure environment. We refer to [Björk et al. \(1997\)](#) for the Heath-Jarrow-Morton framework with jumps and to [Glasserman & Kou \(2003\)](#) for a LIBOR market model with jumps, see also Bruti-Liberati, Nikitopoulos-Sklibosios & Platen ([\(2009\)](#)) for term structure models with jumps under the benchmark approach.

For example, let us mention a specific LIBOR market model with jumps proposed in [Samuelides & Nahum \(2001\)](#) for pricing short-term interest rate derivatives. Given a set of equidistant tenor dates T_1, \dots, T_{d+1} , with $T_{i+1} - T_i = \delta$ for $i \in \{1, \dots, d\}$, the components of the vector $\mathbf{X}_t = (X_t^1, \dots, X_t^d)^\top$ represent discrete compounded forward rates at time t , maturing at tenor dates T_1, \dots, T_d , respectively. In this case the model is driven by one Wiener process, $m = 1$, and two Poisson processes, $r = 2$. The diffusion coefficient is specified as $b(t, x) = \sigma x$, with σ a d -dimensional vector of positive numbers, and the jump coefficient is given by $c(t, x) = \beta x$, where β is a $d \times 2$ -matrix with $\beta^{i,1} > 0$ and $\beta^{i,2} < 0$, for $i \in \{1, \dots, d\}$. In this way the first jump process generates upward jumps, while the second one creates downward jumps. Moreover, the marks are set to $\xi_i = 1$ so that the two driving jump processes are standard Poisson processes. A classical no-arbitrage restriction on the evolution of forward rates under the T_{d+1} -forward measure, see [Björk et al. \(1997\)](#) and [Glasserman & Kou \(2003\)](#), imposes a particular nonlinear form on the drift coefficient $a(t, x)$. Its i th component is given by

$$\begin{aligned} a^i(t, x^1, \dots, x^d) = & - \left(\sum_{j=i+1}^d \frac{\delta x^j}{1 + \delta x^j} \sigma^j + \lambda^1 \prod_{j=i+1}^d \left(1 + \beta^{j,1} \frac{\delta x^j}{1 + \delta x^j} \right) \right. \\ & \left. + \lambda^2 \prod_{j=i+1}^d \left(1 + \beta^{j,2} \frac{\delta x^j}{1 + \delta x^j} \right) \right), \end{aligned} \quad (1.8.9)$$

for $i \in \{1, \dots, d\}$, where λ^1 and λ^2 denote the intensities of the two Poisson processes. A complex nonlinear drift coefficient, as the one in (1.8.9), is a typical feature of LIBOR market models. This makes the application of numerical techniques essential for the pricing and hedging of complex interest rate derivatives. We refer to [Glasserman & Merener \(2003b\)](#) for numerical approximations of jump diffusion LIBOR market models and to Chap. 3 for more details on models with jumps.

1.9 Existence and Uniqueness of Solutions of SDEs

Strong Solution

For any model that uses an SDE it is essential that it has a solution. Furthermore, it is important that it has a unique solution according to some appropriate criterion. One such criterion is described below in detail, which is based on a notion of strong uniqueness. Usually one can only formulate sufficient conditions to establish uniqueness. The techniques presented in the literature for proving existence and uniqueness of a solution of an SDE are rather similar. They typically assume Lipschitz continuity of the drift, diffusion and jump coefficients. We aim to provide here some insight into typical issues that arise when ensuring the existence and uniqueness of a solution of an SDE.

Assume that we have given a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. For the equation (1.8.3) to make sense \mathbf{X} needs to be $\underline{\mathcal{A}}$ -adapted. This leads to the following definition.

Definition 1.9.1. *We call a triplet $(\mathbf{X}, \mathbf{W}, p_\varphi)$, consisting of a stochastic process $\mathbf{X} = \{\mathbf{X}_t, t \in [0, T]\}$, an $\underline{\mathcal{A}}$ -adapted standard Wiener process \mathbf{W} and an $\underline{\mathcal{A}}$ -adapted Poisson measure p_φ , a strong solution of the Itô integral equation (1.8.3) if \mathbf{X} is $\underline{\mathcal{A}}$ -adapted, the integrals on the right hand side are well-defined and the equality in (1.8.3) holds almost surely.*

For fixed coefficient functions \mathbf{a} and \mathbf{b} , any solution \mathbf{X} will usually depend on the particular initial value \mathbf{X}_0 and the sample path of the Wiener process \mathbf{W} and Poisson measure p_φ under consideration. For a specified initial value \mathbf{X}_0 the *uniqueness* of strong solutions of the SDE (1.8.3) refers to the following notion of *indistinguishability* of the solution processes.

Definition 1.9.2. If any two strong solutions \mathbf{X} and $\tilde{\mathbf{X}}$ are indistinguishable on $[0, T]$, that is if

$$\mathbf{X}_t = \tilde{\mathbf{X}}_t \quad (1.9.1)$$

almost surely for all $t \in [0, T]$, then we say that the solution of (1.8.3) on $[0, T]$ is a unique strong solution.

Existence and Uniqueness Theorem

Let us now state a standard theorem on the existence and uniqueness of strong solutions of SDEs with jumps. This ensures that the objects we model are well defined. For details on the definition of strong solutions of jump diffusion SDEs, we can refer, for instance, to [Ikeda & Watanabe \(1989\)](#) or [Protter \(2005\)](#).

We assume that the coefficient functions of the SDE (1.8.2) satisfy the Lipschitz conditions

$$\begin{aligned} |\mathbf{a}(t, \mathbf{x}) - \mathbf{a}(t, \mathbf{y})| &\leq C_1 |\mathbf{x} - \mathbf{y}|, \quad |\mathbf{b}(t, \mathbf{x}) - \mathbf{b}(t, \mathbf{y})| \leq C_2 |\mathbf{x} - \mathbf{y}|, \\ \int_{\mathcal{E}} |\mathbf{c}(t, \mathbf{x}, v) - \mathbf{c}(t, \mathbf{y}, v)|^2 \varphi(dv) &\leq C_3 |\mathbf{x} - \mathbf{y}|^2, \end{aligned} \quad (1.9.2)$$

for every $t \in [0, T]$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, as well as the linear growth conditions

$$\begin{aligned} |\mathbf{a}(t, \mathbf{x})| &\leq K_1(1 + |\mathbf{x}|), \quad |\mathbf{b}(t, \mathbf{x})| \leq K_2(1 + |\mathbf{x}|), \\ \int_{\mathcal{E}} |\mathbf{c}(t, \mathbf{x}, v)|^2 \varphi(dv) &\leq K_3(1 + |\mathbf{x}|^2), \end{aligned} \quad (1.9.3)$$

for all $t \in [0, T]$ and $\mathbf{x} \in \mathbb{R}^d$. Note that the linear growth conditions can usually be derived from the corresponding Lipschitz conditions.

Moreover, we assume that the initial value \mathbf{X}_0 is \mathcal{A}_0 -measurable with

$$E(|\mathbf{X}_0|^2) < \infty. \quad (1.9.4)$$

Theorem 1.9.3. Suppose that the coefficient functions $\mathbf{a}(\cdot)$, $\mathbf{b}(\cdot)$ and $\mathbf{c}(\cdot)$ of the SDE (1.8.2) satisfy the Lipschitz conditions (1.9.2), the linear growth conditions (1.9.3) and the initial condition (1.9.4). Then the SDE (1.8.2) admits a unique strong solution. Moreover, the solution \mathbf{X}_t of the SDE (1.8.2) satisfies the estimate

$$E \left(\sup_{0 \leq s \leq T} |\mathbf{X}_s|^2 \right) \leq C (1 + E(|\mathbf{X}_0|^2)) \quad (1.9.5)$$

with $T < \infty$, where C is a finite positive constant.

The proof of Theorem 1.9.3 can be found in [Ikeda & Watanabe \(1989\)](#) or [Situ \(2005\)](#).

For certain applications, the Lipschitz condition (1.9.2) on the jump coefficient c is too restrictive. For instance, for modeling state-dependent intensities, as discussed in Sect. 1.8, it is convenient to use jump coefficients that are not Lipschitz continuous. Athreya, Kliemann & Koch (1988) provide some results on the existence and uniqueness of the solution of the SDE (1.8.2) when the jump coefficient c is not Lipschitz continuous. The Yamada condition, see Ikeda & Watanabe (1989), provides another condition for SDEs that do not satisfy Lipschitz type conditions. In Cherny (2000) it is shown how the uniqueness of strong solutions of SDEs may break down when Lipschitz conditions are violated. However, if one specifies at certain boundaries the behavior of the solution of an SDE it can typically be made unique. In the mentioned literature one finds also the notion of a weak solution of an SDE. We note that if an SDE has a strong solution, then it has also a weak solution. It is beyond the scope of this book to go deeper into these issues.

Moment Estimates

It is convenient to have the following moment estimates for SDEs with jumps when estimating errors in discrete-time approximations.

Theorem 1.9.4. *Suppose that the coefficient functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ of the SDE (1.8.2) satisfy the Lipschitz conditions (1.9.2) and the linear growth conditions (1.9.3). Moreover, let for $n \in \{1, 2, \dots\}$*

$$E(|X_0|^{2n}) < \infty. \quad (1.9.6)$$

Then the solution X_t of (1.8.2) satisfies

$$E \left(\sup_{0 \leq s \leq T} |X_s|^{2n} \right) \leq C \left(1 + E(|X_0|^{2n}) \right) \quad (1.9.7)$$

for $t \in [0, T]$ with $T < \infty$, where C is a positive constant depending only on T , n and the linear growth bound.

The proof of this result follows from Ikeda & Watanabe (1989) or Protter (2005).

1.10 Exercises

1.1. Determine for a geometric Brownian motion $Z_t = Z_0 \exp\{\mu t + \sigma W_t\}$ the Itô differential for Z_t and $\ln(Z_t)$ by the use of the Itô formula, where W is a standard Wiener process.

1.2. Consider two transformed Wiener processes with $Y_t^1 = a_1 t + b_1 W_t^1$ and $Y_t^2 = a_2 t + b_2 W_t^2$, where W^1 and W^2 are two independent standard Wiener processes. What is the Itô differential for $Y_t^1 Y_t^2$?

1.3. Calculate the covariation between a standard Wiener process and its square.

1.4. For a process $X = \{X_t, t \in [0, \infty)\}$ with $X_t = \sigma W_t + \xi N_t$, where W is a standard Wiener process and N a Poisson process with intensity $\lambda > 0$, characterize the stochastic differential of its exponential when $\sigma, \xi > 0$.

1.5. For the sum $X_t = a N_t^1 + b N_t^2$, where N^1 and N^2 are two independent Poisson processes with intensity $\lambda > 0$, compute the stochastic differential of the exponential.

1.6. Compute the mean and variance as functions of time of the standard Ornstein-Uhlenbeck process with initial value $X_{t_0} = 1$.

1.7. Solve explicitly the scalar SDE

$$dX_t = -\frac{1}{2} X_t dt + X_t dW_t^1 + X_t dW_t^2,$$

where W^1 and W^2 are independent standard Wiener processes.

1.8. Let $N = \{N_t, t \in [0, \infty)\}$ be a Poisson process with intensity $\lambda > 0$ and W a standard Wiener process. For the process $Z = \{Z_t, t \in [0, \infty)\}$ with

$$Z_t = a t + b W_t + c N_t$$

compute the SDE for $X = \{X_t = \exp\{k Z_t\}, t \in [0, \infty)\}$, where $a, b, c, k > 0$.

1.9. For the process X in Exercise 1.8 compute the expectation

$$\mu(t) = E(X_t).$$

1.10. For a Poisson process with intensity $\lambda > 0$ derive a formula for its second moment at time $t > 0$.

1.11. Compute the first moment for a compound Poisson process with intensity $\lambda > 0$ and $U(0, 1)$ distributed i.i.d. marks.

1.12. What is the probability for a compound Poisson process with intensity $\lambda > 0$ and $U(0, 1)$ distributed i.i.d. marks to have no jump until time $t > 0$?

Exact Simulation of Solutions of SDEs

Accurate scenario simulation methods for solutions of multi-dimensional stochastic differential equations find applications in the statistics of stochastic processes and many applied areas, in particular in finance. They play a crucial role when used in standard models in various areas. These models often dominate the communication and thinking in a particular field of application, even though they may be too simple for advanced tasks. Various simulation techniques have been developed over the years. However, the simulation of solutions of some stochastic differential equations can still be problematic. Therefore, it is valuable to identify multi-dimensional stochastic differential equations with solutions that can be simulated exactly. This avoids several of the theoretical and practical problems of those simulation methods that use discrete-time approximations. This chapter follows closely [Platen & Rendek \(2009a\)](#) and provides methods for the exact simulation of paths of multi-dimensional solutions of stochastic differential equations, including Ornstein-Uhlenbeck, square root, squared Bessel, Wishart and Lévy type processes. Other papers that could be considered to be related with exact simulation include [Lewis & Shedler \(1979\)](#), [Beskos & Roberts \(2005\)](#), [Broadie & Kaya \(2006\)](#), [Kahl & Jäckel \(2006\)](#), [Smith \(2007\)](#), [Andersen \(2008\)](#), [Burq & Jones \(2008\)](#) and [Chen \(2008\)](#).

2.1 Motivation of Exact Simulation

Avoiding any error in the simulation of the path of a given process can only be achieved in exceptional cases. However, when it is possible it makes the numerical results very accurate and reliable.

Accurate scenario simulation of solutions of SDEs is widely applicable in stochastic analysis itself and many applied areas, in particular, in quantitative finance and for dynamic financial analysis in insurance, see [Kaufmann, Gadmer & Klett \(2001\)](#). Monographs in this direction include, for instance, [Kloeden & Platen \(1999\)](#), [Kloeden, Platen & Schurz \(2003\)](#), [Milstein](#)

(1995a), Jäckel (2002) and Glasserman (2004). Many discrete-time simulation methods have been developed over the years. However, some SDEs can be problematic in terms of discrete-time approximation via simulation. Therefore, it is necessary to understand and avoid the problems that may arise during the simulation of solutions of such SDEs. For illustration, let us consider a family of SDEs of the form

$$dX_t = a(X_t)dt + \sqrt{2X_t}dW_t \quad (2.1.1)$$

with some given drift coefficient function $a(x)$. Note that the diffusion coefficient function $f(x) = \sqrt{2x}$ is here non-Lipschitz. Its derivative becomes infinite as x tends to 0. The standard convergence theorems, derived in the previously mentioned literature and presented later in this book, do not easily cover such cases. It is, therefore, of interest to identify approximate simulation methods for various types of nonlinear SDEs and also for multi-dimensional SDEs. We will emphasize the fact that the problem of non-Lipschitz coefficients is circumvented for SDEs where we can simulate exact solutions. For squared Bessel processes of integer dimension, see Revuz & Yor (1999) and Platen & Heath (2006), we will explain how to simulate such solutions. Exact solutions can be simulated for a range of diffusion processes by sampling from their explicitly available transition density for some special cases of nonlinear SDEs where the drift function $a(\cdot)$ in (2.1.1) takes a particular form, see Craddock & Platen (2004). These will also include squared Bessel processes of noninteger dimensions, see Sect. 2.2.

Another problem with the simulation of SDEs may be the lack of sufficient numerical stability of the chosen scheme. As we will discuss later in Chap. 14, numerical stability is understood as the ability of a scheme to control the propagation of initial and roundoff errors. Numerical stability may be lost for some parameter ranges of a given SDE when using certain simulation schemes with a particular time step size. The issue of numerical stability can be circumvented when it is possible to simulate exact solutions.

Moreover, for theoretically strictly positive processes it is often not sufficient to use simulation methods that may generate negative values. This problem however can, in some cases, be solved by a transformation of the initial SDE, by use of the Itô formula, to a process which lives on the entire real axis. This is, in particular, useful for geometric Brownian motion, the dynamics of the Black-Scholes model, where one can take the logarithm to obtain a linearly transformed Wiener process. One may try such an approach to transform the *square root process* of the form

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t, \quad (2.1.2)$$

where $t \in \mathbb{R}^+$. This process remains strictly positive for dimension $\delta = \frac{4\kappa\theta}{\sigma^2} > 2$. Suppose that we simulate for $\delta > 2$ the process $Y_t = \sqrt{X_t}$ using a standard explicit numerical scheme such as the Euler scheme, see Kloeden & Platen (1999). The SDE of the corresponding stochastic process $Y = \{Y_t = \sqrt{X_t}, t \geq 0\}$ has then additive noise and is given by

$$dY_t = \left(\frac{\kappa\theta - \sigma^2/4}{2Y_t} - \frac{\kappa}{2}Y_t \right) dt + \frac{\sigma}{2}dW_t. \quad (2.1.3)$$

Theoretically, by squaring the resulting trajectory of Y we should obtain an approximate trajectory of the square root process X . However, note that the drift coefficient $\frac{\kappa\theta - \sigma^2/4}{2y} - \frac{\kappa}{2}y$ is non-Lipschitz and may almost explode for small y . Even though we have additive noise this feature will most likely produce simulation problems near zero. This kind of problem becomes even more serious for small dimension $\delta < 2$ of the square root process. It would be very valuable to have an exact solution, which avoids this kind of problems.

After the Wiener process and its direct transformations, including geometric Brownian motion and the Ornstein-Uhlenbeck process, the family of square root and squared Bessel processes are probably the most frequently used diffusion models in applications. In general, it is a challenging task to obtain, efficiently, a reasonably accurate trajectory of a square root process using simulation, as is documented in an increasing literature on this topic. Here we refer to the work of [Deelstra & Delbaen \(1998\)](#), [Diop \(2003\)](#), [Bossy & Diop \(2004\)](#), [Berkaoui, Bossy & Diop \(2005\)](#), [Alfonsi \(2005\)](#), [Broadie & Kaya \(2006\)](#), [Lord, Koekkoek & van Dijk \(2006\)](#), [Smith \(2007\)](#) and [Andersen \(2008\)](#). We will also study the simulation of multi-dimensional square root and squared Bessel processes below.

In various areas of application of stochastic analysis one has to model vectors or even matrices of evolving dependent stochastic quantities. This is typically the case, for instance, when modeling related asset prices in a financial market. All the above mentioned numerical problems can arise in a complex manner when simulating the trajectories of such multi-dimensional models. For instance, different time scales in the dynamics of certain components can create stiff SDEs in the sense of [Kloeden & Platen \(1999\)](#), which are almost impossible to handle by standard discrete-time schemes. This makes it worthwhile to identify classes of multi-dimensional SDEs with exact solutions. In addition, we will see that almost exact approximations will be of particular interest.

2.2 Sampling from Transition Distributions

In the following we will consider some multi-dimensional diffusion processes with explicitly known transition distributions. Since, it is rare that one has an exact formula for a transition distribution, these examples are of particular interest. We will show how to use explicitly available multivariate transition distributions for the simulation of exact solutions of SDEs. In fact, for certain multi-dimensional diffusions, given by some system of SDEs, one may need extra information about their behavior at finite boundaries to obtain a complete description of the modeled dynamics. The description of the diffusions via transition distributions contains this information. One needs to keep this

in mind since most of the literature is primarily concerned with the modeling via SDEs.

Inverse Transform Method

The following well-known *inverse transform method* can be applied for the generation of a continuous random variable Y with given probability distribution function F_Y . From a uniformly distributed random variable $0 < U < 1$, we obtain an F_Y distributed random variable $y(U)$ by realizing that

$$U = F_Y(y(U)), \quad (2.2.1)$$

so that

$$y(U) = F_Y^{-1}(U). \quad (2.2.2)$$

Here F_Y^{-1} denotes the inverse function of F_Y . More generally, one can still set

$$y(U) = \inf\{y : U \leq F_Y(y)\} \quad (2.2.3)$$

in the case when F_Y is no longer continuous, where $\inf\{y : U \leq F_Y(y)\}$ denotes the lower limit of the set $\{y : U \leq F_Y(y)\}$. If U is a $U(0, 1)$ random variable, then the random variable $y(U)$ in (2.2.2) will be F_Y -distributed. The above calculation in (2.2.2) may need to apply a root finding method, for instance, a Newton method, see [Press, Teukolsky, Vetterling & Flannery \(2002\)](#). Obviously, given an explicit transition distribution function for the solution of a one-dimensional SDE we can sample a trajectory directly from this transition law at given time instants. One simply starts with the initial value, generates the first increment and sequentially the subsequent random increments of the simulated trajectory, using the inverse transform method for the respective transition distributions that emerge.

Also in the case of a two-dimensional SDE we can simulate by sampling from the bivariate transition distribution. We first identify the marginal transition distribution function F_{Y_1} of the first component. Then we use the inverse transform method, as above, for the exact simulation of an outcome of the first component of the two-dimensional random variable based on its marginal distribution function. Afterwards, we exploit the conditional transition distribution function $F_{Y_2|Y_1}$ of the second component Y_2 , given the simulated first component Y_1 , and use again the inverse transform method to simulate also the second component of the considered SDE. This simulation method is exact as long as the root finding procedure involved can be interpreted as being exact. It exploits a well-known basic result on multi-variate distribution functions, see for instance [Rao \(1973\)](#).

It is obvious that this simulation technique can be generalized to the exact simulation of increments of solutions of some d -dimensional SDEs. Based on a given d -variate transition distribution function one needs to find the marginal distribution F_{Y_1} and the conditional distributions $F_{Y_2|Y_1}, F_{Y_3|Y_1, Y_2},$

$\dots, F_{Y_d|Y_1, Y_2, \dots, Y_{d-1}}$. Then the inverse transform method can be applied to each conditional transition distribution function one after the other. This also shows that it is sufficient to characterize explicitly in a model just the marginal and conditional transition distribution functions.

Note also that nonparametrically described transition distribution functions are sufficient for application of the inverse transform method. Of course, explicitly known parametric distributions are preferable for a number of practical reasons. They certainly reduce the complexity of the problem itself by splitting it into a sequence of problems. Further, we will list below important examples of explicitly known multi-variate transition densities and distribution functions.

Copulas

Each multi-variate distribution function has its, so called *copula*, which characterizes the dependence structure between the components. Roughly speaking, the copula is the joint density of the components when they are each transformed into $U(0, 1)$ distributed random variables. Essentially, every multi-variate distribution has a corresponding copula. Conversely, each copula can be used together with some given marginal distributions to obtain a corresponding multi-variate distribution function. This is a consequence of Sklar's theorem, see for instance McNeil et al. (2005).

If, for instance, $\mathbf{Y} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Omega})$ is a Gaussian random vector, then the copula of \mathbf{Y} is the same as the copula of $\mathbf{X} \sim N_d(\mathbf{0}, \boldsymbol{\Omega})$, where $\mathbf{0}$ is the zero vector and $\boldsymbol{\Omega}$ is the correlation matrix of \mathbf{Y} . By the definition of the d -dimensional Gaussian copula we obtain

$$C_{\boldsymbol{\Omega}}^{Ga} = P(N(X_1) \leq u_1, \dots, N(X_d) \leq u_d) = N_{\boldsymbol{\Omega}}(N^{-1}(u_1), \dots, N^{-1}(u_d)), \quad (2.2.4)$$

where N denotes the standard univariate normal distribution function and $N_{\boldsymbol{\Omega}}$ denotes the joint distribution function of \mathbf{X} . Hence, in two dimensions we obtain

$$\begin{aligned} C_{\boldsymbol{\Omega}}^{Ga}(u_1, u_2) &= \int_{-\infty}^{N^{-1}(u_1)} \int_{-\infty}^{N^{-1}(u_2)} \frac{1}{2\pi(1-\varrho^2)^{1/2}} \exp \left\{ \frac{-(s_1^2 - 2\varrho s_1 s_2 + s_2^2)}{2(1-\varrho^2)} \right\} \\ &\quad \times ds_1 ds_2, \end{aligned} \quad (2.2.5)$$

where $\varrho \in [-1, 1]$ is the correlation parameter in $\boldsymbol{\Omega}$.

Another example of a copula is the Clayton copula. This copula can be expressed in the d -dimensional case as

$$C_{\theta}^{Cl} = (u_1^{-\theta} + \dots + u_d^{-\theta} - d + 1)^{-1/\theta}, \quad \theta \geq 0, \quad (2.2.6)$$

where the limiting case $\theta = 0$ is the d -dimensional independence copula. Moreover, d -dimensional Archimedean copulas can be expressed in terms of

Laplace-Stieltjes transforms of distribution functions on \mathbb{R}^+ . If F is a distribution function on \mathbb{R}^+ satisfying $F(0) = 0$, then the Laplace-Stieltjes transform can be expressed by

$$\hat{F}(t) = \int_0^\infty e^{-tx} dF(x), \quad t \geq 0. \quad (2.2.7)$$

Using the Laplace-Stieltjes transform the d -dimensional Archimedian copula has the form

$$C^{Ar}(u_1, \dots, u_d) = E \left(\exp \left\{ -V \sum_{i=1}^d \hat{F}^{-1}(u_i) \right\} \right) \quad (2.2.8)$$

for strictly positive random variables V with Laplace-Stieltjes transform \hat{F} . A simulation method follows directly from this representation, see Marshall & Olkin (1988). More examples of multi-dimensional copulas can be found in McNeil, Frey & Embrechts (2005).

Transition Density of a Multi-dimensional Wiener Process

As an alternative to copulas one can express the dependence structure of the components of a stochastic process by its transition densities. Of course, for any given transition density there exists a corresponding copula. As a first example of a continuous multi-dimensional stochastic process, whose transition density can be expressed explicitly, we focus on the d -dimensional Wiener process, see Sect. 1.1. This fundamental stochastic process has a multivariate Gaussian transition density of the form

$$p(s, \mathbf{x}; t, \mathbf{y}) = \frac{1}{(2\pi(t-s))^{d/2} \sqrt{\det \boldsymbol{\Sigma}}} \exp \left\{ \frac{(\mathbf{y} - \mathbf{x})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{x})}{2(t-s)} \right\}, \quad (2.2.9)$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. Here $\boldsymbol{\Sigma}$ is a normalized covariance matrix. Its copula is the Gaussian copula (2.2.4), which is simply derived from the corresponding multi-variate Gaussian density. In the bivariate case with correlated Wiener processes this transition probability simplifies to

$$p(s, x_1, x_2; t, y_1, y_2) = \frac{1}{2\pi(t-s)\sqrt{1-\varrho^2}} \times \exp \left\{ -\frac{(y_1 - x_1)^2 - 2(y_1 - x_1)(y_2 - x_2)\varrho + (y_2 - x_2)^2}{2(t-s)(1-\varrho^2)} \right\}, \quad (2.2.10)$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $x_1, x_2, y_1, y_2 \in \mathbb{R}$. Here the correlation parameter ϱ varies in the interval $[-1, 1]$. In the case of correlated Wiener processes one can first simulate independent Wiener processes and then form out of these, by linear transforms, correlated ones. Alternatively, one can follow the

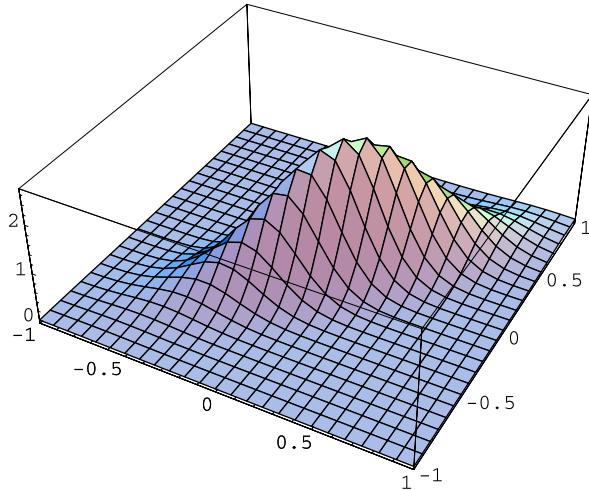


Fig. 2.2.1. Bivariate transition density of the two-dimensional Wiener process for fixed time step $\Delta = 0.1$, $x_1 = x_2 = 0.1$ and $\varrho = 0.8$

above inverse transform method by first using the Gaussian distribution for generating the increments of the first Wiener process component. Then one can condition the Gaussian distribution for the second component on these outcomes.

In Fig. 2.2.1 we illustrate the bivariate transition density of the two-dimensional Wiener process for the time increment $\Delta = t - s = 0.1$, initial values $x_1 = x_2 = 0.1$ and correlation $\varrho = 0.8$. One can also generate dependent Wiener processes that have a joint distribution with a given copula.

Transition Density of a Multi-dimensional Geometric Brownian Motion

The multi-dimensional geometric Brownian motion is a componentwise exponential of a linearly transformed Wiener process. Given a vector of correlated Wiener processes \mathbf{W} with the transition density (2.2.9) we consider the following transformation

$$\mathbf{S}_t = \mathbf{S}_0 \exp\{\mathbf{a}t + \mathbf{B}\mathbf{W}_t\}, \quad (2.2.11)$$

for $t \in [0, \infty)$, where the exponential is taken componentwise. Here \mathbf{a} is a vector of length d , while the elements of the matrix \mathbf{B} are as follows

$$B^{i,j} = \begin{cases} b^j & \text{for } i = j \\ 0 & \text{otherwise,} \end{cases} \quad (2.2.12)$$

where $i, j \in \{1, 2, \dots, d\}$.

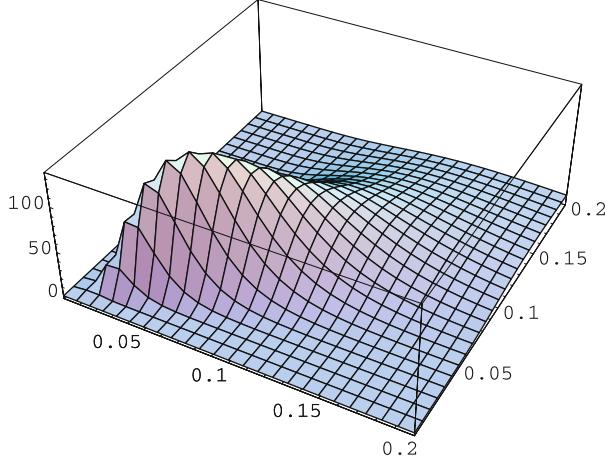


Fig. 2.2.2. Bivariate transition density of the two-dimensional geometric Brownian motion for $\Delta = 0.1$, $x_1 = x_2 = 0.1$, $\varrho = 0.8$, $b^1 = b^2 = 2$ and $a_1 = a_2 = 0.1$

Then the transition density of the above defined geometric Brownian motion has the following form

$$p(s, \mathbf{x}; t, \mathbf{y}) = \frac{1}{(2\pi(t-s))^{d/2} \sqrt{\det \boldsymbol{\Sigma}} \prod_{i=1}^d b^i y_i} \times \exp \left\{ -\frac{(\ln(\mathbf{y}) - \ln(\mathbf{x}) - \mathbf{a}(t-s))^{\top} \mathbf{B}^{-1} \boldsymbol{\Sigma}^{-1} \mathbf{B}^{-1} (\ln(\mathbf{y}) - \ln(\mathbf{x}) - \mathbf{a}(t-s))}{2(t-s)} \right\} \quad (2.2.13)$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^d$. Here the logarithm is understood componentwise. In the bivariate case this transition density takes the particular form

$$\begin{aligned} p(s, x_1, x_2; t, y_1, y_2) &= \frac{1}{2\pi(t-s)\sqrt{1-\varrho^2}b^1b^2y_1y_2} \\ &\times \exp \left\{ -\frac{(\ln(y_1) - \ln(x_1) - a^1(t-s))^2}{2(b^1)^2(t-s)(1-\varrho^2)} \right\} \\ &\times \exp \left\{ -\frac{(\ln(y_2) - \ln(x_2) - a^2(t-s))^2}{2(b^2)^2(t-s)(1-\varrho^2)} \right\} \\ &\times \exp \left\{ \frac{(\ln(y_1) - \ln(x_1) - a^1(t-s))(\ln(y_2) - \ln(x_2) - a^2(t-s))\varrho}{b^1b^2(t-s)(1-\varrho^2)} \right\}, \end{aligned}$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $x_1, x_2, y_1, y_2 \in \mathbb{R}^+$, where $\varrho \in [-1, 1]$.

In Fig. 2.2.2 we illustrate the bivariate transition density of the two-dimensional geometric Brownian motion for the time increment $\Delta = t - s = 0.1$, initial values $x_1 = x_2 = 0.1$, correlation $\varrho = 0.8$, volatilities $b^1 = b^2 = 2$ and growth parameters $a^1 = a^2 = 0.1$.

Transition Density of a Multi-dimensional OU-Process

Another example is the standard d -dimensional *Ornstein-Uhlenbeck* (OU)-process. This process has a Gaussian transition density of the form

$$\begin{aligned} p(s, \mathbf{x}; t, \mathbf{y}) &= \frac{1}{(2\pi(1 - e^{-2(t-s)}))^{d/2} \sqrt{\det \boldsymbol{\Sigma}}} \\ &\times \exp \left\{ -\frac{(\mathbf{y} - \mathbf{x}e^{-(t-s)})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{x}e^{-(t-s)})}{2(1 - e^{-2(t-s)})} \right\}, \end{aligned} \quad (2.2.14)$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $\mathbf{x}, \mathbf{y} \in \Re^d$, with mean $\mathbf{x}e^{-(t-s)}$ and covariance matrix $\boldsymbol{\Sigma}(1 - e^{-2(t-s)})$, $d \in \{1, 2, \dots\}$. In the bivariate case the transition density of the standard OU-process is expressed by

$$\begin{aligned} p(s, x_1, x_2; t, y_1, y_2) &= \frac{1}{2\pi(1 - e^{-2(t-s)}) \sqrt{1 - \varrho^2}} \\ &\times \exp \left\{ -\frac{(y_1 - x_1 e^{-(t-s)})^2 + (y_2 - x_2 e^{-(t-s)})^2}{2(1 - e^{-2(t-s)})(1 - \varrho^2)} \right\} \\ &\times \exp \left\{ \frac{(y_1 - x_1 e^{-(t-s)})(y_2 - x_2 e^{-(t-s)})\varrho}{(1 - e^{-2(t-s)})(1 - \varrho^2)} \right\}, \end{aligned} \quad (2.2.15)$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $x_1, x_2, y_1, y_2 \in \Re$, where $\varrho \in [-1, 1]$.

Transition Density of a Multi-dimensional Geometric OU-Process

The transition density of a d-dimensional *geometric OU-process* can be obtained from the transition density of the multi-dimensional OU-process by applying the exponential transformation. Therefore, it can be expressed as

$$\begin{aligned} p(s, \mathbf{x}; t, \mathbf{y}) &= \frac{1}{(2\pi(1 - e^{-2(t-s)}))^{d/2} \sqrt{\det \boldsymbol{\Sigma} \prod_{i=1}^d y_i}} \\ &\times \exp \left\{ -\frac{(\ln(\mathbf{y}) - \ln(\mathbf{x})e^{-(t-s)})^\top \boldsymbol{\Sigma}^{-1} (\ln(\mathbf{y}) - \ln(\mathbf{x})e^{-(t-s)})}{2(1 - e^{-2(t-s)})} \right\}, \end{aligned} \quad (2.2.16)$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $\mathbf{x}, \mathbf{y} \in \Re_+^d$, $d \in \{1, 2, \dots\}$. In the bivariate case the transition density of the multi-dimensional OU-process is of the form

$$\begin{aligned} p(s, x_1, x_2; t, y_1, y_2) &= \frac{1}{2\pi(1 - e^{-2(t-s)}) \sqrt{1 - \varrho^2} y_1 y_2} \\ &\times \exp \left\{ -\frac{(\ln(y_1) - \ln(x_1)e^{-(t-s)})^2 + (\ln(y_2) - \ln(x_2)e^{-(t-s)})^2}{2(1 - e^{-2(t-s)})(1 - \varrho^2)} \right\} \\ &\times \exp \left\{ \frac{(\ln(y_1) - \ln(x_1)e^{-(t-s)})(\ln(y_2) - \ln(x_2)e^{-(t-s)})\varrho}{(1 - e^{-2(t-s)})(1 - \varrho^2)} \right\}, \end{aligned} \quad (2.2.17)$$

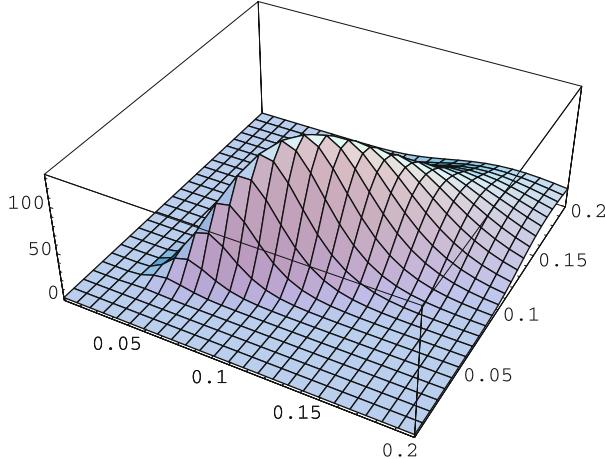


Fig. 2.2.3. Bivariate transition density of the two-dimensional geometric OU-process for $\Delta = 0.1$, $x_1 = x_2 = 0.1$ and $\varrho = 0.8$

for $t \in [0, \infty)$, $s \in [0, t]$ and $x_1, x_2, y_1, y_2 \in \mathfrak{R}_+$, where $\varrho \in [-1, 1]$.

In Fig. 2.2.3 we illustrate the bivariate transition density of the two-dimensional geometric OU-process for the time increment $\Delta = t - s = 0.1$, initial values $x_1 = x_2 = 0.1$ and correlation $\varrho = 0.8$. It is now obvious how to obtain the transition density of the componentwise exponential of other Gaussian vector processes.

Transition Density of a Wishart Process

Let us now study another class of processes that is related to products of Wiener processes. We characterize below the transition density of the matrix valued *Wishart process* for dimension parameter $\delta > 0$, starting at the time $s \in [0, \infty)$, in $\mathbf{X} > 0$ and being at time $t \in (s, \infty)$ in \mathbf{Y} , see [Bru \(1991\)](#) and [Gouriéroux & Sufana \(2004\)](#). Its transition density has the form

$$\begin{aligned}
p_\delta(s, \mathbf{X}; t, \mathbf{Y}) &= \frac{1}{(2(t-s))^{\delta m/2} \Gamma_m(\delta/2)} \text{etr} \left\{ -\frac{\mathbf{X} + \mathbf{Y}}{2(t-s)} \right\} (\det \mathbf{Y})^{(\delta-m-1)/2} \\
&\quad \times {}_0F_1 \left(\frac{\delta}{2}; \frac{\mathbf{XY}}{4(t-s)^2} \right) \\
&= \frac{1}{(2(t-s))^{m(m+1)/2}} \left(\frac{\det(\mathbf{Y})}{\det(\mathbf{X})} \right)^{\frac{\delta-m-1}{4}} \text{etr} \left\{ -\frac{\mathbf{X} + \mathbf{Y}}{2(t-s)} \right\} \\
&\quad \times \tilde{\mathbf{I}}_{(\delta-m-1)/2} \left(\frac{\mathbf{XY}}{4(t-s)^2} \right), \tag{2.2.18}
\end{aligned}$$

see [Donati-Martin, Doumerc, Matsumoto & Yor \(2004\)](#). Here $\text{etr}\{\cdot\}$ denotes the elementwise exponential of the trace of a matrix and $\tilde{\mathbf{I}}_\nu$ is a special function of a matrix argument with index ν defined by

$$\tilde{\mathbf{I}}_\nu(\mathbf{z}) = \frac{(\det \mathbf{z})^{\nu/2}}{\Gamma_m((m+1)/2 + \nu)} {}_0F_1((m+1)/2 + \nu, \mathbf{z}). \quad (2.2.19)$$

It is related to the modified Bessel function of the first kind $I_\nu(\cdot)$ by the relation $\tilde{\mathbf{I}}_\nu(\mathbf{z}) = I_\nu(2\mathbf{z}^{1/2})$. In general, the hypergeometric function ${}_pF_q$ in (2.2.19) can be expressed in terms of zonal polynomials, see [Muirhead \(1982\)](#). Finally, $\Gamma_m(\cdot)$ denotes the multi-dimensional gamma function with

$$\Gamma_m(x) = \int_{\Lambda > 0} \text{etr}\{-\Lambda\} (\det(\Lambda))^{x - \frac{m+1}{2}} d\Lambda. \quad (2.2.20)$$

The transition density of the Wishart process, when it starts at time zero at $\mathbf{X} = \mathbf{0}$ for being at time $t \in (0, \infty)$ in $\mathbf{Y} \geq \mathbf{0}$, can be written as

$$p_\delta(0, \mathbf{0}; t, \mathbf{Y}) = (2t)^{-\delta m/2} \frac{\det(\mathbf{Y})^{(\delta-m-1)/2}}{\Gamma_m(\frac{\delta}{2})} \text{etr}\left\{-\frac{\mathbf{Y}}{2t}\right\}. \quad (2.2.21)$$

[Herz \(1955\)](#) derived a representation of the non-central Wishart density in terms of a Bessel function of matrix argument $A_\nu^{(m)}$. The advantage of this representation is that it also accounts for correlation. It has the form

$$\begin{aligned} p_\delta(s, \mathbf{X}; t, \mathbf{Y}) &= \frac{1}{(2(t-s)\det(\boldsymbol{\Sigma}))^{\delta/2}} \\ &\times \text{etr}\left\{-\frac{\boldsymbol{\Sigma}^{-1}(\mathbf{X} + \mathbf{Y})}{2(t-s)}\right\} \det(\mathbf{Y})^{(\delta-m-1)/2} A_{(\delta-m-1)/2}^{(m)}\left(-\frac{\boldsymbol{\Sigma}^{-1}\mathbf{Y}\boldsymbol{\Sigma}^{-1}\mathbf{X}}{4(t-s)^2}\right), \end{aligned} \quad (2.2.22)$$

where $\boldsymbol{\Sigma}$ is a normalized covariance matrix. [Herz \(1955\)](#) provided also the following representation for the special function

$$A_\nu^{(2)}(\mathbf{z}) = \frac{1}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{1}{j! \Gamma(\nu + j + 1)} A_{\nu+2j+1/2}^{(1)}(\text{tr}(\mathbf{z})) \det(\mathbf{z})^j, \quad (2.2.23)$$

where $A_\nu^{(1)}(z) = \sum_{j=0}^{\infty} (-z)^j / (j! \Gamma(\nu + j + 1))$. Here $\Gamma(\cdot)$ denotes the well-known gamma function.

Hence, with the use of (2.2.22) and (2.2.23) we obtain for $m = 2$ the following transition density for the 2×2 Wishart process of dimension $\delta = 2$:

$$\begin{aligned} p(s, x_{11}, x_{12}, x_{21}, x_{22}; t, y_{11}, y_{12}, y_{21}, y_{22}) &= \frac{{}_0F_3\left(\frac{1}{3}, \frac{2}{3}, 1; K\right)}{2\pi(s-t)^2(1-\varrho^2)\sqrt{y_{11}y_{22} - y_{12}y_{21}}} \\ &\times \exp\left\{-\frac{x_{11} - \varrho x_{12} - \varrho x_{21} + x_{22} + y_{11} - \varrho y_{12} - \varrho y_{21} + y_{22}}{2(s-t)(\varrho^2 - 1)}\right\} \\ &\times \cosh\left(\frac{1}{2}\sqrt{\frac{(x_{12}x_{21} - x_{11}x_{22})(y_{12}y_{21} - y_{11}y_{22})}{(s-t)^4(\varrho^2 - 1)^2}}\right), \end{aligned} \quad (2.2.24)$$

where

$$K = \frac{1}{108(s-t)^2(\varrho^2-1)^2} (x_{22}y_{11}\varrho^2 + x_{12}y_{12}\varrho^2 - x_{12}y_{11}\varrho - x_{22}y_{12}\varrho - x_{22}y_{21}\varrho - x_{12}y_{22}\varrho + x_{12}y_{21} + x_{21}(y_{12} + \varrho(-y_{11} + \varrho y_{21} - y_{22})) + x_{22}y_{22} + x_{11}(y_{11} + \varrho(-y_{12} - y_{21} + \varrho y_{22})))$$

and ${}_pF_q(\cdot, \cdot, \cdot; K)$ is a generalized hypergeometric function explained in Muirhead (1982).

Transition Density of a Square Root Process

Similarly one can characterize the transition density for a *square root* (SR)-process, see (2.1.3).

A scalar square root (SR)-process $X = \{X_t, t \geq 0\}$ of dimension $\delta > 2$ is given by the SDE

$$dX_t = \left(\frac{\delta}{4} c_t^2 + b_t X_t \right) dt + c_t \sqrt{X_t} dW_t \quad (2.2.25)$$

for $t \geq 0$ with $X_0 > 0$. Here $W = \{W_t, t \geq 0\}$ is a scalar Wiener process and c_t and b_t are deterministic functions of time. Let

$$s_t = \exp \left\{ \int_0^t b_u du \right\} \quad (2.2.26)$$

and

$$\varphi_t = \frac{1}{4} \int_0^t \frac{c_u^2}{s_u} du \quad (2.2.27)$$

for $t \geq 0$. Then the transition density of the scalar SR-process X is given in the form

$$p(s, x; t, y) = \frac{1}{2 s_t \varphi_t} \left(\frac{y}{x s_t} \right)^{\frac{\nu}{2}} \exp \left\{ -\frac{x + \frac{y}{s_t}}{2 \varphi_t} \right\} I_\nu \left(\frac{\sqrt{x \frac{y}{s_t}}}{\varphi_t} \right) \quad (2.2.28)$$

for $0 \leq s < t < \infty$ and $x, y \in (0, \infty)$, where $I_\nu(\cdot)$ is the modified Bessel function of the first kind with index $\nu = \frac{\delta}{2} - 1$.

Transition Density of a Matrix SR-process

For the multi-dimensional SR-process, given in terms of a $d \times m$ matrix, we have an analytic transition density that can be derived from (2.2.18) or (2.2.22) in the form

$$p(s, \mathbf{X}; t, \mathbf{Y}) = \frac{p_\delta \left(\varphi(s), \frac{\mathbf{X}}{s_s}; \varphi(t), \frac{\mathbf{Y}}{s_t} \right)}{s_t} \quad (2.2.29)$$

for $0 \leq s < t < \infty$ and nonnegative elements in the $d \times m$ matrices \mathbf{X} and \mathbf{Y} . Here we use the transformed φ -time with

$$\varphi(t) = \varphi(0) + \frac{\bar{b}^2}{4\bar{c}s_0} (1 - \exp\{-\bar{c}t\}) \quad (2.2.30)$$

and $s_t = s_0 \exp\{\bar{c}t\}$ for $t \in [0, \infty)$, $s_0 > 0$, $\bar{c} < 0$ and $\bar{b} \neq 0$.

As an example, we write down the transition density of the 2×2 matrix SR-process of dimension $\delta = 2$. This density follows from (2.2.24) and it can be expressed by

$$\begin{aligned} p(s, x_{11}, x_{12}, x_{21}, x_{22}; t, y_{11}, y_{12}, y_{21}, y_{22}) \\ = \frac{8\bar{c}^2 s_0^2 e^{\bar{c}(2s+t)} {}_0F_3\left(\frac{1}{3}, \frac{2}{3}, 1; K\right)}{\bar{b}^4 (e^{\bar{c}s} - e^{\bar{c}t})^2 \pi (1 - \varrho^2) \sqrt{e^{-2\bar{c}t} (y_{11}y_{22} - y_{12}y_{21})}} \\ \times \exp\left\{\frac{-2\bar{c}e^{\bar{c}t} (x_{11} - \varrho(x_{12} + x_{21}) + x_{22}) - 2\bar{c}e^{\bar{c}s} (y_{11} - \varrho(y_{12} + y_{21}) + y_{22})}{\bar{b}^2 (e^{\bar{c}s} - e^{\bar{c}t})(\varrho^2 - 1)}\right\} \\ \times \cosh\left(8\sqrt{\frac{\bar{c}^4 e^{2\bar{c}(s+t)} (x_{12}x_{21} - x_{11}x_{22})(y_{12}y_{21} - y_{11}y_{22})}{\bar{b}^8 (e^{\bar{c}s} - e^{\bar{c}t})^4 (\varrho^2 - 1)^2}}\right), \end{aligned} \quad (2.2.31)$$

where

$$\begin{aligned} K = \frac{4\bar{c}^2 e^{\bar{c}(s+t)}}{27\bar{b}^4 (e^{\bar{c}s} - e^{\bar{c}t})^2 (\varrho^2 - 1)^2} & (x_{22}y_{11}\varrho^2 + x_{21}y_{21}\varrho^2 - x_{21}y_{11}\varrho - x_{22}y_{12}\varrho \\ & - x_{22}y_{21}\varrho - x_{21}y_{22}\varrho + x_{21}y_{12} + x_{12}(y_{21} + \varrho(-y_{11} + \varrho y_{12} - y_{22})) \\ & + x_{22}y_{22} + x_{11}(y_{22}\varrho^2 - (y_{12} + y_{21})\varrho + y_{11})). \end{aligned}$$

The above transition density looks complex. However, it has the great advantage of providing an explicit formula, which is extremely valuable in many quantitative investigations.

Transition Density of Multi-dimensional Lévy Processes

Lévy processes are a special class of processes which have independent and stationary increments, see Sect. 1.7. It turns out that the distributions of the independent increments of Lévy processes are infinitely divisible, see Cont & Tankov (2004). A widely used class of distributions for the increments with this property are members of the family of generalized hyperbolic (GH) distributions. It is always possible to construct a Lévy process so that the value of the increment of the process over a fixed time interval has a given GH distribution. The GH density can be expressed as

$$p(\mathbf{x}) = c \frac{K_{\lambda - (\frac{d}{2})} \left(\sqrt{(\chi + (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}))(\psi + \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma})} \right) e^{(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}}}{\left(\sqrt{(\chi + (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}))(\psi + \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma})} \right)^{(\frac{d}{2}) - \lambda}}, \quad (2.2.32)$$

where

$$c = \frac{(\chi\psi)^{-\lambda} \psi^\lambda (\psi + \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma})^{(d/2)-\lambda}}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma})^{1/2} K_\lambda (\sqrt{\chi\psi})}, \quad (2.2.33)$$

$\boldsymbol{\Sigma}$ is a correlation type matrix, $\boldsymbol{\mu}$ a drift vector, $\boldsymbol{\gamma}$ a scaling vector, and χ , ψ and λ are shape parameters.

$K_\lambda(\cdot)$ denotes a modified Bessel function of the third kind and $\mathbf{x} \in \mathbb{R}^d$. Additionally, if $\boldsymbol{\gamma} = \mathbf{0}$, then this distribution is symmetric.

Some known special cases of this distribution include the d -dimensional hyperbolic distribution if $\lambda = \frac{1}{2}(d+1)$; the d -dimensional normal inverse Gaussian (NIG) distribution if $\lambda = -\frac{1}{2}$; the d -dimensional variance-gamma (VG) distribution if $\lambda > 0$ and $\chi = 0$ and the d -dimensional skewed Student- t distribution if $\lambda = -\frac{1}{2}\nu$, $\chi = \nu$ and $\psi = 0$. Note that the last two special cases are limiting distributions.

In the two-dimensional case the GH density simplifies to

$$p(x_1, x_2) = c \exp \{A\} \frac{K_{\lambda-1}(\xi)}{\xi^{1-\lambda}}, \quad (2.2.34)$$

where

$$\begin{aligned} A &= \frac{\gamma_1 \mu_1 - \gamma_2 \varrho \mu_1 + \gamma_2 \mu_2 - \gamma_1 \mu_2 \varrho - \gamma_1 x_1 + \gamma_2 \varrho x_1 - \gamma_2 x_2 + \gamma_1 \varrho x_2}{\varrho^2 - 1}, \\ \xi &= \sqrt{(-\gamma_1^2 + 2\gamma_2 \varrho \gamma_1 - \gamma_2^2 + \varrho^2 \psi - \psi) B}, \\ B &= \frac{(\mu_1^2 - 2\mu_2 \varrho \mu_1 + \mu_2^2 - x_1^2 - x_2^2 + \varrho^2 \chi - \chi + 2\varrho x_1 x_2)}{(\varrho^2 - 1)^2} \end{aligned}$$

and

$$c = \frac{(\sqrt{\chi\psi})^{-\lambda} \psi^\lambda \left(\frac{\gamma_1^2 - 2\gamma_2 \varrho \gamma_1 + \gamma_2^2 - \varrho^2 \psi + \psi}{1 - \varrho^2} \right)^{1-\lambda}}{2\pi \sqrt{1 - \varrho^2} K_\lambda (\sqrt{\chi\psi})}. \quad (2.2.35)$$

In Fig. 2.2.4 we illustrate a bivariate GH density with $\mu_1 = \mu_2 = 0.1$, $\gamma_1 = \gamma_2 = 0.2$, $\chi = \psi = 0.4$, $\lambda = -0.5$ and $\varrho = 0.8$. One notes the much fatter tails of the density in Fig. 2.2.4 when compared with those of the Wiener process increment shown in Fig. 2.2.1. Extreme joint events are far more likely. In principle, one can generate paths of multi-dimensional Lévy processes for a given copula and given marginal transition densities as discussed earlier.

Since we do not discuss the simulation of Lévy processes including those with infinite intensity with great detail, we refer the reader for details on this

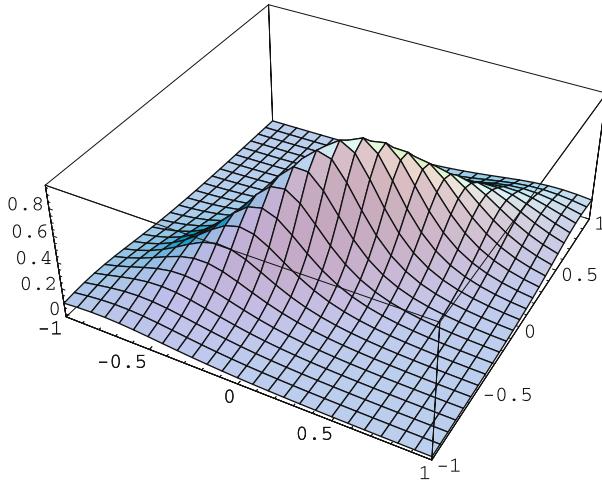


Fig. 2.2.4. Bivariate GH density for $\mu_1 = \mu_2 = 0.1$, $\gamma_1 = \gamma_2 = 0.2$, $\chi = \psi = 0.4$, $\lambda = -0.5$ and $\varrho = 0.8$

subject to Madan & Seneta (1990), Eberlein & Keller (1995), Bertoin (1996), Protter & Talay (1997), Barndorff-Nielsen & Shephard (2001), Geman et al. (2001), Eberlein (2002), Kou (2002), Rubenthaler (2003), Cont & Tankov (2004), Jacod, Kurtz, Méléard & Protter (2005) and Klüppelberg, Lindner & Maller (2006).

Other Explicit Transition Densities

There is still a wider range of one-dimensional Markov processes with explicit transition densities than those mentioned so far. For instance, Cradock & Platen (2004) consider generalized square root processes, see also Platen & Heath (2006). These are diffusion processes $X = \{X_t, t \in [0, \infty)\}$ with a square root function $b(t, x) = \sqrt{2x}$ as diffusion coefficient for all $t \geq 0$ and $x \in [0, \infty)$. By exploiting Lie group symmetries they identified a collection of drift functions $a(t, x) = a(x)$ for which one still has an analytic formula for the corresponding transition density $p(0, x; t, y)$. These include the cases of the following drifts yielding corresponding transition densities that we list below:

(i)

$$a(x) = \alpha > 0,$$

$$p(0, x; t, y) = \frac{1}{t} \left(\frac{x}{y} \right)^{\frac{1-\alpha}{2}} I_{\alpha-1} \left(\frac{2\sqrt{xy}}{t} \right) \exp \left\{ -\frac{(x+y)}{t} \right\};$$

(ii)

$$a(x) = \frac{\mu x}{1 + \frac{\mu}{2}x}, \quad \mu > 0,$$

$$p(0, x; t, y) = \frac{\exp\left\{-\frac{(x+y)}{t}\right\}}{(1 + \frac{\mu}{2}x)t} \left[\left(\sqrt{\frac{x}{y}} + \frac{\mu \sqrt{xy}}{2} \right) I_1\left(\frac{2\sqrt{xy}}{t}\right) + t \delta(y) \right];$$

(iii)

$$a(x) = \frac{1 + 3\sqrt{x}}{2(1 + \sqrt{x})},$$

$$\begin{aligned} p(0, x; t, y) &= \frac{\cosh\left(\frac{2\sqrt{xy}}{t}\right)}{\sqrt{\pi y t}(1 + \sqrt{x})} \left(1 + \sqrt{y} \tanh\left(\frac{2\sqrt{xy}}{t}\right)\right) \\ &\quad \times \exp\left\{-\frac{(x+y)}{t}\right\}; \end{aligned}$$

(iv)

$$a(x) = 1 + \mu \tanh\left(\mu + \frac{1}{2}\mu \ln(x)\right), \quad \mu = \frac{1}{2}\sqrt{\frac{5}{2}},$$

$$\begin{aligned} p(0, x; t, y) &= \left(\frac{x}{y}\right)^{\frac{\mu}{2}} \left[I_{-\mu}\left(\frac{2\sqrt{xy}}{t}\right) + e^{2\mu} y^\mu I_\mu\left(\frac{2\sqrt{xy}}{t}\right) \right] \\ &\quad \times \frac{\exp\left\{-\frac{x+y}{t}\right\}}{(1 + \exp\{2\mu\} x^\mu)t}; \end{aligned}$$

(v)

$$a(x) = \frac{1}{2} + \sqrt{x},$$

$$p(0, x; t, y) = \cosh\left(\frac{(t + 2\sqrt{x})\sqrt{y}}{t}\right) \frac{\exp\{-\sqrt{x}\}}{\sqrt{\pi y t}} \exp\left\{-\frac{(x+y)}{t} - \frac{t}{4}\right\};$$

(vi)

$$a(x) = \frac{1}{2} + \sqrt{x} \tanh(\sqrt{x}),$$

$$p(0, x; t, y) = \frac{\cosh\left(\frac{2\sqrt{xy}}{t}\right)}{\sqrt{\pi y t}} \frac{\cosh(\sqrt{y})}{\cosh(\sqrt{x})} \exp\left\{-\frac{(x+y)}{t} - \frac{t}{4}\right\};$$

(vii)

$$a(x) = \frac{1}{2} + \sqrt{x} \coth(\sqrt{x}),$$

$$p(0, x; t, y) = \frac{\sinh\left(\frac{2\sqrt{xy}}{t}\right)}{\sqrt{\pi y t}} \frac{\sinh(\sqrt{y})}{\sinh(\sqrt{x})} \exp\left\{-\frac{(x+y)}{t} - \frac{t}{4}\right\};$$

(viii)

$$a(x) = 1 + \cot(\ln(\sqrt{x})) \quad \text{for } x \in (\exp\{-2\pi\}, 1),$$

$$p(0, x; t, y) = \frac{\exp\{-\frac{(x+y)}{t}\}}{2\iota t \sin(\ln(\sqrt{x}))} \left(y^{\frac{1}{2}} I_{\iota} \left(\frac{2\sqrt{xy}}{t} \right) - y^{-\frac{1}{2}} I_{-\iota} \left(\frac{2\sqrt{xy}}{t} \right) \right);$$

(ix)

$$a(x) = x \coth \left(\frac{x}{2} \right),$$

$$p(0, x; t, y) = \frac{\sinh(\frac{y}{2})}{\sinh(\frac{x}{2})} \exp \left\{ -\frac{(x+y)}{2 \tanh(\frac{t}{2})} \right\}$$

$$\times \left[\frac{\exp\{\frac{t}{2}\}}{\exp\{t\} - 1} \sqrt{\frac{x}{y}} I_1 \left(\frac{\sqrt{xy}}{\sinh(\frac{t}{2})} \right) + \delta(y) \right];$$

(x)

$$a(x) = x \tanh \left(\frac{x}{2} \right),$$

$$p(0, x; t, y) = \frac{\cosh(\frac{y}{2})}{\cosh(\frac{x}{2})} \exp \left\{ -\frac{(x+y)}{2 \tanh(\frac{t}{2})} \right\}$$

$$\times \left[\frac{\exp\{\frac{t}{2}\}}{\exp\{t\} - 1} \sqrt{\frac{x}{y}} I_1 \left(\frac{\sqrt{xy}}{\sinh(\frac{t}{2})} \right) + \delta(y) \right].$$

Here I_α is the modified Bessel function of the first kind with index α , see [Platen & Heath \(2006\)](#) and [\(2.2.19\)](#).

The first drift above in (i) is that of the well-known squared Bessel process. However, the other drifts are mostly from previously unknown scalar diffusion processes. It is obvious that some multi-dimensional versions of most of these processes can be constructed. It is an area of ongoing research that identifies natural dependence structures between different components of diffusions with explicit transition densities. One can, of course, always start from a given copula and generate dependent diffusions with given marginal transition distributions. An easy case is obtained, when all components of the multi-dimensional diffusion are independent. In general, each component of a multi-dimensional diffusion can come from various types of conditional transition distribution functions. This gives a great variety of multi-dimensional models with paths that can be exactly simulated via the inverse transform method.

Illustration of the Inverse Transform Method

Let us illustrate simulation via the inverse transform method for the following simple example of a two-dimensional standard OU-process. The marginal

transition distribution function of this process for its first component can be represented as

$$\begin{aligned} F_{Y_1}(s, x_1, t, y_1) &= \frac{1}{2} \left(1 - e^{-2(t-s)} \right) \sqrt{1 - \varrho^2} \\ &\times \left[1 - \operatorname{erf} \left\{ x_1 e^{-(t-s)} \sqrt{\frac{1}{2} (1 - e^{-2(t-s)})} \right\} \right. \\ &+ e^{-(t-s)} \sqrt{1 - e^{-2(t-s)}} \left(x_1 \sqrt{\frac{e^{2(t-s)} - 1}{x_1^2}} \operatorname{erf} \left\{ \sqrt{\frac{x_1^2}{2(e^{2(t-s)} - 1)}} \right\} \right. \\ &\left. \left. + (y_1 e^{t-s} - x_1) \sqrt{\frac{1 - e^{-2(t-s)}}{x_1 e^{-(t-s)} - y_1}} \operatorname{erf} \left\{ \frac{x_1 e^{-(t-s)} - y_1}{2(1 - e^{-2(t-s)})} \right\} \right) \right], \end{aligned} \quad (2.2.36)$$

for $x_1, y_1 \in \mathbb{R}$ and $t > s$. Recall that $\operatorname{erf}\{\cdot\}$ denotes the well-known error function, see [Abramowitz & Stegun \(1972\)](#). Additionally, the conditional transition distribution function of the second component given the first component is

$$\begin{aligned} F_{Y_2|Y_1}(s, x_1, x_2, t, y_1, y_2) &= \frac{1}{2} \sqrt{\frac{1}{(1 - e^{-2(t-s)})(1 - \varrho^2)}} \\ &\times \left[\sqrt{(1 - e^{-2(t-s)})(1 - \varrho^2)} - (x_2 - x_1 \varrho - e^{t-s}(y_2 - y_1 \varrho)) \right. \\ &\left. \times \sqrt{\frac{(1 - e^{-2(t-s)})(1 - \varrho^2)}{(x_2 - x_1 \varrho - e^{t-s}(y_2 - y_1 \varrho))^2}} \operatorname{erf} \left\{ \sqrt{\frac{(x_2 - x_1 \varrho - e^{t-s}(y_2 - y_1 \varrho))^2}{2(e^{2(t-s)} - 1)(1 - \varrho^2)}} \right\} \right], \end{aligned} \quad (2.2.37)$$

for $x_1, x_2, y_1, y_2 \in \mathbb{R}$, $\varrho \in [-1, 1]$ and $t > s$. In Fig. 2.2.5 we display the marginal transition distribution function (2.2.36) for fixed $x_1 = 0.1$, while in Fig. 2.2.6 we illustrate the conditional transition distribution function (2.2.37) for fixed initial values $x_1 = x_2 = 0.1$ and the time increment $\Delta = t-s = 0.1$. In both graphs we assume the correlation $\varrho = 0.8$. We use the inverse transform method for the simulation of the two-dimensional OU-process and display a resulting trajectory of the related two components in Fig. 2.2.7. This example illustrates that we need a collection of marginal and conditional transition distribution functions when generating the components of the process one after the other. In the above Gaussian example this involved the use of the $\operatorname{erf}\{\cdot\}$ special function. For squared Bessel processes the non-central chi-square distribution function represents the analogue special function that needs to be employed. Otherwise, the procedure for simulating exact paths is very similar.

2.3 Exact Solutions of Multi-dimensional SDEs

Sometimes, for multi-dimensional distribution functions, as those introduced in Sect. 2.2, it may be more convenient to sample from the known multi-

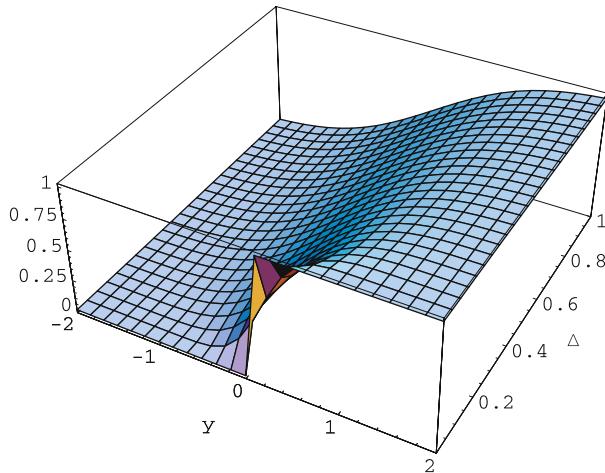


Fig. 2.2.5. Marginal transition distribution function of the first component y_1 of the two-dimensional OU-process for fixed $x_1 = 0.1$ in dependence on the time step Δ

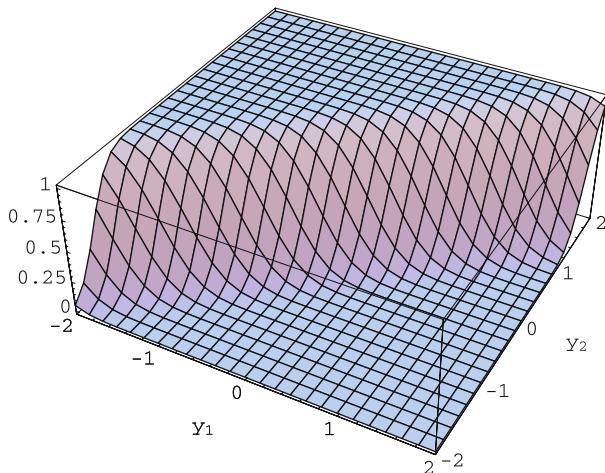


Fig. 2.2.6. Conditional transition distribution function of the second component y_2 of the two-dimensional OU-process given the first component y_1 for fixed $x_1 = x_2 = 0.1$ and $\Delta = 0.1$

dimensional distribution function directly rather than using the inverse-transform method. For instance, the increment of the multi-dimensional Wiener process can be simulated from the following exact relation

$$\mathbf{X}_{t+\Delta} - \mathbf{X}_t \sim N_d(\mathbf{0}, \boldsymbol{\Sigma}\Delta), \quad (2.3.1)$$

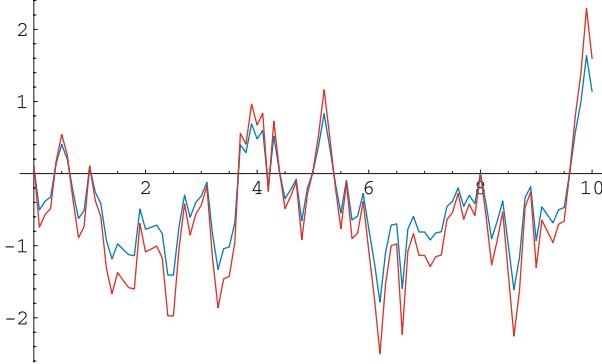


Fig. 2.2.7. Trajectory of the two components of the two-dimensional OU-process for $\Delta = 0.1$, initial value $\mathbf{x} = (0.1, 0.1)$ and $\varrho = 0.8$

where N_d denotes a d -dimensional Gaussian distribution with mean vector $\mathbf{0}$ and covariance matrix $\Sigma\Delta$. Similarly, the value at time $t + \Delta$ of the standard d -dimensional OU-process can be obtained by the relation

$$\mathbf{X}_{t+\Delta} \sim N_d(\mathbf{X}_t e^{-\Delta}, \Sigma(1 - e^{-2\Delta})). \quad (2.3.2)$$

The $m \times m$ Wishart process can be simulated from the non-central Wishart distribution W_m with δ degrees of freedom, covariance matrix $\Sigma\Delta$ and non-centrality matrix $\Sigma^{-1}\mathbf{X}_t\Delta^{-1}$, where

$$\mathbf{X}_{t+\Delta} \sim W_m(\delta, \Sigma\Delta, \Sigma^{-1}\mathbf{X}_t\Delta^{-1}). \quad (2.3.3)$$

For details on how to sample conveniently from the non-central Wishart distribution we refer to [Gleser \(1976\)](#).

The increments of a Lévy processes constructed from a GH distribution can be obtained by

$$\mathbf{X}_{t+\Delta} - \mathbf{X}_t \sim GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma) \quad (2.3.4)$$

for fixed $\Delta > 0$. GH random variables can be simulated by subordination from d -dimensional Gaussian random variables whose mean and covariance matrix are made random in an appropriate way. This yields a mixture of normal increments by the relation

$$\mathbf{X}_{t+\Delta} - \mathbf{X}_t \sim \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}\mathbf{A}\mathbf{Z}. \quad (2.3.5)$$

Here $\mathbf{Z} \sim N_k(\mathbf{0}, \mathbf{I}_k)$, where $W \geq 0$ is a non-negative scalar generalized inverse Gaussian (GIG) random variable that is independent of \mathbf{Z} . Here \mathbf{I}_k denotes a k -dimensional unit matrix. \mathbf{A} is a $d \times k$ matrix and $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are d -dimensional vectors. Hence, the conditional distribution of increment $\mathbf{X}_{t+\Delta} - \mathbf{X}_t$ given $W = w$ is conditionally Gaussian $N_d(\boldsymbol{\mu} + W\boldsymbol{\gamma}, w\Sigma)$, where $\Sigma = \mathbf{A}\mathbf{A}^\top$.

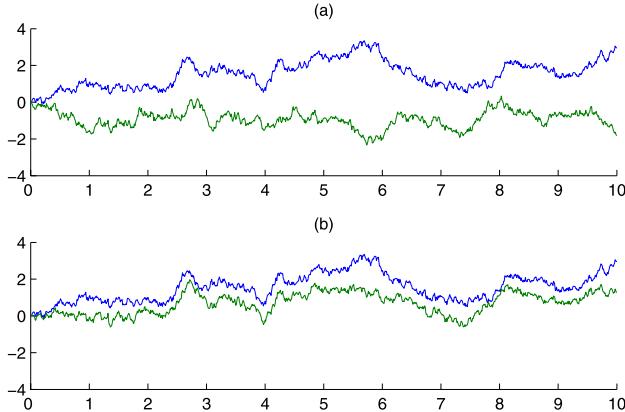


Fig. 2.3.1. (a) The trajectory of the independent components of a two-dimensional standard Wiener process; (b) Trajectories of two correlated Wiener processes

Let us now consider some selected multi-dimensional stochastic processes commonly used for modeling in finance but also popular in other areas of application. Our choice of these processes stems from the fact that they can all be simulated exactly, at least for some special cases.

Wiener Processes

The most important multi-dimensional continuous process with stationary independent increments is the d -dimensional standard Wiener process, see Sect. 1.1. It is a continuous process with independent Gaussian increments. First we assume that the components of the d -dimensional Wiener process \mathbf{W} , are independent. The increments of the Wiener processes $W_t^j - W_s^j$ for $j \in \{1, 2, \dots, d\}$, $t \geq 0$ and $s \leq t$ are then independent Gaussian random variables with mean zero and variance equal to $t - s$. Therefore, one obtains the vector increments of the standard d -dimensional Wiener process $\mathbf{W}_t - \mathbf{W}_s \sim N_d(\mathbf{0}, (t - s)\mathbf{I})$ as a vector of zero mean independent Gaussian random variables with variance $t - s$. \mathbf{I} denotes here the unit matrix. For the values of the trajectory of the standard d -dimensional Wiener process at the discretization times $t_i = i\Delta$, $i \in \{0, 1, \dots\}$, with $\Delta > 0$ we obtain the following iterative formula

$$\begin{aligned} \mathbf{W}_0 &= \mathbf{0} \\ \mathbf{W}_{t_{i+1}} &= \mathbf{W}_{t_i} + \sqrt{\Delta} \mathbf{N}_{i+1}, \end{aligned} \tag{2.3.6}$$

where $\mathbf{N}_{i+1} \sim N_d(\mathbf{0}, \mathbf{I})$ is an independent standard Gaussian random vector and $\mathbf{0}$ denotes the corresponding vector of zeros. We display in Fig. 2.3.1 (a) the trajectories of the independent components of a two-dimensional standard Wiener process.

Correlated Wiener Processes

Let us now define a d -dimensional continuous process

$$\tilde{\mathbf{W}} = \{\tilde{\mathbf{W}}_t = (\tilde{W}_t^1, \tilde{W}_t^2, \dots, \tilde{W}_t^d)^\top, t \in [0, \infty)\} \quad (2.3.7)$$

such that its components $\tilde{W}_t^1, \tilde{W}_t^2, \dots, \tilde{W}_t^d$ are transformed scalar Wiener processes. In vector notation such a d -dimensional transformed Wiener process can be expressed by the linear transform

$$\tilde{\mathbf{W}}_t = \mathbf{a}t + \mathbf{B}\mathbf{W}_t, \quad (2.3.8)$$

where $\mathbf{a} = (a_1, a_2, \dots, a_d)^\top$ is a d -dimensional vector and \mathbf{B} is a $d \times m$ -matrix and $\mathbf{W}_t = (W_t^1, W_t^2, \dots, W_t^m)^\top$ is an m -dimensional standard Wiener process. Note that the k th component of $\tilde{\mathbf{W}}$ is such that

$$\tilde{W}_t^k = a_k t + \sum_{i=1}^m b_{k,i} W_t^i, \quad (2.3.9)$$

for $k \in \{1, 2, \dots, d\}$. This means that \tilde{W}_t^k , $k \in \{1, 2, \dots, d\}$, is constructed as a linear combination of components of the vector \mathbf{W}_t plus some trend. From the properties of Gaussian random variables, the following relation emerges

$$\begin{aligned} \tilde{\mathbf{W}}_0 &= \mathbf{0}, \\ \tilde{\mathbf{W}}_{t_{i+1}} &= \tilde{\mathbf{W}}_{t_i} + \mathbf{a}\Delta + \sqrt{\Delta} \tilde{\mathbf{N}}_{i+1}, \end{aligned} \quad (2.3.10)$$

for $t_i = i\Delta$, $i \in \{0, 1, \dots\}$ with $\Delta > 0$. Here the random vector $\tilde{\mathbf{N}}_{i+1} \sim N_d(\mathbf{0}, \Sigma)$ is a d -dimensional Gaussian vector with $\Sigma = \mathbf{B}\mathbf{B}^\top$, which is independent for each $i \in \{0, 1, \dots\}$.

We display in Fig. 2.3.1 (b) a trajectory of a two-dimensional transformed Wiener process $\tilde{\mathbf{W}} = \{\tilde{\mathbf{W}}_t = (\tilde{W}_t^1, \tilde{W}_t^2)^\top, t \in [0, 10]\}$ with correlated components. The two components of this process are as follows

$$\tilde{W}_t^1 = W_t^1, \quad (2.3.11)$$

$$\tilde{W}_t^2 = \varrho W_t^1 + \sqrt{1 - \varrho^2} W_t^2, \quad (2.3.12)$$

for $t \in [0, 10]$ with correlation $\varrho = 0.8$. In Fig. 2.3.1 (a) we showed the corresponding independent Wiener paths W_t^1 and W_t^2 for $t \in [0, 10]$. We note the expected strong similarity between \tilde{W}_t^1 and \tilde{W}_t^2 in Fig. 2.3.1 (b). The two-dimensional Wiener process $\tilde{\mathbf{W}}$ can also be expressed using matrix multiplication as $\tilde{\mathbf{W}}_t = \mathbf{B}\mathbf{W}_t$, where

$$\mathbf{B} = \begin{pmatrix} 1 & 0 \\ \varrho & \sqrt{1 - \varrho^2} \end{pmatrix} \quad (2.3.13)$$

and $\mathbf{W}_t = (W_t^1, W_t^2)^\top$.

Matrix Wiener Processes

As we observed already with Wishart processes, matrix valued processes may be convenient in a given context. Therefore, let us define a $d \times m$ standard matrix Wiener process $\mathbf{W} = \{\mathbf{W}_t = [W_t^{i,j}]_{i,j=1}^{d,m}, t \in [0, \infty)\}$. This matrix stochastic process can be obtained by the following construction

$$\begin{aligned}\mathbf{W}_0 &= \mathbf{0} \\ \mathbf{W}_{t_{i+1}} &= \mathbf{W}_{t_i} + \sqrt{\Delta} \mathbf{N}_{i+1},\end{aligned}\tag{2.3.14}$$

for $t_i = i\Delta$, $i = \{0, 1, \dots\}$ with $\Delta > 0$ and $d \times m$ -matrix $\mathbf{0}$ of zero elements. Here $\mathbf{N}_{i+1} \sim N_{d \times m}(\mathbf{0}, \mathbf{I}_m \otimes \mathbf{I}_d)$ is a matrix of zero mean Gaussian distributed random variables. The covariance matrix $\mathbf{I}_m \otimes \mathbf{I}_d$ is an $m \times m$ diagonal block matrix of $d \times d$ identity matrices \mathbf{I}_d , that is,

$$\mathbf{I}_m \otimes \mathbf{I}_d = \begin{pmatrix} \mathbf{I}_d & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_d & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_d \end{pmatrix}.\tag{2.3.15}$$

Moreover, similarly to the vector case, we are able to define a transformed matrix Wiener process $\tilde{\mathbf{W}} = \{\tilde{\mathbf{W}}_t, t \in [0, \infty)\}$ using the above matrix Wiener process \mathbf{W} as follows

$$\tilde{\mathbf{W}}_t = \mathbf{M}t + \boldsymbol{\Sigma}_1 \mathbf{W}_t \boldsymbol{\Sigma}_2^\top,\tag{2.3.16}$$

where \mathbf{M} is a $d \times m$ matrix and $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are nonsingular $d \times d$ and $m \times m$ matrices, respectively. Values of such a matrix stochastic process can be obtained at times $t_i = i\Delta$ by the following recursive computation

$$\begin{aligned}\mathbf{W}_0 &= \mathbf{0} \\ \tilde{\mathbf{W}}_{t_{i+1}} &= \tilde{\mathbf{W}}_{t_i} + \mathbf{M}\Delta + \sqrt{\Delta} \tilde{\mathbf{N}}_{i+1},\end{aligned}\tag{2.3.17}$$

for $i \in \{0, 1, \dots\}$ and $\tilde{\mathbf{N}}_{i+1} \sim N_{d \times m}(\mathbf{0}, \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1)$. Here, the covariance matrix $\boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1$ is again an $m \times m$ block matrix of the form

$$\boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma_{1,1}^2 \boldsymbol{\Sigma}_1 & \sigma_{1,2}^2 \boldsymbol{\Sigma}_1 & \dots & \sigma_{1,m}^2 \boldsymbol{\Sigma}_1 \\ \sigma_{2,1}^2 \boldsymbol{\Sigma}_1 & \sigma_{2,2}^2 \boldsymbol{\Sigma}_1 & \dots & \sigma_{2,m}^2 \boldsymbol{\Sigma}_1 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{m,1}^2 \boldsymbol{\Sigma}_1 & \sigma_{m,2}^2 \boldsymbol{\Sigma}_1 & \dots & \sigma_{m,m}^2 \boldsymbol{\Sigma}_1 \end{pmatrix},\tag{2.3.18}$$

where $\boldsymbol{\Sigma}_1 = [\sigma_{i,j}^1]_{i,j}^d$ and $\boldsymbol{\Sigma}_2 = [\sigma_{i,j}^2]_{i,j}^m$.

Let us illustrate this matrix valued stochastic process for a 2×2 matrix case. In Fig. 2.3.2 we display a transformed matrix Wiener process $\tilde{\mathbf{W}}$, which was obtained from a standard 2×2 matrix Wiener process \mathbf{W} by the following transformation

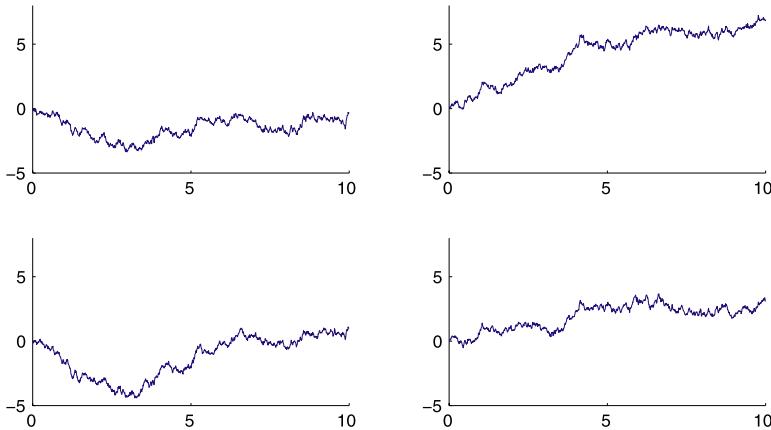


Fig. 2.3.2. 2×2 matrix Wiener process with independent elements in rows and dependent elements in columns

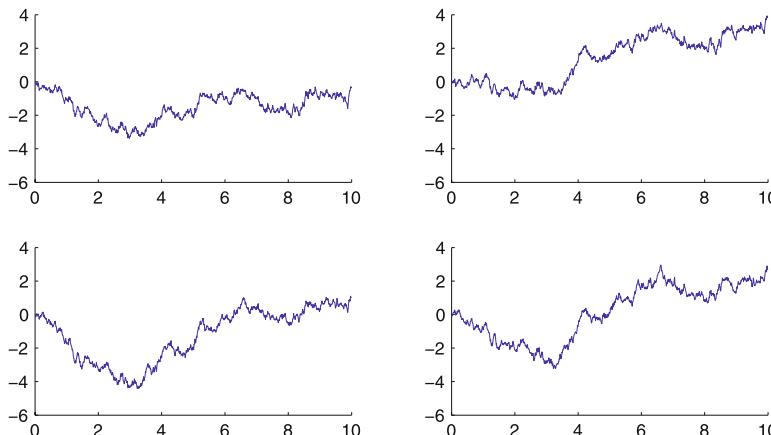


Fig. 2.3.3. 2×2 matrix Wiener process with both correlated rows and columns

$$\tilde{\mathbf{W}}_t = \boldsymbol{\Sigma}_1 \mathbf{W}_t \mathbf{I}^\top, \quad (2.3.19)$$

where $\boldsymbol{\Sigma}_1 = \mathbf{B}$ is as in (2.3.13) with correlation $\varrho = 0.8$. Note that in this case we obtain a matrix stochastic process whose rows have independent elements while its columns are formed by correlated Wiener processes. Similarly, in Fig. 2.3.3 we illustrate a 2×2 matrix transformed Wiener process $\tilde{\mathbf{W}}$, which was obtained from the standard 2×2 matrix Wiener process \mathbf{W} by the transformation of the following form

$$\tilde{\mathbf{W}}_t = \boldsymbol{\Sigma}_1 \mathbf{W}_t \boldsymbol{\Sigma}_2^\top, \quad (2.3.20)$$

where both $\boldsymbol{\Sigma}_1 = \mathbf{B}$ and $\boldsymbol{\Sigma}_2 = \mathbf{B}$ are as in (2.3.13). For $\varrho = 0.8$ we note in Fig. 2.3.3 the correlation effect on the trajectories on both the elements of the columns and rows of such a 2×2 matrix transformed Wiener process.

Time Changed Wiener Processes

Instead of multiplying the time by some constant to scale the fluctuations of the Wiener paths one can introduce a flexible time dependent scaling by a, so called, time change. Let us now consider a vector of time changed standard independent Wiener processes $\mathbf{W}_{\varphi(t)} = (W_{\varphi(t)}^1, \dots, W_{\varphi(t)}^d)^\top$. Given the time discretization $t_i = i\Delta$, $i \in \{0, 1, \dots\}$, with time step size $\Delta > 0$ we obtain this time changed standard Wiener process by the following iterative formula

$$\begin{aligned} \mathbf{W}_{\varphi(0)} &= \mathbf{0} \\ \mathbf{W}_{\varphi(t_{i+1})} &= \mathbf{W}_{\varphi(t_i)} + \sqrt{\varphi(t_{i+1}) - \varphi(t_i)} \mathbf{N}_{i+1}, \end{aligned} \quad (2.3.21)$$

where the vector $\mathbf{N}_{i+1} \sim N_d(\mathbf{0}, \mathbf{I})$ is formed by independent standard Gaussian random variables. Obviously, it is possible to apply different time changes to different elements of the vector \mathbf{W} . For instance to prepare the representation of Ornstein-Uhlenbeck processes, let us define

$$\varphi_j(t) = \frac{b_j^2}{2c_j}(e^{2c_j t} - 1) \quad (2.3.22)$$

for $t \in [0, \infty)$, $b_j > 0$, $c_j > 0$ and $j \in \{1, 2, \dots, d\}$, see (2.2.30). Then the elements of the vector $\mathbf{W}_{\varphi(t)}$ are such that

$$W_{\varphi_j(t_{i+1})}^j - W_{\varphi_j(t_i)}^j \sim N(0, \varphi_j(t_{i+1}) - \varphi_j(t_i)), \quad (2.3.23)$$

where $W_{\varphi_j(0)}^j = 0$, $j \in \{1, 2, \dots, d\}$ and $i \in \{0, 1, \dots\}$.

In order to obtain a time changed vector Wiener process, whose elements are correlated time changed Wiener processes, it is sufficient to define a new vector $\tilde{\mathbf{W}} = \{\tilde{\mathbf{W}}_{\varphi(t)} = (\tilde{W}_{\varphi(t)}^1, \dots, \tilde{W}_{\varphi(t)}^d)^\top, t \in [0, \infty)\}$ by the following transformation

$$\tilde{\mathbf{W}}_{\varphi(t)} = \mathbf{B} \mathbf{W}_{\varphi(t)}, \quad (2.3.24)$$

where \mathbf{B} is a $d \times m$ -matrix of coefficients and $\mathbf{W}_{\varphi(t)} = (W_{\varphi(t)}^1, \dots, W_{\varphi(t)}^m)^\top$ is an m -dimensional time changed Wiener process with independent components as in (2.3.23).

Additionally, let us define a $d \times m$ standard time changed matrix Wiener process $\mathbf{W} = \{\mathbf{W}_{\varphi(t)} = [W_{\varphi(t)}^{j,k}]_{j,k=1}^{d,m}, t \in [0, \infty)\}$. Here, the independent elements of the matrix $\mathbf{W}_{\varphi(t)}$ are such that

$$W_{\varphi_{j,k}(t_{i+1})}^{j,k} - W_{\varphi_{j,k}(t_i)}^{j,k} \sim N(0, \varphi_{j,k}(t_{i+1}) - \varphi_{j,k}(t_i)), \quad (2.3.25)$$

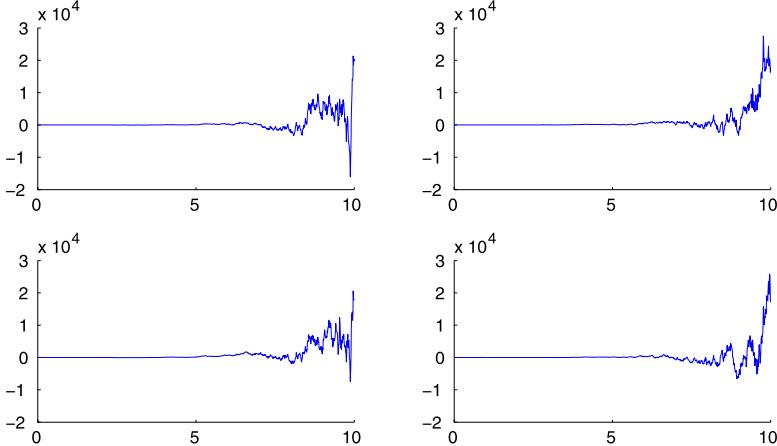


Fig. 2.3.4. Matrix valued time changed Wiener process

where $W_{\varphi_{k,j}(0)}^{k,j} = 0$, $t_i = i\Delta$, $i \in \{0, 1, \dots\}$ and $j \in \{1, 2, \dots, d\}$, $k \in \{1, 2, \dots, m\}$. Again, we may, for instance, define the (j, k) -th time transformation by

$$\varphi_{j,k}(t) = \frac{b_{j,k}^2}{2c_{j,k}} (e^{2c_{j,k}t} - 1) \quad (2.3.26)$$

for $t \in [0, \infty)$, $b_{j,k} > 0$, $c_{j,k} > 0$, and $j \in \{1, 2, \dots, d\}$, $k \in \{1, 2, \dots, m\}$. In order to obtain the time changed matrix Wiener process with correlated elements we can use the formula (2.3.16).

In Fig. 2.3.4 we display a matrix time changed Wiener process for $d = m = 2$ with the covariance matrix $\mathbf{I} \otimes \boldsymbol{\Sigma}_1$, where $\boldsymbol{\Sigma}_1$ is as in (2.3.13), $\varrho = 0.8$ and the parameters in the time change equal $b_{j,k} = \sqrt{2}$ and $c_{j,k} = 1$ for $j, k = \{1, 2\}$. That is, the same time change is applied to each of the elements of this matrix Wiener process. Namely, we construct $\tilde{\mathbf{W}}$ by the relation

$$\tilde{\mathbf{W}}_{\varphi(t)} = \boldsymbol{\Sigma}_1 \mathbf{W}_{\varphi(t)} \mathbf{I}^\top. \quad (2.3.27)$$

In this case we obtain a matrix time changed Wiener process $\tilde{\mathbf{W}}$ whose rows have independent elements, while columns have dependent elements.

Multi-dimensional OU-Processes

Let us now consider multi-dimensional Ornstein-Uhlenbeck (OU)-processes, covering both vector and matrix valued OU-processes, see Sect. 1.7. We will here construct the multi-dimensional OU-process as a time changed and scaled multi-dimensional Wiener process. Note that given the following two functions

$$s_t = \exp\{-ct\} \quad \text{and} \quad \varphi(t) = \frac{b^2}{2c}(e^{2ct} - 1) \quad (2.3.28)$$

for $t \in [0, \infty)$, $b, c > 0$, the scalar OU-process $Y = \{Y_t, t \in \mathfrak{R}\}$ can be represented in terms of a time changed and scaled scalar Wiener process, that is

$$Y_t = s_t W_{\varphi(t)}, \quad (2.3.29)$$

where $W = \{W_\varphi, \varphi \geq 0\}$ is a standard Wiener process in φ -time. By Itô's formula we obtain

$$\begin{aligned} dY_t &= W_{\varphi(t)} ds_t + s_t dW_{\varphi(t)} = -\frac{Y_t}{s_t} cst dt + s_t \frac{b}{s_t} d\tilde{W}_t \\ &= -cY_t dt + b d\tilde{W}_t, \end{aligned} \quad (2.3.30)$$

where $dW_{\varphi(t)} = \frac{b}{s_t} d\tilde{W}_t$, with \tilde{W} denoting a standard Wiener process in t -time. Thus, by a straightforward time change and an application of the Itô formula one obtains a mean-reverting OU-process out of a basic Wiener process.

It is also straightforward to obtain a vector OU-process by

$$\mathbf{Y}_t = s_t \mathbf{W}_{\varphi(t)}, \quad (2.3.31)$$

that is,

$$Y_t^j = s_t^j W_{\varphi_j(t)}^j \quad (2.3.32)$$

for $j \in \{1, 2, \dots, d\}$ and $t \geq 0$. The generalization to a matrix OU-process is obvious. The construction of this process starts by forming a $d \times m$ matrix time changed Wiener process and then scaling each element of this matrix by a function $s_t^{j,k}$ for $j \in \{1, 2, \dots, d\}$ and $k \in \{1, 2, \dots, m\}$. Hence, the elements of such a matrix can be expressed by

$$Y_t^{j,k} = s_t^{j,k} W_{\varphi_{j,k}(t)}^{j,k} \quad (2.3.33)$$

for $j \in \{1, 2, \dots, d\}$ and $k \in \{1, 2, \dots, m\}$.

We illustrate in Fig. 2.3.5 the matrix OU-process obtained from the matrix time changed Wiener process in Fig. 2.3.4 by use of formula (2.3.33). Since, the matrix time changed Wiener process has correlated rows and independent columns, the OU-process in Fig. 2.3.5 shares this feature.

Multi-dimensional SR-Processes via OU-Processes

Now let us consider δ OU-processes, that is

$$dX_t^i = -cX_t^i dt + b dW_t^i \quad (2.3.34)$$

for $t \in [0, \infty)$, with $X_0^i = x_0^i$, $c, b \in \mathfrak{R}$ and independent standard Wiener processes W^i for $i \in \{1, 2, \dots, \delta\}$, $\delta \in \{1, 2, \dots\}$. The square of such an OU-process has the Itô differential

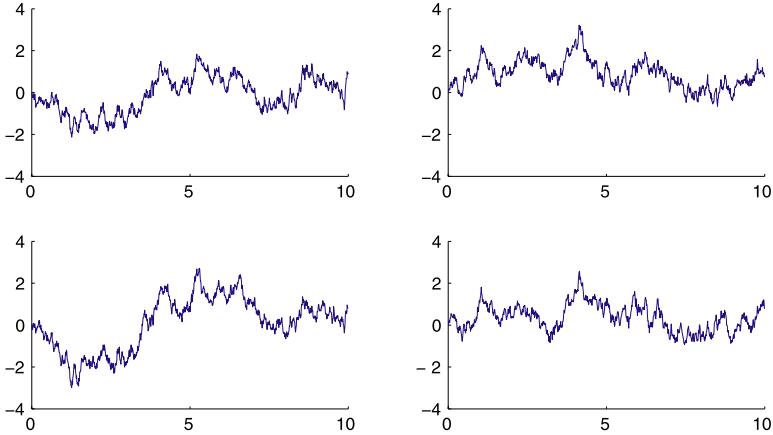


Fig. 2.3.5. Matrix valued Ornstein-Uhlenbeck process

$$d(X_t^i)^2 = (b^2 - 2c(X_t^i)^2) dt + 2bX_t^i dW_t^i, \quad (2.3.35)$$

for $t \in [0, \infty)$ and $i \in \{1, 2, \dots, \delta\}$. Furthermore, we can form the sum of the δ squared OU-processes, that is,

$$Y_t = \sum_{i=1}^{\delta} (X_t^i)^2 \quad (2.3.36)$$

for $t \in [0, \infty)$. The SDE for Y_t is derived to be

$$dY_t = \sum_{i=1}^{\delta} (b^2 - 2c(X_t^i)^2) dt + 2b \sum_{i=1}^{\delta} X_t^i dW_t^i \quad (2.3.37)$$

for $t \in [0, \infty)$. In order to simplify the above SDE we introduce another Wiener process $\bar{W} = \{\bar{W}_t, t \in [0, \infty)\}$ defined as

$$\bar{W}_t = \int_0^t d\bar{W}_s = \sum_{i=1}^{\delta} \int_0^t \frac{X_s^i}{\sqrt{Y_s}} dW_s^i \quad (2.3.38)$$

for $t \in [0, \infty)$. It can be shown that the quadratic variation of \bar{W} equals

$$[\bar{W}]_t = \int_0^t \sum_{i=1}^{\delta} \frac{(X_s^i)^2}{Y_s} ds = t. \quad (2.3.39)$$

Hence, by the Lévy theorem, see Theorem 1.3.3, we see that \bar{W} is a standard Wiener process. Therefore, we obtain an equivalent SDE for the square root process Y in the form

$$dY_t = (\delta b^2 - 2cY_t) dt + 2b\sqrt{Y_t} d\bar{W}_t \quad (2.3.40)$$

for $t \in [0, \infty)$ with $Y_0 = \sum_{i=1}^{\delta} (x_0^i)^2$. Note that this process is a SR-process of dimension $\delta \in \{1, 2, \dots\}$. It is well-known that for $\delta = 1$ the value Y_t can reach zero and is reflected at this boundary. For $\delta \in \{2, 3, \dots\}$ the process never reaches zero for $Y_0 > 0$, see Revuz & Yor (1999).

Matrix Valued Squares of OU-Processes

Kendall (1989) and Bru (1991) studied the matrix generalization for squares of OU-processes. Denote by \mathbf{X}_t a $\delta \times m$ matrix solution of the SDE

$$d\mathbf{X}_t = -c\mathbf{X}_t dt + b d\mathbf{W}_t, \quad (2.3.41)$$

for $t \geq 0$, with $\mathbf{X}_0 = \mathbf{x}_0$. Here \mathbf{W}_t is a $\delta \times m$ matrix Wiener process and \mathbf{x}_0 is a $\delta \times m$ deterministic initial matrix; $b, c \in \Re$. By setting

$$\mathbf{S}_t = \mathbf{X}_t^\top \mathbf{X}_t, \quad \mathbf{s}_0 = \mathbf{x}_0^\top \mathbf{x}_0 \quad (2.3.42)$$

and denoting $d\tilde{\mathbf{W}}_t = \sqrt{\mathbf{S}_t^{-1}} \mathbf{X}_t^\top d\mathbf{W}_t$ we obtain an $m \times m$ matrix Wiener process $\tilde{\mathbf{W}}_t$. Note that the elements of $\tilde{\mathbf{W}}_t$ can be correlated. Then \mathbf{S}_t solves the SDE

$$d\mathbf{S}_t = (\delta b^2 \mathbf{I} - 2c\mathbf{S}_t) dt + b(\sqrt{\mathbf{S}_t} d\tilde{\mathbf{W}}_t + d\tilde{\mathbf{W}}_t^\top \sqrt{\mathbf{S}_t}) \quad (2.3.43)$$

for $t \geq 0$, $\mathbf{S}_0 = \mathbf{s}_0$, where \mathbf{I} is the identity matrix. Here \mathbf{S}_t corresponds to a continuous-time process of stochastic, symmetric, positive definite matrices, while $\sqrt{\mathbf{S}_t}$ is the positive symmetric square root of the matrix \mathbf{S}_t , see Gouriéroux & Sufana (2004). Furthermore, \mathbf{S}_t^{-1} is the inverse of the symmetric positive definite $m \times m$ matrix \mathbf{S}_t and $\sqrt{\mathbf{S}_t^{-1}}$ its square root.

The matrix SR-process \mathbf{S} can be simulated given the above matrix OU-process and using the transform (2.3.42). Note that for $m = 1$ the transform (2.3.42) simplifies to (2.3.36). We illustrate in Fig. 2.3.6 the matrix SR-process obtained from the matrix OU-process from Fig. 2.3.5. Note that not all elements of such a matrix always remain positive. The elements $S^{1,2}$ and $S^{2,1}$ are identical and, in general, need not be positive. Most importantly, the diagonal elements $S^{1,1}$ and $S^{2,2}$ are correlated SR-processes, which are always positive.

Multi-dimensional Squared Bessel Processes

Another important stochastic process in financial and other modeling is the squared Bessel process (BESQ $_{\delta}^{\varphi}$) $X = \{X_{\varphi}, \varphi \in [\varphi_0, \infty)\}$, $\varphi_0 \geq 0$, of dimension $\delta \geq 0$, see Revuz & Yor (1999). We present this scalar process here, since the solution of the corresponding SDE can be simulated exactly in a

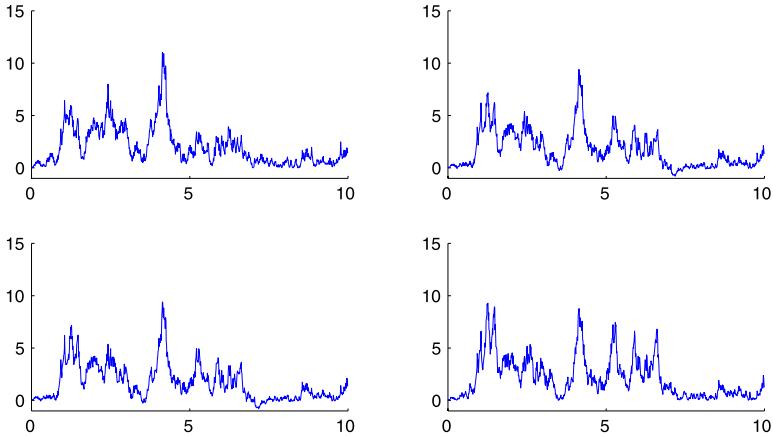


Fig. 2.3.6. Matrix valued square root process

convenient way for the case when the dimension of this process is an integer $\delta \in \{1, 2, \dots\}$. This process can be described by the following SDE

$$dX_\varphi = \delta d\varphi + 2 \sqrt{|X_\varphi|} dW_\varphi \quad (2.3.44)$$

for $\varphi \in [\varphi_0, \infty)$ with $X_{\varphi_0} = x \geq 0$, where $W = \{W_\varphi, \varphi \in [\varphi_0, \infty)\}$ is a standard Wiener process in φ -time starting at the initial φ -time, $\varphi = \varphi_0$, at zero. This means, for $\varphi \in [\varphi_0, \infty)$ one has the increment in the quadratic variation of W as

$$[W]_\varphi - [W]_{\varphi_0} = \varphi - \varphi_0$$

for all $\varphi \in [\varphi_0, \infty)$. Furthermore, if we fix the behavior of X_φ at zero as reflection, then the absolute sign under the square root in (2.3.44) can be removed, and X_φ remains nonnegative and has a unique strong solution, see Revuz & Yor (1999).

It is a useful fact that for $\delta \in \{1, 2, \dots\}$ and $x \geq 0$ the dynamics of a BESQ_x^δ process X can be expressed as the sum of the squares of δ independent Wiener processes $W^1, W^2, \dots, W^\delta$ in φ -time, which start at time $\varphi = 0$ in $w^1 \in \mathbb{R}, w^2 \in \mathbb{R}, \dots, w^\delta \in \mathbb{R}$, respectively, such that

$$x = \sum_{k=1}^{\delta} (w^k)^2. \quad (2.3.45)$$

We can now construct a solution of (2.3.44) as follows

$$X_\varphi = \sum_{k=1}^{\delta} (w^k + W_\varphi^k)^2 \quad (2.3.46)$$

for $\varphi \in [0, \infty)$. Applying the Itô formula we obtain

$$dX_\varphi = \delta d\varphi + 2 \sum_{k=1}^{\delta} (w^k + W_\varphi^k) dW_\varphi^k \quad (2.3.47)$$

for $\varphi \in [0, \infty)$ with

$$X_0 = \sum_{k=1}^{\delta} (w^k)^2 = x. \quad (2.3.48)$$

Furthermore, by setting

$$dW_\varphi = |X_\varphi|^{-\frac{1}{2}} \sum_{k=1}^{\delta} (w^k + W_\varphi^k) dW_\varphi^k \quad (2.3.49)$$

we satisfy with (2.3.46) the SDE (2.3.44). Note that we have for W_φ the quadratic variation

$$[W]_\varphi = \int_0^\varphi \frac{1}{X_s} \sum_{k=1}^{\delta} (w^k + W_s^k)^2 ds = \varphi. \quad (2.3.50)$$

Hence, by the Lévy theorem, see Theorem 1.3.3, W_φ is a Wiener process in φ -time.

Wishart Process

The matrix generalization of a squared Bessel process is a Wishart process, see Bru (1991). The $m \times m$ matrix valued Wishart process with dimension $\delta \in \{1, 2, \dots\}$ is the matrix process $\mathbf{S} = \{\mathbf{S}_t, t \geq 0\}$ with

$$\mathbf{S}_t = \mathbf{W}_t^\top \mathbf{W}_t \quad (2.3.51)$$

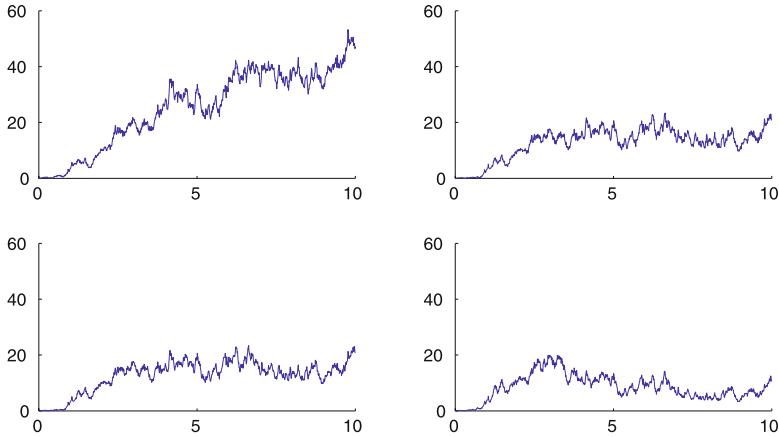
and initial matrix $\mathbf{s}_0 = \mathbf{W}_0^\top \mathbf{W}_0$ for $t \in \mathbb{R}^+$, where \mathbf{W}_t is the value at time $t \geq 0$ of a $\delta \times m$ matrix Wiener process. Itô calculus applied to the relation (2.3.51) results in the following SDE

$$d\mathbf{S}_t = \delta \mathbf{I} dt + d\mathbf{W}_t^\top \mathbf{W}_t + \mathbf{W}_t^\top d\mathbf{W}_t, \quad (2.3.52)$$

where \mathbf{I} is the $m \times m$ identity matrix. It can be shown that $\tilde{\mathbf{W}}_t$ expressed by

$$d\tilde{\mathbf{W}}_t = \left(\sqrt{\mathbf{S}_t} \right)^{-1} \mathbf{W}_t^\top d\mathbf{W}_t \quad (2.3.53)$$

is an $m \times m$ matrix Wiener process. Here $\sqrt{\mathbf{S}_t}$ represents the symmetric positive square root of \mathbf{S}_t , while $(\sqrt{\mathbf{S}_t})^{-1}$ is the inverse of the matrix $\sqrt{\mathbf{S}_t}$. Note also that

**Fig. 2.3.7.** Wishart process

$$\begin{aligned} d\tilde{\mathbf{W}}_t^\top &= d\mathbf{W}_t^\top \mathbf{W}_t \left((\sqrt{\mathbf{S}_t})^{-1} \right)^\top = d\mathbf{W}_t^\top \mathbf{W}_t \left((\sqrt{\mathbf{S}_t})^\top \right)^{-1} \quad (2.3.54) \\ &= d\mathbf{W}_t^\top \mathbf{W}_t \left(\sqrt{\mathbf{S}_t^\top} \right)^{-1} = d\mathbf{W}_t^\top \mathbf{W}_t \left(\sqrt{\mathbf{S}_t} \right)^{-1}, \end{aligned}$$

since \mathbf{S}_t is a symmetric matrix. Therefore, (2.3.52) can be rewritten in the following form

$$d\mathbf{S}_t = \delta \mathbf{I} dt + \sqrt{\mathbf{S}_t} d\tilde{\mathbf{W}}_t + d\tilde{\mathbf{W}}_t^\top \sqrt{\mathbf{S}_t} \quad (2.3.55)$$

for $t \in \mathbb{R}^+$.

In Fig. 2.3.2 we showed the trajectories of the elements of a 2×2 matrix Wiener process. Now, in Fig. 2.3.7 we plot a 2×2 Wishart process of dimension $\delta = 2$ obtained from the Wiener process in Fig. 2.3.2. Recall that the matrix Wiener process in this example was obtained by assuming the covariance matrix $\mathbf{I} \otimes \boldsymbol{\Sigma}_1$, where $\boldsymbol{\Sigma}_1$ is as in (2.3.13).

SR-Processes via Squared Bessel Processes

Using squared Bessel processes one can derive SR-processes by certain transformations. For this reason let $c : [0, \infty) \rightarrow \mathbb{R}$ and $b : [0, \infty) \rightarrow \mathbb{R}$ be given deterministic functions of time. We introduce the exponential

$$s_t = s_0 \exp \left\{ \int_0^t c_u du \right\} \quad (2.3.56)$$

and the φ -time

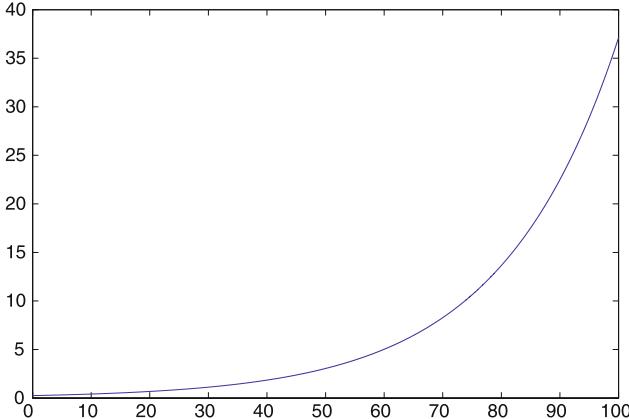


Fig. 2.3.8. Time $\varphi(t)$ against time t

$$\varphi(t) = \varphi(0) + \frac{1}{4} \int_0^t \frac{b_u^2}{s_u} du \quad (2.3.57)$$

for $t \in [0, \infty)$ and $s_0 > 0$ dependent on t -time. Note that we have an explicit representation for the function $\varphi(t)$ in the case of constant parameters $b_t = \bar{b} \neq 0$ and $c_t = \bar{c} \neq 0$, where

$$\varphi(t) = \varphi(0) + \frac{\bar{b}^2}{4\bar{c}s_0} (1 - \exp\{-\bar{c}t\}) \quad (2.3.58)$$

for $t \in [0, \infty)$ and $s_0 > 0$. Furthermore, if $\varphi(0) = -\frac{\bar{b}^2}{4\bar{c}s_0}$, this function simply equals

$$\varphi(t) = -\frac{\bar{b}^2}{4\bar{c}s_0} \exp\{-\bar{c}t\} \quad (2.3.59)$$

for $t \in [0, \infty)$, $s_0 > 0$, $\bar{b} \neq 0$ and $\bar{c} \neq 0$. We show the function $\varphi(t)$ in Fig. 2.3.8 for the choice of $\bar{b} = 1$, $\bar{c} = -0.05$, $s_0 = 20$ and $\varphi(0) = -\frac{\bar{b}^2}{4\bar{c}s_0} = 0.25$. The function $\varphi(t)$ is a time transformation, which when applied to the BESQ $^\delta_x$ yields the following expected value of the time transformed squared Bessel process

$$E(X_{\varphi(t)} | \mathcal{A}_{\varphi(0)}) = X_{\varphi(0)} + \delta(\varphi(t) - \varphi(0)), \quad (2.3.60)$$

for $t \in [0, \infty)$. Note also that for constant parameters $b_t = \bar{b} \neq 0$, $c_t = \bar{c} \neq 0$ and $X_{\varphi(0)} = -\frac{\delta\bar{b}^2}{4\bar{c}s_0}$ this expected value simplifies to

$$E(X_{\varphi(t)} | \mathcal{A}_{\varphi(0)}) = -\frac{\delta\bar{b}^2}{4\bar{c}s_0} \exp\{-\bar{c}t\} \quad (2.3.61)$$

for $t \in [0, \infty)$.

Given a squared Bessel process X of dimension $\delta > 0$ and using our previous notation we introduce the SR-process $Y = \{Y_t, t \geq 0\}$ of dimension $\delta > 0$ with

$$Y_t = s_t X_{\varphi(t)} \quad (2.3.62)$$

in dependence on time $t \geq 0$, see also [Delbaen & Shirakawa \(2002\)](#).

Further, by [\(2.3.44\)](#), [\(2.3.62\)](#), [\(2.3.56\)](#) and [\(2.3.57\)](#) and the Itô formula we can express [\(2.3.62\)](#) in terms of the SDE

$$dY_t = \left(\frac{\delta}{4} b_t^2 + c_t Y_t \right) dt + b_t \sqrt{Y_t} dU_t \quad (2.3.63)$$

for $t \in [0, \infty)$, $Y_0 = s_0 X_{\varphi(0)}$ and

$$dU_t = \sqrt{\frac{4s_t}{b_t^2}} dW_{\varphi(t)}.$$

Note that U_t forms by the Lévy theorem, see [Theorem 1.3.3](#), a Wiener process, since

$$[U]_t = \int_0^t \frac{4s_z}{b_z^2} d\varphi(z) = t. \quad (2.3.64)$$

The same time-change formula applies in the more general matrix case. Given a Wishart process \mathbf{X} it can be shown, that the matrix square root process can be obtained from the Wishart process by the following transformation

$$\mathbf{Y}_t = s_t \mathbf{X}_{\varphi(t)}, \quad (2.3.65)$$

where s_t and $\varphi(t)$ are as in [\(2.3.56\)](#) and [\(2.3.57\)](#), respectively. By [\(2.3.55\)](#), [\(2.3.65\)](#), [\(2.3.56\)](#) and [\(2.3.57\)](#) and the Itô formula we can express [\(2.3.65\)](#) in terms of the matrix SDE

$$d\mathbf{Y}_t = \left(\frac{\delta}{4} b_t^2 \mathbf{I} + c_t \mathbf{Y}_t \right) dt + \frac{b_t}{2} \left(\sqrt{\mathbf{Y}_t} d\mathbf{U}_t + d\mathbf{U}_t^\top \sqrt{\mathbf{Y}_t} \right) \quad (2.3.66)$$

for $t \in [0, \infty)$, $\mathbf{Y}_0 = s_0 \mathbf{X}_{\varphi(0)}$ and where $d\mathbf{U}_t = \sqrt{\frac{4s_t}{b_t^2}} d\mathbf{W}_{\varphi(t)}$ is the differential of a matrix Wiener process.

In [Fig. 2.3.9](#) we display the trajectory of the elements of a 2×2 matrix time changed Wishart process $\mathbf{X}_{\varphi(t)}$ in log-scale. Here the off-diagonal elements do not show any value for the time periods when the argument of the logarithm becomes negative. Indeed we see in [Fig. 2.3.9](#) that the off-diagonal elements have such negative values near the time $t = 7$. This is not the case for the diagonal elements which are of main interest in most applications. In [Fig. 2.3.10](#) we show the corresponding trajectory of a 2×2 matrix SR-process obtained from the time changed Wishart process by the use of formula [\(2.3.65\)](#). Note that this matrix SR-process is identical to the matrix SR-process in [Fig. 2.3.6](#) obtained via squares of OU-processes.

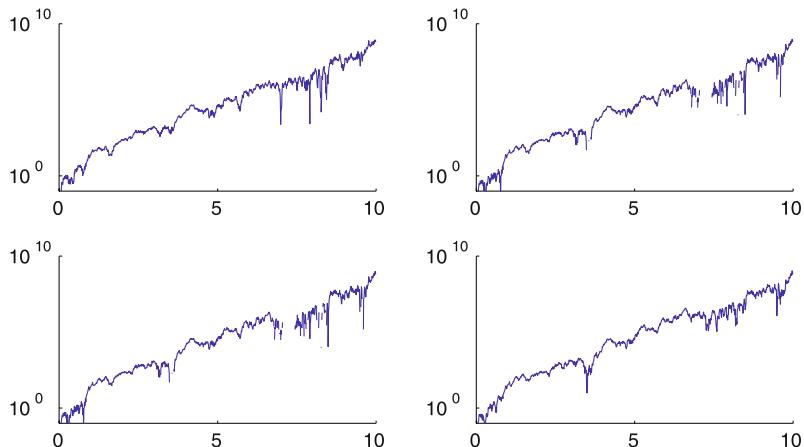


Fig. 2.3.9. Time changed Wishart process in log-scale

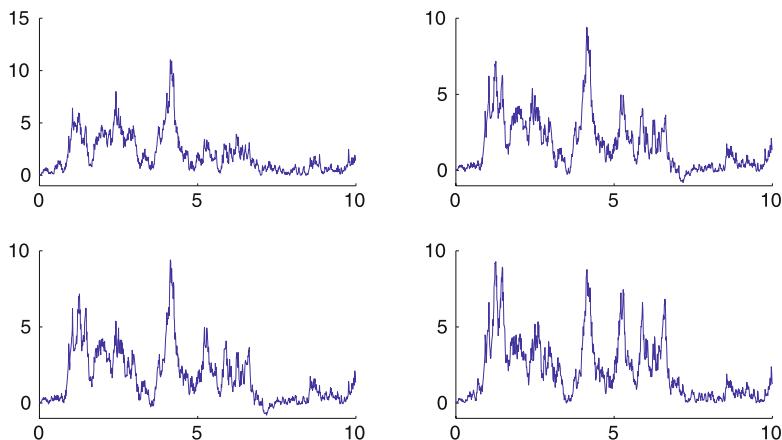


Fig. 2.3.10. Matrix valued square root process

Multi-dimensional Affine Processes

Let us now further transform the above obtained multi-dimensional SR-process in order to obtain multi-dimensional affine processes, see Sect. 1.6. These processes have affine, that is linear drift and linear squared diffusion coefficients. In order to obtain members of this class of multi-dimensional processes we can simply shift the multi-dimensional SR-process by a nonnegative, differentiable function of time $a : [0, \infty) \rightarrow [0, \infty)$ characterized by its derivative

$$a'_t = \frac{da_t}{dt} \quad (2.3.67)$$

for $t \in [0, \infty)$ with $a_0 \in [0, \infty)$. More precisely, we define the process $\mathbf{R} = \{\mathbf{R}_t, t \in [0, \infty)\}$ such that

$$\mathbf{R}_t = \mathbf{Y}_t + a_t \mathbf{I} \quad (2.3.68)$$

for $t \in [0, \infty)$. It is also possible to obtain more general affine processes by shifting the matrix valued SR-process by a matrix \mathbf{A}_t of nonnegative differentiable functions of the type (2.3.67). That is

$$\mathbf{R}_t = \mathbf{Y}_t + \mathbf{A}_t \quad (2.3.69)$$

for $t \in [0, \infty)$. In this case \mathbf{R}_t solves the following matrix SDE

$$d\mathbf{R}_t = \left(\frac{\delta}{4} b_t^2 \mathbf{I} + \mathbf{A}'_t - c_t \mathbf{A}_t + c_t \mathbf{R}_t \right) dt + \frac{b_t}{2} \left(\sqrt{\mathbf{R}_t - \mathbf{A}_t} d\tilde{\mathbf{W}}_t + d\tilde{\mathbf{W}}_t^\top \sqrt{\mathbf{R}_t - \mathbf{A}_t} \right), \quad (2.3.70)$$

for $t \in [0, \infty)$. Here \mathbf{A}'_t denotes the matrix of the derivatives of the type (2.3.67) for the shifts of each element. In principle, we applied here the Itô formula to the equation (2.3.69).

Multi-dimensional Geometric Ornstein-Uhlenbeck Processes

The Itô formula provides a general tool to generate a world of exact solutions of SDEs based on functions of the solutions of those processes we have already considered. As an example, let us generate explicit solutions for a geometric OU-process. Here each element of a matrix valued OU-process is simply exponentiated. That is, denoting by $\mathbf{Y}_t = [Y_t^{j,k}]_{j,k=1}^{d,m}$ the corresponding $d \times m$ matrix geometric OU-process value at time t and by $\mathbf{X}_t = [X_t^{j,k}]_{j,k=1}^{d,m}$ the $d \times m$ matrix OU-process value. We obtain the elements of the matrix \mathbf{Y}_t by

$$Y_t^{j,k} = \exp\{X_t^{j,k}\} \quad (2.3.71)$$

for $t \in [0, \infty)$ and $j \in \{1, 2, \dots, d\}$ and $k \in \{1, 2, \dots, m\}$.

In Fig. 2.3.11 we illustrate a 2×2 matrix geometric OU-process obtained from the matrix OU-process in Fig. 2.3.5 by application of (2.3.71) to each of its elements. More complex applications of the Itô formula generating exact solutions will be considered in the next section.

Multi-dimensional SDEs Driven by Lévy Processes

So far in this section we have considered the exact simulation of solutions of multi-dimensional SDEs driven by vector or matrix Wiener processes. However, the simulation methods described here can be adapted to multi-dimensional SDEs which are driven by more general vector or matrix valued

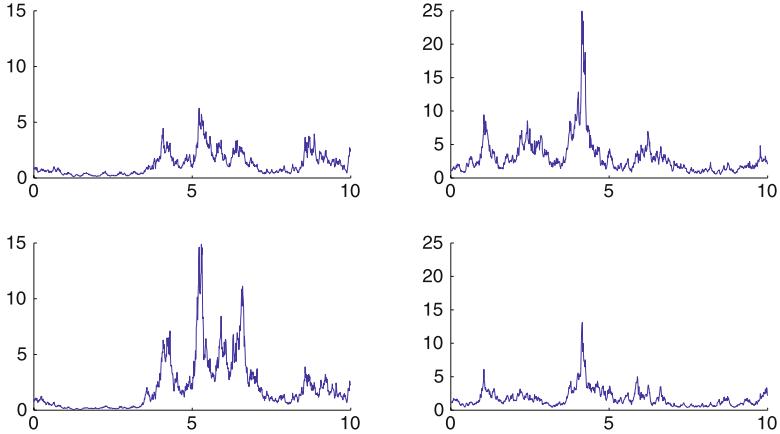


Fig. 2.3.11. Matrix valued geometric OU-process

Lévy processes, see Sect. 1.7. In principle, one can substitute the Wiener processes by some Lévy processes.

Since Lévy processes have independent stationary increments, it is possible to construct paths of a wide range of d -dimensional Lévy processes $\mathbf{L} = \{\mathbf{L}_t, t \geq 0\}$ at given discretization times $t_i = i\Delta$, $i \in \{0, 1, \dots\}$, with fixed time step size $\Delta > 0$. The distribution of the Lévy increments $\mathbf{L}_{t_{i+1}} - \mathbf{L}_{t_i}$, however, must be infinitely divisible for the process \mathbf{L} to be the transition distribution of a Lévy process, see Sect. 2.2. One example of a family of infinitely divisible distributions is the generalized hyperbolic (GH) distribution, see for instance [McNeil et al. \(2005\)](#). This family of distributions yields variance gamma (VG) and normal inverse Gaussian (NIG) processes as special cases.

Simulation of the d -dimensional VG and NIG processes results from their representation as subordinated vector Wiener processes with drift. That is,

$$\mathbf{L}_t = \mathbf{a}V_t + \mathbf{B}\mathbf{W}_{V_t}, \quad (2.3.72)$$

for $t \in [0, \infty)$. Here $\mathbf{a} = (a_1, a_2, \dots, a_d)^\top$ is a d -dimensional vector, \mathbf{B} is a $d \times m$ -matrix and $\mathbf{W} = \{\mathbf{W}_V = (W_V^1, W_V^2, \dots, W_V^m)^\top, V \in [0, \infty)\}$ is a standard m -dimensional vector Wiener process in V time. When V is the gamma process or the inverse Gaussian process, we obtain the d -dimensional VG process and the NIG process, respectively.

We also define $d \times m$ matrix VG and NIG processes by

$$\mathbf{L}_t = \mathbf{M}V_t + \boldsymbol{\Sigma}_1\mathbf{W}_{V_t}\boldsymbol{\Sigma}_2, \quad (2.3.73)$$

where \mathbf{M} is a $d \times m$ matrix and $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are nonsingular $d \times d$ and $m \times m$ matrices, respectively. Here $\mathbf{W} = \{\mathbf{W}_V = [W_V^{i,j}]_{i,j=1}^{d,m}, V \in [0, \infty)\}$ is

a standard $d \times m$ matrix Wiener process and V a gamma or inverse Gaussian process, respectively.

Processes of type (2.3.72) and (2.3.73) possess a number of useful properties because they are conditionally Gaussian. In particular, if one knows how to simulate the increments of the subordinator V , the values of \mathbf{L} in (2.3.73) can be obtained at the discrete times $t_i = i\Delta$ by the following recursive computation

$$\begin{aligned}\mathbf{L}_0 &= \mathbf{0} \\ \mathbf{L}_{t_{i+1}} &= \mathbf{L}_{t_i} + \mathbf{M} \Delta V_{i+1} + \sqrt{\Delta V_{i+1}} \tilde{\mathbf{N}}_{i+1},\end{aligned}\quad (2.3.74)$$

for $i \in \{0, 1, \dots\}$ and $\tilde{\mathbf{N}}_{i+1} \sim \mathcal{N}_{d \times m}(\mathbf{0}, \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1)$. Here the covariance matrix $\boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2$ is as in (2.3.18).

The VG process is obtained by (2.3.74), where $\Delta V_{i+1} \sim \kappa \text{Ga}(\frac{\Delta}{\kappa}, 1)$ are gamma random variables for $i \in \{1, 2, \dots\}$, while the NIG process is obtained by (2.3.74) when $\Delta V_{i+1} \sim \text{IGaussian}(\frac{\Delta^2}{\kappa}, \Delta)$ are inverse Gaussian random variables for $i \in \{1, 2, \dots\}$. Here the parameter κ is the variance of the subordinator V . See also [Cont & Tankov \(2004\)](#) who describe exact simulation of scalar VG and NIG processes. They describe also convenient algorithms for generators of gamma and inverse Gaussian random variables.

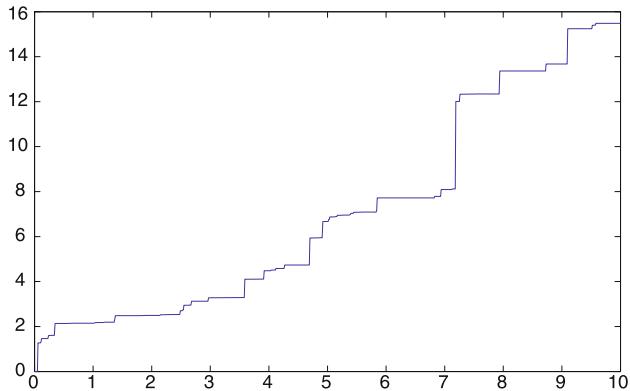
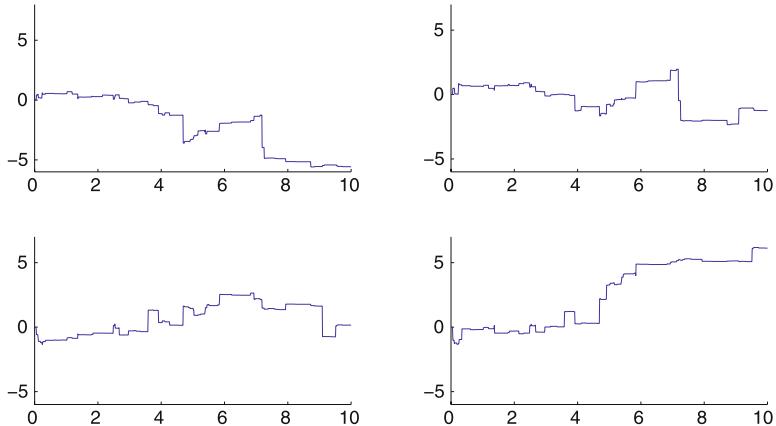
Since we can simulate the paths of such driving Lévy processes exactly it is possible to simulate solutions for the type of the above introduced SDEs when driven by Lévy noise. For instance, let us consider a Wishart process of dimension δ driven by a VG-process. That is, we consider the multi-dimensional SDE of the form

$$d\mathbf{S}_t = \delta \mathbf{I} dt + \sqrt{\mathbf{S}_t} d\mathbf{L}_t + d\mathbf{L}_t^\top \sqrt{\mathbf{S}_t} \quad (2.3.75)$$

for $t \in \mathbb{R}^+$. In order to simulate this Wishart process, which may be driven by the VG-process \mathbf{L} , we first need to simulate a $\delta \times m$ matrix VG-process. Afterwards we obtain the $m \times m$ VG-Wishart process of dimension $\delta \in \{1, 2, \dots\}$ by setting $\mathbf{S}_t = \mathbf{L}_t^\top \mathbf{L}_t$, for $t \in \mathbb{R}^+$.

In Fig. 2.3.12 we show a trajectory of a gamma process, which is always nondecreasing. Here we have chosen $\kappa = 1$. Moreover, in Fig. 2.3.13 we display a 2×2 matrix VG-process, with parameters $\mathbf{M} = \mathbf{0}$ and the covariance matrix $\mathbf{I} \otimes \boldsymbol{\Sigma}_1$, where $\boldsymbol{\Sigma}_1 = B$ is as in (2.3.13) with $\varrho = 0.8$. The subordinator is here chosen to be the gamma process illustrated in Fig 2.3.12. Additionally, in Fig. 2.3.14 we display the corresponding trajectory of the resulting 2×2 Wishart process of dimension $\delta = 2$.

The subordination methodology can be widely applied to generate trajectories of other matrix Lévy processes, for instance, matrix Lévy OU-processes and matrix Lévy-affine processes.

**Fig. 2.3.12.** Gamma process**Fig. 2.3.13.** Matrix VG-process

2.4 Functions of Exact Solutions

Another possibility to obtain a multi-dimensional SDE which is explicitly solvable, is by application of the Itô formula, see also [Kloeden & Platen \(1999\)](#). Let us illustrate this below for linear SDEs driven by a vector of standard Wiener processes.

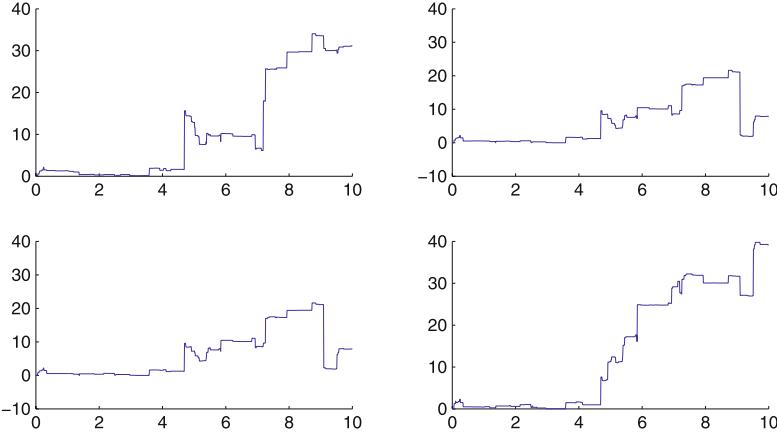


Fig. 2.3.14. Wishart process driven by the matrix VG-process in Fig. 2.3.13

Multi-dimensional Itô Formula Application

We consider an m -dimensional Wiener process $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \in [0, \infty)\}$, a d -dimensional drift coefficient vector function $\mathbf{a} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and a $d \times m$ -matrix diffusion coefficient function $\mathbf{b} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$. In this framework we assume that we have already a general family of explicitly solvable d -dimensional SDEs given as

$$d\mathbf{X}_t = \mathbf{a}(t, \mathbf{X}_t)dt + \mathbf{b}(t, \mathbf{X}_t)d\mathbf{W}_t, \quad (2.4.1)$$

for $t \in [0, \infty)$, $\mathbf{X}_0 \in \mathbb{R}^d$. This means that the k th component of (2.4.1) equals

$$dX_t^k = a^k(t, \mathbf{X}_t)dt + \sum_{j=1}^m b^{k,j}(t, \mathbf{X}_t)dW_t^j. \quad (2.4.2)$$

For a sufficiently smooth vector function $\mathbf{U} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^k$ of the solution \mathbf{X}_t of (2.4.1) we obtain a k -dimensional process

$$\mathbf{Y}_t = \mathbf{U}(t, \mathbf{X}_t). \quad (2.4.3)$$

The expression for its p th component, resulting from the application of the Itô formula, satisfies the SDE

$$\begin{aligned} dY_t^p &= \left(\frac{\partial U^p}{\partial t} + \sum_{i=1}^d a^i \frac{\partial U^p}{\partial x^i} + \frac{1}{2} \sum_{i,j=1}^d \sum_{l=1}^m b^{i,l} b^{j,l} \frac{\partial^2 U^p}{\partial x_i \partial x_j} \right) dt \\ &\quad + \sum_{l=1}^m \sum_{i=1}^d b^{i,l} \frac{\partial U^p}{\partial x_i} dW_t^l, \end{aligned} \quad (2.4.4)$$

for $p \in \{1, 2, \dots, k\}$, where the terms on the right-hand side of (2.4.4) are evaluated at (t, \mathbf{X}_t) . Obviously, also the paths of the solution of the SDE (2.4.4) can be exactly simulated since \mathbf{X}_{t_i} can be obtained at all discretization points and by (2.4.3) the vector \mathbf{Y}_{t_i} is just a function of the components of the vector \mathbf{X}_{t_i} .

Multi-dimensional Linear SDEs

In practice, the multi-dimensional Itô formula turns out to be a useful tool if one wants to construct solutions of certain multi-dimensional SDEs in terms of known solutions of other SDEs. Let us illustrate this with an example involving linear SDEs, where we rely on the results of Sect. 1.7. Recall from (1.7.20) that a d -dimensional linear SDE can be expressed in the form

$$d\mathbf{X}_t = (\mathbf{A}_t \mathbf{X}_t + \boldsymbol{\alpha}_t) dt + \sum_{l=1}^m \left(\mathbf{B}_t^l \mathbf{X}_t + \boldsymbol{\beta}_t^l \right) dW_t^l \quad (2.4.5)$$

for $t \geq 0$ with $\mathbf{X}_0 \in \mathbb{R}^d$. Here $\mathbf{A}, \mathbf{B}^1, \mathbf{B}^2, \dots, \mathbf{B}^m$ are deterministic $d \times d$ matrix functions of time and $\boldsymbol{\alpha}, \boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \dots, \boldsymbol{\beta}^m$ are deterministic d -dimensional vector functions of time. It is possible to express according to (1.7.21) the solution of (2.4.5) in the form

$$\mathbf{X}_t = \boldsymbol{\Psi}_t \left(\mathbf{X}_0 + \int_0^t \boldsymbol{\Psi}_s^{-1} \left(\boldsymbol{\alpha}_s - \sum_{l=1}^m \mathbf{B}_s^l \boldsymbol{\beta}_s^l \right) ds + \sum_{l=1}^m \int_0^t \boldsymbol{\Psi}_s^{-1} \boldsymbol{\beta}_s^l dW_s^l \right). \quad (2.4.6)$$

Here $\boldsymbol{\Psi}_t$ is the $d \times d$ fundamental matrix satisfying $\boldsymbol{\Psi}_0 = \mathbf{I}$ and the homogeneous matrix SDE

$$d\boldsymbol{\Psi}_t = \mathbf{A}_t \boldsymbol{\Psi}_t dt + \sum_{l=1}^m \mathbf{B}_t^l \boldsymbol{\Psi}_t dW_t^l, \quad (2.4.7)$$

see (1.7.22). Unfortunately, it is not possible to solve (2.4.6) explicitly for its fundamental solution in its general form. However, if the matrices $\mathbf{A}, \mathbf{B}^1, \mathbf{B}^2, \dots, \mathbf{B}^m$ are constant and commute, that is if

$$\mathbf{AB}^l = \mathbf{B}^l \mathbf{A} \quad \text{and} \quad \mathbf{B}^l \mathbf{B}^k = \mathbf{B}^k \mathbf{B}^l \quad (2.4.8)$$

for all $k, l \in \{1, 2, \dots, m\}$, then the explicit expression for the fundamental matrix solution is given as

$$\boldsymbol{\Psi}_t = \exp \left\{ \left(\mathbf{A} - \frac{1}{2} \sum_{l=1}^m (\mathbf{B}^l)^2 \right) t + \sum_{l=1}^m \mathbf{B}^l W_t^l \right\}, \quad (2.4.9)$$

where the exponential is interpreted elementwise, see (1.7.26). This allows to cover interesting multi-dimensional models that are relevant to finance and other areas of application.

Multi-dimensional Black-Scholes Model

We will now describe the multi-dimensional Black-Scholes model, which is the standard asset price model in finance. This model emerges from (2.4.5) by assuming that α and β^l equal zero for all $l \in \{1, 2, \dots, m\}$, see also (1.7.27)–(1.7.31). Denote by \mathbf{S}_t a diagonal matrix with j th diagonal element S_t^j , $j \in \{1, 2, \dots, d\}$, representing the j th asset price at time $t \in [0, \infty)$. Then the SDE for the j th Black-Scholes asset price S_t^j is defined by

$$dS_t^j = S_t^j \left(a_t^j dt + \sum_{k=1}^d b_t^{j,k} dW_t^k \right) \quad (2.4.10)$$

for $t \in [0, \infty)$ and $j \in \{1, 2, \dots, d\}$, where a^j and $b^{j,k}$ are deterministic functions of time, see (1.7.13) and (1.7.31). Here W^k , $k \in \{1, 2, \dots, d\}$, denote independent standard Wiener processes. To represent this SDE in matrix form we introduce the diagonal matrix $\mathbf{A}_t = [A_t^{i,j}]_{i,j=1}^d$ with

$$A_t^{i,j} = \begin{cases} a_t^j & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \quad (2.4.11)$$

and diagonal matrix $\mathbf{B}_t^k = [B_t^{k,i,j}]_{i,j=1}^d$ with

$$B_t^{k,i,j} = \begin{cases} b_t^{j,k} & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \quad (2.4.12)$$

for $k, i, j \in \{1, 2, \dots, d\}$ and $t \in [0, \infty)$. All these diagonal matrices commute in the sense of (2.4.8), therefore, we can write the SDE (2.4.10) as the matrix SDE

$$d\mathbf{S}_t = \mathbf{A}_t \mathbf{S}_t dt + \sum_{k=1}^d \mathbf{B}_t^k \mathbf{S}_t dW_t^k \quad (2.4.13)$$

for $t \in [0, \infty)$. Consequently, we obtain for the j th asset price the explicit solution

$$S_t^j = S_0^j \exp \left\{ \int_0^t \left(a_s^j - \frac{1}{2} \sum_{k=1}^d (b_s^{j,k})^2 \right) ds + \sum_{k=1}^d \int_0^t b_s^{j,k} dW_s^k \right\} \quad (2.4.14)$$

for $t \in [0, \infty)$ and $j \in \{1, 2, \dots, d\}$. When taking the following exponential elementwise, the explicit solution of (2.4.10) can be expressed as

$$\mathbf{S}_t = \mathbf{S}_0 \exp \left\{ \int_0^t \left(\mathbf{A}_s - \frac{1}{2} \sum_{k=1}^d (\mathbf{B}_s^k)^2 \right) ds + \sum_{k=1}^d \int_0^t \mathbf{B}_s^k dW_s^k \right\} \quad (2.4.15)$$

for $t \geq 0$, see (1.7.31). If the appreciation rates and volatilities are piecewise constant, then we can simulate exact solutions. In the case where these parameters are time dependent, one can generate in a straightforward manner

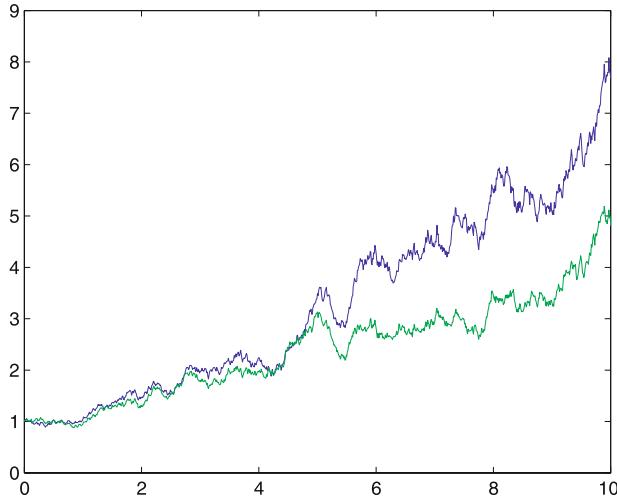


Fig. 2.4.1. Trajectory of a two-dimensional Black-Scholes model with parameters $S_0^1 = S_0^2 = 1$, $a_1 = a_2 = 0.1$, $b_1 = b_2 = 0.2$ and $\varrho = 0.8$

almost exact solutions by using, for instance, the trapezoidal rule. The main advantage of the multi-dimensional Black-Scholes model, which made it so popular, is that it has an explicit solution for the entire market dynamics and allows a wide range of easy calculations.

For a two-dimensional Black-Scholes model with \mathbf{B}^1 and \mathbf{B}^2 we obtain the following exact solution

$$S_t^1 = S_0^1 \exp \left\{ \left(a_1 - \frac{1}{2} b_1^2 \right) t + b_1 W_t^1 \right\}, \quad (2.4.16)$$

$$S_t^2 = S_0^2 \exp \left\{ \left(a_2 - \frac{1}{2} b_2^2 \right) t + b_2 (\varrho W_t^1 + \sqrt{1 - \varrho^2} W_t^2) \right\}, \quad (2.4.17)$$

for $t \in [0, \infty)$. The two components of the trajectory of this two-dimensional model are illustrated in Fig. 2.4.1 for the parameter choice $S_0^1 = S_0^2 = 1$, $a_1 = a_2 = 0.1$, $b_1 = b_2 = 0.2$ and $\varrho = 0.8$.

Multi-dimensional Linear Diffusion Model

The continuous time limits of some popular time series models in finance can be described by a multi-dimensional ARCH diffusion model, see Sect. 1.7. The squared volatilities of this particular model emerge by assuming β^l to be zero in (2.4.5) for all $l \in \{1, 2, \dots, m\}$.

We obtain an exact solution for the above multi-dimensional linear diffusion process using (2.4.6) in the following form

$$\mathbf{X}_t = \boldsymbol{\Psi}_t \left(\mathbf{X}_0 + \int_0^t \boldsymbol{\Psi}_s^{-1} \boldsymbol{\alpha}_s ds \right), \quad (2.4.18)$$

where, as before, $\boldsymbol{\Psi}_t$ satisfies (2.4.7). The multi-dimensional linear diffusion process can be simulated almost exactly provided that the matrices \mathbf{A} , \mathbf{B}^1 , $\mathbf{B}^2, \dots, \mathbf{B}^m$ commute. It simply remains to approximate the time integral in (2.4.18), which can be achieved with high accuracy for small time step size $\Delta > 0$ by using a quadrature formula, for instance, the trapezoidal rule when integrating between time discretization points. More precisely, we substitute for $t = t_i = i\Delta$, $i \in \{0, 1, \dots\}$, the expression in (2.4.18) by the approximation

$$\mathbf{X}_{t_i}^\Delta = \boldsymbol{\Psi}_{t_i} \left(\mathbf{X}_0 + \sum_{k=0}^{i-1} \frac{\Delta}{2} \left(\boldsymbol{\Psi}_{t_{k+1}}^{-1} \boldsymbol{\alpha}_{t_{k+1}} + \boldsymbol{\Psi}_{t_k}^{-1} \boldsymbol{\alpha}_{t_k} \right) \right). \quad (2.4.19)$$

Other approximations are also possible. We remark that as a consequence of the methods described in Kloeden & Platen (1999) and those we will present later in this book, it is straightforward to show that the almost exact solution which uses the trapezoidal rule converges with strong order one in the sense of Kloeden & Platen (1999) and as we will define later in this book. By making the time step size sufficiently small any desired level of accuracy can be achieved. What is important in formula (2.4.19) is that errors do not propagate here, as can be the case with more general discrete-time approximations that will be discussed later in the book. In this sense, the almost exact solutions we simulate by the above method are still very accurate also over long periods of time. Numerical stability issues do not play any role in this context.

Let us now simulate a two-dimensional linear diffusion process with the following parameter matrices

$$\mathbf{A} = \begin{pmatrix} -\kappa_1 & 0 \\ 0 & -\kappa_2 \end{pmatrix}, \quad \mathbf{B}^1 = \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \varrho \end{pmatrix}, \quad \mathbf{B}^2 = \begin{pmatrix} 0 & 0 \\ 0 & \gamma_2 \sqrt{1 - \varrho^2} \end{pmatrix} \quad (2.4.20)$$

and vector

$$\boldsymbol{\alpha} = \begin{pmatrix} \kappa_1 \bar{x}_1 \\ \kappa_2 \bar{x}_2 \end{pmatrix}. \quad (2.4.21)$$

Then the vector SDE (2.4.5) simplifies to

$$dX_t^1 = \kappa_1 (\bar{x}_1 - X_t^1) dt + \gamma_1 X_t^1 dW_t^1, \quad (2.4.22)$$

$$dX_t^2 = \kappa_2 (\bar{x}_2 - X_t^2) dt + \gamma_2 X_t^2 \left(\varrho dW_t^1 + \sqrt{1 - \varrho^2} dW_t^2 \right), \quad (2.4.23)$$

for $t \in [0, \infty)$ with $X_0^1, X_0^2 > 0$. Note that the matrices \mathbf{A} , \mathbf{B}^1 and \mathbf{B}^2 commute, hence we can use the expression (2.4.9) for determining the fundamental matrix $\boldsymbol{\Psi}_t$. In the considered two-dimensional example the logarithm of this matrix equals

$$\ln(\boldsymbol{\Psi}_t) = \begin{pmatrix} \left(-\kappa_1 - \frac{1}{2}\gamma_1^2 \right) t + \gamma_1 W_t^1 & 0 \\ 0 & \left(-\kappa_2 - \frac{1}{2}\gamma_2^2 \right) t + \gamma_2 \left(\varrho W_t^1 + \sqrt{1 - \varrho^2} W_t^2 \right) \end{pmatrix}, \quad (2.4.24)$$

for $t_0 = 0$ and $t \in [0, \infty)$. Here the logarithm is interpreted elementwise. With this fundamental matrix the almost exact solution of (2.4.22) and (2.4.23) equals

$$\begin{aligned} X_{t_i}^{\Delta,1} &= \exp \left\{ \left(-\kappa_1 - \frac{1}{2}\gamma_1^2 \right) t_i + \gamma_1 W_{t_i}^1 \right\} \\ &\times \left(X_0^1 + \kappa_1 \bar{x}_1 \sum_{k=0}^{i-1} \frac{\Delta}{2} \left[\exp \left\{ \left(\kappa_1 + \frac{1}{2}\gamma_1^2 \right) t_{k+1} - \gamma_1 W_{t_{k+1}}^1 \right\} \right. \right. \\ &\quad \left. \left. + \exp \left\{ \left(\kappa_1 + \frac{1}{2}\gamma_1^2 \right) t_k - \gamma_1 W_{t_k}^1 \right\} \right] \right), \end{aligned} \quad (2.4.25)$$

$$\begin{aligned} X_{t_i}^{\Delta,2} &= \exp \left\{ \left(-\kappa_2 - \frac{1}{2}\gamma_2^2 \right) t_i + \gamma_2 (\varrho W_{t_i}^1 + \sqrt{1-\varrho^2} W_{t_i}^2) \right\} \\ &\times \left(X_0^2 + \kappa_2 \bar{x}_2 \sum_{k=0}^{i-1} \frac{\Delta}{2} \left[\exp \left\{ \left(\kappa_2 + \frac{1}{2}\gamma_2^2 \right) t_{k+1} - \gamma_2 (\varrho W_{t_{k+1}}^1 + \sqrt{1-\varrho^2} W_{t_{k+1}}^2) \right\} \right. \right. \\ &\quad \left. \left. + \exp \left\{ \left(\kappa_2 + \frac{1}{2}\gamma_2^2 \right) t_k - \gamma_2 (\varrho W_{t_k}^1 + \sqrt{1-\varrho^2} W_{t_k}^2) \right\} \right] \right), \end{aligned} \quad (2.4.26)$$

for $t_i = \Delta i, i \in \{0, 1, \dots\}$. Given that we use a highly accurate numerical approximation for calculating the integrals in (2.4.25) and (2.4.26) we can simulate trajectories of X^1 and X^2 almost exactly, that means with any desired accuracy. This can be done by using the trapezoidal rule when approximating the time integrals in (2.4.18) and a sufficiently fine time discretization.

In Fig. 2.4.2 we display a trajectory of a two-dimensional linear diffusion process with the correlation parameter $\varrho = 0.8$, initial values $X_0^1 = X_0^2 = 1$ and $\bar{x}_1 = \bar{x}_2 = 0.5$, $\kappa_1 = \gamma_1 = 1$ and $\kappa_2 = \gamma_2 = 2$. One notes in this figure the similarities. On the other hand there are, of course differences in the paths of the two processes.

We refer to Kloeden & Platen (1999) for a list of other specific examples of explicitly solvable multi-dimensional SDEs. The above models can be further generalized by using some time changes and applying time changed Lévy processes. For instance, the above Black-Scholes model can be generalized by subordination to multi-dimensional SDEs driven by Lévy processes, yielding exponential Lévy market models, see Barndorff-Nielsen & Shephard (2001), Geman et al. (2001) and Eberlein (2002).

2.5 Almost Exact Solutions by Conditioning

For some multi-dimensional SDEs it is possible to simulate one component using its marginal distribution or other exact or almost exact methods. The

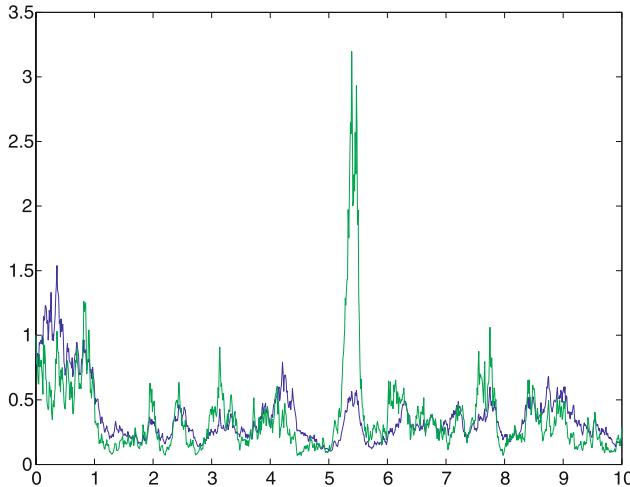


Fig. 2.4.2. Two correlated linear diffusion processes obtained by setting $\varrho = 0.8$, $X_0^1 = X_0^2 = 1$, $\bar{x}_1 = \bar{x}_2 = 0.5$, $\kappa_1 = \gamma_1 = 1$ and $\kappa_2 = \gamma_2 = 2$

second component can be conditioned on the first one, and then simulated exactly or almost exactly. This can eventually be continued to further components. Let us illustrate and describe this concept by applying it to the multi-dimensional Heston model, see [Heston \(1993\)](#) and Sect. 1.6.

One-dimensional Heston Model

The Heston model is probably the most popular stochastic volatility model used in finance. The reason why it is used so widely in practice is that [Heston \(1993\)](#) was able to derive an exact formula for European call and put option prices under this model. This underlines the relevance of the current chapter with its focus on exact solutions.

The Heston model can be expressed by a two-dimensional SDE. Its first component models the stock price S_t while its second component characterizes the squared volatility V_t of the underlying stock. This system of SDEs is typically written in the form

$$dS_t = rS_t dt + \sqrt{V_t} S_t \left[\varrho dW_t^1 + \sqrt{1 - \varrho^2} dW_t^2 \right], \quad (2.5.1)$$

$$dV_t = \kappa(\theta - V_t)dt + \sigma\sqrt{V_t} dW_t^1, \quad (2.5.2)$$

for $t \in [0, \infty)$. Here W^1 and W^2 are two independent Wiener processes, and ϱ represents the correlation parameter. Furthermore, r denotes the short rate,

κ the speed of adjustment, θ the average squared volatility and σ the volatility of the squared volatility.

There is a wide literature on approximating the Heston model by simulation. Recent studies include [Broadie & Kaya \(2006\)](#), [Smith \(2007\)](#) and [Andersen \(2008\)](#) who discuss exact or almost exact simulation techniques for the Heston model. For the purpose of simulation it is convenient to employ the logarithmic transformation $X_t = \ln(S_t)$ instead of using (2.5.1) directly. Therefore, by the Itô formula we obtain the SDE

$$\begin{aligned} dX_t &= \left(r - \frac{1}{2}V_t \right) dt + \sqrt{V_t} \left[\varrho dW_t^1 + \sqrt{1 - \varrho^2} dW_t^2 \right], \\ dV_t &= \kappa(\theta - V_t) dt + \sigma \sqrt{V_t} dW_t^1, \end{aligned} \quad (2.5.3)$$

for $t \in [0, \infty)$.

In [Broadie & Kaya \(2006\)](#), given V_{t_i} , $i \in \{0, 1, \dots\}$, the value $V_{t_i+\Delta}$ is sampled directly from the non-central chi-square distribution with δ degrees of freedom and non-centrality parameter

$$\lambda = \frac{4\kappa e^{-\kappa\Delta}}{\sigma^2(1 - e^{-\kappa\Delta})} V_{t_i}.$$

That is

$$V_{t_i+\Delta} = \frac{\sigma^2(1 - e^{-\kappa\Delta})}{4\kappa} \chi_{(\delta, \lambda)}^2, \quad (2.5.4)$$

where $\delta = \frac{4\theta\kappa}{\sigma^2}$. Here $\chi^2(\delta, \lambda)$ is sampled from the non-central chi-square distribution function

$$F_X(x) = \sum_{k=0}^{\infty} \frac{\exp\left\{-\frac{\lambda}{2}\right\} \left(\frac{\lambda}{2}\right)^k}{k!} \left(1 - \frac{\Gamma\left(\frac{x}{2}; \frac{\delta+2k}{2}\right)}{\Gamma\left(\frac{\delta+2k}{2}\right)}\right), \quad (2.5.5)$$

where

$$\Gamma(u; a) = \int_u^{\infty} t^{a-1} \exp\{-t\} dt$$

is the incomplete gamma function for $u \geq 0$, $a > -1$. Note that the non-central chi-square distribution arises as a scalar version of the non-central Wishart distribution. Sampling from the non-central chi-square distribution is discussed, for instance, in [Glasserman \(2004\)](#). The resulting simulation method for V is exact.

In order to obtain an exact scheme for the simulation of the asset price process in the Heston model we note that for $t_i = \Delta i$, $i \in \{0, 1, \dots\}$, we obtain

$$V_{t_{i+1}} = V_{t_i} + \int_{t_i}^{t_{i+1}} \kappa(\theta - V_u) du + \sigma \int_{t_i}^{t_{i+1}} \sqrt{V_u} dW_u^1. \quad (2.5.6)$$

Hence,

$$\int_{t_i}^{t_{i+1}} \sqrt{V_u} dW_u^1 = \sigma^{-1} \left(V_{t_{i+1}} - V_{t_i} - \kappa\theta\Delta + \kappa \int_{t_i}^{t_{i+1}} V_u du \right). \quad (2.5.7)$$

Additionally,

$$X_{t_{i+1}} = X_{t_i} + \int_{t_i}^{t_{i+1}} \left(r - \frac{1}{2} V_u \right) du + \varrho \int_{t_i}^{t_{i+1}} \sqrt{V_u} dW_u^1 + \sqrt{1 - \varrho^2} \int_{t_i}^{t_{i+1}} \sqrt{V_u} dW_u^2. \quad (2.5.8)$$

Substituting (2.5.7) into (2.5.8) we obtain

$$\begin{aligned} X_{t_{i+1}} &= X_{t_i} + r\Delta + \frac{\varrho}{\sigma} (V_{t_{i+1}} - V_{t_i} - \kappa\theta\Delta) + \left(\frac{\varrho\kappa}{\sigma} - \frac{1}{2} \right) \int_{t_i}^{t_{i+1}} V_u du \\ &\quad + \sqrt{1 - \varrho^2} \int_{t_i}^{t_{i+1}} \sqrt{V_u} dW_u^2. \end{aligned} \quad (2.5.9)$$

Furthermore, the distribution of $\int_{t_i}^{t_{i+1}} \sqrt{V_u} dW_u^2$, given the path generated by V_t , is conditionally normal with mean zero and variance $\int_{t_i}^{t_{i+1}} V_u du$, because V is independent of the Brownian motion W^2 , that is

$$\int_{t_i}^{t_{i+1}} \sqrt{V_u} dW_u^2 \sim N \left(0, \int_{t_i}^{t_{i+1}} V_u du \right). \quad (2.5.10)$$

[Broadie & Kaya \(2006\)](#) discuss in detail how to obtain explicitly $\int_{t_i}^{t_{i+1}} V_u du$. However, it is also possible to approximate $\int_{t_i}^{t_{i+1}} V_u du$ given the path of the process V . This approximation can be achieved with high accuracy by a quadrature formula such as the trapezoidal rule, as previously discussed, which results in an efficient almost exact simulation technique by conditioning for the Heston model.

Multi-dimensional Heston Model with Independent Prices

Let us now define a particular version of the multi-dimensional Heston model given via the following SDE

$$d\mathbf{S}_t = \mathbf{A}_t \left(\mathbf{r} dt + \sqrt{\mathbf{B}_t} d\mathbf{W}_t \right). \quad (2.5.11)$$

Here $\mathbf{S} = \{\mathbf{S}_t = (S_t^1, S_t^2, \dots, S_t^d)^\top, t \in [0, \infty)\}$ is a vector process and $\mathbf{A}_t = [A_t^{i,j}]_{i,j=1}^d$ is a diagonal matrix process with elements

$$A_t^{i,j} = \begin{cases} S_t^i & \text{for } i = j \\ 0 & \text{otherwise.} \end{cases} \quad (2.5.12)$$

Additionally, $\mathbf{r} = (r_1, r_2, \dots, r_d)^\top$ is a d -dimensional vector and $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, W_t^2, \dots, W_t^d)^\top, t \in [0, \infty)\}$ is a d -dimensional vector of correlated Wiener processes. Moreover, $\mathbf{B}_t = [B_t^{i,j}]_{i,j=1}^d$ is a matrix process with elements

$$B_t^{i,j} = \begin{cases} \Sigma_t^{i,i} & \text{for } i=j \\ 0 & \text{otherwise.} \end{cases} \quad (2.5.13)$$

Note that \mathbf{B} is the generalization of V in the one-dimensional case. Here, the matrix process $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_t = \Sigma_t^{i,j}, t \geq 0\}$ is a matrix SR-process given by the SDE (2.3.66). Therefore, \mathbf{B}_t can be constructed from the diagonal elements of $\boldsymbol{\Sigma}_t$. Recall, that these elements $\Sigma_t^{1,1}, \Sigma_t^{2,2}, \dots, \Sigma_t^{d,d}$ form correlated SR-processes. For simplicity, let us first assume that \mathbf{B} is independent of \mathbf{W} .

The simulation of such a multi-dimensional version of the Heston model is straightforward and simply generalizes the simulation of the one-dimensional Heston model. Let us illustrate this on a two-dimensional example. The corresponding two-dimensional SDE for the two asset prices can be represented as

$$\begin{aligned} dS_t^1 &= S_t^1 r_1 dt + \sqrt{\Sigma_t^{1,1}} dW_t^1 \\ dS_t^2 &= S_t^2 r_2 dt + \sqrt{\Sigma_t^{2,2}} [\varrho dW_t^1 + \sqrt{1 - \varrho^2} dW_t^2], \end{aligned} \quad (2.5.14)$$

for $t \in [0, \infty)$. Here $\Sigma^{1,1}$ and $\Sigma^{2,2}$ are diagonal elements of the 2×2 matrix square root process given by the SDE (2.3.66). Note that $\Sigma^{1,1}$ and $\Sigma^{2,2}$ are always positive, hence $\sqrt{\Sigma^{1,1}}$ and $\sqrt{\Sigma^{2,2}}$ are well defined. Furthermore, the logarithmic transformation $\mathbf{X}_t = \ln(\mathbf{S}_t)$ yields the following SDE

$$\begin{aligned} dX_t^1 &= \left(r_1 - \frac{1}{2} \Sigma_t^{1,1} \right) dt + \sqrt{\Sigma_t^{1,1}} dW_t^1 \\ dX_t^2 &= \left(r_2 - \frac{1}{2} \Sigma_t^{2,2} \right) dt + \sqrt{\Sigma_t^{2,2}} [\varrho dW_t^1 + \sqrt{1 - \varrho^2} dW_t^2], \end{aligned} \quad (2.5.15)$$

for $t \in [0, \infty)$. This results in the following representations

$$\begin{aligned} X_{t_{i+1}}^1 &= X_{t_i}^1 + r_1 \Delta - \frac{1}{2} \int_{t_i}^{t_{i+1}} \Sigma_u^{1,1} du + \int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{1,1}} dW_u^1 \\ X_{t_{i+1}}^2 &= X_{t_i}^2 + r_2 \Delta - \frac{1}{2} \int_{t_i}^{t_{i+1}} \Sigma_u^{2,2} du + \varrho \int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{2,2}} dW_u^1 \\ &\quad + \sqrt{1 - \varrho^2} \int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{2,2}} dW_u^2, \end{aligned} \quad (2.5.16)$$

where

$$\int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{1,1}} dW_u^1 \sim N \left(0, \int_{t_i}^{t_{i+1}} \Sigma_u^{1,1} du \right) \quad (2.5.17)$$

$$\int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{2,2}} dW_u^1 \sim N \left(0, \int_{t_i}^{t_{i+1}} \Sigma_u^{2,2} du \right) \quad (2.5.18)$$

$$\int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{2,2}} dW_u^2 \sim N \left(0, \int_{t_i}^{t_{i+1}} \Sigma_u^{2,2} du \right). \quad (2.5.19)$$

Given that we can approximate almost exactly $\int_{t_i}^{t_{i+1}} \Sigma_u^{1,1} du$ and $\int_{t_i}^{t_{i+1}} \Sigma_u^{2,2} du$ it is possible to simulate this two-dimensional Heston model very accurately. We have discussed in Sect. 2.2 and 2.3 how to simulate the matrix SR-process Σ , which applies here. The terms in (2.5.17) and (2.5.18) can be generated as corresponding Gaussian random variables.

Multidimensional Heston Model with Correlated Prices

Let us now consider another multidimensional version of the Heston model, which allows correlation of the volatility vector Σ with the vector asset price process S . We define this generalization by the system of SDEs

$$dS_t = A_t \left[r dt + \sqrt{B_t} (C dW_t^1 + D dW_t^2) \right] \quad (2.5.20)$$

$$d\Sigma_t = (\alpha - E\Sigma_t) dt + F \sqrt{B_t} dW_t^1, \quad (2.5.21)$$

for $t \in [0, \infty)$. Here, $S = \{S_t = (S_t^1, S_t^2, \dots, S_t^d)^\top, t \in [0, \infty)\}$ and $r = (r_1, r_2, \dots, r_n)^\top$. $A_t = [A_t^{i,j}]_{i,j=1}^d$ is a matrix with elements as in (2.5.12) and $B_t = [B_t^{i,j}]_{i,j=1}^d$ is a matrix with elements as in (2.5.13). Additionally, $C = [C^{i,j}]_{i,j=1}^d$ is a diagonal matrix with elements

$$C^{i,j} = \begin{cases} \varrho_i & \text{for } i = j \\ 0 & \text{otherwise,} \end{cases} \quad (2.5.22)$$

and $D = [D^{i,j}]_{i,j=1}^d$ is a diagonal matrix with elements

$$D^{i,j} = \begin{cases} \sqrt{1 - \varrho_i^2} & \text{for } i = j \\ 0 & \text{otherwise,} \end{cases} \quad (2.5.23)$$

where $\varrho \in [-1, 1]$, $i \in \{1, 2, \dots, d\}$. Moreover, $\Sigma = \{\Sigma_t = (\Sigma_t^{1,1}, \Sigma_t^{2,2}, \dots, \Sigma_t^{d,d})^\top, t \in [0, \infty)\}$ and $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d)^\top$. $E = [E^{i,j}]_{i,j=1}^d$ is a diagonal matrix with elements

$$E^{i,j} = \begin{cases} b_i & \text{for } i = j \\ 0 & \text{otherwise,} \end{cases} \quad (2.5.24)$$

and $F = [F^{i,j}]_{i,j=1}^d$ is a diagonal matrix with elements

$$F^{i,j} = \begin{cases} \sigma_i & \text{for } i = j \\ 0 & \text{otherwise.} \end{cases} \quad (2.5.25)$$

Furthermore, $W^1 = \{W_t^1 = (W_t^{1,1}, W_t^{1,2}, \dots, W_t^{1,d})^\top, t \in [0, \infty)\}$ and $W^2 = \{W_t^2 = (W_t^{2,1}, W_t^{2,2}, \dots, W_t^{2,d})^\top, t \in [0, \infty)\}$ are vectors of correlated Wiener processes which are independent of each other.

Let us now illustrate the simulation of a two-dimensional Heston model of this type. The two-dimensional squared volatility process can be described by the following SDE

$$d\Sigma_t^{1,1} = \left(a_1 - b_1 \Sigma_t^{1,1} \right) dt + \sigma_1 \sqrt{\Sigma_t^{1,1}} dW_t^{1,1}, \quad (2.5.26)$$

$$d\Sigma_t^{2,2} = \left(a_2 - b_2 \Sigma_t^{2,2} \right) dt + \sigma_2 \sqrt{\Sigma_t^{2,2}} dW_t^{2,2}, \quad (2.5.27)$$

for $t \in [0, \infty)$. Moreover, the 2-dimensional asset price process is given by

$$dS_t^1 = S_t^1 r_1 dt + S_t^1 \sqrt{\Sigma_t^{1,1}} \left(\varrho_1 dW_t^{1,1} + \sqrt{1 - \varrho_1^2} dW_t^{2,1} \right), \quad (2.5.28)$$

$$dS_t^2 = S_t^2 r_2 dt + S_t^2 \sqrt{\Sigma_t^{2,2}} \left(\varrho_2 dW_t^{1,2} + \sqrt{1 - \varrho_2^2} dW_t^{2,2} \right), \quad (2.5.29)$$

for $t \in [0, \infty)$. Here the elements of the Σ process can be simulated as diagonal elements of a matrix SR-process as described in Sects. 2.2 and 2.3. Given these paths we may simulate the path of $\mathbf{X}_t = \ln(\mathbf{S}_t)$ similar as in the algorithm (2.5.9). That is

$$\begin{aligned} X_{t_{i+1}}^1 &= X_{t_i}^1 + r_1 \Delta + \frac{\varrho_1}{\sigma_1} \left(\Sigma_{t_{i+1}}^{1,1} - \Sigma_{t_i}^{1,1} - a_1 \Delta \right) + \left(\frac{\varrho_1 b_1}{\sigma_1} - \frac{1}{2} \right) \int_{t_i}^{t_{i+1}} \Sigma_u^{1,1} du \\ &\quad + \sqrt{1 - \varrho_1^2} \int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{1,1}} dW_u^{2,1}, \end{aligned}$$

$$\begin{aligned} X_{t_{i+1}}^2 &= X_{t_i}^2 + r_2 \Delta + \frac{\varrho_2}{\sigma_2} \left(\Sigma_{t_{i+1}}^{2,2} - \Sigma_{t_i}^{2,2} - a_2 \Delta \right) + \left(\frac{\varrho_2 b_2}{\sigma_2} - \frac{1}{2} \right) \int_{t_i}^{t_{i+1}} \Sigma_u^{2,2} du \\ &\quad + \sqrt{1 - \varrho_2^2} \int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{2,2}} dW_u^{2,2}. \end{aligned}$$

Here,

$$\int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{1,1}} dW_u^{2,1} \sim N \left(0, \int_{t_i}^{t_{i+1}} \Sigma_u^{1,1} du \right), \quad (2.5.30)$$

$$\int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{2,2}} dW_u^{2,2} \sim N \left(0, \int_{t_i}^{t_{i+1}} \Sigma_u^{2,2} du \right) \quad (2.5.31)$$

have now to be simulated as two correlated Gaussian random variables with zero mean and variances $\int_{t_i}^{t_{i+1}} \Sigma_u^{1,1} du$ and $\int_{t_i}^{t_{i+1}} \Sigma_u^{2,2} du$. The correlation parameters are $\varrho_1, \varrho_2 \in [-1, 1]$. This simulation method is almost exact, since the only approximation used here is the approximation of the time integrals $\int_{t_i}^{t_{i+1}} \Sigma_u^{1,1} du$ and $\int_{t_i}^{t_{i+1}} \Sigma_u^{2,2} du$. Numerical errors are not propagated.

Another Heston Model

There exist alternative ways to construct multi-dimensional Heston type models. Two possible generalizations are discussed in Gouriéroux & Sufana (2004)

and [Da Fonseca, Grasselli & Tebaldi \(2008\)](#). For instance, Gouriéroux & Sufana (2004) consider a d -dimensional stochastic volatility model of the form

$$d\mathbf{S}_t = \mathbf{A}_t \left(r\mathbf{1} dt + \sqrt{\boldsymbol{\Sigma}_t} d\mathbf{W}_t^2 \right), \quad (2.5.32)$$

$$d\boldsymbol{\Sigma}_t = \left(\boldsymbol{\Omega}\boldsymbol{\Omega}^\top + \mathbf{M}\boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t\mathbf{M}^\top \right) dt + \sqrt{\boldsymbol{\Sigma}_t} d\mathbf{W}_t^1 \mathbf{Q} + \mathbf{Q}^\top d(\mathbf{W}_t^1)^\top \sqrt{\boldsymbol{\Sigma}_t},$$

for $t \in [0, \infty)$. Here, $\mathbf{S} = \{\mathbf{S}_t = (S_t^1, S_t^2, \dots, S_t^d)^\top, t \in [0, \infty)\}$ is a vector asset price process and $\boldsymbol{\Sigma}_t = [\Sigma_t^{i,j}]_{i,j=1}^d$ is a matrix squared volatility process. Moreover, elements of the matrix $\mathbf{A}_t = [A_t^{i,j}]_{i,j=1}^d$ are defined by (2.5.12) and $\mathbf{1} = (1, 1, \dots, 1)^\top$. Here, $\mathbf{W}^2 = \{\mathbf{W}_t^2 = (W_t^{2,1}, W_t^{2,2}, \dots, W_t^{2,d})^\top, t \in [0, \infty)\}$ is a vector of independent Wiener processes. In the second equation $\boldsymbol{\Omega}$, \mathbf{M} and \mathbf{Q} are $d \times d$ parameter matrices with $\boldsymbol{\Omega}$ being an invertible matrix. In order to preserve strict positivity and the mean reverting feature of the volatility, the matrix \mathbf{M} is assumed to be negative semi-definite, while $\boldsymbol{\Omega}$ satisfies

$$\boldsymbol{\Omega}\boldsymbol{\Omega}^\top = \beta\boldsymbol{\Omega}^\top\boldsymbol{\Omega} \quad (2.5.33)$$

with real valued parameter $\beta = d - 1$, see [Bru \(1991\)](#). Additionally, $\mathbf{W}_t^1 = [W_t^{1,i,j}]_{i,j=1}^d$ is a matrix Wiener process independent of \mathbf{W}_t^2 .

In the two-dimensional case this model is represented by the SDEs

$$dS_t^1 = rS_t^1 dt + S_t^1 \sqrt{\Sigma_t^{1,1}} dW_t^{2,1} + S_t^1 \sqrt{\Sigma_t^{1,2}} dW_t^{2,2}, \quad (2.5.34)$$

$$dS_t^2 = rS_t^2 dt + S_t^2 \sqrt{\Sigma_t^{2,1}} dW_t^{2,1} + S_t^2 \sqrt{\Sigma_t^{2,2}} dW_t^{2,2}, \quad (2.5.35)$$

for $t \in [0, \infty)$. The SDE for $\mathbf{X}_t = \ln(\mathbf{S}_t)$ is given by

$$dX_t^1 = \left(r - \frac{1}{2}\Sigma_t^{1,1} - \frac{1}{2}\Sigma_t^{1,2} \right) dt + \sqrt{\Sigma_t^{1,1}} dW_t^{2,1} + \sqrt{\Sigma_t^{1,2}} dW_t^{2,2}, \quad (2.5.36)$$

$$dX_t^2 = \left(r - \frac{1}{2}\Sigma_t^{2,1} - \frac{1}{2}\Sigma_t^{2,2} \right) dt + \sqrt{\Sigma_t^{2,1}} dW_t^{2,1} + \sqrt{\Sigma_t^{2,2}} dW_t^{2,2}, \quad (2.5.37)$$

for $t \in [0, \infty)$. Therefore, since $\boldsymbol{\Sigma}$ is independent of \mathbf{W}^2 we obtain an almost exact simulation algorithm for \mathbf{X} constructed as

$$X_{t_{i+1}}^1 = X_{t_i}^1 + r\Delta - \frac{1}{2} \int_{t_i}^{t_{i+1}} \Sigma_u^{1,1} du - \frac{1}{2} \int_{t_i}^{t_{i+1}} \Sigma_u^{1,2} du + \int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{1,1}} dW_u^{2,1}$$

$$+ \int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{1,2}} dW_u^{2,2},$$

$$X_{t_{i+1}}^2 = X_{t_i}^2 - \frac{1}{2} \int_{t_i}^{t_{i+1}} \Sigma_u^{2,1} du - \frac{1}{2} \int_{t_i}^{t_{i+1}} \Sigma_u^{2,2} du + \int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{2,1}} dW_u^{2,1}$$

$$+ \int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{2,2}} dW_u^{2,2}.$$

Here,

$$\int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{1,1}} dW_u^{2,1} \sim N\left(0, \int_{t_i}^{t_{i+1}} \Sigma_u^{1,1} du\right) \quad (2.5.38)$$

$$\int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{1,2}} dW_u^{2,2} \sim N\left(0, \int_{t_i}^{t_{i+1}} \Sigma_u^{1,2} du\right) \quad (2.5.39)$$

$$\int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{2,1}} dW_u^{2,1} \sim N\left(0, \int_{t_i}^{t_{i+1}} \Sigma_u^{2,1} du\right) \quad (2.5.40)$$

$$\int_{t_i}^{t_{i+1}} \sqrt{\Sigma_u^{2,2}} dW_u^{2,2} \sim N\left(0, \int_{t_i}^{t_{i+1}} \Sigma_u^{2,2} du\right) \quad (2.5.41)$$

are independent Gaussian random variables with zero means and variance as indicated. Given that we can approximate $\int_{t_i}^{t_{i+1}} \Sigma_u^{k,j} du$ for $k, j \in \{1, 2\}$ accurately via the trapezoidal rule this simulation is again almost exact.

2.6 Almost Exact Simulation by Time Change

Another useful technique, which can be applied in the simulation of multi-dimensional SDEs, is the almost exact simulation by *stochastic time change*. This method can be used, for instance, for the simulation of solutions of SDEs which can be written as time changed solutions of SDEs that permit almost exact simulation. The time change used can be stochastic and needs to be simulated exactly or almost exactly. We will illustrate this technique by the simulation of two models, a multi-dimensional ARCH diffusion model and a multi-dimensional generalized minimal market model (MMM).

One-dimensional ARCH Diffusion Model

An interesting stochastic volatility model is the ARCH diffusion model, which is represented by a two-dimensional SDE, see (1.7.32)–(1.7.33). Again the first SDE models the stock price while the second one the squared volatility of the underlying stock. The only difference between the Heston model and the ARCH diffusion model is in the specification of the diffusion coefficient function of the squared volatility process, which is multiplicative. That is,

$$dS_t = rS_t dt + \sqrt{V_t} S_t \left[\varrho dW_t^1 + \sqrt{1 - \varrho^2} dW_t^2 \right], \quad (2.6.1)$$

$$dV_t = \kappa(\theta - V_t) dt + \sigma V_t dW_t^1, \quad (2.6.2)$$

for $t \in [0, \infty)$. Here W^1 and W^2 are two independent Wiener processes, and ϱ represents the correlation between the noises driving the return process and the volatility process. In the given case we can simulate the squared volatility process V almost exactly by approximating again some time integral via the trapezoidal rule. Here the exact representation is of the form

$$V_{t_i} = \exp \left\{ \left(-\kappa - \frac{1}{2}\sigma^2 \right) t_i + \sigma W_{t_i}^1 \right\} \\ \times \left(V_{t_0} + \kappa \theta \sum_{k=0}^{i-1} \int_{t_k}^{t_{k+1}} \exp \left\{ \left(\kappa + \frac{1}{2}\sigma^2 \right) s - \sigma W_s^1 \right\} ds \right), \quad (2.6.3)$$

which leads to the recursive almost exact approximation

$$V_{t_i}^\Delta = \exp \left\{ \left(-\kappa - \frac{1}{2}\sigma^2 \right) t_i + \sigma W_{t_i}^1 \right\} \\ \times \left(V_{t_0} + \kappa \theta \sum_{k=0}^{i-1} \frac{\Delta}{2} \left[\exp \left\{ \left(\kappa + \frac{1}{2}\sigma^2 \right) t_k - \sigma W_{t_k}^1 \right\} \right. \right. \\ \left. \left. + \exp \left\{ \left(\kappa + \frac{1}{2}\sigma^2 \right) t_{k+1} - \sigma W_{t_{k+1}}^1 \right\} \right] \right) \quad (2.6.4)$$

for $t_i = \Delta i, i \in \{0, 1, \dots\}$. Note that this simulation requires sampling from a Gaussian law in contrast to the simulation of the squared volatility process in the Heston model, where we sampled from the non-central chi-square distribution. It is now straightforward to correlate the squared volatility process V with the asset price process S .

In order to simulate the asset price process S we consider its discounted version $\bar{S}_t = \frac{S_t}{S_t^0}$, where $S_t^0 = \exp\{rt\}$ is the savings account, for $t \in [0, \infty)$. This results in the following SDE

$$d\bar{S}_t = \sqrt{V_t} \bar{S}_t \left[\varrho dW_t^1 + \sqrt{1 - \varrho^2} dW_t^2 \right], \quad (2.6.5)$$

for $t \in [0, \infty)$. Additionally, we obtain by the Itô formula

$$d \ln(\bar{S}_t) = -\frac{1}{2} V_t dt + \sqrt{V_t} \left[\varrho dW_t^1 + \sqrt{1 - \varrho^2} dW_t^2 \right], \quad (2.6.6)$$

for $t \in [0, \infty)$. Equation (2.6.6) for $\ln(\bar{S}_t) = \ln(\bar{S}(\tau(t)))$ turns out to be that of a time changed, drifted Wiener process $\ln(\bar{S}) = \{\ln(\bar{S}(\tau(t))), t \in [0, \infty)\}$ with SDE of the form

$$d \ln(\bar{S}(\tau(t))) = -\frac{1}{2} d\tau(t) + \left[\varrho d\tilde{W}_{\tau(t)}^1 + \sqrt{1 - \varrho^2} d\tilde{W}_{\tau(t)}^2 \right], \quad (2.6.7)$$

for $t \in [0, \infty)$. Note that we make a difference between \bar{S}_\cdot and $\bar{S}(\cdot)$. Here the time change is stochastic and can be obtained by approximation of the following time integral

$$\tau(t) = \tau(0) + \int_0^t V_s ds, \quad (2.6.8)$$

for $t \in [0, \infty)$. Furthermore, in τ -time the processes \tilde{W}^1 and \tilde{W}^2 are independent Wiener processes. Under the above ARCH diffusion model the asset price is then obtained as $S_t = S_t^0 \bar{S}(\tau(t)) = S_t^0 \bar{S}_t$.

Multi-dimensional ARCH Diffusion

Let us now define a multi-dimensional ARCH diffusion model as follows

$$d\mathbf{S}_t = \mathbf{A}_t \left[\mathbf{r} dt + \sqrt{\mathbf{B}_t} (\mathbf{C} d\mathbf{W}_t^1 + \mathbf{D} d\mathbf{W}_t^2) \right] \quad (2.6.9)$$

$$d\mathbf{V}_t = (\mathbf{a} - \mathbf{E}\mathbf{V}_t) dt + \mathbf{F}\mathbf{B}_t d\mathbf{W}_t^1, \quad (2.6.10)$$

for $t \in [0, \infty)$. Here, $\mathbf{S} = \{\mathbf{S}_t = (S_t^1, S_t^2, \dots, S_t^d)^\top, t \in [0, \infty)\}$, $\mathbf{A}_t = [A_t^{i,j}]_{i,j=1}^d$ is a diagonal matrix with elements as in (2.5.12), $\mathbf{r} = (r_1, r_2, \dots, r_n)^\top$. Furthermore, $\mathbf{B}_t = [B_t^{i,j}]_{i,j=1}^d$ is a diagonal matrix with elements

$$B_t^{i,j} = \begin{cases} V_t^i & \text{for } i = j \\ 0 & \text{otherwise.} \end{cases} \quad (2.6.11)$$

Additionally, $\mathbf{C} = [C^{i,j}]_{i,j=1}^d$ is a diagonal matrix with elements as in (2.5.22) and $\mathbf{D} = [D^{i,j}]_{i,j=1}^d$ is a diagonal matrix with elements as in (2.5.23).

Moreover, $\mathbf{V} = \{\mathbf{V}_t = (V_t^1, V_t^2, \dots, V_t^d)^\top, t \in [0, \infty)\}$, $\mathbf{a} = (a_1, a_2, \dots, a_d)^\top$ and $\mathbf{E} = [E^{i,j}]_{i,j=1}^d$ is a diagonal matrix with elements

$$E^{i,j} = \begin{cases} \kappa_i & \text{for } i = j \\ 0 & \text{otherwise.} \end{cases} \quad (2.6.12)$$

Furthermore, $\mathbf{F} = [F^{i,j}]_{i,j=1}^d$ is a diagonal matrix with elements

$$F^{i,j} = \begin{cases} \sigma_i & \text{for } i = j \\ 0 & \text{otherwise.} \end{cases} \quad (2.6.13)$$

Finally, $\mathbf{W}^1 = \{\mathbf{W}_t^1 = (W_t^{1,1}, W_t^{1,2}, \dots, W_t^{1,d})^\top, t \in [0, \infty)\}$ and $\mathbf{W}^2 = \{\mathbf{W}_t^2 = (W_t^{2,1}, W_t^{2,2}, \dots, W_t^{2,d})^\top, t \in [0, \infty)\}$ are vectors of correlated Wiener processes but independent of each other. This setup then yields, with the previous methodology, almost exact approximate solutions for the multi-dimensional ARCH diffusion given in (2.6.9)–(2.6.10).

One-dimensional Stylized MMM

Let us now describe the minimal market model (MMM), introduced in Platen (2001), see (1.6.9)–(1.6.10). It has been designed for modeling the long term dynamics of the growth optimal portfolio (GOP), see Platen & Heath (2006), when denominated in a given currency. Under this model the discounted GOP, denoted by \bar{S}^{δ_*} , forms a time transformed squared Bessel process of dimension four. This model reflects well various empirical properties of the long term dynamics of a diversified world stock index. In its stylized version it can be described by the SDE

$$d\bar{S}_t^{\delta_*} = \alpha_t^{\delta_*} dt + \sqrt{\bar{S}_t^{\delta_*} \alpha_t^{\delta_*}} dW_t, \quad (2.6.14)$$

for $t \in [0, \infty)$, with $\alpha_t^{\delta_*} = \alpha_0 \exp\{\eta t\}$, for initial scaling parameter $\alpha_0 > 0$ and net growth rate $\eta > 0$. This is equivalent to writing

$$d\bar{S}_t^{\delta_*} = 4d\varphi(t) + 2\sqrt{\bar{S}_t^{\delta_*}} d\tilde{W}_{\varphi(t)}, \quad (2.6.15)$$

for $t \in [0, \infty)$, which is the SDE of a squared Bessel process $X = \{X_{\varphi(t)} = \bar{S}_t^{\delta_*}, t \geq 0\}$ of dimension $\delta = 4$ in φ -time. Here, one has

$$\varphi(t) = \frac{\alpha_0}{4\eta} (\exp\{\eta t\} - 1), \quad (2.6.16)$$

and the process $\tilde{W} = \{\tilde{W}_\varphi, \varphi \geq 0\}$ is a standard Wiener process in φ -time.

The normalized GOP under the stylized MMM is given by the ratio

$$Y_t = \frac{\bar{S}_t^{\delta_*}}{\alpha_t^{\delta_*}}. \quad (2.6.17)$$

By the Itô formula it follows that the process $Y = \{Y_t, t \in [0, T]\}$ is a square root process of dimension four with SDE

$$dY_t = (1 - \eta Y_t) dt + \sqrt{Y_t} dW_t \quad (2.6.18)$$

for $t \in \mathbb{R}^+$ and initial value $Y_0 = \bar{S}_0^{\delta_*}/\alpha_0^{\delta_*}$. Note also that the quadratic variation of the square root of Y_t equals

$$[\sqrt{Y}]_t = \frac{1}{4}t, \quad (2.6.19)$$

for $t \in [0, \infty)$. Additionally, the volatility under the stylized MMM is of the form

$$\sigma_t = \frac{1}{\sqrt{Y_t}}, \quad (2.6.20)$$

for $t \in [0, \infty)$.

Hence, the value \bar{S}^{δ_*} at time t of the discounted GOP can be expressed in the form

$$\bar{S}_t^{\delta_*} = Y_t \alpha_t^{\delta_*} \quad (2.6.21)$$

for $t \in [0, \infty)$. Therefore, the GOP when expressed in units of the domestic currency can be represented by the product

$$S_t^{\delta_*} = S_t^0 \bar{S}_t^{\delta_*} = S_t^0 Y_t \alpha_t^{\delta_*} \quad (2.6.22)$$

for $t \in [0, \infty)$, where $S_t^0 = \exp\{\int_0^t r_s ds\}$ is the domestic savings account with $r = \{r_t, t \in [0, \infty)\}$ denoting the short rate process.

Multi-dimensional Stylized MMM

Let us now consider the MMM for a market with $d + 1$ currencies, similar to the one described in [Platen \(2001\)](#) and [Heath & Platen \(2005\)](#). We can describe the value of the GOP in the i th currency denomination according to [\(2.6.22\)](#) by the expression

$$S_i^{\delta*}(t) = \alpha_t^i Y_t^i S_i^i(t), \quad (2.6.23)$$

where $\alpha_t^i = \alpha_0^i \exp\{\eta^i t\}$ and $S_i^i(t) = \exp\{r^i t\}$, $i \in \{0, 1, \dots, d\}$. Here η^i is the i th net growth rate and r^i is the short rate for the i th currency for $i \in \{0, 1, \dots, d\}$. The i th normalized GOP satisfies then the SDE

$$dY_t^i = (1 - \eta^i Y_t^i) dt + \sqrt{Y_t^i} dW_t^i \quad (2.6.24)$$

for $t \in [0, \infty)$, with $Y_0^i > 0$. Here $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, W_t^2, \dots, W_t^{d+1})^\top, t \in [0, \infty)\}$ is a vector of correlated Wiener processes.

Let us now illustrate the simulation of a two-currency market based on the stylized MMM. We start with the simulation of two time changed 4×2 matrix Wiener processes $\mathbf{W}_{\varphi_1(t)}^1 = [W_{\varphi_1(t)}^{1,i,j}]_{i,j=1}^{4,2}$ and $\mathbf{W}_{\varphi_2(t)}^2 = [W_{\varphi_2(t)}^{2,i,j}]_{i,j=1}^{4,2}$, resulting from the same set of Gaussian random numbers. The elements of such a matrix are as in [\(2.3.25\)](#). The time changes are here

$$\varphi_i(t) = \frac{\alpha_0^i}{4\eta^i} (\exp\{\eta^i t\} - 1), \quad (2.6.25)$$

for $i \in \{1, 2\}$. In the next step we construct two 4×2 matrix Wiener processes $\tilde{\mathbf{W}}_{\varphi_1(t)}^1 = \mathbf{I}_{4 \times 4} \mathbf{W}_{\varphi_1(t)}^1 \boldsymbol{\Sigma}_{2 \times 2}^\top$ and $\tilde{\mathbf{W}}_{\varphi_2(t)}^2 = \mathbf{I}_{4 \times 4} \mathbf{W}_{\varphi_2(t)}^2 \boldsymbol{\Sigma}_{2 \times 2}^\top$, for $t \in [0, \infty)$ where the elements of $\boldsymbol{\Sigma}$ are as in [\(2.3.13\)](#). As a result of this transformation we obtain the 4×2 matrix processes $\tilde{\mathbf{W}}^1$ and $\tilde{\mathbf{W}}^2$ with correlated columns and independent rows. Additionally, the 4×2 matrix process $\tilde{\mathbf{W}}^1$ is φ_1 -time changed and the matrix process $\tilde{\mathbf{W}}^2$ is φ_2 -time changed. Now we construct a new 4×2 matrix process $\bar{\mathbf{W}}$ with elements

$$\bar{W}_t^{i,j} = \begin{cases} w^{i,j} + \tilde{W}_{\varphi_1(t)}^{1,i,j} & \text{for } j = 1 \\ w^{i,j} + \tilde{W}_{\varphi_2(t)}^{2,i,j} & \text{for } j = 2, \end{cases} \quad (2.6.26)$$

where $i \in \{1, 2, 3, 4\}$. The trajectory of the eight elements of this 4×2 matrix process are displayed in Fig. [2.6.1](#). Here we assumed the starting values of the components to be $w^{i,j} = \frac{1}{2}$ for $i \in \{1, 2, 3, 4\}$, $j = 1, 2$ and $\alpha_0^1 = 0.04$, $\alpha_0^2 = 0.05$, $\eta^1 = 0.05$ and $\eta^2 = 0.06$.

The trajectory of a time changed Wishart process is then obtained by the formula $\bar{\mathbf{S}}_t = \bar{\mathbf{W}}_t^\top \bar{\mathbf{W}}_t$, for $t \in [0, \infty)$, see [\(2.3.51\)](#). We illustrate this 2×2 matrix process in Fig. [2.6.2](#). Note that the diagonal elements of this matrix are

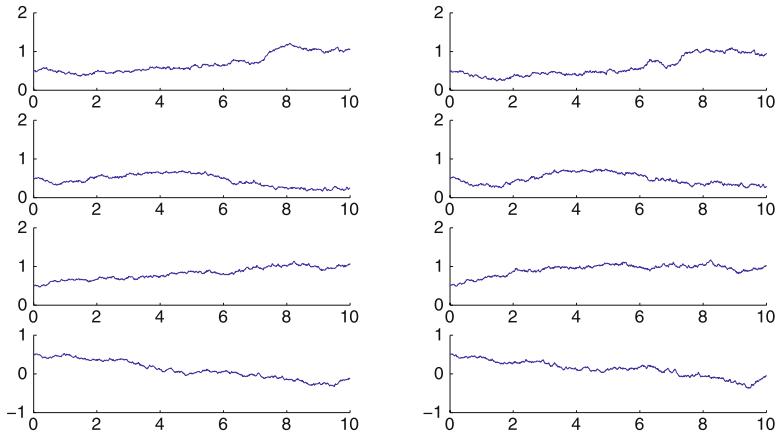


Fig. 2.6.1. Trajectory of the 4×2 matrix time changed Wiener process with independent components in the columns and correlated components in the rows

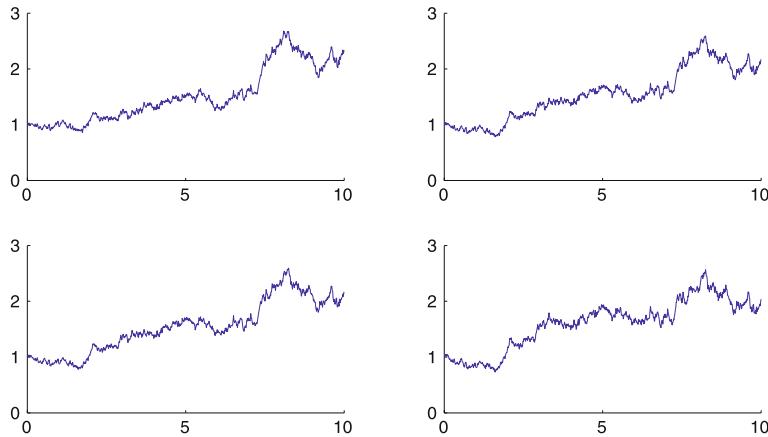


Fig. 2.6.2. Trajectory of the 2×2 time changed Wishart matrix

correlated time changed squared Bessel processes constructed in the following way

$$\begin{aligned} \bar{S}_t^{1,1} &= \left(w^{1,1} + \tilde{W}_{\varphi_1(t)}^{1,1,1} \right)^2 + \left(w^{2,1} + \tilde{W}_{\varphi_1(t)}^{1,2,1} \right)^2 \\ &\quad + \left(w^{3,1} + \tilde{W}_{\varphi_1(t)}^{1,3,1} \right)^2 + \left(w^{4,1} + \tilde{W}_{\varphi_1(t)}^{1,4,1} \right)^2, \end{aligned}$$

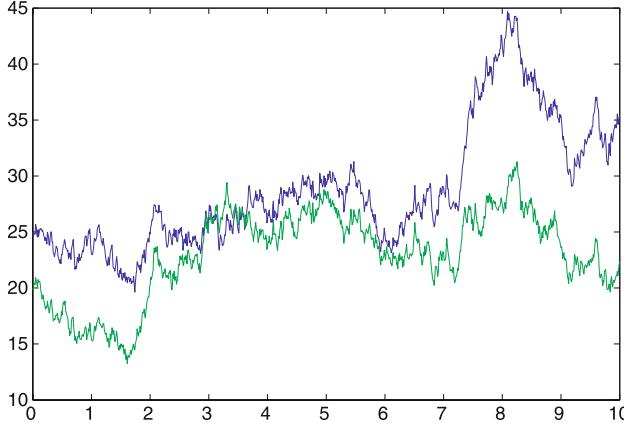


Fig. 2.6.3. Trajectory of two related SR-processes modeling the normalized GOP for two different currencies

$$\begin{aligned}
 \bar{S}_t^{2,2} = & \left(w^{1,2} + \varrho \tilde{W}_{\varphi_2(t)}^{2,1,1} + \sqrt{1 - \varrho^2} \tilde{W}_{\varphi_2(t)}^{2,1,2} \right)^2 \\
 & + \left(w^{2,2} + \varrho \tilde{W}_{\varphi_2(t)}^{2,2,1} + \sqrt{1 - \varrho^2} \tilde{W}_{\varphi_2(t)}^{2,2,2} \right)^2 \\
 & + \left(w^{3,2} + \varrho \tilde{W}_{\varphi_2(t)}^{2,3,1} + \sqrt{1 - \varrho^2} \tilde{W}_{\varphi_2(t)}^{2,3,2} \right)^2 \\
 & + \left(w^{4,2} + \varrho \tilde{W}_{\varphi_2(t)}^{2,4,1} + \sqrt{1 - \varrho^2} \tilde{W}_{\varphi_2(t)}^{2,4,2} \right)^2, \quad (2.6.27)
 \end{aligned}$$

where $S_0^{1,1} = \sum_{i=1}^4 (w^{i,1})^2$ and $S_0^{2,2} = \sum_{i=1}^4 (w^{i,2})^2$. These two processes model the discounted GOP in units of the two respective currencies. Let us also illustrate the normalized GOP for these two currencies. For this part of the simulation it is sufficient to work with the above two squared Bessel processes. We construct, according to (2.6.23), two related SR-processes by setting

$$Y_t^1 = \frac{\bar{S}_t^{1,1}}{\alpha_t^1} \quad \text{and} \quad Y_t^2 = \frac{\bar{S}_t^{2,2}}{\alpha_t^2}, \quad (2.6.28)$$

for $t \in [0, \infty)$. These two processes are illustrated with their trajectories in Fig. 2.6.3.

Multi-dimensional Generalized MMM

Let us now generalize the MMM with a time change by introducing a *market activity* process $m = \{m_t, t \in [0, \infty)\}$. That is, we will use time $\tau = \{\tau(t), t \in$

$[0, \infty)$ }, which we call *market time*, to model random variations in market activity. The market time process τ is given by the relation

$$\tau(t) = \tau(0) + \int_0^t m_s ds, \quad (2.6.29)$$

where $\tau_0 > 0$, see [Breymann, Kelly & Platen \(2006\)](#).

In this generalized setting the discounted GOP value $\bar{S}_t^{\delta_*} = \bar{S}^{\delta_*}(\tau(t))$ is modeled in market time τ via the SDE

$$d\bar{S}^{\delta_*}(\tau(t)) = \alpha^{\delta_*}(\tau(t)) d\tau(t) + \sqrt{\bar{S}^{\delta_*}(\tau(t)) \alpha^{\delta_*}(\tau(t))} dW(\tau(t)), \quad (2.6.30)$$

for $t \in [0, \infty)$, where $W(\cdot)$ is a standard Wiener process in τ -time. The SDE for the normalized GOP $Y(\tau(t)) = \frac{\bar{S}^{\delta_*}(\tau(t))}{\alpha^{\delta_*}(\tau(t))}$ is then

$$dY(\tau(t)) = (1 - \eta Y(\tau(t))) d\tau(t) + \sqrt{Y(\tau(t))} dW(\tau(t)), \quad (2.6.31)$$

for $t \in [0, \infty)$. This SDE can be rewritten for $Y_t = Y(\tau(t))$ in t -time in the following way

$$dY_t = (1 - \eta Y_t) m_t dt + \sqrt{Y_t m_t} dW_t, \quad (2.6.32)$$

for $t \in [0, \infty)$, where W is a standard Wiener process in t -time.

Now, let the market activity process m be modeled by a linear diffusion process as considered in Sect. 2.4, that is

$$dm_t = \kappa(\bar{m} - m_t) dt + \gamma m_t d\tilde{W}_t, \quad (2.6.33)$$

for $t \in [0, \infty)$. Here \tilde{W}_t is a standard Wiener process in t -time. Note that the above market activity resembles the squared volatility of an ARCH diffusion asset price, see (1.7.33).

Let us now consider a market with $d + 1$ currencies. We can describe the value of the GOP in τ_i -time in the i th currency denomination by the expression

$$S_i^{\delta_*}(\tau_i(t)) = \alpha^i(\tau_i(t)) Y_t^i S_i^i(\tau_i(t)), \quad (2.6.34)$$

where $\alpha^i(\tau_i(t)) = \alpha_0^i \exp\{\eta^i \tau_i(t)\}$ and $S_i^i(\tau_i(t)) = \exp\{r^i \tau_i(t)\}$. Here η^i is the i th net growth rate and r^i the i th short rate as in the case of the stylized MMM. The i th normalized GOP satisfies the SDE

$$dY_t^i = (1 - \eta^i Y_t^i) m_t^i dt + \sqrt{Y_t^i m_t^i} dW_t^i \quad (2.6.35)$$

for $t \in [0, \infty)$, with $Y^i(\tau_i(0)) > 0$, where $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, W_t^2, \dots, W_t^{d+1})^\top, t \in [0, \infty)\}$ is a vector of correlated Wiener processes. Additionally, the market activity process $\mathbf{m} = \{\mathbf{m}_t = (m_t^1, m_t^2, \dots, m_t^{d+1})^\top, t \in [0, \infty)\}$ is a $d + 1$ -dimensional linear diffusion process as discussed in Sect. 2.4.

Let us now illustrate the simulation of a two currency market under the generalized MMM. Here we use the market activity given by the correlated

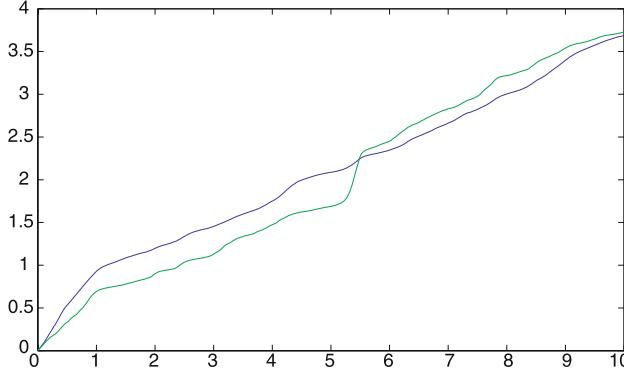


Fig. 2.6.4. Two related market times τ_1 and τ_2 obtained from the linear diffusions in Fig. 2.4.2

linear diffusion processes in Fig. 2.4.2. By approximation of the integral in (2.6.29) we obtain two correlated market times which we display in Fig. 2.6.4. Given these two nondecreasing processes we then proceed with the simulation as in the case of the stylized MMM. However, we use now the time change

$$\bar{\varphi}_i(t) = \varphi_i(\tau_i(t)), \quad (2.6.36)$$

for $i \in \{1, 2\}$ to include the effect of random market activity. This leads to two correlated squared Bessel processes in two correlated times analogous to (2.6.26)–(2.6.27) given by

$$\begin{aligned} \bar{S}_t^{1,1} &= \left(w^{1,1} + \tilde{W}_{\bar{\varphi}_1(t)}^{1,1,1} \right)^2 + \left(w^{2,1} + \tilde{W}_{\bar{\varphi}_1(t)}^{1,2,1} \right)^2 \\ &\quad + \left(w^{3,1} + \tilde{W}_{\bar{\varphi}_1(t)}^{1,3,1} \right)^2 + \left(w^{4,1} + \tilde{W}_{\bar{\varphi}_1(t)}^{1,4,1} \right)^2, \end{aligned} \quad (2.6.37)$$

$$\begin{aligned} \bar{S}_t^{2,2} &= \left(w^{1,2} + \varrho \tilde{W}_{\bar{\varphi}_2(t)}^{2,1,1} + \sqrt{1 - \varrho^2} \tilde{W}_{\bar{\varphi}_2(t)}^{2,1,2} \right)^2 \\ &\quad + \left(w^{2,2} + \varrho \tilde{W}_{\bar{\varphi}_2(t)}^{2,2,1} + \sqrt{1 - \varrho^2} \tilde{W}_{\bar{\varphi}_2(t)}^{2,2,2} \right)^2 \\ &\quad + \left(w^{3,2} + \varrho \tilde{W}_{\bar{\varphi}_2(t)}^{2,3,1} + \sqrt{1 - \varrho^2} \tilde{W}_{\bar{\varphi}_2(t)}^{2,3,2} \right)^2 \\ &\quad + \left(w^{4,2} + \varrho \tilde{W}_{\bar{\varphi}_2(t)}^{2,4,1} + \sqrt{1 - \varrho^2} \tilde{W}_{\bar{\varphi}_2(t)}^{2,4,2} \right)^2, \end{aligned} \quad (2.6.38)$$

where $S_0^{1,1} = \sum_{i=1}^4 (w^{i,1})^2$ and $S_0^{2,2} = \sum_{i=1}^4 (w^{i,2})^2$. We display the entire Wishart matrix process in τ -time in Fig. 2.6.5, whose diagonal elements are as

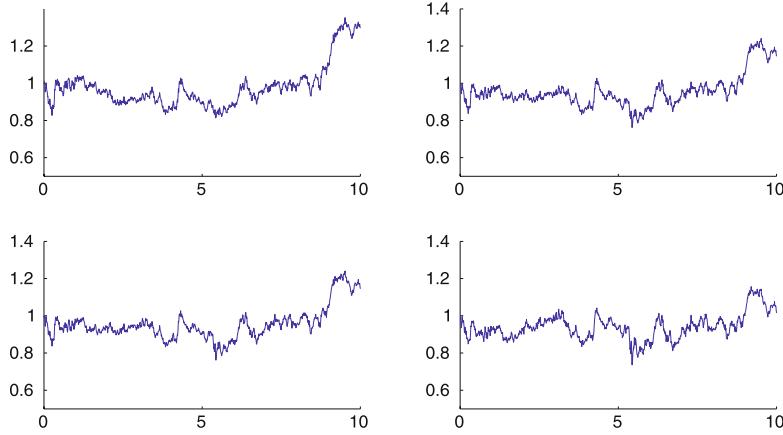


Fig. 2.6.5. Trajectory of the 2×2 time changed Wishart process

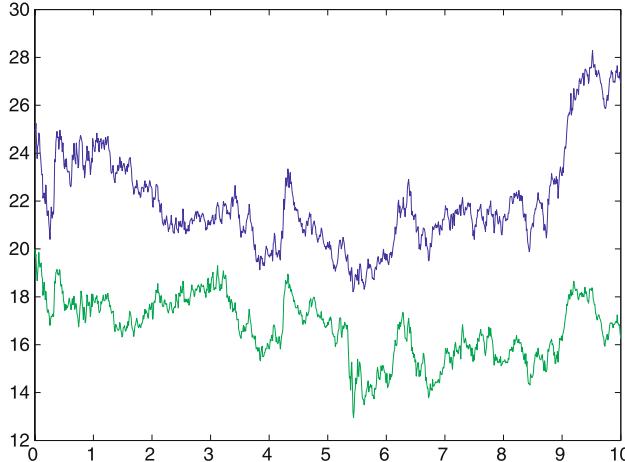


Fig. 2.6.6. Trajectory of two correlated SR-processes modeling the normalized GOP for two different currencies under the generalized MMM

in (2.6.37) and (2.6.38). Additionally, we construct two related SR-processes as

$$Y_t^1 = \frac{\bar{S}_t^{1,1}}{\bar{\alpha}^1(\tau_1(t))} \quad \text{and} \quad Y_t^2 = \frac{\bar{S}_t^{2,2}}{\bar{\alpha}^2(\tau_2(t))}, \quad (2.6.39)$$

for $t \in [0, \infty)$. These two processes are illustrated in Fig. 2.6.6. Moreover, in Fig. 2.6.7 we display the quadratic variations when multiplied by 4 of the

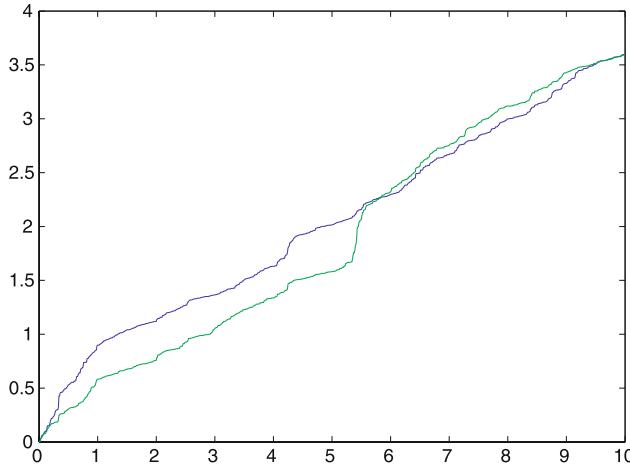


Fig. 2.6.7. Quadratic variation $\times 4$ of the square root of the SR-process from Fig. 2.6.6

square root of the SR-processes from Fig. 2.6.6. These should approximately be equal to the corresponding τ -times already displayed in Fig. 2.6.4, since

$$[\sqrt{Y^i}]_t = \frac{1}{4}\tau_i(t) \quad (2.6.40)$$

for $i \in \{1, 2\}$. The visualization confirms the usefulness of the presented almost exact simulation method for the generalized MMM.

The above presented exact and almost exact simulation methods for multi-dimensional diffusions lead to highly accurate scenario simulations. Since no recursive scheme is involved the minor errors are not propagated and the methods are also reliable over long periods of time. The numerical stability problems that we will highlight in Chap. 14 for recursive discrete-time approximations of SDEs are avoided. We remark that the effect of jumps on the above considered type of dynamics can be introduced via a jump-adapted time discretization, see Platen (1982a) and will be discussed in Sect. 8.6.

2.7 Functionals of Solutions of SDEs

Explicit formulas for functionals of solutions of SDEs are of particular interest. As we will see in Chap. 3, pricing rules are usually expressed via pricing formulas that have the form of conditional expectations. For Markovian state variables these conditional expectations lead to pricing functions that satisfy certain *partial differential equations* (PDEs), or more generally, *partial*

integro differential equations (PIDEs) when jumps are involved. The link between the conditional expectations and respective PDEs can be interpreted as an application of the, so-called, *Feynman-Kac formula*. Below we formulate the Feynman-Kac formula in various ways. For a wide range of models this formula provides the PDEs and PIDEs that characterize corresponding pricing functions of derivatives. Furthermore, this section presents some results on transition probability densities, changes of measures, the Bayes rule, the Girsanov transformation and the Black-Scholes option pricing formula. These are useful when searching for analytic formulas of derivative prices and other functionals.

SDE for Some Factor Process

We consider a fixed time horizon $T \in (0, \infty)$ and a d -dimensional Markovian factor process $\mathbf{X}^{t,\mathbf{x}} = \{\mathbf{X}_s^{t,\mathbf{x}}, s \in [t, T]\}$, which satisfies the vector SDE

$$d\mathbf{X}_s^{t,\mathbf{x}} = \mathbf{a}(s, \mathbf{X}_s^{t,\mathbf{x}}) ds + \sum_{k=1}^m \mathbf{b}^k(s, \mathbf{X}_s^{t,\mathbf{x}}) dW_s^k \quad (2.7.1)$$

for $s \in [t, T]$ with initial value $\mathbf{X}_t^{t,\mathbf{x}} = \mathbf{x} \in \mathbb{R}^d$ at time $t \in [0, T]$. The process $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \in [0, T]\}$ is an m -dimensional standard Wiener process on a filtered probability space $(\Omega, \mathcal{A}, \mathcal{A}, P)$. The process $\mathbf{X}^{t,\mathbf{x}}$ has a drift coefficient $\mathbf{a}(\cdot, \cdot)$ and diffusion coefficients $\mathbf{b}^k(\cdot, \cdot)$, $k \in \{1, 2, \dots, m\}$. In general, $\mathbf{a} = (a^1, \dots, a^d)^\top$ and $\mathbf{b}^k = (b^{1,k}, \dots, b^{d,k})^\top$, $k \in \{1, 2, \dots, m\}$, represent vector valued functions on $[0, T] \times \mathbb{R}^d$ into \mathbb{R}^d , and we assume that a pathwise unique solution of the SDE (2.7.1) exists. The components of the SDE (2.7.1) can be the factors in a financial market model. In the following it will not matter which pricing approach one chooses. The task will be to evaluate simply some conditional expectations.

Terminal Payoff

Let us discuss the case of a European option, where we have a terminal payoff $H(\mathbf{X}_T^{t,\mathbf{x}})$ at the maturity date T with some given payoff function $H : \mathbb{R}^d \rightarrow [0, \infty)$ such that

$$E(|H(\mathbf{X}_T^{t,\mathbf{x}})|) < \infty. \quad (2.7.2)$$

We can then introduce the pricing function $u : [0, T] \times \mathbb{R}^d \rightarrow [0, \infty)$ as the conditional expectation

$$u(t, \mathbf{x}) = E(H(\mathbf{X}_T^{t,\mathbf{x}}) | \mathcal{A}_t) \quad (2.7.3)$$

for $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d$. The *Feynman-Kac formula* for this payoff refers to the fact that under sufficient regularity on $\mathbf{a}, \mathbf{b}^1, \dots, \mathbf{b}^m$ and H the function $u : (0, T) \times \mathbb{R}^d \rightarrow [0, \infty)$ satisfies the PDE

$$\begin{aligned}
L^0 u(t, \mathbf{x}) &= \frac{\partial u(t, \mathbf{x})}{\partial t} + \sum_{i=1}^d a^i(t, \mathbf{x}) \frac{\partial u(t, \mathbf{x})}{\partial x^i} \\
&\quad + \frac{1}{2} \sum_{i,k=1}^d \sum_{j=1}^m b^{i,j}(t, \mathbf{x}) b^{k,j}(t, \mathbf{x}) \frac{\partial^2 u(t, \mathbf{x})}{\partial x^i \partial x^k} \\
&= 0
\end{aligned} \tag{2.7.4}$$

for $(t, \mathbf{x}) \in (0, T) \times \mathbb{R}^d$ with terminal condition

$$u(T, \mathbf{x}) = H(\mathbf{x}) \tag{2.7.5}$$

for $\mathbf{x} \in \mathbb{R}^d$. Equation (2.7.4) is also called the *Kolmogorov backward equation*.

Note that in general, one needs also to specify the behavior of the solution of the PDE at its boundaries. In many cases this is not adding any extra information as is, for instance, the case under the Merton model in Sect. 3.5. However, in some cases, for instance under the minimal market model, as will be discussed in Sect. 3.6, it can be crucial. In financial applications it is important to set the boundaries such that strong arbitrage, in the sense as will be described in Definition 3.3.2, is excluded. This means that a derivative price needs to have an absorbing boundary condition at zero, where when it reaches zero, it keeps its value afterwards at zero.

The above type of European payoff will be covered by a version of the Feynman-Kac formula that we will present below. Under real world pricing the above version of the Feynman-Kac formula allows the calculation of the benchmarked pricing function under the real world probability measure. For instance, it can also be applied to determine the discounted pricing function under risk neutral pricing when the expectation can be taken of the discounted payoff under an equivalent risk neutral probability measure.

Discounted Payoff

We now generalize the above payoff function by discounting it, using a given *discount rate process* r , which is obtained as a function of the given vector diffusion process $\mathbf{X}^{t,\mathbf{x}}$, that is $r : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$. For instance, in a risk neutral setting the discount rate is given by the short term interest rate.

Over the period $[t, T]$ we consider for the *discounted payoff*

$$\exp \left\{ - \int_t^T r(s, \mathbf{X}_s^{t,\mathbf{x}}) ds \right\} H(\mathbf{X}_T^{t,\mathbf{x}})$$

the pricing function

$$u(t, \mathbf{x}) = E \left(\exp \left\{ - \int_t^T r(s, \mathbf{X}_s^{t,\mathbf{x}}) ds \right\} H(\mathbf{X}_T^{t,\mathbf{x}}) \middle| \mathcal{A}_t \right) \tag{2.7.6}$$

for $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d$. As will be shown below, it follows rather generally that the pricing function u satisfies the PDE

$$L^0 u(t, \mathbf{x}) = r(t, \mathbf{x}) u(t, \mathbf{x}) \quad (2.7.7)$$

for $(t, \mathbf{x}) \in (0, T) \times \mathbb{R}^d$ with terminal condition

$$u(T, \mathbf{x}) = H(\mathbf{x}) \quad (2.7.8)$$

for $\mathbf{x} \in \mathbb{R}^d$. Here the PDE operator L^0 is given as in (2.7.4).

Terminal Payoff and Payoff Rate

Now, we add to the above discounted payoff structure some payoff stream that continuously pays with a *payoff rate* $g : [0, T] \times \mathbb{R}^d \rightarrow [0, \infty)$ some amount per unit of time. This allows to model, for instance, continuous dividend payments for a share or continuous interest payments in a savings account. Also fees and insurance premia can be captured in this way. The corresponding *discounted payoff with payoff rate* is then of the form

$$\exp \left\{ - \int_t^T r(s, \mathbf{X}_s^{t, \mathbf{x}}) ds \right\} H(\mathbf{X}_T^{t, \mathbf{x}}) + \int_t^T \exp \left\{ - \int_t^s r(z, \mathbf{X}_z^{t, \mathbf{x}}) dz \right\} g(s, \mathbf{X}_s^{t, \mathbf{x}}) ds,$$

which leads to the pricing function

$$\begin{aligned} u(t, \mathbf{x}) &= E \left(\exp \left\{ - \int_t^T r(s, \mathbf{X}_s^{t, \mathbf{x}}) ds \right\} H(\mathbf{X}_T^{t, \mathbf{x}}) \right. \\ &\quad \left. + \int_t^T \exp \left\{ - \int_t^s r(z, \mathbf{X}_z^{t, \mathbf{x}}) dz \right\} g(s, \mathbf{X}_s^{t, \mathbf{x}}) ds \middle| \mathcal{A}_t \right) \end{aligned} \quad (2.7.9)$$

for $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d$. As will follow below, this pricing function satisfies the PDE

$$L^0 u(t, \mathbf{x}) + g(t, \mathbf{x}) = r(t, \mathbf{x}) u(t, \mathbf{x}) \quad (2.7.10)$$

for $(t, \mathbf{x}) \in (0, T) \times \mathbb{R}^d$ with terminal condition

$$u(T, \mathbf{x}) = H(\mathbf{x}) \quad (2.7.11)$$

for $\mathbf{x} \in \mathbb{R}^d$. As mentioned earlier, for certain dynamics boundary conditions may have to be added for completeness.

SDE with Jumps

We consider now an SDE with jumps describing the dynamics of the underlying factor process. Let Γ denote an open, connected subset of \mathbb{R}^d and

$T \in (0, \infty)$ a fixed time horizon. We consider for a d -dimensional process $\mathbf{X}^{t, \mathbf{x}} = \{\mathbf{X}_s^{t, \mathbf{x}}, s \in [t, T]\}$, see (1.5.19), the vector SDE

$$\begin{aligned} d\mathbf{X}_s^{t, \mathbf{x}} &= \mathbf{a}(s, \mathbf{X}_s^{t, \mathbf{x}}) ds + \sum_{k=1}^m \mathbf{b}^k(s, \mathbf{X}_s^{t, \mathbf{x}}) dW_s^k \\ &+ \sum_{j=1}^{\ell} \int_{\mathcal{E}} \mathbf{c}^j(v, s-, \mathbf{X}_{s-}^{t, \mathbf{x}}) p_{\varphi_j}^j(dv, ds) \end{aligned} \quad (2.7.12)$$

for $t \in [0, T]$ and $s \in [t, T]$ with value

$$\mathbf{X}_t^{t, \mathbf{x}} = \mathbf{x} \quad (2.7.13)$$

at time t and for $\mathbf{x} \in \Gamma$, see (1.6.16). Here $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \in [0, T]\}$ is again an m -dimensional standard Wiener process on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. Furthermore, $p_{\varphi_j}^j(\cdot, \cdot)$ denotes a Poisson measure, $j \in \{1, 2, \dots, \ell\}$, as introduced in Sect. 1.1, satisfying condition (1.1.33). Here $\mathbf{a} = (a^1, \dots, a^d)^\top$ and $\mathbf{b}^k = (b^{1,k}, \dots, b^{d,k})^\top$, $k \in \{1, 2, \dots, m\}$, are vector valued functions from $[0, T] \times \Gamma$ into \Re^d and $\mathbf{c}^j = (c^{1,j}, \dots, c^{d,j})^\top$, $j \in \{1, 2, \dots, \ell\}$, is a vector valued function on $\mathcal{E} \times [0, T] \times \Gamma$, where $\mathcal{E} = \Re \setminus \{0\}$.

For the above payoff structure, with discounted terminal payoff and a given payoff rate, the pricing function is of the form (2.7.9). As will be detailed below, under appropriate conditions u satisfies the PIDE

$$L^0 u(t, \mathbf{x}) + g(t, \mathbf{x}) = r(t, \mathbf{x}) u(t, \mathbf{x}) \quad (2.7.14)$$

for $(t, \mathbf{x}) \in (0, T)$ with terminal condition

$$u(T, \mathbf{x}) = H(\mathbf{x}) \quad (2.7.15)$$

for $\mathbf{x} \in \Re^d$. The corresponding operator L^0 is more general than the one in (2.7.4) and given in the form

$$\begin{aligned} L^0 u(t, \mathbf{x}) &= \sum_{i=1}^d a^i(t, \mathbf{x}) \frac{\partial u(t, \mathbf{x})}{\partial x^i} + \frac{1}{2} \sum_{i,k=1}^d \sum_{j=1}^m b^{i,j}(t, \mathbf{x}) b^{k,j}(t, \mathbf{x}) \frac{\partial^2 u(t, \mathbf{x})}{\partial x^i \partial x^k} \\ &+ \frac{\partial u(t, \mathbf{x})}{\partial t} + \sum_{j=1}^{\ell} \int_{\mathcal{E}} [u(s, x^1 + c^{1,j}(v, s, \mathbf{x}), \dots, x^d + c^{d,j}(v, s, \mathbf{x})) \\ &\quad - u(s, x^1, \dots, x^d)] \varphi_j(dv). \end{aligned} \quad (2.7.16)$$

Here we abuse slightly our notation by writing for $u(s, (x^1, \dots, x^d)^\top)$ also $u(s, x^1, \dots, x^d)$. Note that due to the jumps an extra integral term appears in (2.7.16) as a consequence of the Itô formula with jumps, see (1.5.15) and (1.5.20). Of course, also here some boundary conditions may have to be added to complete the characterization of the function u via the PIDE.

Payoff with First Exit Time

Derivatives like barrier options have a, so-called, *continuation region* Φ , which is an open connected subset of $[0, T] \times \Gamma$. The holder of an option continues to receive payments as long as the process $\mathbf{X}^{t, \mathbf{x}}$ stays in the continuation region Φ . For instance, in the case of a, so-called, *knock-out-barrier option* this would mean that $\mathbf{X}_s^{t, \mathbf{x}}$ has to stay below a given critical barrier to receive the terminal payment. To make this precise, we define the *first exit time* τ_Φ^t from Φ after t as

$$\tau_\Phi^t = \inf\{s \in [t, T] : (s, \mathbf{X}_s^{t, \mathbf{x}}) \notin \Phi\}, \quad (2.7.17)$$

which is a stopping time, see (1.2.11).

Consider now a general payoff structure with *terminal payoff function* $H : (0, T] \times \Gamma \rightarrow [0, \infty)$ for payments at time τ_Φ^t , a *payoff rate* $g : [0, T] \times \Gamma \rightarrow [0, \infty)$ for incremental payments during the time period $[t, \tau_\Phi^t]$ and a *discount rate* $r : [0, T] \times \Gamma \rightarrow \mathbb{R}$. Assume that the process $\mathbf{X}^{t, \mathbf{x}}$ does not explode or leave Γ before the terminal time T . We then define the *pricing function* $u : \Phi \rightarrow [0, \infty)$ by

$$\begin{aligned} u(t, \mathbf{x}) = & E \left(H(\tau_\Phi^t, \mathbf{X}_{\tau_\Phi^t}^{t, \mathbf{x}}) \exp \left\{ - \int_t^{\tau_\Phi^t} r(s, \mathbf{X}_s^{t, \mathbf{x}}) ds \right\} \right. \\ & \left. + \int_t^{\tau_\Phi^t} g(s, \mathbf{X}_s^{t, \mathbf{x}}) \exp \left\{ - \int_t^s r(z, \mathbf{X}_z^{t, \mathbf{x}}) dz \right\} ds \middle| \mathcal{A}_t \right) \end{aligned} \quad (2.7.18)$$

for $(t, \mathbf{x}) \in \Phi$.

For the formulation of the resulting PIDE of the function u we use the operator L^0 given in (2.7.16). Under sufficient regularity of Φ , $\mathbf{a}, \mathbf{b}^1, \dots, \mathbf{b}^m$, $\mathbf{c}^1, \dots, \mathbf{c}^\ell$, H , g , $\varphi_1, \dots, \varphi_\ell$ and r one can show by application of the Itô formula (1.5.15) and some martingale argument that the pricing function u satisfies the PIDE

$$L^0 u(t, \mathbf{x}) + g(t, \mathbf{x}) = r(t, \mathbf{x}) u(t, \mathbf{x}) \quad (2.7.19)$$

for $(t, \mathbf{x}) \in \Phi$ with boundary condition

$$u(t, \mathbf{x}) = H(t, \mathbf{x}) \quad (2.7.20)$$

for $(t, \mathbf{x}) \in ((0, T] \times \Gamma) \setminus \Phi$. This result links the functional (2.7.18) to the PIDE (2.7.19)–(2.7.20) and is often called a *Feynman-Kac formula*.

Generalized Feynman-Kac Formula

For a rather general situation, where $\Phi = (0, T) \times \Gamma$ and assuming no jumps, that is $\mathbf{c}^1 = \dots = \mathbf{c}^\ell = 0$ and $\tau_\Phi^t = T$, let us now formulate sufficient conditions that ensure that the Feynman-Kac formula holds, see [Heath & Schweizer \(2000\)](#) and [Platen & Heath \(2006\)](#).

- (A) The drift coefficient \mathbf{a} and diffusion coefficients \mathbf{b}^k , $k \in \{1, 2, \dots, m\}$, are assumed to be on $[0, T] \times \Gamma$ locally Lipschitz-continuous in \mathbf{x} , uniformly in t . That is, for each compact subset Γ^1 of Γ there exists a constant $K_{\Gamma^1} < \infty$ such that

$$|\mathbf{a}(t, \mathbf{x}) - \mathbf{a}(t, \mathbf{y})| + \sum_{k=1}^m |\mathbf{b}^k(t, \mathbf{x}) - \mathbf{b}^k(t, \mathbf{y})| \leq K_{\Gamma^1} |\mathbf{x} - \mathbf{y}| \quad (2.7.21)$$

for all $t \in [0, T]$ and $\mathbf{x}, \mathbf{y} \in \Gamma^1$.

- (B) For all $(t, \mathbf{x}) \in [0, T] \times \Gamma$ the solution $\mathbf{X}^{t, \mathbf{x}}$ of (2.7.12) neither explodes nor leaves Γ before T , that is

$$P\left(\sup_{t \leq s \leq T} |\mathbf{X}_s^{t, \mathbf{x}}| < \infty\right) = 1 \quad (2.7.22)$$

and

$$P(\mathbf{X}_s^{t, \mathbf{x}} \in \Gamma \text{ for all } s \in [t, T]) = 1. \quad (2.7.23)$$

- (C) There exists an increasing sequence $(\Gamma_n)_{n \in \mathcal{N}}$ of bounded, open and connected domains of Γ such that $\cup_{n=1}^{\infty} \Gamma_n = \Gamma$, and for each $n \in \mathcal{N}$ the PDE

$$L^0 u_n(t, \mathbf{x}) + g(t, \mathbf{x}) = r(t, \mathbf{x}) u_n(t, \mathbf{x}) \quad (2.7.24)$$

has a unique solution u_n in the sense of Friedman (1975) on $(0, T) \times \Gamma_n$ with boundary condition

$$u_n(t, \mathbf{x}) = u(t, \mathbf{x}) \quad (2.7.25)$$

on $((0, T) \times \partial\Gamma_n) \cup (\{T\} \times \Gamma_n)$, where $\partial\Gamma_n$ denotes the boundary of Γ_n .

- (D) The process $b^{i,k}(\cdot, \mathbf{X}_.) \frac{\partial u(\cdot, \mathbf{X}_.)}{\partial x^i}$ is measurable and square integrable on $[0, T]$ for all $i \in \{1, 2, \dots, d\}$ and $k \in \{1, 2, \dots, m\}$.

The proof of the following theorem is given in Platen & Heath (2006).

Theorem 2.7.1. *In the case without jumps and under the conditions (A), (B), (C) and (D), the function u given by (2.7.18) is the unique solution of the PDE (2.7.19) with boundary condition (2.7.20), where u is differentiable with respect to t and twice differentiable with respect to the components of \mathbf{x} .*

Condition (A) is satisfied if, for instance, \mathbf{a} and $\mathbf{b} = (\mathbf{b}^1, \dots, \mathbf{b}^m)$ are differentiable in \mathbf{x} on the open set $(0, T) \times \Gamma$ with derivatives that are continuous on $[0, T] \times \Gamma$.

To establish condition (B) one needs to exploit specific properties of the process $\mathbf{X}^{t, \mathbf{x}}$ given by the SDE (2.7.12).

Condition (C) can be shown to be implied by the following assumptions:

- (C1) There exists an increasing sequence $(\Gamma_n)_{n \in \mathcal{N}}$ of bounded, open and connected subdomains of Γ with $\Gamma_n \cup \partial\Gamma_n \subset \Gamma$ such that $\cup_{n=1}^{\infty} \Gamma_n = \Gamma$, and each Γ_n has a twice differentiable boundary $\partial\Gamma_n$.

- (C2) For each $n \in \mathcal{N}$ the functions \mathbf{a} and $\mathbf{b}\mathbf{b}^\top$ are uniformly Lipschitz-continuous on $[0, T] \times (\Gamma_n \cup \partial\Gamma_n)$.
- (C3) For each $n \in \mathcal{N}$ the function $\mathbf{b}(t, \mathbf{x})\mathbf{b}(t, \mathbf{x})^\top$ is uniformly elliptic on \Re^d for $(t, \mathbf{x}) \in [0, T] \times \Gamma_n$, that is, there exists a $\delta_n > 0$ such that

$$\mathbf{y}^\top \mathbf{b}(t, \mathbf{x}) \mathbf{b}(t, \mathbf{x})^\top \mathbf{y} \geq \delta_n |\mathbf{y}|^2 \quad (2.7.26)$$

for all $\mathbf{y} \in \Re^d$.

- (C4) For each $n \in \mathcal{N}$ the functions r and g are uniformly Hölder-continuous on $[0, T] \times (\Gamma_n \cup \partial\Gamma_n)$, that is, there exists a constant \bar{K}_n and an exponent $q_n > 0$ such that

$$|r(t, \mathbf{x}) - r(t, \mathbf{y})| + |g(t, \mathbf{x}) - g(t, \mathbf{y})| \leq \bar{K}_n |\mathbf{x} - \mathbf{y}|^{q_n} \quad (2.7.27)$$

for $t \in [0, T]$ and $\mathbf{x}, \mathbf{y} \in (\Gamma_n \cup \partial\Gamma_n)$.

- (C5) For each $n \in \mathcal{N}$ the function u is finite and continuous on $([0, T] \times \partial\Gamma_n) \cup (\{T\} \times (\Gamma_n \cup \partial\Gamma_n))$.

Condition (D) is satisfied, for instance, when

$$\int_0^T E \left(\left(b^{i,k}(t, \mathbf{X}_t) \frac{\partial u(t, \mathbf{X}_t)}{\partial x^i} \right)^2 \right) dt < \infty$$

for all $i \in \{1, 2, \dots, d\}$ and $k \in \{1, 2, \dots, m\}$. This condition ensures that the process $u(\cdot, \mathbf{X}_\cdot)$ is a martingale and that the PDE (2.7.19)–(2.7.20) has a unique solution.

Note that in the case when local Lipschitz continuity is not guaranteed, one may have to specify particular boundary conditions to obtain an appropriate description of the pricing function. This is a consequence of the fact that strict local martingales may drive the factor dynamics. These need extra care when defining the behavior of PDE solutions at boundaries.

Kolmogorov Equations

When the drift coefficient $a(\cdot)$ and diffusion coefficient $b(\cdot)$ of the solution of a scalar SDE without jumps are appropriate functions, then the corresponding transition probability density $p(s, x; t, y)$ satisfies a certain PDE. This is the *Kolmogorov forward equation* or *Fokker-Planck equation*

$$\frac{\partial p(s, x; t, y)}{\partial t} + \frac{\partial}{\partial y} \{a(t, y) p(s, x; t, y)\} - \frac{1}{2} \frac{\partial^2}{\partial y^2} \{b^2(t, y) p(s, x; t, y)\} = 0, \quad (2.7.28)$$

for (s, x) fixed. However, $p(s, x; t, y)$ satisfies also the *Kolmogorov backward equation*

$$\frac{\partial p(s, x; t, y)}{\partial s} + a(s, x) \frac{\partial p(s, x; t, y)}{\partial x} + \frac{1}{2} b^2(s, x) \frac{\partial^2 p(s, x; t, y)}{\partial x^2} = 0, \quad (2.7.29)$$

for (t, y) fixed. Obviously, the *initial condition* for both PDEs is given by the *Dirac delta function*

$$p(s, x; s, y) = \delta(y - x) = \begin{cases} \infty & \text{for } y = x \\ 0 & \text{for } y \neq x, \end{cases} \quad (2.7.30)$$

where

$$\int_{-\infty}^{\infty} \delta(y - x) dy = 1 \quad (2.7.31)$$

for given x . In the case of an SDE with jumps (2.7.28) and (2.7.29) become corresponding PIDEs. Of course, the Kolmogorov equations have their obvious multi-dimensional counterparts. In Sect. 2.2 we have given various examples of explicit transition probability densities. The PDEs and PIDEs that we obtain via the Feynman-Kac formula are often called Kolmogorov backward equations.

Stationary Densities

Let us consider solutions of SDEs which provide models for financial quantities that can evolve into some equilibrium. Obviously, we restrict then the class of processes that we consider. For example, such equilibria can be modeled by the Ornstein-Uhlenbeck process or the square root process, see Chap. 1. The corresponding transition probability densities converge over long periods of time towards corresponding stationary densities, see Sect. 2.2.

More precisely, for a diffusion process that permits some equilibrium its *stationary density* $\bar{p}(y)$ is defined as the solution of the integral equation

$$\bar{p}(y) = \int_{-\infty}^{\infty} p(s, x; t, y) \bar{p}(x) dx$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $y \in \mathfrak{R}$. This means, if one starts with the stationary density, then one obtains again the stationary density as the probability density of the process after any given time period. A stationary solution of an SDE is, therefore, obtained when the corresponding process starts with its stationary density.

It is important to know when the solution of an SDE has a stationary density. One can identify the stationary density \bar{p} by noting that it satisfies the corresponding stationary, or time-independent, Kolmogorov forward equation, see (2.7.28). In the case without jumps this *stationary Kolmogorov forward equation* reduces to the ordinary differential equation (ODE)

$$\frac{d}{dy} (a(y) \bar{p}(y)) - \frac{1}{2} \frac{d^2}{dy^2} (b^2(y) \bar{p}(y)) = 0 \quad (2.7.32)$$

with drift $a(x) = a(s, x)$ and diffusion coefficient $b(x) = b(s, x)$. Consequently, it is necessary that equation (2.7.32) is satisfied to ensure that a diffusion has

a stationary density. There is an obvious multi-dimensional PDE generalizing (2.7.32). In the case with jumps a corresponding integral term needs to be added in (2.7.32). We assume in the following that a unique stationary density exists for the process considered.

Note that since \bar{p} is a probability density it must satisfy the relation

$$\int_{-\infty}^{\infty} \bar{p}(y) dy = 1. \quad (2.7.33)$$

One can identify for a large class of scalar stationary diffusion processes the analytic form of their stationary density $\bar{p}(y)$. It is straightforward to check that the explicit expression

$$\bar{p}(y) = \frac{C}{b^2(y)} \exp \left\{ 2 \int_{y_0}^y \frac{a(u)}{b^2(u)} du \right\} \quad (2.7.34)$$

satisfies the equation (2.7.32) for $y \in \mathbb{R}$ with some fixed value $y_0 \in \mathbb{R}$. Here y_0 is some appropriately chosen point so that (2.7.34) makes sense. The constant C can be obtained from the normalization condition (2.7.33).

Ergodicity of a Process

Ergodicity is a concept that is somehow related to stationarity. It does not require the process to start with an initial random variable that already has the stationary density of the process. Ergodicity can be conveniently defined for processes with stationary densities. A process $X = \{X_t, t \in [0, \infty)\}$ is called *ergodic* if it has a stationary density \bar{p} and

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(X_t) dt = \int_{-\infty}^{\infty} f(x) \bar{p}(x) dx, \quad (2.7.35)$$

for all bounded measurable functions $f : \mathbb{R} \rightarrow \mathbb{R}$. That is, the limit as $T \rightarrow \infty$ of the random time average, specified on the left hand side of relation (2.7.35), equals the spatial average with respect to \bar{p} , as given on the right hand side of (2.7.35).

For a scalar SDE with drift $a(\cdot)$ and diffusion coefficient $b(\cdot)$ let us introduce the *scale measure* $s : \mathbb{R} \rightarrow \mathbb{R}^+$ given by

$$s(x) = \exp \left\{ -2 \int_{y_0}^x \frac{a(y)}{b^2(y)} dy \right\} \quad (2.7.36)$$

for $x \in \mathbb{R}$ with y_0 as in (2.7.34). For instance, in Borodin & Salminen (2002) one can find the following result:

A diffusion process with scale measure $s(\cdot)$ satisfying the following two properties:

$$\int_{y_0}^{\infty} s(x) dx = \int_{-\infty}^{y_0} s(x) dx = \infty \quad (2.7.37)$$

and

$$\int_{-\infty}^{\infty} \frac{1}{s(x) b^2(x)} dx < \infty \quad (2.7.38)$$

is ergodic and its stationary density \bar{p} is given by the expression (2.7.34). Of course, one can define the notion of ergodicity also for multi-dimensional processes and in the presence of jumps.

Change of Probability Measure

The ability to change measures before applying the Feynman-Kac formula can be powerful, not only in derivative pricing but also for many other tasks, including estimation, filtering and variance reduction, as we will see later in the book.

We denote by $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \in [0, T]\}$ an m -dimensional standard Wiener process on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ with \mathcal{A}_0 being the trivial σ -algebra, augmented by the sets of zero probability. For an $\underline{\mathcal{A}}$ -predictable m -dimensional stochastic process $\boldsymbol{\theta} = \{\boldsymbol{\theta}_t = (\theta_t^1, \dots, \theta_t^m)^\top, t \in [0, T]\}$ with

$$\int_0^T \sum_{i=1}^m (\theta_t^i)^2 dt < \infty \quad (2.7.39)$$

almost surely, let us assume that the strictly positive *Radon-Nikodym derivative process* $\Lambda_{\boldsymbol{\theta}} = \{\Lambda_{\boldsymbol{\theta}}(t), t \in [0, T]\}$, where

$$\Lambda_{\boldsymbol{\theta}}(t) = \exp \left\{ - \int_0^t \boldsymbol{\theta}_s^\top d\mathbf{W}_s - \frac{1}{2} \int_0^t \boldsymbol{\theta}_s^\top \boldsymbol{\theta}_s ds \right\} < \infty \quad (2.7.40)$$

almost surely for $t \in [0, T]$, is an $(\underline{\mathcal{A}}, P)$ -martingale. By the Itô formula (1.5.8) it follows from (2.7.40) that

$$\Lambda_{\boldsymbol{\theta}}(t) = 1 - \sum_{i=1}^m \int_0^t \Lambda_{\boldsymbol{\theta}}(s) \theta_s^i dW_s^i \quad (2.7.41)$$

for $t \in [0, T]$. Since $\Lambda_{\boldsymbol{\theta}}$ is assumed here to be an $(\underline{\mathcal{A}}, P)$ -martingale we have for $t \in [0, T]$

$$E(\Lambda_{\boldsymbol{\theta}}(t) | \mathcal{A}_0) = \Lambda_{\boldsymbol{\theta}}(0) = 1. \quad (2.7.42)$$

We can now define a measure $P_{\boldsymbol{\theta}}$ via the Radon-Nikodym derivative

$$\frac{dP_{\boldsymbol{\theta}}}{dP} = \Lambda_{\boldsymbol{\theta}}(T) \quad (2.7.43)$$

by setting

$$P_{\boldsymbol{\theta}}(A) = E(\Lambda_{\boldsymbol{\theta}}(T) \mathbf{1}_A) = E_{\boldsymbol{\theta}}(\mathbf{1}_A) \quad (2.7.44)$$

for $A \in \mathcal{A}_T$. Here $\mathbf{1}_A$ is the indicator function for A and $E_{\boldsymbol{\theta}}$ means expectation with respect to $P_{\boldsymbol{\theta}}$.

Note that P_{θ} is not just a measure but also a probability measure because

$$P_{\theta}(\Omega) = E(\Lambda_{\theta}(T)) = E(\Lambda_{\theta}(T) \mid \mathcal{A}_0) = \Lambda_{\theta}(0) = 1 \quad (2.7.45)$$

as a result of the martingale property of Λ_{θ} .

Bayes Rule

It is useful to be able to change the probability measure when taking conditional expectations. The following *Bayes rule* establishes a relationship between conditional expectations with respect to different equivalent probability measures.

Assume for an equivalent probability measure P_{θ} that the corresponding strictly positive Radon-Nikodym derivative process Λ_{θ} is an $(\underline{\mathcal{A}}, P)$ -martingale. Then for any given stopping time $\tau \in [0, T]$ and any \mathcal{A}_{τ} -measurable random variable Y , satisfying the integrability condition

$$E_{\theta}(|Y|) < \infty, \quad (2.7.46)$$

one can apply the *Bayes rule*

$$E_{\theta}(Y \mid \mathcal{A}_s) = \frac{E(\Lambda_{\theta}(\tau) Y \mid \mathcal{A}_s)}{E(\Lambda_{\theta}(\tau) \mid \mathcal{A}_s)} \quad (2.7.47)$$

for $s \in [0, \tau]$. This formula then allows us to change an expectation with respect to the real world probability to one with respect to, say, the risk neutral probability, under appropriate conditions as discussed in Sect. 3.3.

Girsanov Transformation

The following Girsanov transformation allows us to perform a measure transformation, which transforms a drifted Wiener process into a Wiener process under a new probability measure P_{θ} . More precisely, if for $T \in (0, \infty)$ a given strictly positive Radon-Nikodym derivative process Λ_{θ} is an $(\underline{\mathcal{A}}, P)$ -martingale, then the m -dimensional process $\mathbf{W}_{\theta} = \{\mathbf{W}_{\theta}(t), t \in [0, T]\}$, given by

$$\mathbf{W}_{\theta}(t) = \mathbf{W}_t + \int_0^t \theta_s ds \quad (2.7.48)$$

for all $t \in [0, T]$, is an m -dimensional standard Wiener process on the filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P_{\theta})$.

Note that certain assumptions need to be satisfied before one can apply the above Girsanov transformation. The key assumption is that Λ_{θ} must be a strictly positive $(\underline{\mathcal{A}}, P)$ -martingale. For instance, if the Radon-Nikodym derivative process is only a strictly positive local martingale, then this does not guarantee that P_{θ} is a probability measure, see [Platen & Heath \(2006\)](#).

A sufficient condition for the Radon-Nikodym derivative process Λ_{θ} to be an (\mathcal{A}, P) -martingale is the *Novikov condition*, see Novikov (1972a), which requires that

$$E \left(\exp \left\{ \frac{1}{2} \int_0^T \boldsymbol{\theta}_s^\top \boldsymbol{\theta}_s ds \right\} \right) < \infty. \quad (2.7.49)$$

This condition is fulfilled, for instance, for the Black-Scholes model. However, it is not satisfied for the minimal market model, which will be outlined in Sect. 3.6.

Black-Scholes Formula

The following famous Black-Scholes pricing formula for a European call option with geometric Brownian motion as the dynamics for the underlying risky security can be obtained in various ways. These include applications of the Feynman-Kac formula, the Bayes rule and the Girsanov transformation.

Let the underlying risky security S_t satisfy the SDE

$$dS_t = r_t S_t dt + \sigma_t S_t dW_{\theta}(t) \quad (2.7.50)$$

for $t \in [0, T]$ with initial value $S_0 > 0$. Here W_{θ} denotes a standard Wiener process under the risk neutral probability measure P_{θ} . The deterministic functions of time r_t and σ_t denote the interest rate and volatility, respectively. Furthermore, in this simple two-asset market there exists a savings account B_t , which accrues the deterministic interest rate such that

$$dB_t = r_t B_t dt \quad (2.7.51)$$

for $t \in [0, T]$ and $B_0 = 1$. The European call payoff at maturity $T \in (0, \infty)$ is of the form

$$H = (S_T - K)^+ \quad (2.7.52)$$

with $K > 0$ denoting the strike price.

The pricing function $V(t, S_t)$ satisfies, by the risk neutral pricing formula, see (3.3.7), the conditional expectation

$$V(t, S_t) = E_{\theta} \left(\frac{(S_T - K)^+}{B_T} \middle| \mathcal{A}_t \right) \quad (2.7.53)$$

for $t \in [0, T]$, where E_{θ} denotes the expectation under the risk neutral probability measure P_{θ} with W_{θ} denoting a standard Wiener process under P_{θ} . The process W_{θ} would be a drifted Wiener process under the real world probability, see (2.7.48). Via the Feynman-Kac formula (2.7.6)–(2.7.8) it is straightforward to confirm that the pricing function $V(t, S)$ satisfies the following Black-Scholes PDE

$$\frac{\partial V(t, S)}{\partial t} + r_t S \frac{\partial V(t, S)}{\partial S} + \frac{1}{2} \sigma_t^2 S^2 \frac{\partial^2 V(t, S)}{\partial S^2} - r_t V(t, S) = 0 \quad (2.7.54)$$

for $t \in (0, T)$ and $S \in (0, \infty)$, with terminal condition

$$V(T, S) = (S - K)^+ \quad (2.7.55)$$

for $S \in (0, \infty)$. Furthermore, it can be shown that the option pricing function $V(t, S_t)$ satisfies the seminal *Black-Scholes formula*

$$V(t, S_t) = S_t N(d_1(t)) - K \frac{B_t}{B_T} N(d_2(t)) \quad (2.7.56)$$

with

$$d_1(t) = \frac{\ln\left(\frac{S_t}{K}\right) + \int_t^T (r_s + \frac{1}{2} \sigma_s^2) ds}{\sqrt{\int_t^T \sigma_s^2 ds}}$$

and

$$d_2(t) = d_1(t) - \sqrt{\int_t^T \sigma_s^2 ds}$$

for $t \in [0, T]$, see [Black & Scholes \(1973\)](#). Here $N(\cdot)$ denotes the standard Gaussian distribution function.

2.8 Exercises

2.1. Derive the joint transition density for two correlated drifted Wiener processes with constant correlation $\varrho \in (-1, 1)$ and constant drifts $\mu_1, \mu_2 \in \Re$.

2.2. For the two-dimensional process in Exercise 2.1, write down a corresponding SDE driven by two independent Wiener processes W^1 and W^2 .

2.3. Write down the transition density of a two-dimensional geometric Brownian motion that is a martingale, where the driving Wiener processes are correlated with parameter $\varrho \in (-1, 1)$ and each of them having the volatility $b > 0$.

2.4. For the two-dimensional geometric Brownian motion in Exercise 2.3 write down the SDEs for its components driven by two independent Wiener processes W^1 and W^2 .

2.5. Consider the following Merton model SDE with jumps

$$dX_t = a X_t dt + b X_t dW_t + c X_{t-} dN_t,$$

for $t \geq 0$ with $X_0 > 0$, where W is a standard Wiener process independent of the Poisson process N , which has intensity $\lambda > 0$. Verify the Feynman-Kac formula for the functional

$$u(t, x) = E \left(H(X_T) \mid X_t = x \right)$$

for $(t, x) \in [0, T] \times \Re^+$.

2.6. Derive the first moment of a square root process with constant parameters $c > 0$, $b < 0$ and dimension $\delta > 2$ satisfying the SDE

$$dY_t = \left(\frac{\delta}{4} c^2 + b Y_t \right) dt + c \sqrt{Y_t} dW_t$$

for $t \in [0, \infty)$ and $Y_0 > 0$, where W is a Wiener process.

2.7. Prove that the ARCH diffusion model for squared volatility

$$d|\theta_t|^2 = \kappa (\bar{\theta}^2 - |\theta_t|^2) dt + \gamma |\theta_t|^2 dW_t$$

has an inverse gamma density as stationary density.

2.8. Show that the squared volatility of the model

$$d|\theta_t|^2 = \kappa |\theta_t|^2 (\bar{\theta}^2 - |\theta_t|^2) dt + \gamma |\theta_t|^3 dW_t$$

has an inverse gamma density.

2.9. Compute the stationary density for the squared volatility for the Heston model

$$d|\theta_t|^2 = \kappa (\bar{\theta}^2 - |\theta_t|^2) dt + \gamma |\theta_t| dW_t.$$

Benchmark Approach to Finance and Insurance

This chapter introduces a unified continuous time framework for financial and insurance modeling. It is applicable to portfolio optimization, derivative pricing, actuarial pricing and risk measurement when security price processes are modeled via SDEs with jumps. It follows the benchmark approach developed in [Platen & Heath \(2006\)](#). The jumps allow for the modeling of event risk in finance, insurance and other areas. The natural benchmark for asset allocation and the natural numéraire for pricing is represented by the best performing, strictly positive portfolio. This is shown to be the growth optimal portfolio (GOP) which maximizes expected growth. Any nonnegative portfolio, when expressed in units of the GOP, turns out to be a supermartingale. This fundamental property leads to real world pricing which identifies the minimal replicating price. An equivalent risk neutral probability measure need not exist under the benchmark approach, which provides significant freedom for modeling when compared to the classical approach.

3.1 Market Model

Continuous and Event Driven Uncertainty

We consider a market containing $d \in \mathcal{N}$ sources of traded uncertainty. *Continuous traded uncertainty* is represented by $m \in \{1, 2, \dots, d\}$ independent standard Wiener processes $\tilde{W}^k = \{\tilde{W}_t^k, t \geq 0\}$, $k \in \{1, 2, \dots, m\}$. These are defined on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. We also model events of certain types, for instance, corporate defaults, credit rating changes, operational failures or specified insured events, when these are reflected in the price moves of traded securities. Events of the k th type are counted by the $\underline{\mathcal{A}}$ -adapted k th *counting process* $p^k = \{p_t^k, t \geq 0\}$, whose *intensity* $h^k = \{h_t^k, t \geq 0\}$ is a given, predictable, strictly positive process with

$$h_t^k > 0 \quad (3.1.1)$$

and

$$\int_0^t h_s^k ds < \infty \quad (3.1.2)$$

almost surely for $t \geq 0$ and $k \in \{1, 2, \dots, d-m\}$. The k th counting process p^k leads to the k th *jump martingale* $q^k = \{q_t^k, t \geq 0\}$ with stochastic differential

$$dq_t^k = (dp_t^k - h_t^k dt) (h_t^k)^{-\frac{1}{2}} \quad (3.1.3)$$

for $k \in \{1, 2, \dots, d-m\}$ and $t \geq 0$. It is assumed that the above jump martingales do not jump at the same time. They represent the compensated, normalized sources of *event driven traded uncertainty*. The conditional variance of the increment of the k th source of event driven traded uncertainty over a time interval of length Δ equals

$$E \left((q_{t+\Delta}^k - q_t^k)^2 \mid \mathcal{A}_t \right) = \Delta \quad (3.1.4)$$

for all $t \geq 0$, $k \in \{1, 2, \dots, d-m\}$ and $\Delta \in [0, \infty)$. This is similar to what we obtain for a Wiener process. Note that we normalize the compensated counting process by $\sqrt{h_t^k}$ to make the drivers of the event risk similar to Wiener processes when the intensity is high. Then, in the latter case, there is not much difference between modeling via event driven uncertainty with high intensity or via some Wiener process. In this manner we can avoid the modeling of infinite intensity jump martingales.

The evolution of traded uncertainty is modeled by the vector process of independent (\mathcal{A}, P) -martingales $\mathbf{W} = \{\mathbf{W}_t = (\tilde{W}_t^1, \dots, \tilde{W}_t^m, q_t^1, \dots, q_t^{d-m})^\top, t \geq 0\}$. Note that $W^1 = \tilde{W}^1, \dots, W^m = \tilde{W}^m$ are Wiener processes, while $W^{m+1} = q^1, \dots, W^d = q^{d-m}$ are compensated, normalized counting processes. The filtration $\mathcal{A} = (\mathcal{A}_t)_{t \geq 0}$ satisfies the usual conditions and \mathcal{A}_0 is the trivial initial σ -algebra. Note that in addition to traded uncertainties the market typically involves also other uncertainties that impact jump intensities, short rates, appreciation rates, volatilities and other quantities. The modeling of these uncertainties needs not to be specified at this stage.

Primary Security Accounts

A *primary security account* is an investment account, consisting of only one kind of security. The j th risky primary security account value at time t is denoted by S_t^j , for $j \in \{1, 2, \dots, d\}$ and $t \geq 0$. These primary security accounts model the evolution of wealth due to the ownership of primary securities, with all dividends and income reinvested. The 0th primary security account $S^0 = \{S_t^0, t \geq 0\}$ is the domestic locally riskless savings account, which continuously accrues at the short term interest rate r_t .

Without loss of generality, we assume that the nonnegative j th primary security account value S_t^j satisfies the jump diffusion SDE

$$dS_t^j = S_{t-}^j \left(a_t^j dt + \sum_{k=1}^d b_t^{j,k} dW_t^k \right) \quad (3.1.5)$$

for $t \geq 0$ with initial value $S_0^j > 0$ and $j \in \{1, 2, \dots, d\}$. Recall that S_{t-}^j denotes the value of the process S^j just before time t , which is defined as the left hand limit at time t , see (1.3.9). We emphasize that this SDE is driven by the Wiener processes $W_t^1 = \tilde{W}_t^1, \dots, W_t^m = \tilde{W}_t^m$ and the jump martingales $W_t^k = q_t^{k-m}$ for $k \in \{m+1, \dots, d\}$, $t \geq 0$.

We assume that the short rate process r , the *appreciation rate processes* a^j , the *generalized volatility processes* $b^{j,k}$ and the intensity processes h^k take almost surely finite values and are predictable, $j \in \{1, 2, \dots, d\}$, $k \in \{1, 2, \dots, d-m\}$. They are assumed to be such that a unique strong solution of the system of SDEs (3.1.5) exists. To model the limited liability of companies, which means to ensure nonnegativity for each primary security account, we need to make the following assumption.

Assumption 3.1.1 *The condition*

$$b_t^{j,k} \geq -\sqrt{h_t^{k-m}} \quad (3.1.6)$$

holds for all $t \geq 0$, $j \in \{1, 2, \dots, d\}$ and $k \in \{m+1, \dots, d\}$.

Taking into account (3.1.3), it can be seen from the SDE (3.1.5) that this assumption excludes jumps that would lead to negative values for S_t^j , as discussed in Chap. 1.

Market Price of Risk

To securitize the different sources of traded uncertainty properly, we introduce the *generalized volatility matrix* $\mathbf{b}_t = [b_t^{j,k}]_{j,k=1}^d$ for all $t \geq 0$ and make the following assumption.

Assumption 3.1.2 *The generalized volatility matrix \mathbf{b}_t is invertible for Lebesgue-almost-every $t \geq 0$.*

Assumption 3.1.2 allows us to introduce the *market price of risk* vector

$$\boldsymbol{\theta}_t = (\theta_t^1, \dots, \theta_t^d)^\top = \mathbf{b}_t^{-1} [\mathbf{a}_t - r_t \mathbf{1}] \quad (3.1.7)$$

for $t \geq 0$. Here $\mathbf{a}_t = (a_t^1, \dots, a_t^d)^\top$ is the *appreciation rate vector* and $\mathbf{1} = (1, \dots, 1)^\top$ the *unit vector*. Using (3.1.7), we can rewrite the SDE (3.1.5) in the form

$$dS_t^j = S_{t-}^j \left(r_t dt + \sum_{k=1}^d b_t^{j,k} (\theta_t^k dt + dW_t^k) \right) \quad (3.1.8)$$

for $t \geq 0$ and $j \in \{0, 1, \dots, d\}$. For $k \in \{1, 2, \dots, m\}$, the quantity θ_t^k denotes the *market price of risk* with respect to the k th Wiener process W^k . If $k \in \{m+1, \dots, d\}$, then θ_t^k can be interpreted as the *market price of event risk* with respect to the counting process p^{k-m} . The market prices of risk play a central role and determine the risk premia that risky securities attract. For $j = 0$ in (3.1.8) we denote by S_t^0 the savings account, which is locally riskless with $b_t^{0,k} = 0$ for all $k \in \{1, 2, \dots, d\}$ and $t \geq 0$.

Portfolios

The vector process $\mathbf{S} = \{\mathbf{S}_t = (S_t^0, \dots, S_t^d)^\top, t \geq 0\}$ characterizes the evolution of all primary security accounts. We say that a predictable stochastic process $\boldsymbol{\delta} = \{\boldsymbol{\delta}_t = (\delta_t^0, \dots, \delta_t^d)^\top, t \geq 0\}$ is a strategy if the Itô integral $I_{\delta,W}(t)$ of the corresponding gains from trade exists, see Sect. 1.4. The j th component δ_t^j of $\boldsymbol{\delta}_t$ denotes the number of units of the j th primary security account held at time $t \geq 0$ in the portfolio S^δ , $j \in \{0, 1, \dots, d\}$. For a strategy $\boldsymbol{\delta}$ we denote by S_t^δ the value of the corresponding *portfolio process* at time t , when measured in units of the domestic currency. Thus, we set

$$S_t^\delta = \sum_{j=0}^d \delta_t^j S_t^j \quad (3.1.9)$$

for $t \geq 0$. A strategy $\boldsymbol{\delta}$ and the corresponding portfolio process $S^\delta = \{S_t^\delta, t \geq 0\}$ are *self-financing* if

$$dS_t^\delta = \sum_{j=0}^d \delta_t^j dS_t^j \quad (3.1.10)$$

for all $t \geq 0$. This means, instantaneous changes in the portfolio value are only due to instantaneous changes in the primary security accounts. We emphasize that $\boldsymbol{\delta}$ is assumed to be a predictable process and we consider only self-financing portfolios.

3.2 Best Performing Portfolio

Portfolios with Fractions

For a given strategy $\boldsymbol{\delta}$ with strictly positive portfolio process S^δ denote by $\pi_{\delta,t}^j$ the fraction of wealth that is invested in the j th primary security account at time t , that is,

$$\pi_{\delta,t}^j = \delta_t^j \frac{S_t^j}{S_t^\delta} \quad (3.2.1)$$

for $t \geq 0$ and $j \in \{0, 1, \dots, d\}$. These fractions sum to one, that is

$$\sum_{j=0}^d \pi_{\delta,t}^j = 1. \quad (3.2.2)$$

In terms of the vector of fractions $\boldsymbol{\pi}_{\delta,t} = (\pi_{\delta,t}^1, \dots, \pi_{\delta,t}^d)^\top$ we obtain from (3.1.10), (3.1.8) and (3.2.1) the SDE

$$dS_t^\delta = S_{t-}^\delta \{ r_t dt + \boldsymbol{\pi}_{\delta,t-}^\top \mathbf{b}_t (\boldsymbol{\theta}_t dt + d\mathbf{W}_t) \} \quad (3.2.3)$$

for $t \geq 0$, where $d\mathbf{W}_t = (d\tilde{W}_t^1, \dots, d\tilde{W}_t^m, dq_t^1, \dots, dq_t^{m-d})^\top$. Note by (3.1.3) that a portfolio process S^δ remains strictly positive if and only if

$$\sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} > -\sqrt{h_t^{k-m}} \quad (3.2.4)$$

almost surely for all $k \in \{m+1, \dots, d\}$ and $t \geq 0$.

Assumption on Market Price of Event Risk

For a strictly positive portfolio S^δ we obtain for its logarithm, by application of Ito's formula, the SDE

$$\begin{aligned} d \ln(S_t^\delta) &= g_t^\delta dt + \sum_{k=1}^m \sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} dW_t^k \\ &\quad + \sum_{k=m+1}^d \ln \left(1 + \sum_{j=1}^d \pi_{\delta,t-}^j \frac{b_t^{j,k}}{\sqrt{h_t^{k-m}}} \right) \sqrt{h_t^{k-m}} dW_t^k \end{aligned} \quad (3.2.5)$$

for $t \geq 0$. The *growth rate* at time t for S_t^δ equals

$$\begin{aligned} g_t^\delta &= r_t + \sum_{k=1}^m \left[\sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} \theta_t^k - \frac{1}{2} \left(\sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} \right)^2 \right] \\ &\quad + \sum_{k=m+1}^d \left[\sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} \left(\theta_t^k - \sqrt{h_t^{k-m}} \right) + \ln \left(1 + \sum_{j=1}^d \pi_{\delta,t}^j \frac{b_t^{j,k}}{\sqrt{h_t^{k-m}}} \right) h_t^{k-m} \right] \end{aligned} \quad (3.2.6)$$

for $t \geq 0$. Note that for the first sum on the right hand side of (3.2.6) a unique maximum exists because it is a quadratic form with respect to the fractions. Careful inspection of the terms in the second sum reveals that, in general, a unique maximum growth rate only exists if the market prices of event risks are less than the square roots of the corresponding jump intensities, see Christensen & Platen (2005). This important insight leads to the following crucial assumption on the market prices of event risk.

Assumption 3.2.1 *The intensities and market prices of event risk satisfy the inequality*

$$\sqrt{h_t^{k-m}} > \theta_t^k \quad (3.2.7)$$

for all $t \geq 0$ and $k \in \{m+1, \dots, d\}$.

Growth Optimal Portfolio

We shall see that Assumption 3.2.1 guarantees that there are no strictly positive portfolios in the market that could explode. Otherwise, we would have some obvious form of arbitrage. One notes from (3.2.6), if the risk premium for some events is too high, then the growth rate can become infinite. This condition allows us to introduce the predictable vector process $\mathbf{c}_t = (c_t^1, \dots, c_t^d)^\top$ with components

$$c_t^k = \begin{cases} \theta_t^k & \text{for } k \in \{1, 2, \dots, m\} \\ \frac{\theta_t^k}{1 - \theta_t^k (h_t^{k-m})^{-\frac{1}{2}}} & \text{for } k \in \{m+1, \dots, d\} \end{cases} \quad (3.2.8)$$

for $t \geq 0$. Note that a very large jump intensity with $\frac{\theta_t^k}{\sqrt{h_t^{k-m}}} \ll 1$ causes the corresponding component c_t^k to approach the market price of event risk θ_t^k for given $t \geq 0$ and $k \in \{m+1, \dots, d\}$.

We now define for $t \geq 0$ the fractions

$$\boldsymbol{\pi}_{\delta_*, t} = (\pi_{\delta_*, t}^1, \dots, \pi_{\delta_*, t}^d)^\top = (\mathbf{c}_t^\top \mathbf{b}_t^{-1})^\top \quad (3.2.9)$$

of a particular strictly positive portfolio S^{δ_*} , which we will identify below as the *growth optimal portfolio* (GOP). By (3.2.3) and (3.2.8) it follows that $S_t^{\delta_*}$ satisfies the SDE

$$\begin{aligned} dS_t^{\delta_*} &= S_{t-}^{\delta_*} \left(r_t dt + \mathbf{c}_t^\top (\boldsymbol{\theta}_t dt + d\mathbf{W}_t) \right) \\ &= S_{t-}^{\delta_*} \left(r_t dt + \sum_{k=1}^m \theta_t^k (\theta_t^k dt + dW_t^k) \right. \\ &\quad \left. + \sum_{k=m+1}^d \frac{\theta_t^k}{1 - \theta_t^k (h_t^{k-m})^{-\frac{1}{2}}} (\theta_t^k dt + dW_t^k) \right) \quad (3.2.10) \end{aligned}$$

for $t \geq 0$, with $S_0^{\delta_*} > 0$. Inspection of (3.2.10) shows that Assumption 3.2.1 keeps the portfolio process S^{δ_*} strictly positive. Let us now define a GOP in the given market.

Definition 3.2.2. *A strictly positive portfolio process S^δ that maximizes the growth rate g_t^δ , see (3.2.6), of all strictly positive portfolio processes S^δ such that $g_t^\delta \leq g_t^{\delta_*}$ almost surely for all $t \geq 0$ is called a GOP.*

The proof of the following result is given in [Platen & Heath \(2006\)](#).

Corollary 3.2.3 *Under the Assumptions 3.1.1, 3.1.2 and 3.2.1 the portfolio process $S^{\delta*} = \{S_t^{\delta*}, t \geq 0\}$ satisfying (3.2.10) is a GOP.*

As shown in [Platen & Heath \(2006\)](#) the GOP is in many ways the best performing portfolio of the given investment universe. Let us briefly mention one of its most striking properties, which is the fact that the path of the GOP outperforms in the long run the path of any other portfolio almost surely.

Theorem 3.2.4. *The GOP $S^{\delta*}$ has almost surely the largest long term growth rate in comparison with that of any other strictly positive portfolio S^δ , that is,*

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \ln \left(\frac{S_T^{\delta*}}{S_0^{\delta*}} \right) \geq \lim_{T \rightarrow \infty} \frac{1}{T} \ln \left(\frac{S_T^\delta}{S_0^\delta} \right) \quad (3.2.11)$$

almost surely.

The proof of this result is given in [Platen & Heath \(2006\)](#).

We mention that it will be shown in Sect. 3.4 that a global, well diversified portfolio appears to be a good proxy for the GOP. For instance, the MSCI world accumulation index is such a tradable portfolio that one can use as a reasonable proxy for the GOP for various purposes, as we will demonstrate later.

3.3 Supermartingale Property and Pricing

Benchmarked Portfolios

From now on we use the best performing portfolio of our market, the GOP, as *benchmark* for various risk management tasks. Obviously, it is a benchmark in the classical sense for portfolio investing. Additionally, it will also serve as natural numéraire in pricing. Furthermore, it shall be mentioned that in the area of risk measurement the GOP is ideally suited to play the role of the required well diversified portfolio that should be used when measuring general market risk in the regulatory sense, see [Platen & Stahl \(2003\)](#) and [Basle \(1995\)](#). We call prices, when expressed in units of $S^{\delta*}$, benchmarked prices. By the Itô formula and relations (3.2.3) and (3.2.10), a benchmarked portfolio process $\hat{S}^\delta = \{\hat{S}_t^\delta, t \geq 0\}$, with

$$\hat{S}_t^\delta = \frac{S_t^\delta}{S_t^{\delta*}} \quad (3.3.1)$$

for $t \geq 0$, satisfies the SDE

$$\begin{aligned} d\hat{S}_t^\delta &= \sum_{k=1}^m \left(\sum_{j=1}^d \delta_t^j \hat{S}_t^j b_t^{j,k} - \hat{S}_t^\delta \theta_t^k \right) dW_t^k \\ &\quad + \sum_{k=m+1}^d \left(\left(\sum_{j=1}^d \delta_t^j \hat{S}_{t-}^j b_t^{j,k} \right) \left(1 - \frac{\theta_t^k}{\sqrt{h_t^{k-m}}} \right) - \hat{S}_{t-}^\delta \theta_t^k \right) dW_t^k \end{aligned} \quad (3.3.2)$$

for $t \geq 0$.

The SDE (3.3.2) governs the dynamics of a benchmarked portfolio. For example, by (3.1.8) and (3.2.10) the benchmarked savings account \hat{S}_t^0 satisfies the SDE

$$d\hat{S}_t^0 = -\hat{S}_{t-}^0 \sum_{k=1}^d \theta_t^k dW_t^k \quad (3.3.3)$$

for $t \geq 0$.

Supermartingale Property

Note that the right hand side of (3.3.2) is driftless. Thus, a nonnegative benchmarked portfolio \hat{S}^δ forms an $(\underline{\mathcal{A}}, P)$ -local martingale, see [Ansel & Stricker \(1994\)](#). This provides for nonnegative \hat{S}^δ the most important property of our market, which is the following *supermartingale property*.

Theorem 3.3.1. *Any nonnegative benchmarked portfolio process \hat{S}^δ is an $(\underline{\mathcal{A}}, P)$ -supermartingale, that is,*

$$\hat{S}_t^\delta \geq E \left(\hat{S}_\tau^\delta \mid \mathcal{A}_t \right) \quad (3.3.4)$$

for all bounded stopping times $\tau \in [0, \infty)$ and $t \in [0, \tau]$.

A proof of this theorem can be found in [Platen & Heath \(2006\)](#). It essentially states that the GOP is truly best performing so that any other nonnegative portfolio when denominated in units of the GOP can trend only downwards or has at most no trend. The benchmark that satisfies the supermartingale property (3.3.4) is also called the *numéraire portfolio*, see [Long \(1990\)](#) and [Becherer \(2001\)](#). We emphasize that the fundamental fact that nonnegative benchmarked portfolios are supermartingales, holds even in general semimartingale markets as long as a finite numéraire portfolio exists, see [Platen \(2004a\)](#).

Strong Arbitrage

The classical idea of pricing by the exclusion of some standard form of arbitrage dominated, for several decades, theory and practice in derivative pricing, see [Ross \(1976\)](#) and [Harrison & Kreps \(1979\)](#).

Based on the above supermartingale property of nonnegative benchmarked portfolios and a strong notion of arbitrage that will be introduced below, it will become clear that pricing based on excluding strong arbitrage makes little sense in our general setting.

Definition 3.3.2. *A nonnegative portfolio that starts at zero and reaches at some later time a strictly positive value with strictly positive probability is called a strong arbitrage.*

By the supermartingale property (3.3.4) we know that all nonnegative benchmarked portfolios are supermartingales. Applying the well-known fact that a nonnegative supermartingale that starts at zero always remains at zero, simply yields the following conclusion.

Corollary 3.3.3 *There is no nonnegative portfolio that is a strong arbitrage.*

Consequently, we make the crucial observation that pricing on the basis of excluding strong arbitrage makes no sense in our general market. Essentially, by assuming in our market that a GOP exists we obtain a far richer modeling world than the classical theory can cover. This also means that there may be free lunches with vanishing risk in our market in the sense of Delbaen & Schachermayer (1998) and also some other weak forms of classical arbitrage. As we will see, this does not create any problem in reality. On the contrary, it extends significantly the universe of potential models allowing us to better reflect real market behavior, in particular, in the long run.

Law of the Minimal Price

As a consequence of the possible existence of classical forms of arbitrage we cannot rely on the classical no-arbitrage pricing methodology as in, for example, the risk neutral paradigm, see Harrison & Kreps (1979).

For preparing a general pricing concept we introduce the notion of a *fair security*, for which its benchmarked price is simply an (\mathcal{A}, P) -martingale. It is a basic and well-known fact that in a family of supermartingales, which all share the same nonnegative value at some future time, it is the martingale that attains now the minimal possible price, see Platen & Heath (2006). This insight leads us to the following *Law of the Minimal Price*:

Corollary 3.3.4 *Consider a bounded stopping time $\tau \in (0, \infty)$ and a given future \mathcal{A}_τ -measurable payoff H to be paid at time τ , where $E(\frac{H}{S_\tau^\delta} | \mathcal{A}_0) < \infty$. If there exists a fair nonnegative portfolio S^δ with $S_\tau^\delta = H$ almost surely, then this is the minimal possible nonnegative portfolio that replicates the payoff.*

This means a fair portfolio provides the least expensive choice for an investor to reach some payoff H at time τ .

As we will see, there exist in our general market setting self-financing portfolios that are not fair. This means the classical Law of One Price needs to be substituted by the above Law of the Minimal Price.

Real World Pricing

Let H denote an \mathcal{A}_τ -measurable payoff, with $E(\frac{H}{S_t^{\delta_*}}) < \infty$, to be paid at a stopping time $\tau \in [0, \infty)$. By the Law of the Minimal Price applied to the payoff H , its fair price $U_H(t)$ at time $t \in [0, \tau]$ is then the minimal possible price, which is given by the *real world pricing formula*

$$U_H(t) = S_t^{\delta_*} E \left(\frac{H}{S_\tau^{\delta_*}} \mid \mathcal{A}_t \right). \quad (3.3.5)$$

This formula represents an absolute pricing rule. It makes pricing an investment decision and values the payoff relative to the best performing portfolio under the real world expectation. We emphasize that its numéraire is the GOP, which is the best performing portfolio. The expectation is taken under the real world probability measure. By the supermartingale property we know that any other replicating self-financing portfolio process can only be more expensive than the fair price process. In a competitive market the real world pricing formula provides the economically correct price.

Risk Neutral Pricing

As shown in [Platen & Heath \(2006\)](#), real world pricing is equivalent to risk neutral pricing as long as, the candidate Radon-Nikodym derivative

$$\Lambda_\theta(t) = \frac{S_t^0}{S_t^{\delta_*}} \frac{S_0^{\delta_*}}{S_0^0} \quad (3.3.6)$$

for the putative risk neutral probability measure forms an (\mathcal{A}, P) -martingale, see also Sect. 2.7. In this case the *risk neutral pricing formula* results from (3.3.5) by the Bayes rule in the form

$$U_H(t) = E \left(\frac{\Lambda_\theta(T)}{\Lambda_\theta(t)} \frac{S_t^0}{S_T^0} H \mid \mathcal{A}_t \right) = S_t^0 E_\theta \left(\frac{H}{S_T^0} \mid \mathcal{A}_t \right). \quad (3.3.7)$$

Here E_θ denotes the expectation under the risk neutral pricing measure P_θ , with $\frac{dP_\theta}{dP}|_{\mathcal{A}_T} = \Lambda_\theta(T)$.

In the case when the benchmarked savings account has a downward trend, which means it is a strict supermartingale, $\Lambda_\theta(t)$ does not form a martingale and risk neutral prices can be more expensive than fair prices. For more details on these issues we refer to [Platen & Heath \(2006\)](#).

A simple derivative is the fair zero coupon bond. It pays one unit of the domestic currency at the given maturity date $T \in [0, \infty)$. By the real world

pricing formula (3.3.5) the price $P(t, T)$ at time t for the fair zero coupon bond is given by the conditional expectation

$$P(t, T) = E \left(\frac{S_t^{\delta_*}}{S_T^{\delta_*}} \mid \mathcal{A}_t \right) \quad (3.3.8)$$

for $t \in [0, T]$, $T \in [0, \infty)$. In the case of a deterministic short rate r_t , the supermartingale property of the benchmarked savings account \hat{S}_t^0 yields by (3.3.8) that

$$P(t, T) = \exp \left\{ - \int_t^T r_s ds \right\} E \left(\frac{\hat{S}_T^0}{\hat{S}_t^0} \mid \mathcal{A}_t \right) \leq P^*(t, T) \quad (3.3.9)$$

with $P^*(t, T) = \exp \left\{ - \int_t^T r_s ds \right\}$ denoting the savings bond. The latter would be the zero coupon bond price under classical risk neutral pricing.

In reality one observes that the savings account, when benchmarked by a global equity market index trends systematically downwards in the long run and is more realistically modeled by a strict supermartingale than a martingale. The above benchmark approach can accommodate this fact. The classical risk neutral approach ignores any trends arising in reality and represents a form of relative pricing as opposed to the absolute pricing of the real world pricing formula. As soon as a derivative has been priced and traded, risk neutral pricing allows us to calibrate models consistently with other derivatives that may be perfectly hedged. Still, all these risk neutral prices may be significantly higher than the corresponding minimal prices obtained through absolute pricing via the real world pricing formula, as we will demonstrate in Sect. 3.6. If the derivatives industry follows risk neutral pricing, then it is possible that large classes of derivatives can be overpriced on a macro-economic scale.

Finally, we remark that the *actuarial pricing formula*, see Bühlmann & Platen (2003), arises from (3.3.7) when H is independent of $S_T^{\delta_*}$, yielding

$$U_H(t) = P(t, T) E(H \mid \mathcal{A}_t), \quad (3.3.10)$$

with the fair zero coupon bond $P(t, T)$ as discount factor. This can be interpreted as a form of absolute pricing. Thus, real world pricing unifies actuarial and derivative pricing and makes both an investment decision.

3.4 Diversification

We have seen that there is no strong arbitrage in our general market setting. The only significant free benefit that an investment strategy can bring is through diversification, as will be explained below.

This section considers diversified portfolios in a sequence of markets. It presents the Diversification Theorem, see Platen (2005b), which is model independent and states that diversified portfolios approximate the GOP.

Sequence of Markets

We continue to rely on a filtered probability space $(\Omega, \mathcal{A}, \mathcal{A}, P)$. Continuous traded uncertainty is represented by independent standard Wiener processes $\tilde{W}^k = \{\tilde{W}_t^k, t \in [0, \infty)\}$ for $k \in \mathcal{N}$. Event driven traded uncertainty is modeled by counting processes $p^k = \{p_t^k, t \geq 0\}$ characterized by corresponding predictable, strictly positive intensity processes $h^k = \{h_t^k, t \geq 0\}$ for $k \in \mathcal{N}$. We define the k th jump martingale $q^k = \{q_t^k, t \geq 0\}$ as in (3.1.3), for $k \in \mathcal{N}$.

In what follows, we consider a sequence of markets indexed by the number $d \in \mathcal{N}$ of risky primary security accounts. For a given integer d , the corresponding market comprises $d + 1$ primary security accounts. To be precise we write for the j th primary security account in the j th market $S_{(d)}^j(t)$ satisfying an SDE as given in (3.1.8). The primary security accounts include a savings account $S_{(d)}^0 = \{S_{(d)}^0(t), t \geq 0\}$, whose value at time t is given by the exponential $S_{(d)}^0(t) = \exp \left\{ \int_0^t r_s ds \right\}$ for $t \geq 0$. Here $r = \{r_t, t \geq 0\}$ denotes an adapted short rate process, which we assume, for simplicity, to be the same in each market. We include d nonnegative, risky primary security account processes $S_{(d)}^j = \{S_{(d)}^j(t), t \geq 0\}$, $j \in \{1, 2, \dots, d\}$, in the d th market, each of which can be driven by the Wiener processes $\tilde{W}^1, \tilde{W}^2, \dots, \tilde{W}^m$ and the jump martingales q^1, q^2, \dots, q^{d-m} . Here $m = [\mu d]$ denotes the largest integer not exceeding μd with $\mu \in [0, 1]$ a fixed real number. In the d th market we have then the traded uncertainty driven by the d -dimensional vector process $\mathbf{W} = \{\mathbf{W}_t = (\tilde{W}_t^1, \dots, \tilde{W}_t^m, q_t^1, \dots, q_t^{d-m})^\top, t \geq 0\}$. Obviously, if μ equals one, then we model no jumps in our markets. Of course, there may exist many other sources of nontraded uncertainty in our market, which we do not need to specify for our purposes.

In the d th market a given strategy δ and the volatilities and market prices of risk depend typically on d . For simplicity, we shall suppress these dependencies in our notation and only mention it when required.

For the d th market we assume that there exists a unique GOP $S_{(d)}^{\delta_*} = \{S_{(d)}^{\delta_*}(t), t \geq 0\}$, satisfying the SDE (3.2.10) where we fix the initial value to

$$S_{(d)}^{\delta_*}(0) = 1. \quad (3.4.1)$$

Any portfolio $S_{(d)}^\delta$ in the d th market, when expressed in units of $S_{(d)}^{\delta_*}$, yields a corresponding benchmarked portfolio $\hat{S}_{(d)}^\delta = \{\hat{S}_{(d)}^\delta(t), t \geq 0\}$ defined by

$$\hat{S}_{(d)}^\delta(t) = \frac{S_{(d)}^\delta(t)}{S_{(d)}^{\delta_*}(t)} \quad (3.4.2)$$

at time $t \geq 0$, which by (3.3.2) satisfies a driftless SDE.

General and Specific Market Risk

To obtain a compact formulation of the SDE (3.3.2), let us define in the d th market the (j, k) th *specific generalized volatility* $\sigma_{(d)}^{j,k}(t)$ by setting

$$\sigma_{(d)}^{0,k}(t) = \theta_t^k \quad (3.4.3)$$

for $j = 0$ and $k \in \{1, 2, \dots, d\}$, and

$$\sigma_{(d)}^{j,k}(t) = \begin{cases} \theta_t^k - b_t^{j,k} & \text{for } k \in \{1, 2, \dots, m\} \\ \theta_t^k - b_t^{j,k} \left(1 - \frac{\theta_t^k}{\sqrt{h_t^{k-m}}}\right) & \text{for } k \in \{m+1, \dots, d\} \end{cases} \quad (3.4.4)$$

for $t \geq 0$ and $j \in \{1, 2, \dots, d\}$. Here the notations of θ_t^k , $b_t^{j,k}$ and h_t^{k-m} suppress the dependencies on d . One can say that the market prices of risk model the *general market risk* which relates to the movements of the market as a whole. The specific generalized volatilities of the j th primary security account model then the *specific market risk* of the j th primary security account. This idiosyncratic risk measures its movements against the market as a whole. This interpretation allows us to accommodate, conveniently, regulatory requirements and guidelines on risk measurement, see [Basle \(1995\)](#) and [Platen & Stahl \(2003\)](#). By using (3.4.4) and (3.4.3) one can rewrite the SDE (3.3.2) in the form

$$d\hat{S}_{(d)}^\delta(t) = - \sum_{k=1}^d \sum_{j=0}^d \delta_t^j \hat{S}_{(d)}^j(t-) \sigma_{(d)}^{j,k}(t) dW_t^k, \quad (3.4.5)$$

and for strictly positive $S_{(d)}^\delta(t)$ as

$$d\hat{S}_{(d)}^\delta(t) = -\hat{S}_{(d)}^\delta(t-) \sum_{k=1}^d \sum_{j=0}^d \pi_{\delta,t-}^j \sigma_{(d)}^{j,k}(t) dW_t^k \quad (3.4.6)$$

for $t \geq 0$, using the notation (3.2.1) for fractions, where we suppress the dependence on d .

For the above SDEs and the market model to make sense, the specific generalized volatilities need to be finite and the benchmarked primary security accounts have to remain nonnegative. To secure these properties we make the following realistic assumption.

Assumption 3.4.1 *For all $d \in \mathcal{N}$, $T \in [0, \infty)$ and $j \in \{0, 1, \dots, d\}$ suppose that*

$$\int_0^T \sum_{k=1}^d \left(\sigma_{(d)}^{j,k}(t)\right)^2 dt \leq \bar{K}_T < \infty \quad (3.4.7)$$

almost surely, where $\bar{K}_T < \infty$ denotes some finite \mathcal{A}_T -measurable random variable which does not depend on d . Furthermore, the inequalities

$$\sigma_{(d)}^{j,k}(t) < \sqrt{h_t^{k-m}} \quad (3.4.8)$$

and

$$\theta_t^k < \sqrt{h_t^{k-m}} \quad (3.4.9)$$

hold almost surely for all $t \geq 0$, $k \in \{m+1, m+2, \dots, d\}$ and $j \in \{0, 1, \dots, d\}$ and the generalized volatility matrix is invertible for each of the markets.

Sequence of Diversified Portfolios

Now, we can form a sequence of markets indexed by the number d of risky primary security accounts. For such a sequence we will identify sequences of portfolios that approximate the corresponding sequence of GOPs.

First, let us introduce the definition for a sequence of *diversified portfolios*.

Definition 3.4.2. For a sequence of markets we call a corresponding sequence $(S_{(d)}^\delta)_{d \in \mathcal{N}}$ of strictly positive portfolio processes $S_{(d)}^\delta$ a sequence of diversified portfolios if some constants $K_1, K_2 \in (0, \infty)$ and $K_3 \in \mathcal{N}$ exist, independently of d , such that for $d \in \{K_3, K_3 + 1, \dots\}$ the inequality

$$|\pi_{\delta,t}^j| \leq \frac{K_2}{d^{\frac{1}{2}+K_1}} \quad (3.4.10)$$

holds almost surely for all $j \in \{0, 1, \dots, d\}$ and $t \geq 0$.

Note that in (3.4.10) the strategy δ depends on d . Essentially, we require that the fraction that a diversified portfolio holds in a given primary security account vanishes slightly faster than the order $d^{-\frac{1}{2}}$ as d tends to infinity.

We introduce for all $t \geq 0$, $d \in \mathcal{N}$ and $k \in \{1, 2, \dots, d\}$ the k th total specific volatility for the d th market in the form

$$\hat{\sigma}_{(d)}^k(t) = \sum_{j=0}^d |\sigma_{(d)}^{j,k}(t)|. \quad (3.4.11)$$

Depending on k , the k th total specific volatility represents the sum of the absolute values of the specific generalized volatilities with respect to the k th traded uncertainty.

The following *regularity property* of a sequence of markets ensures that each of the independent sources of traded uncertainty influences only a restricted set of benchmarked primary security accounts.

Definition 3.4.3. A sequence of markets is called regular if there exists a constant $K_4 \in (0, \infty)$, independent of d , such that

$$E \left(\left(\hat{\sigma}_{(d)}^k(t) \right)^2 \right) \leq K_4 \quad (3.4.12)$$

for all $t \geq 0$, $d \in \mathcal{N}$ and $k \in \{1, 2, \dots, d\}$.

Note that in the special case when each benchmarked risky primary security account is independent from the other ones, then the market is regular.

Diversification Theorem

Consider for given $d \in \mathcal{N}$ in the d th market a strictly positive portfolio process $S_{(d)}^\delta$ with strategy $\boldsymbol{\delta} = \{\boldsymbol{\delta}_t = (\delta_t^0, \delta_t^1, \dots, \delta_t^d)^\top, t \geq 0\}$. Let us introduce the *tracking rate*

$$R_{(d)}^\delta(t) = \sum_{k=1}^d \left(\sum_{j=0}^d \pi_{\delta,t}^j \sigma_{(d)}^{j,k}(t) \right)^2 \quad (3.4.13)$$

at time $t \geq 0$ for the portfolio $S_{(d)}^\delta$. By (3.4.6) one notes that the benchmarked portfolio $\hat{S}_{(d)}^\delta$ is constant with

$$\hat{S}_{(d)}^\delta(t) = \hat{S}_{(d)}^\delta(0) \quad (3.4.14)$$

if and only if the tracking rate vanishes, that is,

$$R_{(d)}^\delta(t) = 0 \quad (3.4.15)$$

for all $t \geq 0$. Obviously, in the case of a constant benchmarked portfolio $\hat{S}_{(d)}^\delta$, the portfolio

$$S_{(d)}^\delta(t) = \hat{S}_{(d)}^\delta(0) S_{(d)}^{\delta_*}(t) \quad (3.4.16)$$

equals, by relation (3.4.2), a multiple of the GOP for all $t \geq 0$. Therefore, a given portfolio process $S_{(d)}^\delta$ equals, up to a constant factor, the GOP if the tracking rate $R_{(d)}^\delta(t)$ vanishes for all $t \geq 0$. We formalize this simple fact by the following definition.

Definition 3.4.4. *For a sequence of markets we call a sequence of strictly positive portfolio processes $(S_{(d)}^\delta)_{d \in \mathcal{N}}$ a sequence of approximate GOPs if for all $t \geq 0$ the corresponding sequence of tracking rates vanishes in probability. That is, for each $\varepsilon > 0$ we have*

$$\lim_{d \rightarrow \infty} P \left(R_{(d)}^\delta(t) > \varepsilon \right) = 0 \quad (3.4.17)$$

for all $t \geq 0$.

Now, we can state the *Diversification Theorem* proved in [Platen \(2005b\)](#).

Theorem 3.4.5. *For a regular sequence of markets, each sequence $(S_{(d)}^\delta)_{d \in \mathcal{N}}$ of diversified portfolios is a sequence of approximate GOPs.*

This result is highly relevant for the practical applicability of the benchmark approach. In particular, it allows us to approximate the GOP by a diversified market index, without the need of an exact parameter estimation and subsequent calculation of the fractions of the GOP. In practice it is not a realistic task to estimate risk premia, since even under the simple Black-Scholes model it takes several hundred years to obtain any reasonable estimate for its

drift parameter, see DeMiguel, Garlappi & Uppal (2009). We emphasize that the above statement on diversification is very robust since it is model independent. The following examples, which use exact or almost exact simulations, illustrate the robustness of the diversification effect stated in Theorem 3.4.5 and are taken from Platen & Rendek (2009b).

Diversification in Various Markets

We simulate over $T = 35$ years, similar as in Platen & Rendek (2009b), the driftless benchmarked trajectories of $d + 1 = 50$ primary security accounts according to some given dynamics. For simplicity, we set the interest rate to zero. Thus, the inverse of the benchmarked savings account $\hat{S}_{(d)}^0(t)$ equals the GOP when denominated in domestic currency, that is $S_{(d)}^{\delta_*}(t) = (\hat{S}_{(d)}^0(t))^{-1}$. The product of the GOP with the j th benchmarked primary security account yields the value of the primary security account in domestic currency, that is

$$S_{(d)}^j(t) = \hat{S}_{(d)}^j(t) S_{(d)}^{\delta_*}(t), \quad (3.4.18)$$

$j \in \{1, 2, \dots, d\}$. For simplicity we start each of the above mentioned price processes at $S_{(d)}^j(0) = 1$ at time zero. We will plot the first twenty primary security account prices in a first figure. In a separate second figure we will compare the GOP, the market capitalization weighted index (MCI) containing one stock of each primary security account, with the equi-value weighted index (EWI), which always keeps the same fraction for each stock.

We study below these three portfolios for three different market models. These examples will demonstrate that a diversified portfolio does not need an extremely large number of primary security accounts to become a reasonable approximate GOP. In practice, the number of investable primary security accounts is very large and reaches more than 10,000. This allows us to expect that a broadly diversified market index is likely to be a good proxy of the GOP.

First, let us illustrate the fundamental phenomenon of diversification by simulating, in a Black-Scholes market, the mentioned diversified portfolios. By following our discussion in Sect. 2.4 this can be performed exactly without any error. For simplicity, the benchmarked primary security accounts are assumed to be independent with constant volatilities of magnitude 0.2. In Fig. 3.4.1 we display the resulting values of the first twenty asset prices. Note that we first simulated driftless benchmarked primary security accounts. From the benchmarked savings account we obtained the GOP. This then gave us the primary security accounts, see (3.4.18).

In Fig. 3.4.2 we show the corresponding simulated GOP, the MCI and the EWI. In this case the EWI approximates rather well the GOP. The MCI seems to be still reasonably diversified and represents a good proxy for the GOP. The Diversification Theorem appears to be applicable in the given situation.

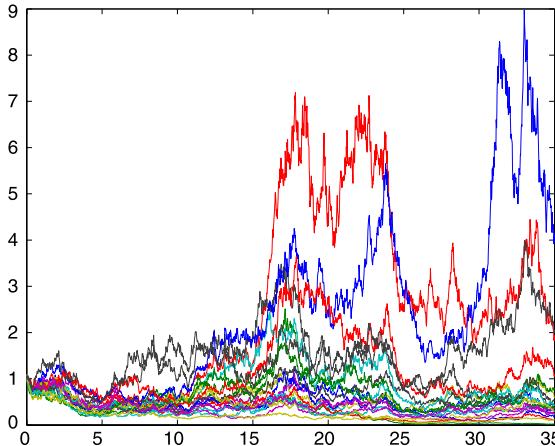


Fig. 3.4.1. Primary security accounts under the Black-Scholes model

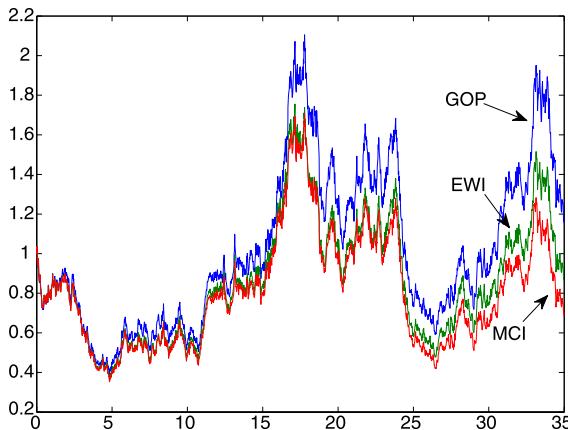


Fig. 3.4.2. GOP, MCI and EWI for the Black-Scholes model

As second market model we use for the benchmarked primary security accounts the multi-dimensional Heston model with independent prices, as described in Sect. 2.5. The square root processes in (2.5.13) are chosen such that an initial and average squared volatility emerge with a value of 0.04. The dimension of the independent square root process is set to $\delta = 4$, see (2.3.66). The constant parameter process b is set to $b_t = b = 0.2$. We plot in Fig. 3.4.3 twenty of the resulting primary security accounts. In Fig. 3.4.4 we display the resulting simulated GOP, EWI and MCI when we use the 49 risky primary

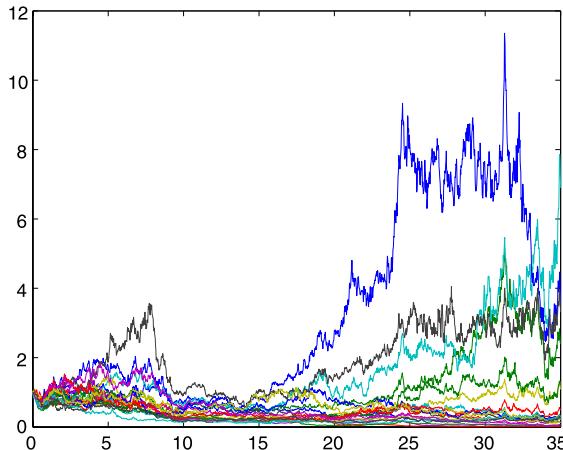


Fig. 3.4.3. Primary security accounts under the Heston model

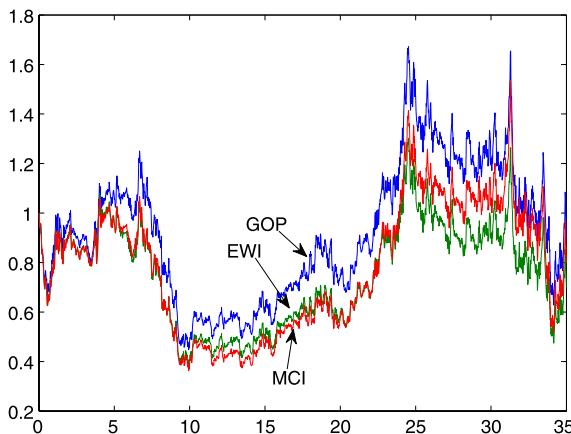


Fig. 3.4.4. GOP, MCI and EWI for the Heston model

security accounts. Here the EWI provides a good proxy for the GOP. Also the MCI represents a reasonable approximation of the GOP.

As the third example we simulate the benchmarked primary security accounts via a multi-dimensional ARCH diffusion as described in Sect. 2.6. We use the same squared volatility process for all benchmarked primary security accounts, where we set in (2.6.2) $\kappa = 3.4$, $\theta = 0.04$, $\sigma = 2$ and $V_0 = 0.04$. For simplicity, the driving noises of the benchmarked asset prices are chosen to be independent from each other and from that of the volatility process.

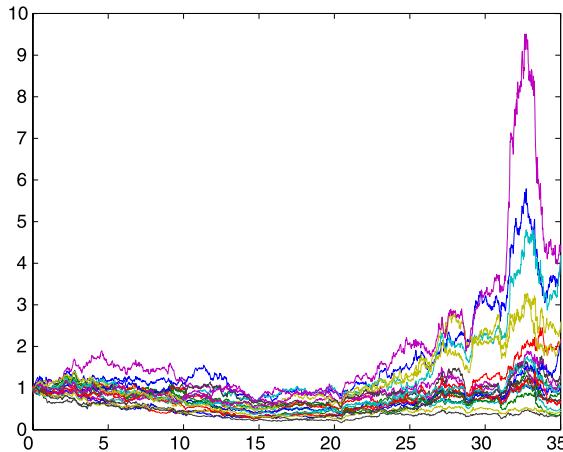


Fig. 3.4.5. Primary security accounts under the ARCH diffusion model

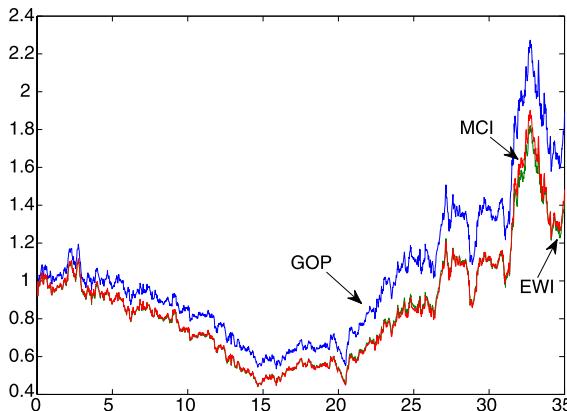


Fig. 3.4.6. GOP, MCI and EWI under the ARCH diffusion model

In Fig. 3.4.5 we show the first twenty of the resulting risky primary security accounts. Fig. 3.4.6 displays the corresponding GOP, EWI and MCI based on 49 risky primary security accounts. Also here the EWI appears to be a good proxy of the GOP. Again in this case the MCI comes reasonably close to the GOP.

In all the above examples we note that the primary security accounts have realistic common fluctuations. These model the general market risk and are generated by the GOP. The benchmarked risky primary security accounts

model the specific market risk of the respective risky asset. The EWI is the simplest diversified portfolio and approximates well in all our examples the GOP. If we would double the number of risky primary security accounts in the above simulations, then the EWI would become an even better proxy of the GOP. The MCI is still a good proxy of the GOP. However, its fractions become in some cases quite large, which can work against the diversification effect.

The above simulation experiments and more extensive simulation studies in [Platen & Rendek \(2009c\)](#) illustrate under different models the power and robustness of the Diversification Theorem, which identifies diversified portfolios as proxies for the GOP without any particular modeling assumptions on the market dynamics. The diversification phenomenon, known since ancient times, turns out to be very robust. One simply has to make sure that only small fractions are invested in each of the primary security accounts to exploit the diversification effect. Only the minor regularity property [\(3.4.12\)](#) is required to ensure that specific market risks diversify properly when forming a diversified portfolio to approximate the GOP.

3.5 Real World Pricing Under Some Models

To illustrate how pricing works under the benchmark approach, this and the following section consider two examples where real world pricing for common payoffs is applied along the lines of results in [Hulley, Miller & Platen \(2005\)](#) and [Platen & Heath \(2006\)](#). The resulting explicit formulas are of importance for variance reduction techniques, see Chap. 16. Another aim is to illustrate how real world pricing retrieves classical risk neutral prices for certain instruments, and so is consistent with the classical literature. Of course, one could apply the standard risk neutral theory directly to obtain the pricing formulas under the MM, but this would defeat our purpose of illustrating real world pricing under the benchmark approach.

Specifying a Continuous GOP

Let us interpret the GOP as a well-diversified global portfolio that is expressed in units of, say, US dollars, in a market with $d \in \mathcal{N}$ risky primary security accounts. One may think of a diversified equity portfolio or market index, the MSCI world index for example. The aggregating of all the jumps in the underlying primary security accounts is assumed to produce only noise which is approximately continuous. In other words, we would expect the jumps to be invisible to an observer of the GOP, say, in US dollar denomination. According to the SDE [\(3.2.10\)](#), the only way to eliminate jumps from the GOP dynamics is by setting the market prices of event risk equal to zero. This is a key assumption that has been made in [Merton \(1976\)](#) for the *Merton model* (MM), see [\(1.7.36\)](#). Of course, jumps of the GOP with high intensity are typically

small and are asymptotically modeled by Wiener processes. Henceforth, the following simplifying assumption will be used in this and the following section, which means for both models we consider.

Assumption 3.5.1 *The market prices of event risks are zero, that is*

$$\theta_t^k = 0, \quad (3.5.1)$$

for each $k \in \{m+1, \dots, d\}$ and all $t \geq 0$.

There is technically no problem to extend the following pricing of derivatives to the case of nonzero market prices of event risk. The exposition becomes only more complex. Substitution of (3.5.1) into (3.2.10) produces the following SDE for the resulting continuous GOP

$$dS_t^{\delta_*} = S_t^{\delta_*} \left(r_t dt + \sum_{k=1}^m \theta_t^k (\theta_t^k dt + dW_t^k) \right), \quad (3.5.2)$$

in US dollar denomination for all $t \geq 0$, with

$$S_0^{\delta_*} = 1. \quad (3.5.3)$$

As before, W^1, \dots, W^m denote independent standard Wiener processes. The explicit solution to (3.5.2) is given by

$$S_t^{\delta_*} = \exp \left\{ \int_0^t \left(r_s + \frac{1}{2} \sum_{k=1}^m (\theta_s^k)^2 \right) ds + \sum_{k=1}^m \int_0^t \theta_s^k dW_s^k \right\}, \quad (3.5.4)$$

for all $t \in [0, T]$.

Benchmarked Primary Security Accounts

The SDE for the j th benchmarked primary security account is derived from (3.4.6) by setting $\pi_{\delta,t}^i = 1$ for $i = j$ and $\pi_{\delta,t}^i = 0$ otherwise, yielding

$$d\hat{S}_t^j = -\hat{S}_{t-}^j \sum_{k=1}^d \sigma_t^{j,k} dW_t^k, \quad (3.5.5)$$

for all $j \in \{0, 1, \dots, d\}$ and $t \geq 0$, with $\hat{S}_0^j = S_0^j$. Here we have set $\sigma_t^{j,k} = \sigma_{(d)}^{j,k}(t)$, see (3.4.4), for all $j \in \{0, 1, \dots, d\}$, $k \in \{1, 2, \dots, d\}$ and $t \geq 0$. As before, we choose W^{m+1}, \dots, W^d to be compensated, normalized jump martingales with corresponding intensity processes h^1, \dots, h^{d-m} , respectively. From (3.5.5) we obtain, via the Itô formula for the j th benchmarked primary security account, the explicit expression

$$\begin{aligned}\hat{S}_t^j &= S_0^j \exp \left\{ -\frac{1}{2} \int_0^t \sum_{k=1}^m (\sigma_s^{j,k})^2 ds - \sum_{k=1}^m \int_0^t \sigma_s^{j,k} dW_s^k \right\} \\ &\quad \times \exp \left\{ \int_0^t \sum_{k=m+1}^d \sigma_s^{j,k} \sqrt{h_s^{k-m}} ds \right\} \prod_{k=m+1}^d \prod_{l=1}^{p_t^{k-m}} \left(1 - \frac{\sigma_{\tau_l^{k-m}}^{j,k}}{\sqrt{h_{\tau_l^{k-m}}^{k-m}}} \right) \end{aligned} \quad (3.5.6)$$

for each $j \in \{0, 1, \dots, d\}$ and all $t \geq 0$. Here $(\tau_l^k)_{l \in \mathcal{N}}$ denotes the sequence of jump times of the k th counting process p^k for events of k th type, $k \in \{m+1, \dots, d\}$.

Under the benchmark approach the benchmarked primary security accounts \hat{S}_t^j , $j \in \{0, 1, \dots, d\}$, are the pivotal objects of study. The savings account S_t^0 together with the benchmarked primary security accounts are sufficient to specify the entire investment universe. More precisely, the ratio $S_t^{\delta_*} = \frac{S_t^0}{S_t^0}$, for all $t \geq 0$, see (3.3.1), derives the GOP in terms of the savings account and the benchmarked savings account. The product $S_t^j = \hat{S}_t^j S_t^{\delta_*}$ expresses each primary security account in terms of the corresponding benchmarked primary security account and the GOP for each $j \in \{1, \dots, d\}$ and all $t \geq 0$.

Define the processes $|\sigma^j| = \{|\sigma_t^j|, t \geq 0\}$ for $j \in \{0, 1, \dots, d\}$, by setting

$$|\sigma_t^j| = \sqrt{\sum_{k=1}^m (\sigma_t^{j,k})^2}. \quad (3.5.7)$$

We also introduce the *aggregate continuous noise processes* $\hat{W}^j = \{\hat{W}_t^j, t \in [0, \infty)\}$ for $j \in \{0, 1, \dots, d\}$, defined by

$$\hat{W}_t^j = \sum_{k=1}^m \int_0^t \frac{\sigma_s^{j,k}}{|\sigma_s^j|} dW_s^k. \quad (3.5.8)$$

By Lévy's Theorem for the characterization of the Wiener process, see Theorem 1.3.3, it follows that \hat{W}^j is a Wiener process for each $j \in \{0, 1, \dots, d\}$. Note that the Wiener processes $\hat{W}^0, \hat{W}^1, \dots, \hat{W}^d$ can be correlated. Furthermore, we enforce Assumption 3.1.2, such that the generalized volatility matrix $\mathbf{b}_t = [b_t^{j,k}]_{j,k=1}^d$ becomes for all $t \geq 0$ invertible. Recall by (3.4.4) that

$$b_t^{j,k} = \theta_t^k - \sigma_t^{j,k} \quad (3.5.9)$$

for $k \in \{1, 2, \dots, m\}$ and by (3.5.1) and (3.4.4) that

$$b_t^{j,k} = -\sigma_t^{j,k} \quad (3.5.10)$$

for $k \in \{m+1, \dots, d\}$, $j \in \{1, 2, \dots, d\}$ and $t \geq 0$.

For simplicity, we assume in this and the next section that the parameters governing the jump behavior are constant. Thus, the counting processes p^k

are, in fact, time homogenous Poisson processes with constant intensities, where

$$h_t^k = h^k > 0 \quad (3.5.11)$$

for each $k \in \{1, 2, \dots, d-m\}$ and all $t \geq 0$. Also, the jump ratios $\sigma_t^{j,k}$ for the benchmarked primary security accounts are assumed to be constant, that is

$$\sigma_t^{j,k} = \sigma^{j,k} \leq \sqrt{h^{k-m}} \quad (3.5.12)$$

for all $j \in \{0, 1, \dots, d\}$, $k \in \{m+1, \dots, d\}$ and $t \geq 0$. Note that Assumption 3.5.1 on zero market prices of event risk ensures that (3.5.11) does not violate Assumption 3.2.1. Furthermore, Assumption 3.5.1 and relation (3.4.4) ensure that (3.5.12) satisfies Assumption 3.1.1.

Using (3.5.7)–(3.5.12), we can rewrite the benchmarked j th primary security account in (3.5.6) as the product

$$\hat{S}_t^j = \hat{S}_t^{j,c} S_t^{j,d} \quad (3.5.13)$$

with its continuous part

$$\hat{S}_t^{j,c} = S_0^j \exp \left\{ -\frac{1}{2} \int_0^t |\sigma_s^j|^2 ds - \int_0^t |\sigma_s^j| d\hat{W}_s^j \right\} \quad (3.5.14)$$

and its compensated jump part

$$S_t^{j,d} = \exp \left\{ t \sum_{k=m+1}^d \sigma^{j,k} \sqrt{h^{k-m}} \right\} \prod_{k=m+1}^d \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}} \right)^{p_t^{k-m}} \quad (3.5.15)$$

for each $j \in \{0, 1, \dots, d\}$ and all $t \geq 0$. The specific model for the benchmarked primary security accounts, which we now discuss, differs from the one we will discuss in the next section in terms of how the continuous processes (3.5.14) are modeled. The jump processes (3.5.15) will be, for simplicity, chosen to be the same in both cases.

The Merton Model

The *Merton model* (MM) has emerged as the standard market model when including event risk in asset dynamics. We describe here a modification of the jump diffusion model introduced in Merton (1976), see (1.7.36). Each benchmarked primary security account can be expressed as the product of a driftless geometric Brownian motion and an independent jump martingale. Therefore, it is itself a martingale. The MM arises if one assumes that all parameter processes, that is, the short rate, the volatilities and the jump intensities, are constant. Therefore, in addition to (3.5.11) and (3.5.12) we set $r_t = r$ and $\sigma_t^{j,k} = \sigma^{j,k}$ for each $j \in \{0, 1, \dots, d\}$, $k \in \{1, 2, \dots, m\}$ and $t \geq 0$. In this case (3.5.14) can be written as

$$\hat{S}_t^{j,c} = S_0^j \exp \left\{ -\frac{t}{2} |\sigma^j|^2 - |\sigma^j| \hat{W}_t^j \right\} \quad (3.5.16)$$

for each $j \in \{0, 1, \dots, d\}$ and all $t \geq 0$. In this special case, the benchmarked primary security accounts are martingales since they are the products of driftless geometric Brownian motions and compensated Poisson processes. The model is similar to the one introduced in [Samuelson \(1965b\)](#), which eventually became the Black-Scholes model with jumps by [Merton \(1976\)](#). The MM is sometimes also called the Merton jump diffusion model.

By Assumption 3.5.1 and relations (3.5.13)–(3.5.15), the benchmarked savings account \hat{S}^0 exhibits no jumps. Furthermore, \hat{S}^0 is a driftless geometric Brownian motion and, thus, a continuous martingale. Consequently, with this specification of the market, the Radon-Nikodym derivative process $\Lambda_\theta(t)$ in (3.3.6) is an $(\underline{\mathcal{A}}, P)$ -martingale. Therefore, the standard risk neutral pricing approach could be used for derivative pricing under the MM with the risk neutral pricing formula (3.3.7). While not advocating the MM as an accurate description of observed market behavior, its familiarity makes it useful for illustrating real world pricing under the benchmark approach instead of risk neutral pricing, which in this case delivers the same results. Furthermore, a great advantage of this model is that one obtains for many derivatives explicit formulas, which is extremely useful in practice.

Zero Coupon Bonds

We first consider a standard default-free zero coupon bond, paying one unit of the domestic currency at its maturity date $T \in [0, \infty)$. According to the real world pricing formula (3.3.5), the value of the zero coupon bond at time t is given by the expression

$$P(t, T) = S_t^{\delta_*} E \left(\frac{1}{S_T^{\delta_*}} \middle| \mathcal{A}_t \right) = \frac{1}{\hat{S}_t^0} E \left(\exp \left\{ - \int_t^T r_s ds \right\} \hat{S}_T^0 \middle| \mathcal{A}_t \right) \quad (3.5.17)$$

for all $t \in [0, T]$. Since \hat{S}^0 is an $(\underline{\mathcal{A}}, P)$ -martingale and r_s is constant we obtain

$$P(t, T) = \exp \{-r(T-t)\} \frac{1}{\hat{S}_t^0} E \left(\hat{S}_T^0 \middle| \mathcal{A}_t \right) = \exp \{-r(T-t)\} \quad (3.5.18)$$

for all $t \in [0, T]$. In other words, we obtain the usual bond pricing formula determined by the deterministic savings account. This is fully in line with the results that one obtains under risk neutral pricing, see [Harrison & Kreps \(1979\)](#) and Sect. 2.7.

Forward Contracts

For the moment let us fix $j \in \{0, 1, \dots, d\}$, $T \in [0, \infty)$ and $t \in [0, T]$. Consider now a *forward contract* with the delivery of one unit of the j th primary security

account at the maturity date T , which is written at time $t \in [0, T]$. The value of the forward contract at the time t when it is written is defined to be zero. According to the real world pricing formula (3.3.5) the *forward price* $F^j(t, T)$ at time $t \in [0, T]$ for this contract is then determined by the relation

$$S_t^{\delta_*} E \left(\frac{F^j(t, T) - S_T^j}{S_T^{\delta_*}} \mid \mathcal{A}_t \right) = 0. \quad (3.5.19)$$

By (3.5.17), solving this equation yields the forward price

$$F^j(t, T) = \frac{S_t^{\delta_*} E \left(\hat{S}_T^j \mid \mathcal{A}_t \right)}{S_t^{\delta_*} E \left(\frac{1}{S_T^{\delta_*}} \mid \mathcal{A}_t \right)} = \frac{S_t^j}{P(t, T) \hat{S}_t^j} E \left(\hat{S}_T^j \mid \mathcal{A}_t \right) \quad (3.5.20)$$

if $S_t^j > 0$ for given $t \in [0, T]$. In the case when $S_t^j = 0$ we have also $F^j(t, T) = 0$.

In the MM case, with reference to (3.5.16), the same argument, which established that the benchmarked savings account is a continuous martingale, also applies to the driftless geometric Brownian motion $\hat{S}^{j,c}$, while the compensated Poisson process $\hat{S}^{j,d}$ is a jump martingale. Consequently, \hat{S}^j is the product of independent martingales, and hence itself an (\underline{A}, P) -martingale. Together with (3.5.18) this enables us to write the forward price (3.5.20) as

$$F^j(t, T) = S_t^j \exp\{r(T - t)\} \quad (3.5.21)$$

for all $t \in [0, T]$. Thus, in the MM case we recover the standard risk neutral formula for the forward price, see for instance [Musiela & Rutkowski \(2005\)](#).

Asset-or-Nothing Binaries

Binary options can be regarded as basic building blocks for complex derivatives. This has been exploited in the valuation of exotic options, when a complex payoff can be decomposed into a series of binaries, see [Ingersoll \(2000\)](#) and [Buchen \(2004\)](#). Let us fix $j \in \{0, 1, \dots, d\}$ and consider a derivative contract on the j th primary security account with maturity T and strike $K \in \mathbb{R}^+$. We fix $k \in \{m + 1, \dots, d\}$ and assume that $\sigma^{j,k} \neq 0$ and $\sigma^{j,l} = 0$, for each $l \in \{m + 1, \dots, d\}$ with $l \neq k$. In other words, we assume that the j th primary security account responds only to the $(k - m)$ th jump process. This does not affect the generality of our calculations below, but it does result in analytic expressions.

The derivative contract under consideration is an *asset-or-nothing binary* on the j th primary security account. At its maturity T it pays its holder one unit of the j th primary security account if this is greater than the strike K , and nothing otherwise. According to the real world pricing formula (3.3.5), its value is given by

$$\begin{aligned}
A^{j,k}(t, T, K) &= S_t^{\delta_*} E \left(\mathbf{1}_{\{S_T^j \geq K\}} \frac{S_T^j}{S_T^{\delta_*}} \middle| \mathcal{A}_t \right) \\
&= \frac{S_t^j}{\hat{S}_t^j} E \left(\mathbf{1}_{\{\hat{S}_T^j \geq K(S_T^0)^{-1}\hat{S}_T^0\}} \hat{S}_T^j \middle| \mathcal{A}_t \right) \\
&= \frac{S_t^j}{\hat{S}_t^{j,c}} E \left(\mathbf{1}_{\{\hat{S}_T^{j,c} \geq g(p_T^{k-m} - p_t^{k-m})\hat{S}_T^0\}} \right. \\
&\quad \times \exp \left\{ \sigma^{j,k} \sqrt{h^{k-m}} (T-t) \right\} \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}} \right)^{p_T^{k-m} - p_t^{k-m}} \hat{S}_T^{j,c} \Big| \mathcal{A}_t \Big) \\
&= \sum_{n=0}^{\infty} \exp \left\{ -h^{k-m}(T-t) \right\} \frac{(h^k(T-t))^n}{n!} \exp \left\{ \sigma^{j,k} \sqrt{h^{k-m}} (T-t) \right\} \\
&\quad \times \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}} \right)^n \frac{S_t^j}{\hat{S}_t^{j,c}} E \left(\mathbf{1}_{\{\hat{S}_T^{j,c} \geq g(n)\hat{S}_T^0\}} \hat{S}_T^{j,c} \middle| \mathcal{A}_t \right) \quad (3.5.22)
\end{aligned}$$

for all $t \in [0, T]$, where

$$g(n) = \frac{K}{S_t^0 S_t^{j,d}} \exp \left\{ - \left(r + \sigma^{j,k} \sqrt{h^{k-m}} \right) (T-t) \right\} \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}} \right)^{-n} \quad (3.5.23)$$

for all $n \in \mathcal{N}$, see Hulley et al. (2005).

Equation (3.5.22) yields the following explicit formula:

$$\begin{aligned}
A^{j,k}(t, T, K) &= \sum_{n=0}^{\infty} \exp \left\{ -\sqrt{h^{k-m}} (T-t) \right\} \frac{(\sqrt{h^{k-m}}(T-t))^n}{n!} \\
&\quad \times \exp \left\{ \sigma^{j,k} \sqrt{h^{k-m}} (T-t) \right\} \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}} \right)^n S_t^j N(d_1(n)) \quad (3.5.24)
\end{aligned}$$

for all $t \in [0, T]$, where

$$d_1(n) = \frac{\ln \left(\frac{S_t^j}{K} \right) + \left(r + \sigma^{j,k} \sqrt{h^{k-m}} + n \frac{\ln \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}} \right)}{T-t} + \frac{1}{2} (\hat{\sigma}^{0,j})^2 \right) (T-t)}{\hat{\sigma}^{0,j} \sqrt{T-t}} \quad (3.5.25)$$

for each $n \in \mathcal{N}$. Here $N(\cdot)$ is the Gaussian distribution function. In (3.5.25) we employ the following notation

$$\hat{\sigma}^{i,j} = \sqrt{|\sigma^i|^2 - 2 \varrho^{i,j} |\sigma^i| |\sigma^j| + |\sigma^j|^2} \quad (3.5.26)$$

for $i, j \in \{0, 1, \dots, d\}$, where $\varrho^{i,j}$ is the correlation between the Wiener processes \hat{W}^i and \hat{W}^j .

Bond-or-Nothing Binaries

Now let us price a *bond-or-nothing binary*, which pays the strike $K \in \Re^+$ at maturity T , in the case that the j th primary security account at time T is not less than K , where $j \in \{0, 1, \dots, d\}$ is still fixed. As before, let us assume that the j th primary security account only responds to the k th jump martingale W^k , where $k \in \{m+1, \dots, d\}$ is fixed.

At its maturity the bond-or-nothing binary under consideration pays its holder the strike K if the value of the j th primary security account is in excess of this amount and nothing otherwise. The real world pricing formula (3.3.5) then yields

$$\begin{aligned}
B^{j,k}(t, T, K) &= S_t^{\delta_*} E \left(\mathbf{1}_{\{S_T^j \geq K\}} \frac{K}{S_T^{\delta_*}} \middle| \mathcal{A}_t \right) \\
&= K P(t, T) - K S_t^{\delta_*} E \left(\mathbf{1}_{\{S_T^j < K\}} \frac{1}{S_T^{\delta_*}} \middle| \mathcal{A}_t \right) \\
&= K P(t, T) - K \frac{S_t^0}{\hat{S}_t^0} E \left(\mathbf{1}_{\{\hat{S}_T^0 > K^{-1} S_T^0 \hat{S}_T^j\}} \frac{\hat{S}_T^0}{S_T^0} \middle| \mathcal{A}_t \right) \\
&= K P(t, T) - K \exp\{-r(T-t)\} \frac{1}{\hat{S}_t^0} E \left(\mathbf{1}_{\{\hat{S}_T^0 > g(p_T^{k-m} - p_t^{k-m})^{-1} \hat{S}_T^{j,c}\}} \hat{S}_T^0 \middle| \mathcal{A}_t \right) \\
&= K P(t, T) - K \exp\{-r(T-t)\} \sum_{n=0}^{\infty} \exp\{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!} \\
&\quad \times \frac{1}{\hat{S}_t^0} E \left(\mathbf{1}_{\{\hat{S}_T^0 > g(n)^{-1} \hat{S}_T^{j,c}\}} \hat{S}_T^0 \middle| \mathcal{A}_t \right) \tag{3.5.27}
\end{aligned}$$

for all $t \in [0, T]$, where $g(n)$ is given by (3.5.23), for each $n \in \mathcal{N}$.

Relation (3.5.27) provides the following explicit formula:

$$\begin{aligned}
B^{j,k}(t, T, K) &= K \exp\{-r(T-t)\} \\
&\quad \times \left(1 - \sum_{n=0}^{\infty} \exp\{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!} N(-d_2(n)) \right) \\
&= \sum_{n=0}^{\infty} \exp\{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!} K \exp\{-r(T-t)\} N(d_2(n))
\end{aligned} \tag{3.5.28}$$

for all $t \in [0, T]$, where

$$\begin{aligned}
d_2(n) &= \frac{\ln\left(\frac{S_t^j}{K}\right) + \left(r + \sigma^{j,k} \sqrt{h^{k-m}} + n \frac{\ln(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}})}{T-t} - \frac{1}{2} (\hat{\sigma}^{0,j})^2\right) (T-t)}{\hat{\sigma}^{0,j} \sqrt{T-t}} \\
&= d_1(n) - \hat{\sigma}^{0,j} \sqrt{T-t}
\end{aligned} \tag{3.5.29}$$

for each $n \in \mathcal{N}$, see [Hulley et al. \(2005\)](#). Again $\hat{\sigma}^{0,j}$ is given by [\(3.5.26\)](#).

European Call Options

We fix again $j \in \{0, 1, \dots, d\}$ and consider a European call option with maturity T and strike $K \in \mathbb{R}^+$ on the j th primary security account. As before, we make the simplifying assumption that the j th primary security account is only sensitive to the $(k-m)$ th jump process, for some fixed $k \in \{m+1, \dots, d\}$. According to the real world pricing formula [\(3.3.5\)](#) the European call option price is given by

$$\begin{aligned}
c_{T,K}^{j,k}(t) &= S_t^{\delta_*} E\left(\frac{(S_T^j - K)^+}{S_T^{\delta_*}} \mid \mathcal{A}_t\right) = S_t^{\delta_*} E\left(\mathbf{1}_{\{S_T^j \geq K\}} \frac{S_T^j - K}{S_T^{\delta_*}} \mid \mathcal{A}_t\right) \\
&= A^{j,k}(t, T, K) - B^{j,k}(t, T, K)
\end{aligned} \tag{3.5.30}$$

for all $t \in [0, T]$. Combining [\(3.5.24\)](#) and [\(3.5.28\)](#) gives

$$\begin{aligned}
c_{T,K}^{j,k}(t) &= \sum_{n=0}^{\infty} \exp\{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!} \left(\exp\left\{\sigma^{j,k} \sqrt{h^{k-m}} (T-t)\right\} \right. \\
&\quad \times \left. \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}}\right)^n S_t^j N(d_1(n)) - K \exp\{-r(T-t)\} N(d_2(n)) \right)
\end{aligned} \tag{3.5.31}$$

for all $t \in [0, T]$, where $d_1(n)$ and $d_2(n)$ are given by [\(3.5.25\)](#) and [\(3.5.29\)](#), respectively, for each $n \in \mathcal{N}$.

The explicit formula [\(3.5.31\)](#) corresponds to the original pricing formula for a European call on a stock whose price follows a jump diffusion, as given in [Merton \(1976\)](#). The only difference is that there the jump ratios were taken to be independent log-normally distributed, while in our case they are constant. Furthermore, this formula can be used to price an option to exchange the j th primary security account for the i th primary security account. In that case, the option pricing formula obtained instead of [\(3.5.31\)](#) is a generalization of that given in [Margrabe \(1978\)](#).

Defaultable Zero Coupon Bonds

We have incorporated default risk in our market model. This allows us to study the pricing of credit derivatives. Here we consider the canonical example of

such a contract, which is a defaultable zero coupon bond with maturity T . To keep the analysis simple, we fix $k \in \{m+1, \dots, d\}$ and assume that the bond under consideration defaults at the first jump time τ_1^{k-m} of p^{k-m} , provided that this time is not greater than T . In other words, default occurs if and only if $\tau_1^{k-m} \leq T$, in which case τ_1^{k-m} is the default time. As a further simplification, we assume zero recovery upon default. According to the real world pricing formula (3.3.5), the price of this instrument is given by

$$\begin{aligned}\tilde{P}^{k-m}(t, T) &= S_t^{\delta_*} E \left(\frac{\mathbf{1}_{\{\tau_1^{k-m} > T\}}}{S_T^{\delta_*}} \mid \mathcal{A}_t \right) = S_t^{\delta_*} E \left(\frac{1}{S_T^{\delta_*}} \mid \mathcal{A}_t \right) E \left(\mathbf{1}_{\{\tau_1^{k-m} > T\}} \mid \mathcal{A}_t \right) \\ &= P(t, T) P(p_T^{k-m} = 0 \mid \mathcal{A}_t)\end{aligned}\quad (3.5.32)$$

for all $t \in [0, T]$. Note that the second equality above follows from the independence of the GOP and the underlying Poisson process, see (3.5.2).

Equation (3.5.32) shows that the price of the defaultable bond can be expressed as the product of the price of the corresponding default-free bond and the conditional probability of survival. In our setup the latter may be further evaluated as

$$\begin{aligned}P(p_T^{k-m} = 0 \mid \mathcal{A}_t) &= E \left(\mathbf{1}_{\{p_t^{k-m} = 0\}} \mathbf{1}_{\{p_T^{k-m} - p_t^{k-m} = 0\}} \mid \mathcal{A}_t \right) \\ &= \mathbf{1}_{\{p_t^{k-m} = 0\}} P \left(p_T^{k-m} - p_t^{k-m} = 0 \mid \mathcal{A}_t \right) \\ &= \mathbf{1}_{\{p_t^{k-m} = 0\}} E \left(\exp \left\{ - \int_t^T h_s^{k-m} ds \right\} \mid \mathcal{A}_t \right)\end{aligned}\quad (3.5.33)$$

for all $t \in [0, T]$. One has to combine (3.5.32) and (3.5.33) with (3.5.18) to obtain an explicit pricing formula for the defaultable bond. Note that the expression obtained by combining (3.5.32) and (3.5.33) is similar to the familiar risk neutral pricing formula for the price of a defaultable zero coupon bond in a simple reduced form model for credit risk, see Schönbucher (2003). In (3.5.32) and (3.5.33), however, only the real world probability measure is playing a role. The crucial advantage of the benchmark approach in such a situation is that one avoids the undesirable challenge of distinguishing between real world default probabilities, as determined by historical data and economic reasoning, and putative risk neutral default probabilities, as calibrated by observed credit spreads and suggested by credit rating agencies. If in reality the existence of an equivalent risk neutral probability measure cannot be reasonably confirmed, then formally computed risk neutral prices can be substantially larger than real world prices, see (3.3.9), which will be demonstrated in the next section. Risk neutral pricing is a form of relative pricing and ignores all trends. An entire class of derivatives can be massively overpriced using classical risk neutral pricing, as it indeed happened in 2007/2008 with subprime credit default instruments. Real world pricing represents a form of absolute

pricing. The following section will apply a realistic and still parsimonious model under the benchmark approach to the same derivatives as discussed above, where explicit formulas will emerge.

3.6 Real World Pricing Under the MMM

A Minimal Market Model with Jumps

The stylized *minimal market model* (MMM), mentioned in Sect. 2.6, will be generalized in this section to include jumps. For simplicity, we suppose the parameters associated with the jump parts to be constant, as was the case for the MM in the previous section. The continuous parts of the dynamics are modeled via inverted time transformed squared Bessel processes of dimension four. Consequently, each benchmarked primary security account is the product of an inverted, time transformed squared Bessel process of dimension four and an independent jump martingale. Since inverted squared Bessel processes of dimension four are strict local martingales, see Revuz & Yor (1999), the benchmarked savings account is not a martingale under the MMM. However, the Radon-Nikodym derivative of the putative risk neutral measure is the benchmarked savings account, normalized to one at its initial time, see Sect. 2.6. Hence, a viable equivalent risk neutral probability measure does not exist under this model, and we will use real world pricing for evaluating derivatives.

A version of the MMM for the continuous part of the benchmarked primary security accounts is obtained by modeling the time transformations as exponential functions. Let us now recall the notation and outline this model as it will be applied in this section. For further details we refer to Hulley et al. (2005) or Platen & Heath (2006).

In the multi-dimensional stylized MMM with jumps, for each $j \in \{0, 1, \dots, d\}$ we define the scaling function $\alpha^j : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ by setting

$$\alpha_t^j = \alpha_0^j \exp\{\eta^j t\} \quad (3.6.1)$$

for all $t \geq 0$ with $\alpha_0^j > 0$. We refer to η^j as the net growth rate of the j th primary security account, for $j \in \{0, 1, \dots, d\}$. The j th primary security account can be interpreted as savings account of the j th currency. Next, we define the j th square root process $Y^j = \{Y_t^j, t \geq 0\}$ for $j \in \{0, 1, \dots, d\}$ through the system of SDEs

$$dY_t^j = \left(1 - \eta^j Y_t^j\right) dt + \sqrt{Y_t^j} dW_t^j \quad (3.6.2)$$

for each $j \in \{0, 1, \dots, d\}$ and all $t \geq 0$, with $Y_0^j = \frac{1}{\alpha_0^j S_0^j}$. Here W^0, W^1, \dots, W^d are, for simplicity, independent Wiener processes. The continuous parts $\hat{S}_t^{j,c}$ of the benchmarked primary security accounts (3.5.14) are modeled in terms of these square root processes by setting

$$\hat{S}_t^{j,c} = \frac{1}{\alpha_t^j Y_t^j} \quad (3.6.3)$$

for each $j \in \{0, 1, \dots, d\}$ and all $t \geq 0$. Since (3.6.3) combined with (3.5.13) and (3.5.14) represents a version of the multi-dimensional stylized MMM for benchmarked primary security accounts, we shall henceforth refer to it as MMM in this section.

As previously mentioned, between jumps the benchmarked primary security accounts are inverted time transformed squared Bessel processes of dimension four. More precisely, define the continuous strictly increasing functions $\varphi_j: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ for $j \in \{0, 1, \dots, d\}$ by setting

$$\varphi_j(t) = \varphi_j(0) + \frac{1}{4} \int_0^t \alpha_s^j ds \quad (3.6.4)$$

for each $j \in \{0, 1, \dots, d\}$ and all $t \geq 0$ with $\varphi_j(0) \in \mathbb{R}^+$. Continuity and monotonicity imply that φ_j possesses an inverse $(\varphi_j)^{-1}: [\varphi_j(0), \infty) \rightarrow \mathbb{R}^+$ for each $j \in \{0, 1, \dots, d\}$. Now define the processes $X^j = \{X_\varphi^j, \varphi \in [\varphi_j(0), \infty)\}$ for each $j \in \{0, 1, \dots, d\}$ by setting

$$X_{\varphi_j(t)}^j = \alpha_t^j Y_t^j = \frac{1}{\hat{S}_t^{j,c}} \quad (3.6.5)$$

for each $j \in \{0, 1, \dots, d\}$ and all $t \geq 0$. It then follows, see Sect. 2.6, that X^j is a squared Bessel process of dimension four in φ -time, so that $\frac{1}{\hat{S}_t^{j,c}}$ is a time transformed squared Bessel process for each $j \in \{0, 1, \dots, d\}$.

Zero Coupon Bond and Forward Price

Under the MMM the benchmarked savings account is driftless and nonnegative and, thus, a local martingale. Hence by Lemma 1.3.2 (i) it is a supermartingale. It is, when normalized, also the candidate for the Radon-Nikodym derivative process for the putative risk neutral measure, see Sect. 3.3. However, according to Revuz & Yor (1999) the candidate Radon-Nikodym derivative is under the MMM a strict supermartingale and not an (\underline{A}, P) -martingale. Consequently, risk neutral derivative pricing is impossible under the MMM, and we shall resort to the more general real world pricing. Platen & Heath (2006) showed that the MMM is attractive for a number of reasons. Its modest number of parameters and explicitly known transition density makes it a practical tool. It conveniently illustrates some differences between the benchmark approach and the classical risk neutral approach.

To simplify the notation let us set

$$\lambda_t^j = \frac{1}{\hat{S}_t^j (\varphi_j(t) - \varphi_j(T))} \quad (3.6.6)$$

for $t \in [0, T]$ and $j \in \{0, 1, \dots, d\}$, where $\lambda_T^j = \infty$. It is argued in Miller & Platen (2005) that the interest rate process and the discounted GOP can be assumed to be independent. If we accept this, and apply it to (3.5.17), while remembering that $\hat{S}_T^0 = \hat{S}_T^{0,c}$, we obtain for the fair zero coupon bond the formula

$$\begin{aligned} P(t, T) &= E \left(\exp \left\{ - \int_t^T r_s ds \right\} \middle| \mathcal{A}_t \right) \frac{1}{\hat{S}_t^0} E \left(\hat{S}_T^0 \middle| \mathcal{A}_t \right) \\ &= E \left(\exp \left\{ - \int_t^T r_s ds \right\} \middle| \mathcal{A}_t \right) \left(1 - \exp \left\{ - \frac{1}{2} \lambda_t^0 \right\} \right) \end{aligned} \quad (3.6.7)$$

for all $t \in [0, T]$, which uses the real world pricing formula (3.3.5) and a moment property of squared Bessel processes, see Platen (2002) and (3.6.6).

According to (3.6.5), $\hat{S}^{j,c}$ is an inverted time transformed squared Bessel process of dimension four, while $S^{j,d}$ is an independent jump martingale, as before in the case of the MM. Thus, we obtain, similar to the previous section, for the j th primary security account

$$\frac{1}{\hat{S}_t^j} E \left(\hat{S}_T^j \middle| \mathcal{A}_t \right) = \frac{1}{\hat{S}_t^{j,c}} E \left(\hat{S}_T^{j,c} \middle| \mathcal{A}_t \right) \frac{1}{S_t^{j,d}} E \left(S_T^{j,d} \middle| \mathcal{A}_t \right) = 1 - \exp \left\{ - \frac{1}{2} \lambda_t^j \right\} \quad (3.6.8)$$

for all $t \in [0, T]$, by using (3.6.6). Putting (3.5.20) together with (3.6.7) and (3.6.8) gives for the forward price the formula

$$F^j(t, T) = S_t^j \frac{1 - \exp \left\{ - \frac{1}{2} \lambda_t^j \right\}}{1 - \exp \left\{ - \frac{1}{2} \lambda_t^0 \right\}} E \left(\exp \left\{ \int_t^T r_s ds \right\} \middle| \mathcal{A}_t \right) \quad (3.6.9)$$

for all $t \in [0, T]$. This demonstrates that the forward price of a primary security account is a tractable quantity under the MMM.

Asset-or-Nothing Binaries

As we have just seen, calculating the price of a payoff written on a primary security account requires the evaluation of a double integral involving the transition density of a two-dimensional process. This is a consequence of choosing the GOP as numéraire. Closed form derivative pricing formulas can be obtained in many cases for the MM, but in the case of the MMM this is more difficult, because the joint transition probability densities of two squared Bessel processes are, in general, difficult to describe, see Bru (1991). A natural response to this is to obtain the derivative price numerically by Monte Carlo simulation, as will be described in Chap. 11. However, to give the reader a feeling for the types of formulas that emerge from applying real world pricing under the MMM, we have assumed, for simplicity, that the processes \hat{S}^0 and $\hat{S}^{j,c}$ are independent, which is also a reasonable assumption in many practical

situations. Combining (3.5.22) and (3.5.23), and remembering that $\hat{S}^0 = \hat{S}^{0,c}$, results in the formula

$$\begin{aligned} A^{j,k}(t, T, K) &= S_t^{\delta_*} E \left(\mathbf{1}_{\{S_T^j \geq K\}} \frac{S_T^j}{S_T^{\delta_*}} \middle| \mathcal{A}_t \right) \\ &= \sum_{n=0}^{\infty} \exp \{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!} \\ &\quad \times \exp \left\{ \sigma^{j,k} \sqrt{h^{k-m}(T-t)} \right\} \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}} \right)^n \\ &\quad \times S_t^j \left(G''_{0,4} \left(\frac{\varphi_0(T) - \varphi_0(t)}{g(n)}; \lambda_t^j, \lambda_t^0 \right) - \exp \left\{ -\frac{1}{2} \lambda_t^j \right\} \right) \quad (3.6.10) \end{aligned}$$

for all $t \in [0, T]$, $k \in \{m+1, \dots, d\}$. Here $G''_{0,4}(x; \lambda, \lambda')$ equals the probability $P(\frac{Z}{Z'} \leq x)$ for the ratio $\frac{Z}{Z'}$ of a non-central chi-square distributed random variable $Z \sim \chi^2(0, \lambda)$ with degrees of freedom zero and non-centrality parameter $\lambda > 0$, and a non-central chi-square distributed random variable $Z' \sim \chi^2(4, \lambda')$ with four degrees of freedom and non-centrality parameter λ' , see Hulle et al. (2005). By implementing this special function one obtains the pricing formula given in (3.6.10), see Johnson, Kotz & Balakrishnan (1995).

Bond-or-Nothing Binaries

Subject to the assumption that \hat{S}_T^0 and $\hat{S}_T^{j,c}$ are independent, we can combine (3.5.27), (3.5.23) and (3.6.7), to obtain the bond-or-nothing binary

$$\begin{aligned} B^{j,k}(t, T, K) &= S_t^{\delta_*} E \left(\mathbf{1}_{\{S_T^j \geq K\}} \frac{K}{S_T^{\delta_*}} \middle| \mathcal{A}_t \right) \\ &= K \exp \{-r(T-t)\} \left(1 - \exp \left\{ -\frac{1}{2} \lambda_t^0 \right\} \right) \\ &\quad - \sum_{n=0}^{\infty} \exp \{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!} K \exp \{-r(T-t)\} \\ &\quad \times \left(G''_{0,4} \left(\frac{\varphi_j(T) - \varphi_j(t)}{g(n)}; \lambda_t^0, \lambda_t^j \right) - \exp \left\{ -\frac{1}{2} \lambda_t^0 \right\} \right) \\ &= \sum_{n=0}^{\infty} \exp \{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!} \\ &\quad \times K \exp \{-r(T-t)\} \left(1 - G''_{0,4} \left((\varphi_j(T) - \varphi_j(t))g(n); \lambda_t^0, \lambda_t^j \right) \right) \quad (3.6.11) \end{aligned}$$

for all $t \in [0, T]$, see Hulley et al. (2005). For the second equality in (3.6.11), we have again used the fact that

$$\sum_{n=0}^{\infty} \exp\{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!}$$

is the total probability of a Poisson random variable with parameter $h^{k-m}(T-t)$.

European Call Option

The European call option pricing formula is obtained by subtracting (3.6.11) from (3.6.10), according to (3.5.30), yielding

$$\begin{aligned} c_{T,K}^{j,k}(t) &= S_t^{\delta_*} E \left(\frac{(S_T^j - K)^+}{S_T^{\delta_*}} \mid \mathcal{A}_t \right) \\ &= \sum_{n=0}^{\infty} \exp\{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!} \left[\exp\left\{ \sigma^{j,k} \sqrt{h^{k-m}} (T-t) \right\} \right. \\ &\quad \times \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}} \right)^n S_t^j \left(G''_{0,4} \left((\varphi_0(T) - \varphi_0(t)) g(n); \lambda_t^j, \lambda_t^0 \right) - \exp\left\{ -\frac{1}{2} \lambda_t^j \right\} \right) \\ &\quad \left. - K \exp\{-r(T-t)\} \left(1 - G''_{0,4} \left(\frac{\varphi_j(T) - \varphi_j(t)}{g(n)}; \lambda_t^0, \lambda_t^j \right) \right) \right] \quad (3.6.12) \end{aligned}$$

for all $t \in [0, T]$, where $g(n)$ is given by (3.5.23), for each $n \in \mathcal{N}$ and λ_t^j in (3.6.6).

Savings Bond and Fair Zero Coupon Bond

Let us now compare the MM and the MMM when pricing an extremely long dated zero coupon bond using historical data to calibrate the models. When we set for a defaultable bond the recovery rate to one and the default intensity to zero we obtain a fair zero coupon bond $P(t, T)$ as priced in (3.5.17) and (3.6.7) for the respective model.

We interpret now the US dollar savings account discounted world equity index, displayed in Fig. 3.6.1, as GOP. Its logarithm is shown in Fig. 3.6.2 with a trendline. Under both models, the MM and the MMM, we assumed deterministic interest rates. Therefore, the historical US three months treasury bill rate is here interpreted as a deterministic interest rate. The resulting benchmarked savings bond

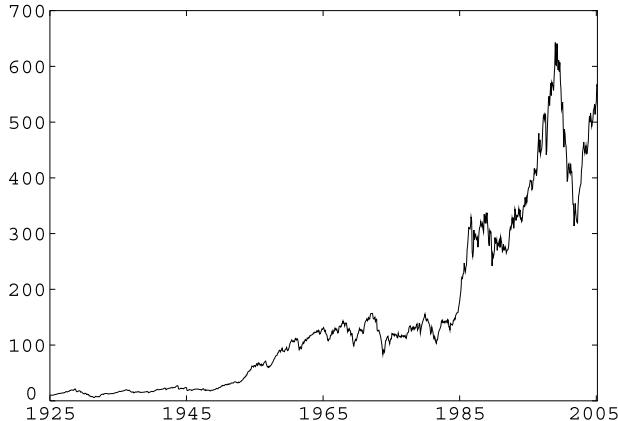


Fig. 3.6.1. Discounted Equity Index

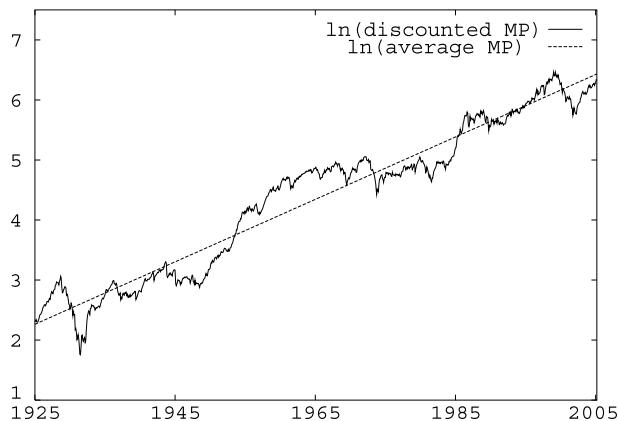


Fig. 3.6.2. Logarithm of discounted equity index

$$\hat{P}^*(t, T) = \frac{1}{S_t^{\delta_*}} \exp \left\{ - \int_t^T r_s ds \right\}$$

we display in Fig. 3.6.3 as the upper graph. One notes its systematic downward trend. This trend is a consequence of the equity premium, which in the long run gives the world equity index a larger expected growth than the savings account. The savings bond invests in the deterministic savings account and delivers \$1 at maturity, as shown in the upper graph of Fig. 3.6.4. This is the bond price that would result under the MM.

We now use the MMM to model the discounted GOP as a time transformed squared Bessel process of dimension four. Here the net growth rate can be estimated from the slope of the trendline in Fig. 3.6.2 with $\eta^0 = 0.052$ and the scaling parameter emerges as $\alpha_0^0 = 0.1828$. Then the resulting fair zero coupon bond $P(t, T)$, see (3.6.7), is shown in Fig. 3.6.4 in the lower graph. Its price is for the long term to maturity significantly cheaper than that of the savings

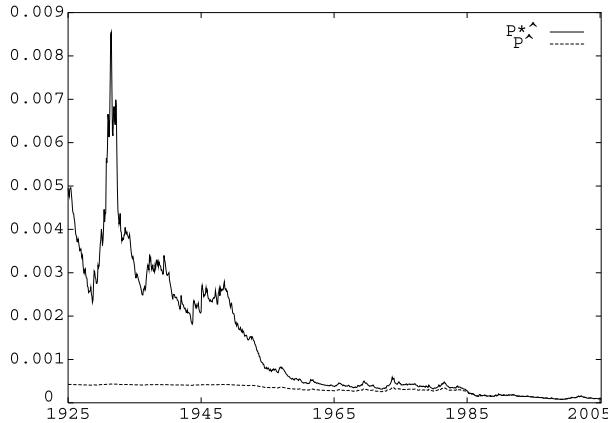


Fig. 3.6.3. Benchmarked savings bond and zero coupon bond

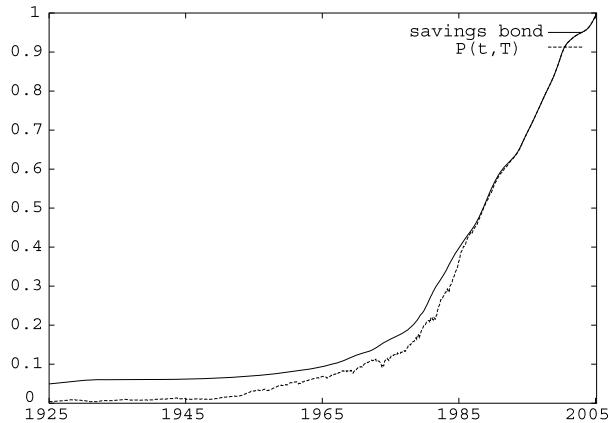


Fig. 3.6.4. Savings bond and fair zero coupon bond

bond. Both their benchmarked values are shown in Fig. 3.6.3. One notes in this figure that the trajectory of the benchmarked fair zero coupon is lower and reflects more realistically that of a martingale, whereas the trajectory of the benchmarked savings bond can be better interpreted as that of a strict supermartingale, see Platen (2009).

Under the MMM the fair zero coupon bond describes the minimal price process that can replicate the payoff of \$1 at maturity. Also, the savings bond is replicating the payoff of \$1 at maturity in a self-financing way. However, this price process is not fair and, therefore, not the minimal price process that can achieve this replication. To illustrate the economic reasoning why the fair zero coupon bond can be cheaper than the savings bond, we display in Fig. 3.6.5 the logarithms of both bonds. One notices that the fair zero coupon bond simply exploits for a long time the superior long term growth of the equity index and only close to maturity it slides finally over into the savings bond. Reasonable financial planning has always been suggesting such kind of behavior in life cy-

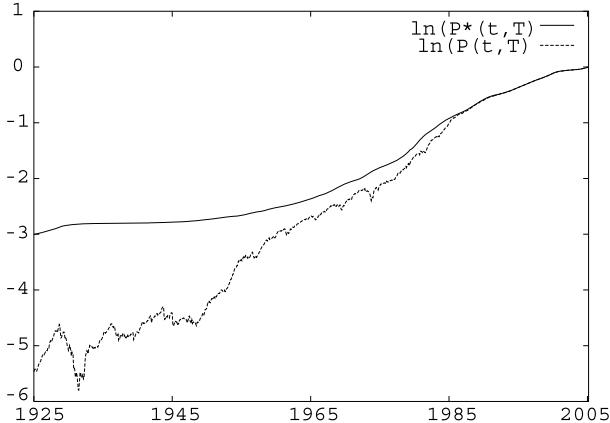


Fig. 3.6.5. Logarithm of savings bond and fair bond

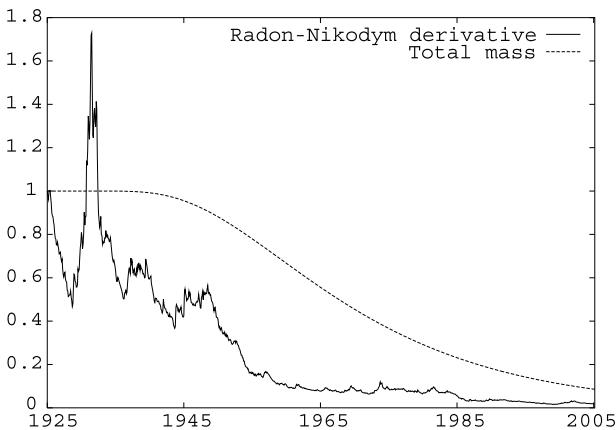


Fig. 3.6.6. Radon-Nikodym derivative and total neutral measure

cle investing. A detailed quantification of such strategy is possible here because the MMM under real world pricing takes real world trends fully into account. Risk neutral pricing ignores these trends and is satisfied by the fact that the savings bond replicates the desired payoff. We emphasize that the cheapest replication of \$1 is achieved by the replicating fair zero coupon bond portfolio.

For illustration, in Fig. 3.6.6 we display the candidate for the Radon-Nikodym derivative of the putative risk neutral measure under the MMM. This figure shows also the total mass of the putative risk neutral measure under the MMM calibrated to historical data. One notes that the risk neutral measure is not a probability since it does not add up to one after some time, see Sect. 2.7.

If one considers a defaultable bond with long term to maturity and some given default intensity and given recovery rate, then the same effect as described above for the zero coupon bond will emerge. Using the classical risk

neutral pricing paradigm under the MM provides a substantially higher price than can be identified under the MMM via real world pricing. We emphasize again, real world pricing is a form of absolute pricing that takes real world trends fully into account. Classical risk neutral pricing, however, is a form of relative pricing which can lead to significantly overpriced derivatives.

The two financial market models considered in this and the previous section highlight some aspects of the benchmark approach relevant to derivative pricing. This methodology can be applied generally to much more general market models.

3.7 Binomial Option Pricing

To prepare more numerically motivated questions, this section studies binomial option pricing in a situation where one wants to obtain numerically real world prices of derivatives for a given payoff. For the underlying security its benchmarked dynamics, which is always that of a supermartingale, is modeled by a binomial tree, and we follow in our exposition [Platen & Heath \(2006\)](#). Binomial trees will be also considered in Sect. 17.1.

Single-Period Binomial Model

In the following, we use a discrete-time version of the benchmark approach, which has been formulated with a view towards insurance in Bühlmann & Platen (2003). Let $S^{\delta_*} = \{S_t^{\delta_*}, t \in [0, T]\}$ denote the growth optimal portfolio (GOP), which shall be interpreted as a broadly diversified index, for example, a world stock index. Furthermore, we introduce a primary security account process $S = \{S_t, t \in [0, T]\}$, where S_t denotes the cum-dividend price of a stock. This means, the dividends are all reinvested. The benchmarked primary security account price \hat{S}_t at time t is defined as

$$\hat{S}_t = \frac{S_t}{S_t^{\delta_*}} \quad (3.7.1)$$

for $t \in [0, T]$. Obviously, the benchmarked GOP equals the constant $\hat{S}_t^{\delta_*} = 1$ for all $t \in [0, T]$. The benchmarked asset price process $\hat{S} = \{\hat{S}_t, t \in [0, T]\}$ and the GOP are modeled on a filtered probability space $(\Omega, \mathcal{A}_T, \mathcal{A}, P)$, which satisfies the usual conditions. Between discretization times we hold benchmarked price processes constant.

A benchmarked portfolio $\hat{S}^\delta = \{\hat{S}_t^\delta, t \in [0, T]\}$ with strategy $\boldsymbol{\delta} = \{\boldsymbol{\delta}_t = (\delta_t^0, \delta_t^1)^\top, t \in [0, T]\}$, which at time t holds δ_t^0 units of the GOP and δ_t^1 units of the stock, has at time t the benchmarked value

$$\hat{S}_t^\delta = \delta_t^0 + \delta_t^1 \hat{S}_t \quad (3.7.2)$$

for $t \in [0, T]$.

In the following we consider the simple *binomial model* to generate randomness. This is probably one of the simplest ways to model randomness evolving over time. We model the uncertain value of the benchmarked security \hat{S}_Δ at time $t = \Delta > 0$ by only allowing the two possible benchmarked values $(1 + u)\hat{S}_0$ and $(1 + d)\hat{S}_0$, such that

$$P(\hat{S}_\Delta = (1 + u)\hat{S}_0) = p \quad (3.7.3)$$

and

$$P(\hat{S}_\Delta = (1 + d)\hat{S}_0) = 1 - p,$$

respectively, for $p \in (0, 1)$ and $-d, u \in (0, \infty)$. In the case of an upward move the value $u = \frac{\hat{S}_\Delta - \hat{S}_0}{\hat{S}_0}$ can be interpreted as positive return and for a downward move $d = \frac{\hat{S}_\Delta - \hat{S}_0}{\hat{S}_0}$ as negative return. We will see that u, d and p have to satisfy some relationship to make \hat{S} a supermartingale.

Real World Pricing

Consider a European call option on the benchmarked stock with benchmarked payoff

$$\hat{H} = (\hat{S}_\Delta - \hat{K})^+, \quad (3.7.4)$$

where, for convenience of the following exposition, the deterministic benchmarked strike \hat{K} lies in the interval $((1 + d)\hat{S}_0, (1 + u)\hat{S}_0)$ and the maturity is set to $T = \Delta$. The option yields at maturity Δ a benchmarked payoff of $((1 + u)\hat{S}_0 - \hat{K})^+$ in the case of an upward move, that is for $\hat{S}_\Delta = (1 + u)\hat{S}_0$, and is zero otherwise. The probability for the upward move of \hat{S}_Δ , and thus some positive payment, equals p . Assuming that the GOP has initial value $S_0^{\delta_*} = 1$ the real world pricing formula (3.3.5) yields the fair call option price

$$\begin{aligned} S_0^{\delta_H} &= S_0^{\delta_*} E(\hat{H} | \mathcal{A}_0) \\ &= ((1 + u)\hat{S}_0 - \hat{K}) p. \end{aligned} \quad (3.7.5)$$

We additionally assume that the cum-dividend stock price process S is a fair price process. Then it must hold that

$$\begin{aligned} \hat{S}_0 &= E(\hat{S}_\Delta | \mathcal{A}_0) \\ &= p(1 + u)\hat{S}_0 + (1 - p)(1 + d)\hat{S}_0 \\ &= (p(1 + u) + (1 - p)(1 + d))\hat{S}_0. \end{aligned} \quad (3.7.6)$$

Consequently, we obtain the condition

$$1 = p(1 + u) + (1 - p)(1 + d)$$

and, thus, the probability

$$p = \frac{-d}{u-d}. \quad (3.7.7)$$

Therefore, we obtain from (3.7.5) and (3.7.7) at time $t = 0$ the *European call option price*

$$S_0^{\delta_H} = \hat{S}_0^{\delta_H} = \left((1+u) \hat{S}_0 - \hat{K} \right) \frac{-d}{u-d}, \quad (3.7.8)$$

where $S_0^{\delta_*} = 1$. Note that the option price increases with increasing u and decreases with increasing d .

Hedging

Let us determine the hedge ratio for the above derivative. Assuming a self-financing fair portfolio, we need by (3.7.2) to hold at time $t = 0$ in benchmarked terms the value

$$\hat{S}_0^{\delta_H} = \delta_0^0 + \delta_0^1 \hat{S}_0 \quad (3.7.9)$$

and at time $t = \Delta$ the value

$$\hat{S}_\Delta^{\delta_H} = \delta_0^0 + \delta_0^1 \hat{S}_\Delta. \quad (3.7.10)$$

This provides a system of equations, where $\hat{S}_0^{\delta_H}$ is given by (3.7.8) and

$$\hat{S}_\Delta^{\delta_H} = \hat{H}_\Delta = \left(\hat{S}_\Delta - \hat{K} \right)^+. \quad (3.7.11)$$

It then follows from (3.7.9)–(3.7.11) the hedge ratio

$$\delta_0^1 = \frac{\hat{S}_\Delta^{\delta_H} - \hat{S}_0^{\delta_H}}{\hat{S}_\Delta - \hat{S}_0},$$

which is satisfied for the case of an upward move when $\hat{S}_\Delta = (1+u)\hat{S}_0$ with

$$\delta_0^1 = \frac{(1+u)\hat{S}_0 - \hat{K} - \left((1+u)\hat{S}_0 - \hat{K} \right) \frac{-d}{u-d}}{u\hat{S}_0} = \frac{(1+u)\hat{S}_0 - \hat{K}}{(u-d)\hat{S}_0} \quad (3.7.12)$$

and for a downward move, that is for $\hat{S}_\Delta = (1+d)\hat{S}_0$, with

$$\delta_0^1 = \frac{-\left((1+u)\hat{S}_0 - \hat{K} \right) \frac{-d}{u-d}}{d\hat{S}_0} = \frac{(1+u)\hat{S}_0 - \hat{K}}{(u-d)\hat{S}_0}. \quad (3.7.13)$$

Important to note is that both values for the hedge ratio in (3.7.12) and (3.7.13) are the same. This shows that the real world price (3.7.5) provides a perfect hedge for both scenarios. Fair pricing is performed in a way that the perfect replication of contingent claims in this market is secured. The

simple market that we consider is therefore *complete*. Each contingent claim can be replicated by investing in the stock and the GOP. Furthermore, it can be shown that the market does not permit strong arbitrage in the sense of Sect. 3.3. This means, one is unable by using a nonnegative portfolio to generate from zero initial capital strictly positive wealth. This is a consequence of the supermartingale property of benchmarked nonnegative portfolios and automatically given through the existence of the GOP. It is, therefore, not appropriate to say that a price is here obtained by excluding strong arbitrage, see Definition 3.3.2. We have simply identified the minimal price that provides a perfect replication of the payoff.

Note that the probability p is a real world probability which measures the likelihood of an upward move of the benchmarked security. It is interesting to see in this simple example how an option can be hedged by using the index and the stock. This is different to the standard approach where one uses the domestic savings account and the underlying security to form a hedge portfolio.

Binomial Volatility

The variability of the benchmarked stock price \hat{S}_Δ , and thus its movements against the GOP or index, can be measured by the variance of the ratio

$$\frac{\hat{S}_\Delta}{\hat{S}_0} = 1 + \eta, \quad (3.7.14)$$

which involves the two-point distributed random variable $\eta \in \{d, u\}$, where

$$P(\eta = u) = p \quad (3.7.15)$$

and

$$P(\eta = d) = 1 - p. \quad (3.7.16)$$

It then follows that the variance of η and thus $\frac{\hat{S}_\Delta}{\hat{S}_0}$ equals

$$E \left(\left(\frac{\hat{S}_\Delta}{\hat{S}_0} - 1 \right)^2 \mid \mathcal{A}_0 \right) = (u - d)^2 p (1 - p).$$

Therefore, we obtain by (3.7.7) the formula

$$E \left(\left(\frac{\hat{S}_\Delta}{\hat{S}_0} - 1 \right)^2 \mid \mathcal{A}_0 \right) = -u d. \quad (3.7.17)$$

Obviously, the variance increases when u or $-d$ increase. We call the value

$$\sigma_\Delta = \sqrt{\frac{1}{\Delta} E \left(\left(\frac{\hat{S}_\Delta}{\hat{S}_0} - 1 \right)^2 \mid \mathcal{A}_0 \right)} = \sqrt{\frac{-u d}{\Delta}} \quad (3.7.18)$$

the *binomial volatility* of the benchmarked security under the given binomial model. Under the BS model we know from the Black-Scholes formula, see Sect. 2.7, that the option price increases for increasing volatility, which appears to be appropriate from a risk management point of view. Due to the formula

$$S_0^{\delta_H} = \left((1+u) \hat{S}_0 - \hat{K} \right) \frac{(\sigma_\Delta)^2 \Delta}{u(u-d)}$$

there is still substantial freedom to choose for increasing binomial volatility σ_Δ the returns u and d such that the binomial option price does not increase. This possibility is somehow not satisfactory and counter intuitive. It results from the simplicity of the binomial model. Thus, one has to be careful and should avoid to read too much into the binomial model.

Multi-Period Binomial Model

We consider now a multi-period binomial model. The benchmarked price of the stock at time $t = n\Delta$ satisfies for an upward move the equation

$$\hat{S}_{n\Delta} = \hat{S}_{(n-1)\Delta} (1+u) \quad (3.7.19)$$

with probability

$$p = \frac{-d}{u-d} \quad (3.7.20)$$

and for a downward move the relation

$$\hat{S}_{n\Delta} = \hat{S}_{(n-1)\Delta} (1+d) \quad (3.7.21)$$

with probability $1-p$ for $n \in \{1, 2, \dots\}$.

We consider a European call option on the benchmarked stock price $\hat{S}_{2\Delta}$ at maturity $T = 2\Delta$ with benchmarked strike \hat{K} . The benchmarked security $\hat{S}_{2\Delta}$ can take at time $T = 2\Delta$ one of the three values $\hat{S}_{uu} = (1+u)^2 \hat{S}_0$, $\hat{S}_{ud} = \hat{S}_{du} = (1+u)(1+d)\hat{S}_0$ and $\hat{S}_{dd} = (1+d)^2 \hat{S}_0$. One can then build for these events a binomial tree as in Fig. 3.7.1. The benchmarked value $\hat{S}_{2\Delta}^\delta$ of a portfolio at time $T = 2\Delta$ is given as

$$\hat{S}_{2\Delta}^\delta = \delta_{2\Delta}^0 + \delta_{2\Delta}^1 \hat{S}_{2\Delta} \quad (3.7.22)$$

with $\hat{S}_{2\Delta}$ taking the above described three possible values. By using our previous result in (3.7.5) we can compute the option price at time $t = \Delta$ for the case when $\hat{S}_\Delta = (1+u) \hat{S}_0$, which follows as

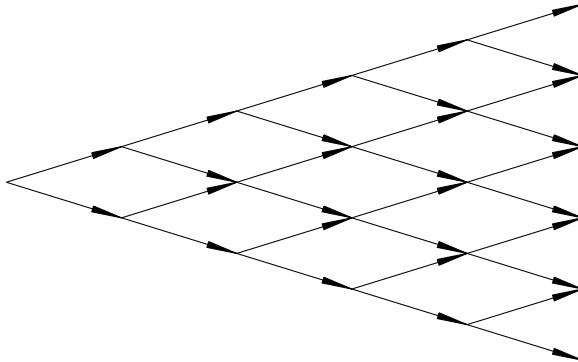


Fig. 3.7.1. Binomial tree

$$\hat{S}_{\Delta}^{\delta_H} = \left((1+u)^2 \hat{S}_0 - \hat{K} \right)^+ \frac{-d}{u-d} + \left((1+u)(1+d) \hat{S}_0 - \hat{K} \right)^+ \frac{u}{u-d}. \quad (3.7.23)$$

Similarly, we obtain for $\hat{S}_{\Delta} = (1+d)\hat{S}_0$ at time $t = \Delta$ the option price

$$\hat{S}_{\Delta}^{\delta_H} = \left((1+u)(1+d) \hat{S}_0 - \hat{K} \right)^+ \frac{-d}{u-d} + \left((1+d)^2 \hat{S}_0 - \hat{K} \right)^+ \frac{u}{u-d}. \quad (3.7.24)$$

One can now interpret the values (3.7.23) and (3.7.24) as those of a benchmarked payoff at time $t = \Delta$. This allows us to obtain, similar as before, the corresponding fair benchmarked price at time $t = 0$ as the expectation

$$\begin{aligned} \hat{S}_0^{\delta_H} &= p \left\{ \left((1+u)^2 \hat{S}_0 - \hat{K} \right)^+ \frac{-d}{u-d} + \left((1+u)(1+d) \hat{S}_0 - \hat{K} \right)^+ \frac{u}{u-d} \right\} \\ &\quad + (1-p) \left\{ \left((1+u)(1+d) \hat{S}_0 - \hat{K} \right)^+ \frac{-d}{u-d} + \left((1+d)^2 \hat{S}_0 - \hat{K} \right)^+ \frac{u}{u-d} \right\}. \end{aligned} \quad (3.7.25)$$

Also for this fair price a corresponding perfect hedging strategy can be described.

More generally, by using backward recursion and the well-known binomial probabilities, see Cox, Ross & Rubinstein (1979), one can show that the following Cox-Ross-Rubinstein (CRR) type *binomial option pricing formula* holds for a benchmarked European call payoff with maturity $T = i\Delta$ and benchmarked strike \hat{K} , where

$$\begin{aligned}
\hat{S}_0^{\delta_H} &= E \left(\left(\hat{S}_T - \hat{K} \right)^+ \mid \mathcal{A}_0 \right) \\
&= \sum_{k=0}^i \frac{i!}{k!(i-k)!} p^k (1-p)^{i-k} \left((1+u)^k (1+d)^{i-k} \hat{S}_0 - \hat{K} \right)^+ \\
&= \hat{S}_0 \sum_{k=\tilde{k}}^i \frac{i!}{k!(i-k)!} p^k (1-p)^{i-k} (1+u)^k (1+d)^{i-k} \\
&\quad - \hat{K} \sum_{k=\tilde{k}}^i \frac{i!}{k!(i-k)!} p^k (1-p)^{i-k}.
\end{aligned} \tag{3.7.26}$$

Here \tilde{k} denotes the first integer k for which $(1+u)^k (1+d)^{i-k} \hat{S}_0 > \hat{K}$.

Note that this is not exactly the CRR binomial option pricing formula, which was originally derived for discounted securities under the risk neutral probability measure. Here we have obtained an option pricing formula that refers to benchmarked securities under the real world probability measure.

Let us denote by

$$t_N(t_k, i, p) = \sum_{k=\tilde{k}}^i \frac{i!}{k!(i-k)!} p^k (1-p)^{i-k} \tag{3.7.27}$$

the, so called, *complementary binomial distribution*. Then we can express the *binomial option pricing formula* (3.7.26) in the form

$$S_0^{\delta_H} = S_0 t_N(\tilde{k}, i, 1-p) - \hat{K} t_N(\tilde{k}, i, p). \tag{3.7.28}$$

As already mentioned, this formula is similar to the Cox-Ross-Rubinstein binomial option pricing formula. The only difference is that here the strike price \hat{K} describes how many units of the GOP or index are exchanged for one unit of the stock.

One can also deduce the corresponding hedging strategy $\boldsymbol{\delta} = \{\delta_{k\Delta} = (\delta_{k\Delta}^0, \delta_{k\Delta}^1)^\top, k \in \{0, 1, \dots, i-1\}\}$ that replicates the given contingent claim \hat{H} . The corresponding self-financing benchmarked hedge portfolio has before and after time $t = k\Delta$ the form

$$\hat{S}_{k\Delta}^{\delta_H} = \delta_{(k-1)\Delta}^0 + \delta_{(k-1)\Delta}^1 \hat{S}_{k\Delta} = \delta_{k\Delta}^0 + \delta_{k\Delta}^1 \hat{S}_{k\Delta} \tag{3.7.29}$$

for $k \in \{1, 2, \dots, i\}$. Similarly, as in (3.7.13) one obtains the hedge ratio

$$\delta_{k\Delta}^1 = \sum_{\ell=\tilde{k}_k}^{i-k} \frac{(i-k)!}{\ell!(i-k-\ell)!} p^{(i-k-\ell)} (1-p)^\ell, \tag{3.7.30}$$

where \tilde{k}_k denotes the smallest integer ℓ for which $(1+u)^\ell (1+d)^{i-k-\ell} \hat{S}_{k\Delta} > \hat{K}$.

From (3.7.29) it follows that

$$\delta_{k\Delta}^0 = \hat{S}_{k\Delta}^{\delta_H} - \delta_{k\Delta}^1 \hat{S}_{k\Delta}. \quad (3.7.31)$$

The binomial model is a very convenient but also a rather simple model. It allows an easy calculation of European and also American option prices. However, due to its simplicity it has severe short comings that we will demonstrate later on. For instance, such trees have numerical stability problems that closely resemble those of explicit weak schemes, as we will discuss in Chap. 17.

Approximating the Black-Scholes Price

In the following, we indicate how the binomial model approximates asymptotically the Black-Scholes model in a weak sense when the time step size Δ converges to zero. In principle, the random walk that is performed on the binomial tree converges in a weak sense to a limiting process as the time step size decreases to zero. This limiting process turns out to be a geometric Brownian motion.

To see the indicated asymptotics of the above binomial model, let us fix u and d by setting

$$\ln(1+u) = \sigma \sqrt{\Delta} \quad (3.7.32)$$

and

$$\ln(1+d) = -\sigma \sqrt{\Delta}, \quad (3.7.33)$$

where the *volatility parameter* $\sigma > 0$ is a fixed value. Thus, the binomial method makes the spatial step size proportional to $\sqrt{\Delta}$ and we obtain the Cox-Ross-Rubinstein return values in the form

$$u = \exp \left\{ \sigma \sqrt{\Delta} \right\} - 1 \approx \sigma \sqrt{\Delta} \quad (3.7.34)$$

and

$$d = \exp \left\{ -\sigma \sqrt{\Delta} \right\} - 1 \approx -\sigma \sqrt{\Delta}. \quad (3.7.35)$$

Note that the binomial volatility σ_Δ was introduced in (3.7.18) and we have for small $\Delta \ll 1$ by (3.7.34)–(3.7.35) approximately

$$\sigma_\Delta = \frac{1}{\sqrt{\Delta}} \sqrt{-u d} \approx \sigma. \quad (3.7.36)$$

Let us now introduce a stochastic process $Y^\Delta = \{Y_t^\Delta, t \in [0, T]\}$ with

$$Y_t^\Delta = \hat{S}_{n\Delta} \quad (3.7.37)$$

for $t \in [n\Delta, (n+1)\Delta], n \in \{0, 1, \dots\}$. Then by (3.7.19)–(3.7.21) it follows that

$$Y_{n\Delta+\Delta}^\Delta = Y_{n\Delta}^\Delta + Y_{n\Delta}^\Delta \eta_n \quad (3.7.38)$$

with some independent random variable η_n , where

$$P(\eta_n = u) = p = \frac{-d}{u - d} \quad (3.7.39)$$

and

$$P(\eta_n = d) = 1 - p = \frac{u}{u - d}, \quad (3.7.40)$$

see (3.7.14)–(3.7.16). By using (3.7.32)–(3.7.33) we realize that one can rewrite (3.7.38) approximately in the form

$$Y_{n\Delta+\Delta}^\Delta \approx Y_{n\Delta}^\Delta + Y_{n\Delta}^\Delta \sigma \sqrt{\Delta} \xi_n, \quad (3.7.41)$$

where we use an independent random variable ξ_n with

$$P(\xi_n = 1) = p = \frac{\exp\{-\sigma \sqrt{\Delta}\} - 1}{\exp\{\sigma \sqrt{\Delta}\} - \exp\{-\sigma \sqrt{\Delta}\}} \quad (3.7.42)$$

and

$$P(\xi_n = -1) = 1 - p.$$

Note that

$$E(Y_{n\Delta+\Delta}^\Delta - Y_{n\Delta}^\Delta | \mathcal{A}_{n\Delta}) = 0 \quad (3.7.43)$$

and

$$E((Y_{n\Delta+\Delta}^\Delta - Y_{n\Delta}^\Delta)^2 | \mathcal{A}_{n\Delta}) = (Y_{n\Delta}^\Delta)^2 \sigma_\Delta^2 \Delta \approx (Y_{n\Delta}^\Delta)^2 \sigma^2 \Delta. \quad (3.7.44)$$

More precisely, it will turn out that the discrete-time approximation Y^Δ converges in a weak sense to a diffusion process $X = \{X_t, t \in [0, T]\}$ as $\Delta \rightarrow 0$, where

$$dX_t = X_t \sigma dW_t \quad (3.7.45)$$

for $t \in [0, T]$ with $X_0 = \hat{S}_0$.

More precisely, the binomial option pricing formula (3.7.28) approximates for $\Delta \rightarrow 0$ the well-known Black-Scholes pricing formula, see Sect. 2.7, where

$$S_0^{\delta_H} = S_0 N(\hat{d}_1) - \hat{K} N(\hat{d}_2) \quad (3.7.46)$$

with

$$\hat{d}_1 = \frac{\ln(\frac{S_0}{\hat{K}}) + \frac{1}{2} \sigma^2 T}{\sigma \sqrt{T}} \quad (3.7.47)$$

and

$$\hat{d}_2 = \hat{d}_1 - \sigma \sqrt{T}. \quad (3.7.48)$$

Here $N(\cdot)$ is the standard Gaussian distribution function. Note that we consider here benchmarked securities, which is different to the standard version of the Black-Scholes formula described in Sect. 2.7.

The important observation here is that the binomial tree method does not involve any simulation. The nodes of the tree reflect all possible paths of the underlying random walk.

Finally, we remark that the pricing of American style options on binomial and other trees can be conveniently performed. At first, one builds the tree in a forward manner as described above. Then one uses a backward algorithm to obtain from the maturity date backward the price of the contingent claim at each of the nodes. In particular, at each node one decides which is the greater, the price that has been obtained under the assumption that one continues to hold the American style derivative to the next period, or the price that results when one exercises it at this node.

3.8 Exercises

3.1. In a continuous financial market derive the form of the growth rate g_t^δ of a strictly positive portfolio S^δ satisfying the SDE

$$dS_t^\delta = S_t^\delta \left(r_t dt + \sum_{k=1}^d \sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} (\theta_t^k dt + dW_t^k) \right).$$

3.2. For a nonnegative portfolio S_t^δ with SDE as in Exercise 3.1 and the growth optimal portfolio $S_t^{\delta*}$ satisfying the SDE

$$dS_t^{\delta*} = S_t^{\delta*} \left(r_t dt + \sum_{k=1}^d \theta_t^k (\theta_t^k dt + dW_t^k) \right),$$

show for the benchmarked portfolio value $\hat{S}_t^\delta = \frac{S_t^\delta}{S_t^{\delta*}}$ its SDE. If \hat{S}^δ is square integrable does this SDE, in general, imply that \hat{S}^δ is a submartingale, martingale or supermartingale?

3.3. Calculate the SDE for the logarithm of the discounted GOP under the MMM.

3.4. Derive the SDE of the square root for the discounted GOP under the MMM.

3.5. Derive the SDE for the normalized GOP $Y_t = \frac{\bar{S}_t^{\delta*}}{\alpha_t^{\delta*}}$ under the MMM if

$$\alpha_t^{\delta*} = \alpha_0 \exp \left\{ \int_0^t \eta_s ds \right\}.$$

Which type of process is this?

3.6. Calculate the SDE for the squared volatility of the discounted GOP under the MMM.

Stochastic Expansions

In this rather demanding chapter we present the *Wagner-Platen expansion* for solutions of SDEs with jumps. This stochastic expansion generalizes the deterministic Taylor formula and the Wagner-Platen expansion for diffusions to the case of SDEs with jumps. It allows expanding the increments of smooth functions of Itô processes in terms of multiple stochastic integrals. As we will see, it is the key tool for the construction of stochastic numerical methods and very convenient for other approximation tasks.

4.1 Introduction to Wagner-Platen Expansions

The deterministic Taylor formula has proven to be an indispensable tool in both theoretical and practical investigations. It allows the approximation of a sufficiently smooth function in a neighborhood of a given expansion point to any desired order of accuracy. We will see that an expansion with similar structure can be obtained for the stochastic case.

Example of a Wagner-Platen Expansion

As in the deterministic case, it is often essential in practical applications to be able to expand the increments of smooth functions of solutions of SDEs. This means, a stochastic expansion with analogous properties to the deterministic Taylor formula can be very useful. Such a stochastic Taylor expansion is the *Wagner-Platen expansion*, which was in its first version derived in [Wagner & Platen \(1978\)](#) and later developed further in [Platen \(1982b\)](#), [Platen & Wagner \(1982\)](#) and [Kloeden & Platen \(1992\)](#). We derive it here in the form of an example for a process $X = \{X_t, t \in [t_0, T]\}$ satisfying the SDE

$$X_t = X_{t_0} + \int_{t_0}^t a(X_s) ds + \int_{t_0}^t b(X_s) dW_s \quad (4.1.1)$$

for $t \in [t_0, T]$. Here W is a standard Wiener process on a filtered probability space $(\Omega, \mathcal{A}, \mathcal{A}, P)$. The coefficients a and b are assumed to be sufficiently smooth, real-valued functions such that a unique strong solution of (4.1.1) exists. For a twice continuously differentiable function $f : \mathbb{R} \rightarrow \mathbb{R}$, the Itô formula provides then the representation

$$\begin{aligned} f(X_t) &= f(X_{t_0}) + \int_{t_0}^t \left(a(X_s) \frac{\partial}{\partial x} f(X_s) + \frac{1}{2} b^2(X_s) \frac{\partial^2}{\partial x^2} f(X_s) \right) ds \\ &\quad + \int_{t_0}^t b(X_s) \frac{\partial}{\partial x} f(X_s) dW_s \\ &= f(X_{t_0}) + \int_{t_0}^t L^0 f(X_s) ds + \int_{t_0}^t L^1 f(X_s) dW_s, \end{aligned} \quad (4.1.2)$$

for $t \in [t_0, T]$. Here we use the operators

$$L^0 = a \frac{\partial}{\partial x} + \frac{1}{2} b^2 \frac{\partial^2}{\partial x^2} \quad (4.1.3)$$

and

$$L^1 = b \frac{\partial}{\partial x}. \quad (4.1.4)$$

Obviously, for the special case $f(x) \equiv x$ we have $L^0 f = a$ and $L^1 f = b$, for which the representation (4.1.2) reduces to (4.1.1). Since the representation (4.1.2) still relies on integrands that are functions of processes we now apply the Itô formula (4.1.2) to the functions $f = a$ and $f = b$ in (4.1.1) to obtain

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^t \left(a(X_{t_0}) + \int_{t_0}^s L^0 a(X_z) dz + \int_{t_0}^s L^1 a(X_z) dW_z \right) ds \\ &\quad + \int_{t_0}^t \left(b(X_{t_0}) + \int_{t_0}^s L^0 b(X_z) dz + \int_{t_0}^s L^1 b(X_z) dW_z \right) dW_s \\ &= X_{t_0} + a(X_{t_0}) \int_{t_0}^t ds + b(X_{t_0}) \int_{t_0}^t dW_s + R_2 \end{aligned} \quad (4.1.5)$$

with remainder term

$$\begin{aligned} R_2 &= \int_{t_0}^t \int_{t_0}^s L^0 a(X_z) dz ds + \int_{t_0}^t \int_{t_0}^s L^1 a(X_z) dW_z ds \\ &\quad + \int_{t_0}^t \int_{t_0}^s L^0 b(X_z) dz dW_s + \int_{t_0}^t \int_{t_0}^s L^1 b(X_z) dW_z dW_s. \end{aligned}$$

This already represents a simple example of a *Wagner-Platen expansion*. We can extend the above expansion, for instance, by applying the Itô formula (4.1.2) to the function $f = L^1 b$ in (4.1.5). In this case we obtain

$$X_t = X_{t_0} + a(X_{t_0}) \int_{t_0}^t ds + b(X_{t_0}) \int_{t_0}^t dW_s + L^1 b(X_{t_0}) \int_{t_0}^t \int_{t_0}^s dW_z dW_s + R_3 \quad (4.1.6)$$

with remainder term

$$\begin{aligned} R_3 = & \int_{t_0}^t \int_{t_0}^s L^0 a(X_z) dz ds + \int_{t_0}^t \int_{t_0}^s L^1 a(X_z) dW_z ds \\ & + \int_{t_0}^t \int_{t_0}^s L^0 b(X_z) dz dW_s + \int_{t_0}^t \int_{t_0}^s \int_{t_0}^z L^0 L^1 b(X_u) du dW_z dW_s \\ & + \int_{t_0}^t \int_{t_0}^s \int_{t_0}^z L^1 L^1 b(X_u) dW_u dW_z dW_s. \end{aligned}$$

We obtained in (4.1.6) an expansion where the leading terms are functions of the value of the solution of the SDE at the expansion point, weighted by corresponding multiple stochastic integrals.

Later we will derive such an expansion for a general multi-factor process X , a general smooth function f and a given expansion level. The main properties of this type of expansion are already apparent in the preceding example. It is of particular importance to note that we obtain expansions which involve multiple Itô integrals of the type

$$\begin{aligned} \int_{t_0}^t ds &= t - t_0, \\ \int_{t_0}^t dW_s &= W_t - W_{t_0}, \\ \int_{t_0}^t \int_{t_0}^s dW_z dW_s &= \frac{1}{2} \left\{ (W_t - W_{t_0})^2 - (t - t_0) \right\}, \end{aligned}$$

see (4.1.6). The remainder term, here R_3 , consists of the next following multiple Itô integrals with nonconstant integrands. A Wagner-Platen expansion can be interpreted as a generalization of both the Itô formula and the classical deterministic Taylor formula. As we have seen, it emerges from an iterated application of the Itô formula.

Generalized Wagner-Platen Expansion

Let us now expand the changes of a function value with respect to the underlying process X itself. From the Itô formula it follows for a twice continuously differentiable function $f : [0, T] \times \mathfrak{R} \rightarrow \mathfrak{R}$ the representation

$$\begin{aligned} f(t, X_t) = & f(t_0, X_{t_0}) + \int_{t_0}^t \frac{\partial}{\partial t} f(s, X_s) ds + \int_{t_0}^t \frac{\partial}{\partial x} f(s, X_s) dX_s \\ & + \frac{1}{2} \int_{t_0}^t \frac{\partial^2}{\partial x^2} f(s, X_s) d[X]_s \end{aligned} \quad (4.1.7)$$

for $t \in [t_0, T]$, where X_t satisfies the SDE (4.1.1). Here the quadratic variation $[X]_t$ has the form

$$[X]_t = [X]_{t_0} + \int_{t_0}^t b^2(X_s) ds \quad (4.1.8)$$

for $t \in [t_0, T]$. Now it is possible to expand further the above representation at the expansion point (t_0, X_{t_0}) by application of the above formula. This leads to the equation

$$\begin{aligned} f(t, X_t) &= f(t_0, X_{t_0}) + \int_{t_0}^t \left\{ \frac{\partial}{\partial t} f(t_0, X_{t_0}) + \int_{t_0}^s \frac{\partial^2}{\partial t^2} f(z, X_z) dz \right. \\ &\quad \left. + \int_{t_0}^s \frac{\partial^2}{\partial x \partial t} f(z, X_z) dX_z + \int_{t_0}^s \frac{1}{2} \frac{\partial^2}{\partial x^2} \frac{\partial}{\partial t} f(z, X_z) d[X]_z \right\} ds \\ &\quad + \int_{t_0}^t \left\{ \frac{\partial}{\partial x} f(t_0, X_{t_0}) + \int_{t_0}^s \frac{\partial^2}{\partial t \partial x} f(z, X_z) dz \right. \\ &\quad \left. + \int_{t_0}^s \frac{\partial^2}{\partial x^2} f(z, X_z) dX_z + \int_{t_0}^s \frac{1}{2} \frac{\partial^3}{\partial x^3} f(z, X_z) d[X]_z \right\} dX_s \\ &\quad + \frac{1}{2} \int_{t_0}^t \left\{ \frac{\partial^2}{\partial x^2} f(t_0, X_{t_0}) + \int_{t_0}^s \frac{\partial^3}{\partial t \partial x^2} f(z, X_z) dz \right. \\ &\quad \left. + \int_{t_0}^s \frac{\partial^3}{\partial x^3} f(z, X_z) dX_z + \int_{t_0}^s \frac{1}{2} \frac{\partial^4}{\partial x^4} f(z, X_z) d[X]_z \right\} d[X]_s \\ &= f(t_0, X_{t_0}) + \frac{\partial}{\partial t} f(t_0, X_{t_0})(t - t_0) + \frac{\partial}{\partial x} f(t_0, X_{t_0})(X_t - X_{t_0}) \\ &\quad + \frac{1}{2} \frac{\partial^2}{\partial x^2} f(t_0, X_{t_0}) ([X]_t - [X]_{t_0}) + R_f(t_0, t), \end{aligned} \quad (4.1.9)$$

where the remainder term $R_f(t_0, t)$ accommodates the terms with nonconstant integrands. The generalized Wagner-Platen expansion is similar to a Taylor expansion but significantly different in that it includes various terms that would not otherwise arise in deterministic calculus. One could continue to expand in (4.1.9) those terms that appear in the remainder term by application of the Itô formula. This would lead, for instance, to an expansion of the form

$$\begin{aligned}
f(t, X_t) &= f(t_0, X_{t_0}) + \frac{\partial}{\partial t} f(t_0, X_{t_0})(t - t_0) + \frac{\partial}{\partial x} f(t_0, X_{t_0})(X_t - X_{t_0}) \\
&\quad + \frac{1}{2} \frac{\partial^2}{\partial x^2} f(t_0, X_{t_0}) ([X]_t - [X]_{t_0}) \\
&\quad + \frac{\partial^2}{\partial x \partial t} f(t_0, X_{t_0}) \int_{t_0}^t \int_{t_0}^s dX_z ds \\
&\quad + \frac{1}{2} \frac{\partial^2}{\partial x^2} \frac{\partial}{\partial t} f(t_0, X_{t_0}) \int_{t_0}^t \int_{t_0}^s d[X]_z ds \\
&\quad + \frac{\partial^2}{\partial t \partial x} f(t_0, X_{t_0}) \int_{t_0}^t \int_{t_0}^s dz dX_s \\
&\quad + \frac{\partial^2}{\partial x^2} f(t_0, X_{t_0}) \int_{t_0}^t \int_{t_0}^s dX_z dX_s \\
&\quad + \frac{1}{2} \frac{\partial^3}{\partial x^3} f(t_0, X_{t_0}) \int_{t_0}^t \int_{t_0}^s d[X]_z dX_s \\
&\quad + \frac{1}{2} \frac{\partial^3}{\partial t \partial x^3} f(t_0, X_{t_0}) \int_{t_0}^t \int_{t_0}^s dz d[X]_s \\
&\quad + \frac{1}{2} \frac{\partial^3}{\partial x^3} f(t_0, X_{t_0}) \int_{t_0}^t \int_{t_0}^s dX_z d[X]_s \\
&\quad + \frac{1}{4} \frac{\partial^4}{\partial x^4} f(t_0, X_{t_0}) \int_{t_0}^t \int_{t_0}^s d[X]_z d[X]_s + \bar{R}_f(t_0, t) \quad (4.1.10)
\end{aligned}$$

for $t \in [t_0, t]$, where $\bar{R}_f(t_0, t)$ is the resulting remainder term. For example, one can expand the increment of an option pricing function by the above Wagner-Platen expansion. This would reveal some sensitivities.

Stratonovich-Taylor Expansion

There are tasks, where it is convenient to rely on the classical rules of calculus. These arise naturally if one uses the symmetric stochastic integral or Stratonovich integral, see (1.4.34) or Kloeden & Platen (1999), instead of the Itô integral. This leads to a so called, Stratonovich stochastic differential equation.

We start from the *Stratonovich equation*

$$X_t = X_{t_0} + \int_{t_0}^t \underline{a}(X_s) ds + \int_{t_0}^t b(X_s) \circ dW_s \quad (4.1.11)$$

for $t \in [t_0, T]$, where the second integral in (4.1.11) is the symmetric stochastic integral, also known as Stratonovich integral, and the coefficients \underline{a} and b

are sufficiently smooth real valued functions, such that a unique solution of (4.1.11) exists. Note that when

$$\underline{a} = a - \frac{1}{2} b b', \quad (4.1.12)$$

the Itô equation (4.1.1) and the Stratonovich equation (4.1.11) are equivalent and have, under appropriate assumptions, the same unique strong solutions. The solution of a Stratonovich SDE transforms formally according to the *classical chain rule*, see Kloeden & Platen (1999). Thus for any continuously differentiable function $f : \mathbb{R} \rightarrow \mathbb{R}$ we have

$$\begin{aligned} f(X_t) &= f(X_{t_0}) + \int_{t_0}^t \underline{a}(X_s) \frac{\partial}{\partial x} f(X_s) ds + \int_{t_0}^t b(X_s) \frac{\partial}{\partial x} f(X_s) \circ dW_s \\ &= f(X_{t_0}) + \int_{t_0}^t \underline{L}^0 f(X_s) ds + \int_{t_0}^t \underline{L}^1 f(X_s) \circ dW_s, \end{aligned} \quad (4.1.13)$$

for $t \in [t_0, T]$, with the operators

$$\underline{L}^0 = \underline{a} \frac{\partial}{\partial x} \quad (4.1.14)$$

and

$$\underline{L}^1 = b \frac{\partial}{\partial x}. \quad (4.1.15)$$

Obviously, for the special case $f(x) \equiv x$ we have $\underline{L}^0 f = \underline{a}$ and $\underline{L}^1 = b$, and (4.1.13) reduces to the original Stratonovich equation (4.1.11).

Analogously, as in the Itô case one can apply (4.1.13) to the functions $f = \underline{a}$ and $f = b$ in (4.1.11). This gives

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^t \left(\underline{a}(X_{t_0}) + \int_{t_0}^s \underline{L}^0 \underline{a}(X_z) dz + \int_{t_0}^s \underline{L}^1 \underline{a}(X_z) \circ dW_z \right) ds \\ &\quad + \int_{t_0}^t \left(b(X_{t_0}) + \int_{t_0}^s \underline{L}^0 b(X_z) dz + \int_{t_0}^s \underline{L}^1 b(X_z) \circ dW_z \right) \circ dW_s \\ &= X_{t_0} + \underline{a}(X_{t_0}) \int_{t_0}^t ds + b(X_{t_0}) \int_{t_0}^t \circ dW_s + R_4, \end{aligned} \quad (4.1.16)$$

with remainder term

$$\begin{aligned} R_4 &= \int_{t_0}^t \int_{t_0}^s \underline{L}^0 \underline{a}(X_z) dz ds + \int_{t_0}^t \int_{t_0}^s \underline{L}^1 \underline{a}(X_z) \circ dW_z ds \\ &\quad + \int_{t_0}^t \int_{t_0}^s \underline{L}^0 b(X_z) dz \circ dW_s + \int_{t_0}^t \int_{t_0}^s \underline{L}^1 b(X_z) \circ dW_z \circ dW_s. \end{aligned}$$

This is an example for a *Stratonovich-Taylor expansion* of $f(X_t)$. We can continue to expand (4.1.16), for instance, by applying (4.1.13) to the integrand $f = \underline{L}^1 b$ to obtain

$$\begin{aligned} X_t &= X_{t_0} + \underline{a}(X_{t_0}) \int_{t_0}^t ds + b(X_{t_0}) \int_{t_0}^t \circ dW_s \\ &\quad + \underline{L}^1 b(X_{t_0}) \int_{t_0}^t \int_{t_0}^s \circ dW_z \circ dW_s + R_5, \end{aligned}$$

with remainder term

$$\begin{aligned} R_5 &= \int_{t_0}^t \int_{t_0}^s \underline{L}^0 \underline{a}(X_z) dz ds + \int_{t_0}^t \int_{t_0}^s \underline{L}^1 \underline{a}(X_z) \circ dW_z ds \\ &\quad + \int_{t_0}^t \int_{t_0}^s \underline{L}^0 b(X_z) dz \circ dW_s + \int_{t_0}^t \int_{t_0}^s \int_{t_0}^z \underline{L}^0 \underline{L}^1 b(X_u) du \circ dW_z \circ dW_s \\ &\quad + \int_{t_0}^t \int_{t_0}^s \int_{t_0}^z \underline{L}^1 \underline{L}^1 b(X_u) \circ dW_u \circ dW_z \circ dW_s. \end{aligned}$$

In Kloeden & Platen (1999) the Stratonovich-Taylor expansion is derived for a general function f and for arbitrary high order. It is similar to the Wagner-Platen expansion, but instead of multiple Itô integrals it involves multiple Stratonovich integrals.

Such expansions can be used in the construction of numerical solutions of SDEs. They may be applied in many ways where the classical calculus provides advantages in handling a problem. Finally, the above expansions can be extended to the case of multi-dimensional processes satisfying stochastic differential equations that are driven by vector Wiener processes.

Expansions for Pure Jump Processes

For certain quantities jumps may occur in many applications, in particular, in finance and insurance. These jumps may be caused by the default of a company, an operational failure, a large claim that hits an insurance company, an information release about extraordinary future earnings of a corporation or other sudden events.

On a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ let $N = \{N_t, t \in [0, T]\}$ denote a *counting process*, which is $\underline{\mathcal{A}}$ -adapted, right continuous, starting at $N_0 = 0$ and nondecreasing with increments of size one at event times. For instance, a Poisson process is such a counting process which counts the arrival of events. For any right-continuous process Z we define the jump ΔZ_t at time t as the difference

$$\Delta Z_t = Z_t - Z_{t-} \tag{4.1.17}$$

for $t \in [0, T]$. Thus we have

$$N_t = \sum_{s \in (0,t]} \Delta N_s \quad (4.1.18)$$

for $t \in [0, T]$. For any measurable function $f : \mathfrak{R} \rightarrow \mathfrak{R}$ we then obtain the equation

$$f(N_t) = f(N_0) + \sum_{s \in (0,t]} \Delta f(N_s) \quad (4.1.19)$$

for $t \in [0, T]$. This is a special case of the Itô formula for semimartingales that exhibit jumps. We can formally write the equation (4.1.19) as the SDE

$$f(N_t) = f(N_0) + \int_{(0,t]} [f(N_{s-} + 1) - f(N_{s-})] dN_s \quad (4.1.20)$$

for $t \in [0, T]$. Obviously, when

$$\tilde{\Delta} f(N) = f(N + 1) - f(N) \quad (4.1.21)$$

is a measurable function in N , then $\tilde{\Delta} f(N_{s-})$ is a measurable function. Therefore, similarly as for the previous case, we can write

$$\begin{aligned} f(N_t) &= f(N_0) + \int_{(0,t]} \tilde{\Delta} f(N_{s-}) dN_s \\ &= f(N_0) + \int_{(0,t]} \tilde{\Delta} f(N_0) dN_s + \int_{(0,t]} \int_{(0,s_2)} \tilde{\Delta} (\tilde{\Delta} f(N_{s_1-})) dN_{s_1} dN_{s_2} \end{aligned} \quad (4.1.22)$$

for $t \in [0, T]$. One notes that multiple stochastic integrals with respect to the counting process N naturally arise. It is straightforward to prove by induction, as in [Engel \(1982\)](#), [Platen \(1984\)](#) and [Studer \(2001\)](#), that

$$\begin{aligned} \int_{(0,t]} dN_s &= N_t \\ \int_{(0,t]} \int_{(0,s_1)} dN_{s_1} dN_{s_2} &= \frac{1}{2!} N_t (N_t - 1), \\ \int_{(0,t]} \int_{(0,s_1)} \int_{(0,s_2)} dN_{s_1} dN_{s_2} dN_{s_3} &= \frac{1}{3!} N_t (N_t - 1) (N_t - 2), \\ \int_{(0,t]} \int_{(0,s_1)} \cdots \int_{(0,s_n)} dN_{s_1} \cdots dN_{s_{n-1}} dN_{s_n} &= \begin{cases} \binom{N_t}{n} & \text{for } N_t \geq n \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (4.1.23)$$

for $t \in [0, T]$, where

$$\binom{i}{n} = \frac{i(i-1)(i-2)\dots(i-n+1)}{1 \cdot 2 \cdot \dots \cdot n} = \frac{i!}{n!(i-n)!} \quad (4.1.24)$$

for $i \geq n$.

This allows us to formulate a stochastic Taylor expansion for a function f of a counting process N . By using the above notation and arguments we obtain by (4.1.22) iteratively

$$f(N_t) = f(N_0) + \int_{(0,t]} \tilde{\Delta} f(N_0) dN_s + \int_{(0,t]} \int_{(0,s_2)} \tilde{\Delta}(\tilde{\Delta} f(N_0)) dN_{s_1} dN_{s_2} + \bar{R}_3(t)$$

with

$$\bar{R}_3(t) = \int_{(0,t]} \int_{(0,s_2)} \int_{(0,s_3)} \tilde{\Delta}(\tilde{\Delta}(\tilde{\Delta} f(N_{s_1-}))) dN_{s_1} dN_{s_2} dN_{s_3}.$$

Thus we have the expansion

$$f(N_t) = f(N_0) + \tilde{\Delta} f(N_0) \binom{N_t}{1} + \tilde{\Delta}(\tilde{\Delta} f(N_0)) \binom{N_t}{2} + \bar{R}_3(t),$$

where

$$\begin{aligned} \tilde{\Delta} f(N_0) &= \tilde{\Delta} f(0) = f(1) - f(0), \\ \tilde{\Delta}(\tilde{\Delta} f(N_0)) &= f(2) - 2f(1) + f(0). \end{aligned}$$

This means, we have for a general function f of a counting process or pure jump process N the expansion

$$f(N_t) = f(0) + (f(1) - f(0)) N_t + (f(2) - 2f(1) + f(0)) \frac{1}{2} N_t (N_t - 1) + \bar{R}_3(t). \quad (4.1.25)$$

The above calculations demonstrate that one can expand elegantly functions of pure jump processes. It is interesting to note that for those $\omega \in \Omega$ with $N_t \leq 3$ the remainder term $\bar{R}_3(t)$ equals zero, and the truncated expansion, when neglecting the remainder term, is then exact. These are important observations that can be very useful in practice.

The above type of expansions can be generalized also for jump diffusions and for functions on processes that involve several counting processes, as we will discuss later.

4.2 Multiple Stochastic Integrals

In this section we introduce a compact notation for multiple stochastic integrals that will appear in corresponding stochastic expansions.

Compensated Poisson Measure

Let us first rewrite the SDE (1.8.2) in a way where the jump part will be expressed as a stochastic integral with respect to the *compensated Poisson measure*

$$\tilde{p}_\varphi(dv, dt) = p_\varphi(dv, dt) - \varphi(dv) dt. \quad (4.2.1)$$

By compensating the Poisson measure in the SDE (1.8.2), we obtain

$$d\mathbf{X}_t = \tilde{a}(t, \mathbf{X}_t)dt + b(t, \mathbf{X}_t)dW_t + \int_{\mathcal{E}} c(t, \mathbf{X}_{t-}, v)\tilde{p}_\varphi(dv, dt), \quad (4.2.2)$$

where the *compensated drift* coefficient is given by

$$\tilde{a}(t, \mathbf{x}) = a(t, \mathbf{x}) + \int_{\mathcal{E}} c(t, \mathbf{x}, v) \varphi(dv), \quad (4.2.3)$$

for $t \in [0, T]$, $T \in (0, \infty)$, with initial value $\mathbf{X}_0 \in \Re^d$ and mark space $\mathcal{E} \subseteq \Re \setminus \{0\}$. Note that by relation (4.2.3), the Cauchy-Schwarz inequality, see (1.2.25), and conditions (1.9.2)–(1.9.3), the compensated drift coefficient \tilde{a} satisfies the Lipschitz condition

$$|\tilde{a}(t, \mathbf{x}) - \tilde{a}(t, \mathbf{y})|^2 \leq K|\mathbf{x} - \mathbf{y}|^2, \quad (4.2.4)$$

for every $t \in [0, T]$ and $\mathbf{x}, \mathbf{y} \in \Re^d$, as well as the linear growth condition

$$|\tilde{a}(t, \mathbf{x})|^2 \leq K(1 + |\mathbf{x}|^2), \quad (4.2.5)$$

for all $t \in [0, T]$ and $\mathbf{x} \in \Re^d$.

To construct discrete-time approximations we will exploit Wagner-Platen expansions of solutions of the SDE (1.8.2). With these formulas we expand smooth functions of solutions of (1.8.2) in terms of multiple stochastic integrals. Note that, in a similar way as in (4.2.2), a multiple stochastic integral involving jumps can be defined by using as integrator either the Poisson measure p_φ or the compensated Poisson measure \tilde{p}_φ . Therefore, we will derive in the case of jumps two different types of stochastic expansions. We will call the former *Wagner-Platen expansions* and the latter *compensated Wagner-Platen expansions*. To express these expansions in a compact form, we introduce below certain notation.

Multi-Indices

We call a row vector $\alpha = (j_1, j_2, \dots, j_l)$, where $j_i \in \{-1, 0, 1, \dots, m\}$ for $i \in \{1, 2, \dots, l\}$, a *multi-index* of length $l = l(\alpha) \in \mathcal{N}$. Here m represents the number of Wiener processes considered in the SDE (1.8.2). Then for $m \in \mathcal{N}$ the set of all multi-indices α is denoted by

$$\mathcal{M}_m = \{(j_1, \dots, j_l) : j_i \in \{-1, 0, 1, 2, \dots, m\}, i \in \{1, 2, \dots, l\} \text{ for } l \in \mathcal{N}\} \cup \{v\},$$

where v is the multi-index of length zero. Later, by a component $j \in \{1, 2, \dots, m\}$ of a multi-index we will denote in a multiple stochastic integral the integration with respect to the j th Wiener process. A component $j = 0$

will denote integration with respect to time. Finally, a component $j = -1$ will denote integration with respect to the compensated Poisson measure \tilde{p}_φ .

We write $n(\alpha)$ for the number of components of a multi-index α that are equal to 0 and $s(\alpha)$ for the number of components of a multi-index α that equal -1 . Moreover, we write $\alpha-$ for the multi-index obtained by deleting the last component of α and $-\alpha$ for the multi-index obtained by deleting the first component of α . For instance, assuming $m = 2$ one obtains

$$\begin{aligned} l((0, -1, 1)) &= 3 & l((0, 1, -1, 0, 2)) &= 5 \\ n((0, -1, 1)) &= 1 & n((0, 1, -1, 0, 2)) &= 2 \\ s((0, -1, 1)) &= 1 & s((0, 1, -1, 0, 2)) &= 1 \\ (0, -1, 1)- &= (0, -1) & (0, 1, -1, 0, 2)- &= (0, 1, -1, 0) \\ -(0, -1, 1) &= (-1, 1) & -(0, 1, -1, 0, 2) &= (1, -1, 0, 2). \end{aligned} \quad (4.2.6)$$

Additionally, given two multi-indices $\alpha_1 = (j_1, \dots, j_k)$ and $\alpha_2 = (i_1, \dots, i_l)$, we introduce the *concatenation operator* \star on \mathcal{M}_m defined by

$$\alpha_1 \star \alpha_2 = (j_1, \dots, j_k, i_1, \dots, i_l). \quad (4.2.7)$$

This operator allows us to combine two multi-indices.

Multiple Integrals

We shall define certain sets of adapted stochastic processes $g = \{g(t), t \in [0, T]\}$ that will be allowed as integrands of multiple stochastic integrals in stochastic expansions. We define

$$\begin{aligned} \mathcal{H}_v &= \left\{ g : \sup_{t \in [0, T]} E(|g(t, \omega)|) < \infty \right\} \\ \mathcal{H}_{(0)} &= \left\{ g : E\left(\int_0^T |g(s, \omega)| ds\right) < \infty \right\} \\ \mathcal{H}_{(-1)} &= \left\{ g : E\left(\int_0^T \int_{\mathcal{E}} |g(s, v, \omega)|^2 \varphi(dv) ds\right) < \infty \right\} \\ \mathcal{H}_{(j)} &= \left\{ g : E\left(\int_0^T |g(s, \omega)|^2 ds\right) < \infty \right\}, \end{aligned} \quad (4.2.8)$$

for $j \in \{1, 2, \dots, m\}$. The corresponding set \mathcal{H}_α for a given multi-index $\alpha \in \mathcal{M}_m$ with $l(\alpha) > 1$ will be defined below.

Let ϱ and τ be two stopping times with $0 \leq \varrho \leq \tau \leq T$ almost surely. For a multi-index $\alpha \in \mathcal{M}_m$ and an adapted process $g(\cdot) \in \mathcal{H}_\alpha$, we define the *multiple stochastic integral* $I_\alpha[g(\cdot)]_{\varrho, \tau}$ recursively as

$$I_\alpha[g(\cdot)]_{\varrho,\tau} = \begin{cases} g(\tau) & \text{when } l=0 \text{ and } \alpha=v \\ \int_\varrho^\tau I_{\alpha-}[g(\cdot)]_{\varrho,z} dz & \text{when } l \geq 1 \text{ and } j_l=0 \\ \int_\varrho^\tau I_{\alpha-}[g(\cdot)]_{\varrho,z} dW_z^{j_l} & \text{when } l \geq 1 \text{ and } j_l \in \{1, \dots, m\} \\ \int_\varrho^\tau \int_{\mathcal{E}} I_{\alpha-}[g(\cdot)]_{\varrho,z} - p_\varphi(dv_{s(\alpha)}, dz) & \text{when } l \geq 1 \text{ and } j_l=-1, \end{cases} \quad (4.2.9)$$

where $g(\cdot) = g(\cdot, v_1, \dots, v_{s(\alpha)})$. As previously, by $z-$ we denote the left-hand limit of z . However, $\alpha-$ for a multi-index α has another meaning as described earlier. For simplicity, when it is not strictly necessary, here and in the sequel we may omit the dependence of the integrand process g on one or more of the variables $v_1, \dots, v_{s(\alpha)}$ that express the dependence on the marks of the Poisson jump measure.

Note that by definition (4.2.9), whenever we have an integration with respect to the Poisson measure p_φ , we take the left-continuous modification of the integrand. Therefore, since we have required that the integrand is adapted, its left-continuous modification will be predictable, see for instance Klebaner (2005). Thus, the corresponding stochastic integral with respect to the Poisson measure p_φ is well defined.

As defined in (4.2.9), in a multi-index α the components that equal 0 refer to an integration with respect to time, the components that equal $j \in \{1, 2, \dots, m\}$ refer to an integration with respect to the j th component of the Wiener process, while the components that equal -1 refer to an integration with respect to the Poisson measure p_φ . For instance, for $m=2$ one has

$$I_{(0,-1,1)}[g(\cdot)]_{\varrho,\tau} = \int_\varrho^\tau \int_\varrho^{z_3} \int_{\mathcal{E}} \int_\varrho^{z_2-} g(z_1, v_1) dz_1 p_\varphi(dv_1, dz_2) dW_{z_3}^1,$$

$$I_{(2,0)}[g(\cdot)]_{\varrho,\tau} = \int_\varrho^\tau \int_\varrho^{z_2} g(z_1) dW_{z_1}^2 dz_2$$

and

$$I_{(-1,-1)}[g(\cdot)]_{\varrho,\tau} = \int_\varrho^\tau \int_{\mathcal{E}} \int_\varrho^{z_2-} \int_{\mathcal{E}} g(z_1-, v_1, v_2) p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2).$$

Now for every multi-index $\alpha = (j_1, \dots, j_l) \in \mathcal{M}_m$ with $l(\alpha) > 1$, we recursively define the sets \mathcal{H}_α as the sets of adapted stochastic processes $g = \{g(t), t \geq 0\}$ such that the integral process $\{I_{\alpha-}[g(\cdot)]_{\varrho,t}, t \in [0, T]\}$ satisfies

$$I_{\alpha-}[g(\cdot)]_{\varrho,\cdot} \in \mathcal{H}_{(j_l)}. \quad (4.2.10)$$

As we shall see later, it is useful to define multiple stochastic integrals where the integrations are defined with respect to the compensated Poisson measure \tilde{p}_φ instead of the Poisson measure p_φ . Therefore, for a multi-index $\alpha \in \mathcal{M}_m$ and an adapted stochastic process $g \in \mathcal{H}_\alpha$, we define the *compensated multiple stochastic integral* $\tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau}$ recursively by

$$\tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau} = \begin{cases} g(\tau) & \text{when } l = 0 \text{ and } \alpha = v \\ \int_\varrho^\tau \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,z} dz & \text{when } l \geq 1 \text{ and } j_l = 0 \\ \int_\varrho^\tau \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,z} dW_z^{j_l} & \text{when } l \geq 1 \text{ and } j_l \in \{1, \dots, m\} \\ \int_\varrho^\tau \int_{\mathcal{E}} \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,z-} \tilde{p}_\varphi(dv_{s(\alpha)}, dz) & \text{when } l \geq 1 \text{ and } j_l = -1, \end{cases} \quad (4.2.11)$$

where $g(\cdot) = g(\cdot, v_1, \dots, v_{s(\alpha)})$. Note that the multiple stochastic integral $I_\alpha[g(\cdot)]_{\varrho,\tau}$, defined previously in (4.2.9), involves integrations with respect to the Poisson jump measure p_φ , while the compensated multiple stochastic integral $\tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau}$ is defined in terms of integrations with respect to the compensated Poisson measure \tilde{p}_φ . For instance, for $m = 2$ one obtains

$$\begin{aligned} \tilde{I}_{(0,-1,1)}[g(\cdot)]_{\varrho,\tau} &= \int_\varrho^\tau \int_\varrho^{z_3} \int_{\mathcal{E}} \int_\varrho^{z_2-} g(z_1, v_1) dz_1 \tilde{p}_\varphi(dv_1, dz_2) dW_{z_3}^1, \\ \tilde{I}_{(2,0)}[g(\cdot)]_{\varrho,\tau} &= I_{(2,0)}[g(\cdot)]_{\varrho,\tau} \end{aligned}$$

and

$$\tilde{I}_{(-1,-1)}[g(\cdot)]_{\varrho,\tau} = \int_\varrho^\tau \int_{\mathcal{E}} \int_\varrho^{z_2-} \int_{\mathcal{E}} g(z_1-, v_1, v_2) \tilde{p}_\varphi(dv_1, dz_1) \tilde{p}_\varphi(dv_2, dz_2).$$

We now illustrate the obvious link between a multiple stochastic integral of the type (4.2.9) with respect to the Poisson jump measure p_φ and a compensated multiple stochastic integral of the type (4.2.11) with respect to the compensated Poisson measure \tilde{p}_φ . This link is governed by the relationship between the Poisson jump measure p_φ and its compensated version \tilde{p}_φ , see (4.2.1).

Remark 4.2.1. Let $\alpha \in \mathcal{M}_m$, and ϱ and τ denote two stopping times with τ being \mathcal{A}_ϱ -measurable and $0 \leq \varrho \leq \tau \leq T$ almost surely. Consider an adapted stochastic process $g \in \mathcal{H}_\alpha$, then

$$\tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau} = I_\alpha[g(\cdot)]_{\varrho,\tau} + \sum_{i=1}^{2^{s(\alpha)}-1} H_{\alpha,i}. \quad (4.2.12)$$

The terms $H_{\alpha,i}$ denote multiple stochastic integrals of the adapted process $g(\cdot)$ that use as integrators the time, the Wiener processes, the Poisson jump measure p_φ and the intensity measure φ . A complete description of these terms could be given by defining recursively a new type of multiple stochastic integral. Then by using the relationship (4.2.1) together with (4.2.9) and (4.2.11), the terms $H_{\alpha,i}$ could be recursively characterized. However, since this characterization requires again rather technical notation, we omit for simplicity this obvious but complex characterization. Instead, we provide the following two illustrative examples.

For $\alpha = (0, -1, 1)$ one has

$$\begin{aligned}\tilde{I}_{(0,-1,1)}[g(\cdot)]_{\varrho,\tau} &= \int_{\varrho}^{\tau} \int_{\varrho}^{z_3} \int_{\mathcal{E}} \int_{\varrho}^{z_2-} g(z_1, v_1) dz_1 \tilde{p}_{\varphi}(dv_1, dz_2) dW_{z_3}^1 \\ &= \int_{\varrho}^{\tau} \int_{\varrho}^{z_3} \int_{\mathcal{E}} \int_{\varrho}^{z_2-} g(z_1, v_1) dz_1 p_{\varphi}(dv_1, dz_2) dW_{z_3}^1 \\ &\quad - \int_{\varrho}^{\tau} \int_{\varrho}^{z_3} \int_{\mathcal{E}} \int_{\varrho}^{z_2} g(z_1, v_1) dz_1 \varphi(dv_1) dz_2 dW_{z_3}^1 \\ &= I_{(0,-1,1)}[g(\cdot)]_{\varrho,\tau} + H_{(0,-1,1),1},\end{aligned}$$

so that

$$H_{(0,-1,1),1} = - \int_{\varrho}^{\tau} \int_{\varrho}^{z_3} \int_{\mathcal{E}} \int_{\varrho}^{z_2} g(z_1, v_1) dz_1 \varphi(dv_1) dz_2 dW_{z_3}^1.$$

For $\alpha = (-1, -1)$ one obtains

$$\begin{aligned}\tilde{I}_{(-1,-1)}[g(\cdot)]_{\varrho,\tau} &= \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^{z_2-} \int_{\mathcal{E}} g(z_1-, v_1, v_2) \tilde{p}_{\varphi}(dv_1, dz_1) \tilde{p}_{\varphi}(dv_2, dz_2) \\ &= \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^{z_2-} \int_{\mathcal{E}} g(z_1-, v_1, v_2) p_{\varphi}(dv_1, dz_1) p_{\varphi}(dv_2, dz_2) \\ &\quad - \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^{z_2} \int_{\mathcal{E}} g(z_1-, v_1, v_2) p_{\varphi}(dv_1, dz_1) \varphi(dv_2) dz_2 \\ &\quad - \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^{z_2-} \int_{\mathcal{E}} g(z_1, v_1, v_2) \varphi(dv_1) dz_1 p_{\varphi}(dv_2, dz_2) \\ &\quad + \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^{z_2} \int_{\mathcal{E}} g(z_1, v_1, v_2) \varphi(dv_1) dz_1 \varphi(dv_2) dz_2 \\ &= I_{(-1,-1)}[g(\cdot)]_{\varrho,\tau} + H_{(-1,-1),1} + H_{(-1,-1),2} + H_{(-1,-1),3},\end{aligned}$$

where each term in the sum expresses the corresponding double integral.

Similarly, we can express a multiple stochastic integral of the type (4.2.9), which uses as integrator for the jumps the Poisson jump measure p_{φ} , as a sum of terms involving multiple stochastic integrals that use as integrators for the jumps the compensated Poisson measure \tilde{p}_{φ} .

Remark 4.2.2. Let $\alpha \in \mathcal{M}_m$, and ϱ and τ denote two stopping times with τ being \mathcal{A}_{ϱ} -measurable and $0 \leq \varrho \leq \tau \leq T$ almost surely. Consider an adapted stochastic process $g \in \mathcal{H}_{\alpha}$. Then

$$I_{\alpha}[g(\cdot)]_{\varrho,\tau} = \tilde{I}_{\alpha}[g(\cdot)]_{\varrho,\tau} + \sum_{i=1}^{2^{s(\alpha)}-1} \tilde{H}_{\alpha,i}. \quad (4.2.13)$$

Here the terms $\tilde{H}_{\alpha,i}$ are multiple stochastic integrals of the process $g(\cdot)$ that use as integrators the time, the Wiener processes, the compensated Poisson measure \tilde{p}_φ and the intensity measure φ . For instance, for $\alpha = (-1, -1)$ one obtains

$$\begin{aligned} I_{(-1,-1)}[g(\cdot)]_{\varrho,\tau} &= \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^{z_2-} \int_{\mathcal{E}} g(z_1-, v_1, v_2) p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2) \\ &= \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^{z_2-} \int_{\mathcal{E}} g(z_1-, v_1, v_2) \tilde{p}_\varphi(dv_1, dz_1) \tilde{p}_\varphi(dv_2, dz_2) \\ &\quad + \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^{z_2} \int_{\mathcal{E}} g(z_1-, v_1, v_2) \tilde{p}_\varphi(dv_1, dz_1) \varphi(dv_2) dz_2 \\ &\quad + \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^{z_2-} \int_{\mathcal{E}} g(z_1, v_1, v_2) \varphi(dv_1) dz_1 \tilde{p}_\varphi(dv_2, dz_2) \\ &\quad + \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^{z_2} \int_{\mathcal{E}} g(z_1, v_1, v_2) \varphi(dv_1) dz_1 \varphi(dv_2) dz_2 \\ &= \tilde{I}_{(-1,-1)}[g(\cdot)]_{\varrho,\tau} + \tilde{H}_{(-1,-1),1} + \tilde{H}_{(-1,-1),2} + \tilde{H}_{(-1,-1),3}. \end{aligned}$$

For instance, when $\alpha = (0, -1, 1)$ one has

$$\begin{aligned} I_{(0,-1,1)}[g(\cdot)]_{\varrho,\tau} &= \int_{\varrho}^{\tau} \int_{\varrho}^{z_3} \int_{\mathcal{E}} \int_{\varrho}^{z_2-} g(z_1, v_1) dz_1 p_\varphi(dv_1, dz_2) dW_{z_3}^1 \\ &= \int_{\varrho}^{\tau} \int_{\varrho}^{z_3} \int_{\mathcal{E}} \int_{\varrho}^{z_2-} g(z_1, v_1) dz_1 \tilde{p}_\varphi(dv_1, dz_2) dW_{z_3}^1 \\ &\quad + \int_{\varrho}^{\tau} \int_{\varrho}^{z_3} \int_{\mathcal{E}} \int_{\varrho}^{z_2} g(z_1, v_1) dz_1 \varphi(dv_1) dz_2 dW_{z_3}^1 \\ &= \tilde{I}_{(0,-1,1)}[g(\cdot)]_{\varrho,\tau} + \tilde{H}_{(0,-1,1),1}, \end{aligned}$$

so that

$$\tilde{H}_{(0,-1,1),1} = \int_{\varrho}^{\tau} \int_{\varrho}^{z_3} \int_{\mathcal{E}} \int_{\varrho}^{z_2} g(z_1, v_1) dz_1 \varphi(dv_1) dz_2 dW_{z_3}^1.$$

Relationships Between Multiple Itô Integrals

Note that there exist recursive relationships between multiple Itô integrals. To give a group of examples in the case without jumps we write for convenience

$$I_{\alpha,t} = I_\alpha[1]_{0,t} \tag{4.2.14}$$

and

$$W_t^0 = t \quad (4.2.15)$$

for $\alpha \in \mathcal{M}_m$ and $t \in [0, T]$. Recall that $\mathbf{1}_A$ denotes the indicator function of the event A . The following result from [Kloeden & Platen \(1999\)](#) provides a set of rules for the case without jumps that allows us to compute certain multiple stochastic integrals by using other stochastic integrals.

Lemma 4.2.3 *Let $j_1, \dots, j_\ell \in \{0, 1, \dots, m\}$ and $\alpha = (j_1, \dots, j_\ell) \in \mathcal{M}_m$, where $\ell \in \mathcal{N}$, then*

$$W_t^j I_{\alpha,t} = \sum_{i=0}^{\ell} I_{(j_1, \dots, j_i, j, j_{i+1}, \dots, j_\ell), t} + \sum_{i=1}^{\ell} \mathbf{1}_{\{j_i=j \neq 0\}} I_{(j_1, \dots, j_{i-1}, 0, j_{i+1}, \dots, j_\ell), t} \quad (4.2.16)$$

for all $t \geq 0$.

It is interesting to note that multiple Itô integrals with respect to one Wiener process are related to Hermite polynomials. In particular, it follows that

$$\begin{aligned} I_{(j), \tau} &= W_\tau^j \\ I_{(j,j), \tau} &= \frac{1}{2} \left((W_\tau^j)^2 - \tau \right) \\ I_{(j,j,j), \tau} &= \frac{1}{3!} \left((W_\tau^j)^3 - 3\tau W_\tau^j \right) \\ I_{(j,j,j,j), \tau} &= \frac{1}{4!} \left((W_\tau^j)^4 - 6\tau (W_\tau^j)^2 + 3\tau^2 \right) \\ I_{(j,j,j,j,j), \tau} &= \frac{1}{5!} \left((W_\tau^j)^5 - 10\tau (W_\tau^j)^3 + 15\tau^2 W_\tau^j \right) \end{aligned} \quad (4.2.17)$$

for $\varrho = 0$, $\tau \in [0, \infty)$, $j \in \{1, 2, \dots, m\}$, when $g(\cdot) = 1$ and we write $I_\alpha[1]_{0,\tau}$ as $I_{\alpha,\tau}$.

In the case with jumps similar relationships hold as shown in [\(4.1.23\)](#) for multiple integrals of counting processes.

4.3 Coefficient Functions

Set of Functions

We need to define some sets of sufficiently smooth and integrable functions to be used as coefficient functions of the stochastic expansions. Let $s \in \{0, 1, \dots\}$ denote some integer that will become clear from the context when the following definitions will be applied. By \mathcal{L}^0 we denote the set of functions $f(t, \mathbf{x}, \mathbf{u}) : [0, T] \times \mathbb{R}^d \times \mathcal{E}^s \rightarrow \mathbb{R}^d$ for which

$$f(t, \mathbf{x} + c(t, \mathbf{x}, \mathbf{v}), \mathbf{u}) - f(t, \mathbf{x}, \mathbf{u}) \quad (4.3.1)$$

is $\varphi(dv)$ -integrable for all $t \in [0, T]$, $\mathbf{x} \in \Re^d$, $\mathbf{u} \in \mathcal{E}^s$ and $f(\cdot, \cdot, \mathbf{u}) \in \mathcal{C}^{1,2}$. Here $\mathcal{C}^{1,2}$ denotes the set of functions that are continuously differentiable with respect to time and twice continuously differentiable with respect to the spatial variables. Note that, according to the notation adopted, c^i denotes the i th component of the jump coefficient vector \mathbf{c} . Moreover, with $\tilde{\mathcal{L}}^0$ we denote the set of functions $f(t, \mathbf{x}, \mathbf{u}) : [0, T] \times \Re^d \times \mathcal{E}^s \rightarrow \Re^d$ for which

$$f(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x}, \mathbf{v}), \mathbf{u}) - f(t, \mathbf{x}, \mathbf{u}) - \sum_{i=1}^d c^i(t, \mathbf{x}, \mathbf{v}) \frac{\partial}{\partial x^i} f(t, \mathbf{x}, \mathbf{u}) \quad (4.3.2)$$

is $\varphi(dv)$ -integrable for all $t \in [0, T]$, $\mathbf{x} \in \Re^d$, $\mathbf{u} \in \mathcal{E}^s$ and $f(\cdot, \cdot, \mathbf{u}) \in \mathcal{C}^{1,2}$. With \mathcal{L}^k , $k \in \{1, 2, \dots, m\}$, we denote the set of functions $f(t, \mathbf{x}, \mathbf{u})$ with partial derivatives $\frac{\partial}{\partial x^i} f(t, \mathbf{x}, \mathbf{u})$, $i \in \{1, 2, \dots, d\}$. With \mathcal{L}^{-1} we denote the set of functions for which

$$|f(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x}, \mathbf{v}), \mathbf{u}) - f(t, \mathbf{x}, \mathbf{u})|^2 \quad (4.3.3)$$

is $\varphi(dv)$ -integrable for all $t \in [0, T]$, $\mathbf{x} \in \Re^d$ and $\mathbf{u} \in \mathcal{E}^s$.

Some Operators

Let us now define the following operators for a function $f(t, \mathbf{x}, \mathbf{u}) \in \mathcal{L}^k$:

$$\begin{aligned} L^0 f(t, \mathbf{x}, \mathbf{u}) &= \frac{\partial}{\partial t} f(t, \mathbf{x}, \mathbf{u}) + \sum_{i=1}^d a^i(t, \mathbf{x}) \frac{\partial}{\partial x^i} f(t, \mathbf{x}, \mathbf{u}) \\ &+ \frac{1}{2} \sum_{i,l=1}^d \sum_{j=1}^m b^{i,j}(t, \mathbf{x}) b^{l,j}(t, \mathbf{x}) \frac{\partial^2}{\partial x^i \partial x^l} f(t, \mathbf{x}, \mathbf{u}) \end{aligned} \quad (4.3.4)$$

$$L^k f(t, \mathbf{x}, \mathbf{u}) = \sum_{i=1}^d b^{i,k}(t, \mathbf{x}) \frac{\partial}{\partial x^i} f(t, \mathbf{x}, \mathbf{u}), \quad (4.3.5)$$

for $k \in \{1, 2, \dots, m\}$, and

$$L_v^{-1} f(t, \mathbf{x}, \mathbf{u}) = f(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x}, \mathbf{v}), \mathbf{u}) - f(t, \mathbf{x}, \mathbf{u}), \quad (4.3.6)$$

for all $t \in [0, T]$, $\mathbf{x} \in \Re^d$ and $\mathbf{u} \in \mathcal{E}^s$. Note that the operator in (4.3.6) adds an extra dependence $v \in \mathcal{E}$ on the mark components. Let us also define the operator

$$\begin{aligned}
\tilde{L}^0 f(t, \mathbf{x}, \mathbf{u}) &= L^0 f(t, \mathbf{x}, \mathbf{u}) + \int_{\mathcal{E}} \{f(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x}, \mathbf{v}), \mathbf{u}) - f(t, \mathbf{x}, \mathbf{u})\} \varphi(d\mathbf{v}) \\
&= \frac{\partial}{\partial t} f(t, \mathbf{x}, \mathbf{u}) + \sum_{i=1}^d \tilde{a}^i(t, \mathbf{x}) \frac{\partial}{\partial x^i} f(t, \mathbf{x}, \mathbf{u}) \\
&\quad + \frac{1}{2} \sum_{i,l=1}^d \sum_{j=1}^m b^{i,j}(t, \mathbf{x}) b^{l,j}(t, \mathbf{x}) \frac{\partial^2}{\partial x^i \partial x^l} f(t, \mathbf{x}, \mathbf{u}) \\
&\quad + \int_{\mathcal{E}} \left\{ f(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x}, \mathbf{v}), \mathbf{u}) - f(t, \mathbf{x}, \mathbf{u}) \right. \\
&\quad \left. - \sum_{i=1}^d c^i(t, \mathbf{x}, \mathbf{v}) \frac{\partial}{\partial x^i} f(t, \mathbf{x}, \mathbf{u}) \right\} \varphi(d\mathbf{v}),
\end{aligned} \tag{4.3.7}$$

for all $t \in [0, T]$, $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{u} \in \mathcal{E}^s$, which allows us to describe conveniently the impact of the compensated Poisson measure.

Itô Coefficient Functions

By using the above definitions, for all $\alpha = (j_1, \dots, j_{l(\alpha)}) \in \mathcal{M}_m$ and a function $f : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, we define recursively the *Itô coefficient functions*

$$f_\alpha(t, x, u) = \begin{cases} f(t, \mathbf{x}) & \text{for } l(\alpha) = 0, \\ L^{j_1} f_{-\alpha}(t, \mathbf{x}, u_1, \dots, u_{s(-\alpha)}) & \text{for } l(\alpha) \geq 1, j_1 \in \{0, 1, \dots, m\} \\ L_{u_{s(\alpha)}}^{-1} f_{-\alpha}(t, \mathbf{x}, u_1, \dots, u_{s(-\alpha)}) & \text{for } l(\alpha) \geq 1, j_1 = -1. \end{cases} \tag{4.3.8}$$

By $\mathbf{u} = (u^1, \dots, u^{s(\alpha)})$ we denote a vector $\mathbf{u} \in \mathcal{E}^{s(\alpha)}$. Note that the dependence on \mathbf{u} in (4.3.8) is introduced by the repeated application of the operator (4.3.6). Additionally, we assume that the coefficients of the SDE (1.8.2) and the function f satisfy the smoothness and integrability conditions needed for the operators in (4.3.8) to be well defined, see also Remark 4.4.2.

If we choose the identity function $f(t, x) = x$, then for the case $d = m = 1$ we can write the examples

$$f_{(-1,0)}(t, x, u) = L_u^{-1} a(t, x) = a(t, x + c(t, x, u)) - a(t, x),$$

$$\begin{aligned}
f_{(0,1)}(t, x) &= L^0 b(t, x) \\
&= \frac{\partial}{\partial t} b(t, x) + a(t, x) \frac{\partial}{\partial x} b(t, x) + \frac{1}{2} (b(t, x))^2 \frac{\partial^2}{\partial x^2} b(t, x)
\end{aligned}$$

and

$$\begin{aligned} f_{(-1,-1)}(t, x, \mathbf{u}) &= L_{u^2}^{-1} c(t, x, u^1) \\ &= c(t, x + c(t, x, u^2), u^1) - c(t, x, u^1), \end{aligned}$$

with $\mathbf{u} = (u^1, u^2) \in \mathcal{E}^2$.

Compensated Itô Coefficient Functions

For all $\alpha = (j_1, \dots, j_{l(\alpha)}) \in \mathcal{M}_m$ and a function $f : [0, T] \times \mathfrak{R}^d \rightarrow \mathfrak{R}^d$, let us define recursively the *compensated Itô coefficient functions*

$$\tilde{f}_\alpha(t, \mathbf{x}, \mathbf{u}) = \begin{cases} f(t, \mathbf{x}) & \text{for } l(\alpha) = 0, \\ \tilde{L}^0 \tilde{f}_{-\alpha}(t, \mathbf{x}, u_1, \dots, u_{s(-\alpha)}) & \text{for } l(\alpha) \geq 1, j_1 = 0 \\ L^{j_1} \tilde{f}_{-\alpha}(t, \mathbf{x}, u_1, \dots, u_{s(-\alpha)}) & \text{for } l(\alpha) \geq 1, j_1 \in \{1, \dots, m\} \\ L_{u_{s(\alpha)}}^{-1} \tilde{f}_{-\alpha}(t, \mathbf{x}, u_1, \dots, u_{s(-\alpha)}) & \text{for } l(\alpha) \geq 1, j_1 = -1. \end{cases} \quad (4.3.9)$$

Here we assume again that the coefficients of the SDE (1.8.2) and the function f satisfy the smoothness and integrability conditions needed for the operators in (4.3.9) to be well defined. For illustration, if we choose the identity function $f(t, x) = x$, then for the case $d = m = 1$ we can formulate the examples

$$\begin{aligned} \tilde{f}_{(-1,0)}(t, x, u) &= L_u^{-1} \tilde{a}(t, x) \\ &= \tilde{a}(t, x + c(t, x, u)) - \tilde{a}(t, x) \\ &= a(t, x + c(t, x, u)) - a(t, x) \\ &\quad + \int_{\mathcal{E}} \{c(t, x + c(t, x, u), v) - c(t, x, v)\} \varphi(dv), \end{aligned}$$

$$\begin{aligned} \tilde{f}_{(0,1)}(t, x) &= \tilde{L}^0 b(t, x) \\ &= \frac{\partial}{\partial t} b(t, x) + a(t, x) \frac{\partial}{\partial x} b(t, x) + \frac{1}{2} (b(t, x))^2 \frac{\partial^2}{\partial x^2} b(t, x) \\ &\quad + \int_{\mathcal{E}} \{b(t, x + c(t, x, v)) - b(t, x)\} \varphi(dv) \\ &= \frac{\partial}{\partial t} b(t, x) + \tilde{a}(t, x) \frac{\partial}{\partial x} b(t, x) + \frac{1}{2} (b(t, x))^2 \frac{\partial^2}{\partial x^2} b(t, x) \\ &\quad + \int_{\mathcal{E}} \left\{ b(t, x + c(t, x, v)) - b(t, x) - c(t, x, v) \frac{\partial}{\partial x} b(t, x) \right\} \varphi(dv) \end{aligned}$$

and

$$\tilde{f}_{(-1,-1)}(t, x, u) = f_{(-1,-1)}(t, x, u). \quad (4.3.10)$$

Hierarchical and Remainder Sets

To define a Wagner-Platen expansion we finally need to select some appropriate sets of multi-indices that characterize its expansion part. A subset $\mathcal{A} \in \mathcal{M}_m$ is a *hierarchical* if \mathcal{A} is non-empty, the multi-indices in \mathcal{A} are uniformly bounded in length, which means $\sup_{\alpha \in \mathcal{A}} l(\alpha) < \infty$, and $-\alpha \in \mathcal{A}$ for each $\alpha \in \mathcal{A} \setminus \{v\}$.

We also define the *remainder* set $\mathcal{B}(\mathcal{A})$ of a hierarchical set \mathcal{A} by

$$\mathcal{B}(\mathcal{A}) = \{\alpha \in \mathcal{M}_m \setminus \mathcal{A} : -\alpha \in \mathcal{A}\}. \quad (4.3.11)$$

Then the remainder set consists of all next following multi-indices with respect to the given hierarchical set.

Let us give an example of a hierarchical set and the corresponding remainder set: When $m = 1$, we will later see that the hierarchical set \mathcal{A}_E , corresponding to the popular Euler scheme, is of the form

$$\mathcal{A}_E = \{v, (-1), (0), (1)\}. \quad (4.3.12)$$

The corresponding remainder set $\mathcal{B}(\mathcal{A}_E)$ equals then

$$\mathcal{B}(\mathcal{A}_E) = \{(-1, -1), (0, -1), (1, -1), (-1, 0), (0, 0), (1, 0), (-1, 1), (0, 1), (1, 1)\}.$$

4.4 Wagner-Platen Expansions

General Wagner-Platen Expansion

We will now present the general form of Wagner-Platen expansions for the solution of the d -dimensional SDE with jumps

$$d\mathbf{X}_t = \mathbf{a}(t, \mathbf{X}_t) dt + \mathbf{b}(t, \mathbf{X}_t) d\mathbf{W}_t + \int_{\mathcal{E}} \mathbf{c}(t, \mathbf{X}_{t-}, v) p_{\varphi}(dv, dt), \quad (4.4.1)$$

for $t \in [0, T]$, with $\mathbf{X}_0 \in \mathbb{R}^d$, as described in (1.8.2) and (4.2.2).

Theorem 4.4.1. *For two given stopping times ϱ and τ with $0 \leq \varrho \leq \tau \leq T$ a.s., a hierarchical set $\mathcal{A} \in \mathcal{M}_m$, and a function $f : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, we obtain the corresponding Wagner-Platen expansion*

$$f(\tau, \mathbf{X}_{\tau}) = \sum_{\alpha \in \mathcal{A}} I_{\alpha}[f_{\alpha}(\varrho, \mathbf{X}_{\varrho})]_{\varrho, \tau} + \sum_{\alpha \in \mathcal{B}(\mathcal{A})} I_{\alpha}[f_{\alpha}(\cdot, \mathbf{X}_{\cdot})]_{\varrho, \tau} \quad (4.4.2)$$

and the compensated Wagner-Platen expansion

$$f(\tau, \mathbf{X}_{\tau}) = \sum_{\alpha \in \mathcal{A}} \tilde{I}_{\alpha}[\tilde{f}_{\alpha}(\varrho, \mathbf{X}_{\varrho})]_{\varrho, \tau} + \sum_{\alpha \in \mathcal{B}(\mathcal{A})} \tilde{I}_{\alpha}[\tilde{f}_{\alpha}(\cdot, \mathbf{X}_{\cdot})]_{\varrho, \tau}, \quad (4.4.3)$$

assuming that the function f and the coefficients of the SDE (4.4.1) are sufficiently smooth and integrable such that the arising coefficient functions f_{α} and \tilde{f}_{α} are well defined and the corresponding multiple stochastic integrals exist.

Note that in (4.4.2) and (4.4.3) we have, for simplicity, neglected that f_α and \tilde{f}_α are vectors and suppressed their dependence on $\mathbf{u} \in \mathcal{E}^{s(\alpha)}$ in our notation. We will also continue to do so in the following when no misunderstanding is possible.

Remark 4.4.2. For instance, if the function f and the coefficients \mathbf{a} , \mathbf{b} and \mathbf{c} of the SDE (4.4.1) are $2(l(\alpha)+1)$ times continuously differentiable, uniformly bounded with uniformly bounded derivatives, then all coefficient functions and multiple stochastic integrals in the expansions (4.4.2) and (4.4.3) are well defined.

The proof of the Wagner-Platen expansion is again based on an iterated application of the Itô formula. Since it is formally the same as in [Platen \(1982a, 1982b\)](#) and analogous to the one in [Kloeden & Platen \(1999\)](#), it is here omitted.

By choosing as function f the identity function $f(t, \mathbf{x}) = \mathbf{x}$, we can represent the solution $\mathbf{X} = \{\mathbf{X}_t, t \in [0, T]\}$ of the SDE (4.4.1) by the *Wagner-Platen expansion*

$$\mathbf{X}_\tau = \sum_{\alpha \in \mathcal{A}} I_\alpha[f_\alpha(\varrho, \mathbf{X}_\varrho)]_{\varrho, \tau} + \sum_{\alpha \in \mathcal{B}(\mathcal{A})} I_\alpha[f_\alpha(\cdot, \mathbf{X}_\cdot)]_{\varrho, \tau} \quad (4.4.4)$$

and also by the *compensated Wagner-Platen expansion*

$$\mathbf{X}_\tau = \sum_{\alpha \in \mathcal{A}} \tilde{I}_\alpha[\tilde{f}_\alpha(\varrho, \mathbf{X}_\varrho)]_{\varrho, \tau} + \sum_{\alpha \in \mathcal{B}(\mathcal{A})} \tilde{I}_\alpha[\tilde{f}_\alpha(\cdot, \mathbf{X}_\cdot)]_{\varrho, \tau}, \quad (4.4.5)$$

where ϱ and τ are two given stopping times with $0 \leq \varrho \leq \tau \leq T$ a.s.

We remark that Wagner-Platen expansions are generalizations of the deterministic Taylor formula. On the other hand, since they are obtained by an iterative application of the Itô formula, they also generalize the Itô formula. As we shall see later, discrete-time approximations of an Itô process will be constructed by using truncated Wagner-Platen expansions and neglecting their remainder terms.

Euler Expansion

For instance, for the hierarchical set \mathcal{A}_E with $m = 1$, see (4.3.12), which generates the *Euler scheme*, we obtain from (4.4.4) the Wagner-Platen expansion

$$X_\tau = X_\varrho + a(\varrho, X_\varrho) \int_\varrho^\tau dz + b(\varrho, X_\varrho) \int_\varrho^\tau dW_z + \int_\varrho^\tau \int_{\mathcal{E}} c(\varrho, X_\varrho, v) p_\varphi(dv, dz) + R$$

with remainder

$$\begin{aligned}
R = & \int_{\varrho}^{\tau} \int_{\varrho}^s L^0 a(z, X_z) dz ds + \int_{\varrho}^{\tau} \int_{\varrho}^s L^1 a(z, X_z) dW_z ds \\
& + \int_{\varrho}^{\tau} \int_{\varrho}^s \int_{\mathcal{E}} L_v^{-1} a(z, X_z) p_{\varphi}(dv, dz) ds \\
& + \int_{\varrho}^{\tau} \int_{\varrho}^s L^0 b(z, X_z) dz dW_s + \int_{\varrho}^{\tau} \int_{\varrho}^s L^1 b(z, X_z) dW_z dW_s \\
& + \int_{\varrho}^{\tau} \int_{\varrho}^s \int_{\mathcal{E}} L_v^{-1} b(z, X_z) p_{\varphi}(dv, dz) dW_s \\
& + \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^s L^0 c(z, X_z, v) dz p_{\varphi}(dv, ds) \\
& + \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^s L^1 c(z, X_z, v) dW_z p_{\varphi}(dv, ds) \\
& + \int_{\varrho}^{\tau} \int_{\mathcal{E}} \int_{\varrho}^s \int_{\mathcal{E}} L_u^{-1} c(z, X_z, v) p_{\varphi}(dv, dz) p_{\varphi}(du, ds). \tag{4.4.6}
\end{aligned}$$

There is an analogous Wagner-Platen expansion that follows from (4.4.5), where the Poisson measure is replaced by the compensated Poisson measure, the drift by the compensated drift and the Itô coefficient functions by their counterparts for the compensated case. Note that in this particular case, for the hierarchical set \mathcal{A}_E , one can show that these two expansions are equivalent. However in general, these two expansions are different.

Wagner-Platen Expansion up to Multiplicity Three

Now, let us consider a complex but useful example for the case without jumps. If we apply the Wagner-Platen expansion in the case $d = m = 1$ for $f(t, x) = x$, $\varrho = 0$, $\tau \in (0, \infty)$ for the hierarchical set

$$\mathcal{A} = \{\alpha \in \mathcal{M}_m : \ell(\alpha) \leq 3\}$$

to the process X with drift $a(t, x) = a(x)$, diffusion coefficient $b(t, x) = b(x)$ and zero jump coefficient $c(t, x, v) = 0$, then we obtain, by setting $I_{\alpha}[1]_{0,\tau} = I_{\alpha}$ and writing for any function $g(\cdot)$ evaluated at X_0 simply g , the expansion

$$\begin{aligned}
X_{\tau} = & X_0 + a I_{(0)} + b I_{(1)} + \left(a a' + \frac{1}{2} b^2 a'' \right) I_{(0,0)} \\
& + \left(a b' + \frac{1}{2} b^2 b'' \right) I_{(0,1)} + b a' I_{(1,0)} + b b' I_{(1,1)}
\end{aligned}$$

$$\begin{aligned}
& + \left[a \left(a a'' + (a')^2 + b b' a'' + \frac{1}{2} b^2 a''' \right) + \frac{1}{2} b^2 (a a''' + 3 a' a'') \right. \\
& \quad \left. + ((b')^2 + b b'') a'' + 2 b b' a''' \right) + \frac{1}{4} b^4 a^{(4)} \Big] I_{(0,0,0,0)} \\
& + \left[a \left(a' b' + a b'' + b b' b'' + \frac{1}{2} b^2 b''' \right) + \frac{1}{2} b^2 (a'' b' + 2 a' b'') \right. \\
& \quad \left. + a b''' + ((b')^2 + b b'') b'' + 2 b b' b''' + \frac{1}{2} b^2 b^{(4)} \right) \Big] I_{(0,0,0,1)} \\
& + \left[a (b' a' + b a'') + \frac{1}{2} b^2 (b'' a' + 2 b' a'' + b a''') \right] I_{(0,1,0,0)} \\
& + \left[a ((b')^2 + b b'') + \frac{1}{2} b^2 (b'' b' + 2 b b'' + b b''') \right] I_{(0,1,1,0)} \\
& + b \left(a a'' + (a')^2 + b b' a'' + \frac{1}{2} b^2 a''' \right) I_{(1,0,0,0)} \\
& + b \left(a b'' + a' b' + b b' b'' + \frac{1}{2} b^2 b''' \right) I_{(1,0,1,0)} \\
& + b (a' b' + a'' b) I_{(1,1,0,0)} + b ((b')^2 + b b'') I_{(1,1,1,0)} + R_6. \quad (4.4.7)
\end{aligned}$$

This is a useful formula since it will allow us to readily establish various numerical schemes later.

The most important property of the Wagner-Platen expansion is that it allows a function of a process to be expanded as the sum of a finite number of multiple Itô integrals with constant integrands. Like the deterministic Taylor expansion, it can be flexibly used for the approximation of increments of solutions of SDEs over small time intervals. If one needs to exploit martingale properties, as it is often the case in finance, then one should use a Wagner-Platen expansion because the Itô stochastic integrals, which are martingales, vanish when taking conditional expectations. However, if there are advantages from a less complex structure under the Stratonovich calculus, then a Stratonovich type stochastic expansion is appropriate, see [Kloeden & Platen \(1999\)](#).

Examples of Wagner-Platen Expansions

Let us give three specific examples for Wagner-Platen expansions:

1. First, we consider a *Vasicek interest rate model*

$$dr_t = \gamma (\bar{r} - r_t) dt + \beta dW_t \quad (4.4.8)$$

for $t \in [0, T]$, $r_0 \geq 0$. Its Wagner-Platen expansion involving all triple integrals, that is, with hierarchical set

$$\mathcal{A} = \{\alpha \in \mathcal{M}_1 : \ell(\alpha) \leq 3\},$$

is of the form

$$\begin{aligned} r_t &= r_0 + \gamma (\bar{r} - r_0) t + \beta W_t - \gamma^2 (\bar{r} - r_0) \frac{t^2}{2} - \beta \gamma \int_0^t W_s ds \\ &\quad + \gamma^3 (\bar{r} - r_0) \frac{t^3}{6} + \beta \gamma^2 \int_0^t \int_0^{s_2} W_{s_1} ds_1 ds_2 + R_6. \end{aligned} \quad (4.4.9)$$

2. As another example consider the *Black-Scholes dynamics*

$$dS_t = S_t (a dt + \sigma dW_t) \quad (4.4.10)$$

for $t \in [0, T]$, $S_0 \geq 0$. Its Wagner-Platen expansion, again with hierarchical set

$$\mathcal{A} = \{\alpha \in \mathcal{M}_1 : \ell(\alpha) \leq 3\},$$

is given by

$$\begin{aligned} S_t &= S_0 \left[1 + a t + \sigma W_t + a^2 \frac{t^2}{2} + a \sigma (I_{(0,1)} + I_{(1,0)}) \right. \\ &\quad + \sigma^2 \frac{1}{2} ((W_t)^2 - t) + a^3 \frac{t^3}{6} + a^2 \sigma (I_{(0,0,1)} + I_{(0,1,0)} + I_{(1,0,0)}) \quad (4.4.11) \\ &\quad \left. + a \sigma^2 (I_{(0,1,1)} + I_{(1,0,1)} + I_{(1,1,0)}) + \sigma^3 \frac{1}{6} ((W_t)^3 - 3t W_t) \right] + R_6. \end{aligned}$$

Using the relationship (4.2.16) between multiple Itô integrals, we have

$$t W_t = I_{(0)} I_{(1)} = I_{(0,1)} + I_{(1,0)}, \quad W_t \frac{t^2}{2} = I_{(1)} I_{(0,0)} = I_{(0,0,1)} + I_{(0,1,0)} + I_{(1,0,0)}$$

and

$$t \frac{1}{2} ((W_t)^2 - t) = I_{(0)} I_{(1,1)} = I_{(0,1,1)} + I_{(1,0,1)} + I_{(1,1,0)}.$$

Therefore, it follows from (4.4.11) that

$$\begin{aligned} S_t &= S_0 \left[1 + a t + \sigma W_t + a^2 \frac{t^2}{2} + a \sigma t W_t + \frac{\sigma^2}{2} ((W_t)^2 - t) + a^3 \frac{t^3}{6} \right. \\ &\quad \left. + a^2 \sigma W_t \frac{t^2}{2} + a \sigma^2 \frac{t}{2} ((W_t)^2 - t) + \sigma^3 \frac{1}{6} ((W_t)^3 - 3t W_t) \right] + R_6. \end{aligned}$$

3. Finally, let us expand a *squared Bessel process* $X = \{X_t, t \in [0, T]\}$ of dimension $\nu > 0$ given by the SDE

$$dX_t = \nu dt + 2\sqrt{X_t} dW_t$$

for $t \in [0, T]$ with $X_0 > 0$. Here the Wagner-Platen expansion with hierarchical set

$$\mathcal{A} = \{\alpha \in \mathcal{M}_1 : \ell(\alpha) \leq 2\},$$

is obtained as

$$X_t = X_0 + (\nu - 1)t + 2\sqrt{X_0} W_t + \frac{\nu - 1}{\sqrt{X_0}} \int_0^t s dW_s + (W_t)^2 + \tilde{R}. \quad (4.4.12)$$

4.5 Moments of Multiple Stochastic Integrals

The lemmas in this section provide estimates of multiple stochastic integrals. These constitute the basis of the proofs of convergence theorems that will be presented later.

First Moment Estimate

We will consider both the case of multiple stochastic integrals $I_\alpha[g(\cdot)]_{\varrho,\tau}$ with respect to the Poisson jump measure and compensated multiple stochastic integrals $\tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau}$ with respect to the compensated Poisson measure. The estimates to be derived in these two cases differ in the values of some finite constants, but show the same structural dependence on the length of the interval of integration $(\tau - \varrho)$. Because of the martingale property of the compensated Poisson measure \tilde{p}_φ , the proofs flow more naturally in the case of the compensated multiple stochastic integral $\tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau}$. Therefore, in the following proofs we will first consider this case in full detail. Then, by using the decomposition

$$p_\varphi(dv, dt) = \tilde{p}_\varphi(dv, dt) + \varphi(dv)dt, \quad (4.5.1)$$

see (4.2.1), we will prove corresponding estimates for the case of the multiple stochastic integrals $I_\alpha[g(\cdot)]_{\varrho,\tau}$.

The first moment of a multiple Itô integral vanishes if it has at least one integration with respect to a component of the Wiener process or the compensated Poisson measure. This is the case when only using compensated Poisson measures and not all of the components of the multi-index α are equal to zero, which is equivalent to the condition $\ell(\alpha) \neq n(\alpha)$, where $n(\alpha)$ denotes the number of zeros in α . The following lemma is a straightforward generalization of a similar statement in [Kloeden & Platen \(1999\)](#).

Lemma 4.5.1 *Let $\alpha \in \mathcal{M}_m \setminus \{v\}$ with $\ell(\alpha) \neq n(\alpha)$, $g \in \mathcal{H}_\alpha$ and ϱ and τ be two stopping times with $0 \leq \varrho \leq \tau \leq T$, almost surely, then*

$$E \left(\tilde{I}_\alpha [g(\cdot)]_{\varrho,\tau} \mid \mathcal{A}_\varrho \right) = 0. \quad (4.5.2)$$

Second Moment Estimates

The following lemma provides a uniform mean-square estimate of multiple stochastic integrals.

Lemma 4.5.2 *Let $\alpha \in \mathcal{M}_m \setminus \{v\}$, $g \in \mathcal{H}_\alpha$, $\Delta > 0$ and ϱ and τ denote two stopping times with τ being \mathcal{A}_ϱ -measurable and $0 \leq \varrho \leq \tau \leq \varrho + \Delta \leq T$ almost surely. Then*

$$\tilde{F}_\tau^\alpha = E \left(\sup_{\varrho \leq s \leq \tau} \left| \tilde{I}_\alpha[g(\cdot)]_{\varrho, s} \right|^2 \middle| \mathcal{A}_\varrho \right) \leq 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_\varrho^\tau V_{\varrho, z, s(\alpha)} dz, \quad (4.5.3)$$

and

$$F_\tau^\alpha = E \left(\sup_{\varrho \leq s \leq \tau} |I_\alpha[g(\cdot)]_{\varrho, s}|^2 \middle| \mathcal{A}_\varrho \right) \leq 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \hat{K}^{s(\alpha)} \int_\varrho^\tau V_{\varrho, z, s(\alpha)} dz, \quad (4.5.4)$$

where

$$V_{\varrho, z, s(\alpha)} = \int_{\mathcal{E}} \dots \int_{\mathcal{E}} E \left(\sup_{\varrho \leq t \leq z} |g(t, v^1, \dots, v^{s(\alpha)})|^2 \middle| \mathcal{A}_\varrho \right) \varphi(dv^1) \dots \varphi(dv^{s(\alpha)}) < \infty \quad (4.5.5)$$

for $z \in [\varrho, \tau]$, and $\hat{K} = \frac{1}{2}(4 + T\lambda)$.

Proof: We will first prove assertion (4.5.3) by induction with respect to $l(\alpha)$.

1. Let us assume that $l(\alpha) = 1$ and $\alpha = (0)$. By the Cauchy-Schwarz inequality we have the estimate

$$\left| \int_\varrho^s g(z) dz \right|^2 \leq (s - \varrho) \int_\varrho^s |g(z)|^2 dz. \quad (4.5.6)$$

Therefore, we obtain

$$\begin{aligned} \tilde{F}_\tau^{(0)} &= E \left(\sup_{\varrho \leq s \leq \tau} \left| \int_\varrho^s g(z) dz \right|^2 \middle| \mathcal{A}_\varrho \right) \\ &\leq E \left(\sup_{\varrho \leq s \leq \tau} (s - \varrho) \int_\varrho^s |g(z)|^2 dz \middle| \mathcal{A}_\varrho \right) \\ &= E \left((\tau - \varrho) \int_\varrho^\tau |g(z)|^2 dz \middle| \mathcal{A}_\varrho \right) \\ &\leq \Delta E \left(\int_\varrho^\tau |g(z)|^2 dz \middle| \mathcal{A}_\varrho \right) \\ &= \Delta \int_\varrho^\tau E(|g(z)|^2 | \mathcal{A}_\varrho) dz \end{aligned}$$

$$\begin{aligned} &\leq \Delta \int_{\varrho}^{\tau} E \left(\sup_{\varrho \leq t \leq z} |g(t)|^2 \middle| \mathcal{A}_{\varrho} \right) dz \\ &= 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{\varrho}^{\tau} V_{\varrho, z, s(\alpha)} dz, \end{aligned} \quad (4.5.7)$$

where the interchange between expectation and integral holds by the \mathcal{A}_{ϱ} -measurability of τ and an application of Fubini's theorem.

2. When $l(\alpha) = 1$ and $\alpha = (j)$ with $j \in \{1, 2, \dots, m\}$, we first observe that the process

$$\left\{ \tilde{I}_{\alpha}[g(\cdot)]_{\varrho, t}, t \in [\varrho, T] \right\} = \left\{ \int_{\varrho}^t g(s) dW_s^j, t \in [\varrho, T] \right\} \quad (4.5.8)$$

is a local martingale. Since $g \in \mathcal{H}_{(j)}$, by (4.2.8) it is also a martingale. Therefore, applying Doob's inequality and Itô's isometry we have

$$\begin{aligned} \tilde{F}_{\tau}^{(j)} &= E \left(\sup_{\varrho \leq s \leq \tau} \left| \int_{\varrho}^s g(z) dW_z^j \right|^2 \middle| \mathcal{A}_{\varrho} \right) \\ &\leq 4 E \left(\left| \int_{\varrho}^{\tau} g(z) dW_z^j \right|^2 \middle| \mathcal{A}_{\varrho} \right) \\ &= 4 E \left(\int_{\varrho}^{\tau} |g(z)|^2 dz \middle| \mathcal{A}_{\varrho} \right) \\ &= 4 \int_{\varrho}^{\tau} E(|g(z)|^2 | \mathcal{A}_{\varrho}) dz \\ &\leq 4 \int_{\varrho}^{\tau} E \left(\sup_{\varrho \leq t \leq z} |g(t)|^2 \middle| \mathcal{A}_{\varrho} \right) dz \\ &= 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{\varrho}^{\tau} V_{\varrho, z, s(\alpha)} dz. \end{aligned} \quad (4.5.9)$$

Here again the interchange between expectation and integral holds by the \mathcal{A}_{ϱ} -measurability of τ and Fubini's theorem.

3. Let us now consider the case with $l(\alpha) = 1$ and $\alpha = (-1)$. The process

$$\left\{ \tilde{I}_{\alpha}[g(\cdot)]_{\varrho, t}, t \in [\varrho, T] \right\} = \left\{ \int_{\varrho}^t \int_{\mathcal{E}} g(s-, v) \tilde{p}_{\varphi}(dv, ds), t \in [\varrho, T] \right\} \quad (4.5.10)$$

is by (4.2.1) and (4.2.8) a martingale. Then, by Doob's inequality and the isometry for Poisson type jump martingales, we obtain

$$\begin{aligned}
\tilde{F}_\tau^{(-1)} &= E \left(\sup_{\varrho \leq s \leq \tau} \left| \int_\varrho^s \int_{\mathcal{E}} g(z-, v) \tilde{p}_\varphi(dv, dz) \right|^2 \middle| \mathcal{A}_\varrho \right) \\
&\leq 4 E \left(\left| \int_\varrho^\tau \int_{\mathcal{E}} g(z-, v) \tilde{p}_\varphi(dv, dz) \right|^2 \middle| \mathcal{A}_\varrho \right) \\
&= 4 E \left(\int_\varrho^\tau \int_{\mathcal{E}} |g(z, v)|^2 \varphi(dv) dz \middle| \mathcal{A}_\varrho \right) \\
&= 4 \int_\varrho^\tau \int_{\mathcal{E}} E(|g(z, v)|^2 | \mathcal{A}_\varrho) \varphi(dv) dz \\
&\leq 4 \int_\varrho^\tau \int_{\mathcal{E}} E \left(\sup_{\varrho \leq t \leq z} |g(t, v)|^2 \middle| \mathcal{A}_\varrho \right) \varphi(dv) dz \\
&= 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_\varrho^\tau V_{\varrho, z, s(\alpha)} dz, \tag{4.5.11}
\end{aligned}$$

since $s(\alpha) = 1$. This shows that the result of Lemma 4.5.2 holds for $l(\alpha) = 1$.

4. Now, let $l(\alpha) = n+1$, where $\alpha = (j_1, \dots, j_{n+1})$ and $j_{n+1} = 0$. By applying the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}
\tilde{F}_\tau^\alpha &= E \left(\sup_{\varrho \leq s \leq \tau} \left| \int_\varrho^s \tilde{I}_{\alpha-[g(\cdot)]_{\varrho,z}} dz \right|^2 \middle| \mathcal{A}_\varrho \right) \\
&\leq E \left(\sup_{\varrho \leq s \leq \tau} (s - \varrho) \int_\varrho^s |\tilde{I}_{\alpha-[g(\cdot)]_{\varrho,z}}|^2 dz \middle| \mathcal{A}_\varrho \right) \\
&= E \left((\tau - \varrho) \int_\varrho^\tau |\tilde{I}_{\alpha-[g(\cdot)]_{\varrho,z}}|^2 dz \middle| \mathcal{A}_\varrho \right) \\
&\leq \Delta E \left(\int_\varrho^\tau |\tilde{I}_{\alpha-[g(\cdot)]_{\varrho,z}}|^2 dz \middle| \mathcal{A}_\varrho \right) \\
&\leq \Delta E \left(\int_\varrho^\tau \sup_{\varrho \leq s \leq \tau} |\tilde{I}_{\alpha-[g(\cdot)]_{\varrho,s}}|^2 dz \middle| \mathcal{A}_\varrho \right) \\
&= \Delta E \left(\int_\varrho^\tau dz \times \sup_{\varrho \leq s \leq \tau} |\tilde{I}_{\alpha-[g(\cdot)]_{\varrho,s}}|^2 \middle| \mathcal{A}_\varrho \right) \\
&\leq \Delta^2 E \left(\sup_{\varrho \leq s \leq \tau} |\tilde{I}_{\alpha-[g(\cdot)]_{\varrho,s}}|^2 \middle| \mathcal{A}_\varrho \right). \tag{4.5.12}
\end{aligned}$$

Then, by the inductive hypothesis it follows that

$$\begin{aligned}\tilde{F}_\tau^\alpha &\leq \Delta^2 4^{l(\alpha)-n(\alpha-)} \Delta^{l(\alpha-)+n(\alpha-)-1} \int_\varrho^\tau V_{\varrho,z,s(\alpha-)} dz \\ &= 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_\varrho^\tau V_{\varrho,z,s(\alpha)} dz,\end{aligned}\quad (4.5.13)$$

where the last line holds since $l(\alpha) = l(\alpha-) + 1$, $n(\alpha) = n(\alpha-) + 1$ and $s(\alpha) = s(\alpha-)$.

5. Let us now consider the case when $l(\alpha) = n + 1$, where $\alpha = (j_1, \dots, j_{n+1})$ and $j_{n+1} \in \{1, 2, \dots, m\}$. The process

$$\left\{ \tilde{I}_\alpha[g(\cdot)]_{\varrho,t}, t \in [\varrho, T] \right\} \quad (4.5.14)$$

is again a martingale. Therefore, by Doob's inequality and Itô's isometry we obtain

$$\begin{aligned}\tilde{F}_\tau^\alpha &= E \left(\sup_{\varrho \leq s \leq \tau} \left| \int_\varrho^s \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,z} dW_z^{j_{n+1}} \right|^2 \middle| \mathcal{A}_\varrho \right) \\ &\leq 4 E \left(\left| \int_\varrho^\tau I_{\alpha-}[g(\cdot)]_{\varrho,z} dW_z^{j_{n+1}} \right|^2 \middle| \mathcal{A}_\varrho \right) \\ &= 4 E \left(\int_\varrho^\tau |\tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,z}|^2 dz \middle| \mathcal{A}_\varrho \right) \\ &\leq 4 E \left(\int_\varrho^\tau \sup_{\varrho \leq s \leq \tau} |\tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,s}|^2 dz \middle| \mathcal{A}_\varrho \right) \\ &= 4 E \left((\tau - \varrho) \sup_{\varrho \leq s \leq \tau} |\tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,s}|^2 \middle| \mathcal{A}_\varrho \right) \\ &\leq 4 \Delta E \left(\sup_{\varrho \leq s \leq \tau} |\tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,s}|^2 \middle| \mathcal{A}_\varrho \right).\end{aligned}\quad (4.5.15)$$

By the inductive hypothesis we have

$$\begin{aligned}\tilde{F}_\tau^\alpha &\leq 4 \Delta 4^{l(\alpha)-n(\alpha-)} \Delta^{l(\alpha-)+n(\alpha-)-1} \int_\varrho^\tau V_{\varrho,z,s(\alpha-)} dz \\ &= \Delta^{l(\alpha)+n(\alpha)-1} 4^{l(\alpha)-n(\alpha)} \int_\varrho^\tau V_{\varrho,z,s(\alpha)} dz,\end{aligned}\quad (4.5.16)$$

since $l(\alpha) = l(\alpha-) + 1$, $n(\alpha) = n(\alpha-)$ and $s(\alpha) = s(\alpha-)$.

6. Finally, let us suppose that $l(\alpha) = n + 1$, where $\alpha = (j_1, \dots, j_{n+1})$ and $j_{n+1} = -1$. The process

$$\left\{ \tilde{I}_\alpha[g(\cdot)]_{\varrho,t}, t \in [\varrho, T] \right\} \quad (4.5.17)$$

is again a martingale. Therefore, by applying Doob's inequality and the isometry for jump martingales, we obtain

$$\begin{aligned}
\tilde{F}_\tau^\alpha &= E \left(\sup_{\varrho \leq s \leq \tau} \left| \int_\varrho^s \int_{\mathcal{E}} \tilde{I}_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{\varrho, z-} \tilde{p}_\varphi(dv^{s(\alpha)}, dz) \right|^2 \middle| \mathcal{A}_\varrho \right) \\
&\leq 4 E \left(\left| \int_\varrho^\tau \int_{\mathcal{E}} \tilde{I}_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{\varrho, z-} \tilde{p}_\varphi(dv^{s(\alpha)}, dz) \right|^2 \middle| \mathcal{A}_\varrho \right) \\
&= 4 E \left(\int_\varrho^\tau \int_{\mathcal{E}} |\tilde{I}_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{\varrho, z}|^2 \varphi(dv^{s(\alpha)}) dz \middle| \mathcal{A}_\varrho \right) \\
&\leq 4 E \left(\int_\varrho^\tau \int_{\mathcal{E}} \sup_{\varrho \leq s \leq \tau} |\tilde{I}_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{\varrho, s}|^2 \varphi(dv^{s(\alpha)}) dz \middle| \mathcal{A}_\varrho \right) \\
&= 4 E \left((\tau - \varrho) \int_{\mathcal{E}} \sup_{\varrho \leq s \leq \tau} |\tilde{I}_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{\varrho, s}|^2 \varphi(dv^{s(\alpha)}) \middle| \mathcal{A}_\varrho \right) \\
&\leq 4 \Delta \int_{\mathcal{E}} E \left(\sup_{\varrho \leq s \leq \tau} |\tilde{I}_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{\varrho, s}|^2 \middle| \mathcal{A}_\varrho \right) \varphi(dv^{s(\alpha)}). \quad (4.5.18)
\end{aligned}$$

By the inductive hypothesis we have

$$\begin{aligned}
\tilde{F}_\tau^\alpha &\leq 4 \Delta 4^{l(\alpha-) - n(\alpha-)} \Delta^{l(\alpha-) + n(\alpha-) - 1} \int_{\mathcal{E}} \int_\varrho^\tau V_{\varrho, z, s(\alpha-)} dz \varphi(dv^{s(\alpha)}) \\
&= 4^{l(\alpha) - n(\alpha)} \Delta^{l(\alpha) + n(\alpha) - 1} \int_\varrho^\tau V_{\varrho, z, s(\alpha)} dz, \quad (4.5.19)
\end{aligned}$$

since $l(\alpha) = l(\alpha-) + 1$, $n(\alpha) = n(\alpha-)$ and $s(\alpha) = s(\alpha-) + 1$, which completes the proof of assertion (4.5.3).

To prove assertion (4.5.4) we can proceed in a similar way by induction with respect to $l(\alpha)$. The idea, which will be used also in other lemmas, is to rewrite the Poisson jump measure as a sum of the compensated Poisson measure and a time integral as follows from (4.5.1).

Let us consider the case of $l(\alpha) = 1$ with $\alpha = (-1)$. By (4.5.1) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
F_\tau^{(-1)} &= E \left(\sup_{\varrho \leq s \leq \tau} \left| \int_\varrho^s \int_{\mathcal{E}} g(z-, v) p_\varphi(dv, dz) \right|^2 \middle| \mathcal{A}_\varrho \right) \\
&= E \left(\sup_{\varrho \leq s \leq \tau} \left| \int_\varrho^s \int_{\mathcal{E}} g(z-, v) \tilde{p}_\varphi(dv, dz) + \int_\varrho^s \int_{\mathcal{E}} g(z, v) \varphi(dv) dz \right|^2 \middle| \mathcal{A}_\varrho \right)
\end{aligned}$$

$$\begin{aligned} &\leq 2 E \left(\sup_{\varrho \leq s \leq \tau} \left| \tilde{I}_{(-1)}[g(\cdot)]_{\varrho, \tau} \right|^2 \middle| \mathcal{A}_\varrho \right) \\ &\quad + 2 E \left(\sup_{\varrho \leq s \leq \tau} \left| \tilde{I}_{(0)} \left[\int_{\mathcal{E}} g(\cdot, v) \varphi(dv) \right]_{\varrho, \tau} \right|^2 \middle| \mathcal{A}_\varrho \right). \end{aligned} \quad (4.5.20)$$

By applying the estimate of the already obtained assertion (4.5.3) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} F_\tau^{(-1)} &\leq 8 \int_\varrho^\tau \int_{\mathcal{E}} E \left(\sup_{\varrho \leq t \leq z} |g(t, v)|^2 \middle| \mathcal{A}_\varrho \right) \varphi(dv) dz \\ &\quad + 2 \Delta \int_\varrho^\tau E \left(\sup_{\varrho \leq t \leq z} \left| \int_{\mathcal{E}} g(t, v) \varphi(dv) \right|^2 \middle| \mathcal{A}_\varrho \right) dz \\ &\leq 4 \hat{K} \int_\varrho^\tau V_{\varrho, z, s(\alpha)} dz \\ &= 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \hat{K}^{s(\alpha)} \int_\varrho^\tau V_{\varrho, z, s(\alpha)} dz, \end{aligned} \quad (4.5.21)$$

with $\hat{K} = \frac{1}{2}(4 + T\lambda)$, $l(\alpha) = 1$ and $\alpha = (-1)$ so that $n(\alpha) = 0$.

To finalize the proof we consider the case of $l(\alpha) = n + 1$ where $\alpha = (j_1, \dots, j_{n+1})$ with $j_{n+1} = -1$. By (4.5.1), the Cauchy-Schwarz inequality, Doob's inequality, the Itô isometry and similar steps as those used in (4.5.12) and (4.5.18), one obtains

$$\begin{aligned} F_\tau^\alpha &= E \left(\sup_{\varrho \leq s \leq \tau} \left| \int_\varrho^s \int_{\mathcal{E}} I_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{\varrho, z-} p_\varphi(dv^{s(\alpha)}, dz) \right|^2 \middle| \mathcal{A}_\varrho \right) \\ &\leq 2 E \left(\sup_{\varrho \leq s \leq \tau} \left| \int_\varrho^s \int_{\mathcal{E}} I_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{\varrho, z-} \tilde{p}_\varphi(dv^{s(\alpha)}, dz) \right|^2 \middle| \mathcal{A}_\varrho \right) \\ &\quad + 2 E \left(\sup_{\varrho \leq s \leq \tau} \left| \int_\varrho^s \int_{\mathcal{E}} I_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{\varrho, z-} \varphi(dv^{s(\alpha)}) dz \right|^2 \middle| \mathcal{A}_\varrho \right) \\ &\leq 8 \Delta \int_{\mathcal{E}} E \left(\sup_{\varrho \leq s \leq \tau} |I_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{\varrho, s}|^2 \middle| \mathcal{A}_\varrho \right) \varphi(dv^{s(\alpha)}) \\ &\quad + 2 \lambda \Delta^2 \int_{\mathcal{E}} E \left(\sup_{\varrho \leq s \leq \tau} |I_{\alpha-}[g(\cdot)]_{\varrho, s}|^2 \middle| \mathcal{A}_\varrho \right) \varphi(dv^{s(\alpha)}). \end{aligned} \quad (4.5.22)$$

By the induction hypothesis, we finally obtain

$$\begin{aligned}
F_\tau^\alpha &\leq 8 \Delta 4^{l(\alpha)-n(\alpha-)} \Delta^{l(\alpha-)+n(\alpha-)-1} \hat{K}^{s(\alpha-)} \int_{\mathcal{E}} \int_{\varrho}^{\tau} V_{\varrho, z, s(\alpha-)} dz \varphi(dv^{s(\alpha)}) \\
&+ 2 \lambda \Delta T 4^{l(\alpha)-n(\alpha-)} \Delta^{l(\alpha-)+n(\alpha-)-1} \hat{K}^{s(\alpha-)} \int_{\mathcal{E}} \int_{\varrho}^{\tau} V_{\varrho, z, s(\alpha-)} dz \varphi(dv^{s(\alpha)}) \\
&\leq 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \hat{K}^{s(\alpha)} \int_{\varrho}^{\tau} V_{\varrho, z, s(\alpha)} dz,
\end{aligned} \tag{4.5.23}$$

which completes the proof of Lemma 4.5.2. \square

Higher Moments of Multiple Integrals

The following lemma provides an estimate for higher moments of multiple stochastic integrals. A similar result is presented in Li & Liu (2000).

Lemma 4.5.3 *Let $\alpha \in \mathcal{M}_m \setminus \{v\}$, $g \in \mathcal{H}_\alpha$, and ϱ and τ denote two stopping times with τ being \mathcal{A}_ϱ -measurable and $0 \leq \varrho \leq \tau \leq T$ almost surely. Then for any $q \in \{1, 2, \dots\}$ there exist positive constants C_1 and C_2 independent of $(\tau - \varrho)$ such that*

$$\begin{aligned}
\tilde{F}_\tau^\alpha &= E \left(\left| \tilde{I}_\alpha[g(\cdot)]_{\varrho, \tau} \right|^{2q} \middle| \mathcal{A}_\varrho \right) \\
&\leq C_1 (\tau - \varrho)^{q(l(\alpha)+n(\alpha)-s(\alpha))+s(\alpha)-1} \int_{\varrho}^{\tau} V_{\varrho, z, s(\alpha)} dz,
\end{aligned} \tag{4.5.24}$$

and

$$\begin{aligned}
F_\tau^\alpha &= E \left(\left| I_\alpha[g(\cdot)]_{\varrho, \tau} \right|^{2q} \middle| \mathcal{A}_\varrho \right) \\
&\leq C_2 (\tau - \varrho)^{q(l(\alpha)+n(\alpha)-s(\alpha))+s(\alpha)-1} \int_{\varrho}^{\tau} V_{\varrho, z, s(\alpha)} dz,
\end{aligned} \tag{4.5.25}$$

where

$$V_{\varrho, z, s(\alpha)} = \int_{\mathcal{E}} \dots \int_{\mathcal{E}} E \left(|g(z, v_1, \dots, v_{s(\alpha)})|^{2q} \middle| \mathcal{A}_\varrho \right) \varphi(dv_1) \dots \varphi(dv_{s(\alpha)}) < \infty$$

for $z \in [\varrho, \tau]$.

Proof: We will first prove the estimate (4.5.24) by induction on $l(\alpha)$.

1. Let us first prove the assertion for a multi-index α of length one, which means $l(\alpha) = 1$. When $l(\alpha) = 1$ and $\alpha = (0)$, by applying the Hölder inequality we obtain

$$\begin{aligned}
\tilde{F}_\tau^{(0)} &= E \left(\left| \int_\varrho^\tau g(z) dz \right|^{2q} \middle| \mathcal{A}_\varrho \right) \\
&\leq (\tau - \varrho)^{2q-1} \int_\varrho^\tau E \left(|g(z)|^{2q} \middle| \mathcal{A}_\varrho \right) dz \\
&= C(\tau - \varrho)^q (l(\alpha) + n(\alpha) - s(\alpha)) + s(\alpha) - 1 \int_\varrho^\tau E \left(|g(z)|^{2q} \middle| \mathcal{A}_\varrho \right) dz.
\end{aligned} \tag{4.5.26}$$

Note that here and in the following we denote by C any finite positive constant independent of $(\tau - \varrho)$.

2. For $l(\alpha) = 1$ and $\alpha = (j)$, with $j \in \{1, 2, \dots, m\}$, we obtain, see [Krylov \(1980\)](#), Corollary 3, page 80,

$$\begin{aligned}
\tilde{F}_\tau^{(j)} &= E \left(\left| \int_\varrho^\tau g(z) dW_z^j \right|^{2q} \middle| \mathcal{A}_\varrho \right) \\
&\leq 2^q (2q-1)^q (\tau - \varrho)^{q-1} \int_\varrho^\tau E \left(|g(z)|^{2q} \middle| \mathcal{A}_\varrho \right) dz \\
&= C(\tau - \varrho)^q (l(\alpha) + n(\alpha) - s(\alpha)) + s(\alpha) - 1 \int_\varrho^\tau E \left(|g(z)|^{2q} \middle| \mathcal{A}_\varrho \right) dz.
\end{aligned} \tag{4.5.27}$$

3. For $l(\alpha) = 1$ and $\alpha = (-1)$, let us define

$$x_\tau = \int_\varrho^\tau \int_{\mathcal{E}} g(z-, v) p_\varphi(dv, dz).$$

By applying Itô's formula to $|x_\tau|^{2q}$ together with the Hölder inequality, we obtain

$$\begin{aligned}
|x_\tau|^{2q} &= \int_\varrho^\tau \int_{\mathcal{E}} \left(|x_{z-} + g(z-, v)|^{2q} - |x_{z-}|^{2q} \right) p_\varphi(dv, dz) \\
&\leq (2^{2q-1} - 1) \int_\varrho^\tau \int_{\mathcal{E}} |x_{z-}|^{2q} p_\varphi(dv, dz) \\
&\quad + 2^{2q-1} \int_\varrho^\tau \int_{\mathcal{E}} |g(z-, v)|^{2q} p_\varphi(dv, dz).
\end{aligned} \tag{4.5.28}$$

Therefore, by the properties of the Poisson jump measure, we have

$$\begin{aligned}
E(|x_\tau|^{2q} | \mathcal{A}_\varrho) &\leq (2^{2q-1} - 1) \lambda \int_\varrho^\tau E(|x_z|^{2q} | \mathcal{A}_\varrho) dz \\
&\quad + 2^{2q-1} \int_\varrho^\tau \int_{\mathcal{E}} E(|g(z, v)|^{2q} | \mathcal{A}_\varrho) \varphi(dv) dz,
\end{aligned} \tag{4.5.29}$$

where we recall that $\lambda = \varphi(\mathcal{E})$ is the total intensity. Moreover, let us note that, by the first line of (4.5.28) and the properties of the Poisson jump measure, we obtain

$$E(|x_\tau|^{2q}|\mathcal{A}_\varrho) = \int_\varrho^\tau \int_{\mathcal{E}} E\left((|x_z + g(z, v)|^{2q} - |x_z|^{2q}) \Big| \mathcal{A}_\varrho\right) \varphi(dv) dz,$$

which proves the continuity of $E(|x_\tau|^{2q}|\mathcal{A}_\varrho)$ as a function of τ . Therefore, by applying the Gronwall inequality (1.2.33) to (4.5.29), we obtain

$$\begin{aligned} E(|x_\tau|^{2q}|\mathcal{A}_\varrho) &= E\left(\left|\int_\varrho^\tau \int_{\mathcal{E}} g(z-, v)p_\varphi(dv, dz)\right|^{2q} \Big| \mathcal{A}_\varrho\right) \\ &\leq \exp\{(2^{2q-1} - 1) \lambda (\tau - \varrho)\} 2^{2q-1} \\ &\quad \times \int_\varrho^\tau \int_{\mathcal{E}} E(|g(z, v)|^{2q}|\mathcal{A}_\varrho) \varphi(dv) dz \\ &\leq \exp\{(2^{2q-1} - 1) \lambda T\} 2^{2q-1} \\ &\quad \times \int_\varrho^\tau \int_{\mathcal{E}} E(|g(z, v)|^{2q}|\mathcal{A}_\varrho) \varphi(dv) dz \\ &= C(\tau - \varrho)^{q(l(\alpha)+n(\alpha)-s(\alpha))+s(\alpha)-1} \\ &\quad \times \int_\varrho^\tau \int_{\mathcal{E}} E\left(|g(z, v_{s(\alpha)})|^{2q} \Big| \mathcal{A}_\varrho\right) \varphi(dv_{s(\alpha)}) dz. \end{aligned} \tag{4.5.30}$$

Finally, by (4.2.1), the Hölder inequality, (4.5.30) and (4.5.26), we obtain

$$\begin{aligned} \tilde{F}_\tau^{(-1)} &= E\left(\left|\int_\varrho^\tau \int_{\mathcal{E}} g(z-, v)p_\varphi(dv, dz) - \int_\varrho^\tau \int_{\mathcal{E}} g(z, v)\varphi(dv) dz\right|^{2q} \Big| \mathcal{A}_\varrho\right) \\ &\leq 2^{2q-1} \left\{ E\left(\left|\int_\varrho^\tau \int_{\mathcal{E}} g(z-, v)p_\varphi(dv, dz)\right|^{2q} \Big| \mathcal{A}_\varrho\right) \right. \\ &\quad \left. + E\left(\left|\int_\varrho^\tau \int_{\mathcal{E}} g(z, v)\varphi(dv) dz\right|^{2q} \Big| \mathcal{A}_\varrho\right) \right\} \\ &\leq C(\tau - \varrho)^{q(l(\alpha)+n(\alpha)-s(\alpha))+s(\alpha)-1} \\ &\quad \times \left\{ \int_\varrho^\tau \int_{\mathcal{E}} E\left(|g(z, v)|^{2q} \Big| \mathcal{A}_\varrho\right) \varphi(dv) dz \right. \\ &\quad \left. + \int_\varrho^\tau E\left(\left|\int_{\mathcal{E}} g(z, v)\varphi(dv)\right|^{2q} \Big| \mathcal{A}_\varrho\right) dz \right\} \\ &\leq C(\tau - \varrho)^{q(l(\alpha)+n(\alpha)-s(\alpha))+s(\alpha)-1} \\ &\quad \times \int_\varrho^\tau \int_{\mathcal{E}} E\left(|g(z, v_{s(\alpha)})|^{2q} \Big| \mathcal{A}_\varrho\right) \varphi(dv_{s(\alpha)}) dz. \end{aligned}$$

4. Let us now consider the case $l(\alpha) = n + 1$ and $\alpha = (j_1, \dots, j_{n+1})$, with $j_{n+1} = 0$. By (4.5.26) and the inductive hypothesis, we get

$$\begin{aligned}
\tilde{F}_\tau^\alpha &= E \left(\left| \int_\varrho^\tau \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,s} ds \right|^{2q} \middle| \mathcal{A}_\varrho \right) \\
&\leq (\tau - \varrho)^{2q-1} \int_\varrho^\tau E \left(|\tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,s}|^{2q} \middle| \mathcal{A}_\varrho \right) ds \\
&\leq C (\tau - \varrho)^{2q-1} (\tau - \varrho)^{q(l(\alpha-) + n(\alpha-) - s(\alpha-)) + s(\alpha-) - 1} \\
&\quad \times \int_\varrho^\tau \int_\varrho^s V_{\varrho,z,s(\alpha-)} dz ds \\
&= C (\tau - \varrho)^{q(l(\alpha) + n(\alpha) - s(\alpha)) + s(\alpha) - 1} \int_\varrho^\tau V_{\varrho,z,s(\alpha)} dz, \quad (4.5.31)
\end{aligned}$$

where the last line holds since $l(\alpha) = l(\alpha-) + 1$, $n(\alpha) = n(\alpha-) + 1$ and $s(\alpha) = s(\alpha-)$.

5. When $l(\alpha) = n + 1$ and $\alpha = (j_1, \dots, j_{n+1})$, with $j_{n+1} \in \{1, 2, \dots, m\}$, by (4.5.27) and the inductive hypothesis, we obtain

$$\begin{aligned}
\tilde{F}_\tau^\alpha &= E \left(\left| \int_\varrho^\tau \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,s} dW_s^{j_{n+1}} \right|^{2q} \middle| \mathcal{A}_\varrho \right) \\
&\leq 2^q (2q-1)^q (\tau - \varrho)^{q-1} \int_\varrho^\tau E \left(|\tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,s}|^{2q} \middle| \mathcal{A}_\varrho \right) ds \\
&\leq C (\tau - \varrho)^{q-1} (\tau - \varrho)^{q(l(\alpha-) + n(\alpha-) - s(\alpha-)) + s(\alpha-) - 1} \\
&\quad \times \int_\varrho^\tau \int_\varrho^s V_{\varrho,z,s(\alpha-)} dz ds \\
&= C (\tau - \varrho)^{q(l(\alpha) + n(\alpha) - s(\alpha)) + s(\alpha) - 1} \int_\varrho^\tau V_{\varrho,z,s(\alpha)} dz,
\end{aligned}$$

where the last line holds since $l(\alpha) = l(\alpha-) + 1$, $n(\alpha) = n(\alpha-)$ and $s(\alpha) = s(\alpha-)$.

6. Finally, let us suppose that $l(\alpha) = n + 1$ and $\alpha = (j_1, \dots, j_{n+1})$, with $j_{n+1} = -1$. By (4.5.30) and the inductive hypothesis, we obtain

$$\begin{aligned}
\tilde{F}_\tau^\alpha &= E \left(\left| \int_\varrho^\tau \int_{\mathcal{E}} \tilde{I}_{\alpha-} [g(\cdot, v_{s(\alpha)})]_{\varrho, s-} \tilde{p}_\varphi(dv_{s(\alpha)}, ds) \right|^{2q} \middle| \mathcal{A}_\varrho \right) \\
&\leq C \int_\varrho^\tau \int_{\mathcal{E}} E \left(|\tilde{I}_{\alpha-} [g(\cdot, v_{s(\alpha)})]_{\varrho, s-}|^{2q} \middle| \mathcal{A}_\varrho \right) \varphi(dv_{s(\alpha)}) ds \\
&\leq C (\tau - \varrho)^{q(l(\alpha-) + n(\alpha-) - s(\alpha-)) + s(\alpha-) - 1} \\
&\quad \times \int_\varrho^\tau \int_\varrho^s \int_{\mathcal{E}} V_{\varrho, z, s(\alpha-)} \varphi(dv_{s(\alpha)}) dz ds \\
&\leq C (\tau - \varrho)^{q(l(\alpha) + n(\alpha) - s(\alpha)) + s(\alpha) - 1} \int_\varrho^\tau V_{\varrho, z, s(\alpha)} dz, \tag{4.5.32}
\end{aligned}$$

where the last line holds since $l(\alpha) = l(\alpha-) + 1$, $n(\alpha) = n(\alpha-)$ and $s(\alpha) = s(\alpha-) + 1$, and this completes the proof of the assertion (4.5.24).

The assertion (4.5.25) can be shown similarly by induction on $l(\alpha)$. The case of $l(\alpha) = 1$ with $\alpha = (-1)$ has already been shown in (4.5.30). The case of $l(\alpha) = n+1$, with $\alpha = (j_1, \dots, j_{n+1})$ and $j_{n+1} = -1$, can be derived by using (4.2.1), the Hölder inequality, the inductive hypothesis, (4.5.31) and (4.5.32). This completes the proof of Lemma 4.5.3. \square

Expectations of Products

Let us introduce some notation needed for the following lemma. For any $z \in \mathbb{R}$, let us denote by $[z]$ the integer part of z . Moreover, for a given integer $p \in \mathcal{N}$ we will use the set \mathcal{A}_p of multi-indices $\alpha = (j_1, \dots, j_l)$ of length $l \leq p$ with components $j_i \in \{-1, 0\}$, for $i \in \{1, \dots, l\}$. For a given function $h : [\varrho, \tau] \times \mathbb{R}^d \rightarrow \mathbb{R}$, to be defined in the lemma below, and a multi-index $\alpha \in \mathcal{A}_p$, we consider the coefficient function \tilde{f}_α defined in (4.3.9), with $f(t, x) = h(t, x)$. For instance, if $\alpha = (-1, 0, 0)$, then $\tilde{f}_\alpha(t, X_t) = L_v^{-1} \tilde{L}^0 h(t, X_t)$.

Furthermore, for $p \in \mathcal{N}$, we denote by $\mathcal{C}^{p,2p}([\varrho, \tau] \times \mathbb{R}^d, \mathbb{R})$ the set of functions that are p times continuously differentiable with respect to time and $2p$ times continuously differentiable with respect to the spatial variables.

By using Lemma 4.5.3 we prove the following result similar to that in Liu & Li (2000).

Lemma 4.5.4 Consider $\alpha \in \mathcal{M}_m$, $p = l(\alpha) - \lfloor \frac{l(\alpha)+n(\alpha)}{2} \rfloor$, and let ϱ and τ be two stopping times with τ being \mathcal{A}_ϱ -measurable and $0 \leq \varrho \leq \tau \leq T$ almost surely. Moreover, define the process $Z = \{Z_t = h(t, \mathbf{X}_t), t \in [\varrho, \tau]\}$, with $h \in \mathcal{C}^{p,2p}([\varrho, \tau] \times \mathbb{R}^d, \mathbb{R})$. Assume that there exist a finite positive constant K such that for any $\alpha \in \mathcal{A}_p$ we have the estimate $E(\tilde{f}_\alpha(t, \mathbf{X}_t)^2 | \mathcal{A}_\varrho) \leq K$ a.s. for $t \in [\varrho, \tau]$. Additionally, consider an adapted process $g(\cdot) \in \mathcal{H}_\alpha$, where

$g = g(\cdot, v)$ with $v \in \mathcal{E}^{s(\alpha)}$. If there exists a positive, $\varphi(dv)$ -integrable function $K(v)$ such that $E(g(t, v)^2 | \mathcal{A}_\varrho) < K(v)$ a.s. for $t \in [\varrho, \tau]$, then we obtain

$$\left| E\left(Z_\tau \tilde{I}_\alpha[g(\cdot)]_{\varrho, \tau} | \mathcal{A}_\varrho \right) \right| \leq C_1 (\tau - \varrho)^{l(\alpha)}, \quad (4.5.33)$$

and

$$\left| E\left(Z_\tau I_\alpha[g(\cdot)]_{\varrho, \tau} | \mathcal{A}_\varrho \right) \right| \leq C_2 (\tau - \varrho)^{l(\alpha)}, \quad (4.5.34)$$

where the positive constants C_1 and C_2 do not depend on $(\tau - \varrho)$.

Proof: Let us first prove the assertion (4.5.33) by induction on $l(\alpha)$. For $l(\alpha) = 0$ we can prove (4.5.33) by applying the Cauchy-Schwarz inequality

$$\left| E\left(Z_\tau g(\tau) | \mathcal{A}_\varrho \right) \right| \leq \sqrt{E\left(Z_\tau^2 | \mathcal{A}_\varrho \right)} \sqrt{E\left(g(\tau)^2 | \mathcal{A}_\varrho \right)} \leq C,$$

where we denote by C any constant that does not depend on $(\tau - \varrho)$. However, it may depend on α, λ, T, h and g .

Now consider the case $l(\alpha) = n + 1$. By Itô's formula we have

$$\begin{aligned} Z_\tau &= Z_\varrho + \int_\varrho^\tau L^0 h(z, \mathbf{X}_z) dz + \sum_{i=1}^m \int_\varrho^\tau L^i h(z, \mathbf{X}_z) dW_z^i \\ &\quad + \int_\varrho^\tau \int_{\mathcal{E}} L_v^{-1} h(z, \mathbf{X}_{z-}) p_\varphi(dv, dz), \end{aligned}$$

where the operators L^i , with $i \in \{0, 1, \dots, m\}$, and L_v^{-1} are defined in (4.3.4)–(4.3.6).

1. If $\alpha = (j_1, \dots, j_{l+1})$ with $j_{l+1} = -1$, we can write

$$\tilde{I}_\alpha[g(\cdot)]_{\varrho, \tau} = \int_\varrho^\tau \int_{\mathcal{E}} \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho, z-} p_\varphi(dv, dz) - \int_\varrho^\tau \int_{\mathcal{E}} \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho, z} \varphi(dv) dz.$$

By the product rule of stochastic calculus we obtain

$$\begin{aligned} Z_\tau \tilde{I}_\alpha[g(\cdot)]_{\varrho, \tau} &= \int_\varrho^\tau \left\{ (L^0 h(z, X_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho, z} \right. \\ &\quad \left. - h(z, X_z) \int_{\mathcal{E}} \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho, z} \varphi(dv) \right\} dz \\ &\quad + \sum_{i=1}^m \int_\varrho^\tau (L^i h(z, \mathbf{X}_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho, z} dW_z^i \\ &\quad + \int_\varrho^\tau \int_{\mathcal{E}} \left\{ h(z, \mathbf{X}_{z-}) \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho, z-} \right. \\ &\quad \left. + (L_v^{-1} h(z, \mathbf{X}_{z-})) \tilde{I}_\alpha[g(\cdot)]_{\varrho, z-} \right. \\ &\quad \left. + (L_v^{-1} h(z, \mathbf{X}_{z-})) \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho, z-} \right\} p_\varphi(dv, dz). \end{aligned}$$

Therefore, by the properties of Itô's integral and the induction hypothesis we have

$$\begin{aligned}
\left| E\left(Z_\tau \tilde{I}_\alpha[g(\cdot)]_{\varrho, \tau} | \mathcal{A}_\varrho \right) \right| &= \left| - \int_\varrho^\tau E\left(h(z, \mathbf{X}_z) \int_{\mathcal{E}} \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho, z} \varphi(dv) | \mathcal{A}_\varrho \right) dz \right. \\
&\quad + \int_\varrho^\tau \int_{\mathcal{E}} E\left(\{h(z, \mathbf{X}_z) + L_v^{-1} h(z, \mathbf{X}_z)\} \right. \\
&\quad \times \left. \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho, z} | \mathcal{A}_\varrho \right) \varphi(dv) dz \\
&\quad + \int_\varrho^\tau E\left((L^0 h(z, \mathbf{X}_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho, z} | \mathcal{A}_\varrho \right) dz \\
&\quad \left. + \int_\varrho^\tau \int_{\mathcal{E}} E\left((L_v^{-1} h(z, \mathbf{X}_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho, z} | \mathcal{A}_\varrho \right) \varphi(dv) dz \right| \\
&\leq C (\tau - \varrho)^{n+1} \\
&\quad + \int_\varrho^\tau \left| E\left((\tilde{L}^0 h(z, \mathbf{X}_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho, z} | \mathcal{A}_\varrho \right) \right| dz, \tag{4.5.35}
\end{aligned}$$

where we have used the equality (4.3.7).

Note that $\tilde{L}^0 h \in \mathcal{C}^{p-1, 2(p-1)}([\varrho, \tau] \times \mathbb{R}^d, \mathbb{R})$ and we have the estimate

$$E\left((\tilde{L}^0(h(t, \mathbf{X}_t)))^2 | \mathcal{A}_\varrho \right) \leq K$$

for $t \in [\varrho, \tau]$. Therefore, by the same steps used so far, we can show that

$$\begin{aligned}
&\left| E\left((\tilde{L}^0 h(\tau, \mathbf{X}_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho, z} | \mathcal{A}_\varrho \right) \right| \tag{4.5.36} \\
&\leq C (z - \varrho)^{n+1} + \int_\varrho^z \left| E\left((\tilde{L}^0(\tilde{L}^0 h(z_1, \mathbf{X}_{z_1}))) \tilde{I}_\alpha[g(\cdot)]_{\varrho, z_1} | \mathcal{A}_\varrho \right) \right| dz_1,
\end{aligned}$$

for $z \in [\varrho, \tau]$.

By using (4.5.36) in (4.5.35), we obtain

$$\begin{aligned}
&\left| E\left(Z_\tau \tilde{I}_\alpha[g(\cdot)]_{\varrho, \tau} | \mathcal{A}_\varrho \right) \right| \leq C (\tau - \varrho)^{n+1} \tag{4.5.37} \\
&\quad + \int_\varrho^\tau \int_\varrho^{z_2} \left| E\left((\tilde{L}^0(\tilde{L}^0 h(z_1, \mathbf{X}_{z_1}))) \tilde{I}_\alpha[g(\cdot)]_{\varrho, z_1} | \mathcal{A}_\varrho \right) \right| dz_1 dz_2.
\end{aligned}$$

By applying again this procedure $p - 2$ times and using the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}
& \left| E\left(Z_\tau \tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau} | \mathcal{A}_\varrho \right) \right| \\
& \leq C (\tau - \varrho)^{n+1} \\
& \quad + \int_\varrho^\tau \dots \int_\varrho^{z_2} \left| E\left(\tilde{f}_{\alpha_p}(z_1, \mathbf{X}_{z_1}) \tilde{I}_\alpha[g(\cdot)]_{\varrho,z_1} | \mathcal{A}_\varrho \right) \right| dz_1 \dots dz_p \\
& \leq C (\tau - \varrho)^{n+1} \\
& \quad + \int_\varrho^\tau \dots \int_\varrho^{z_2} \left[E\left(\tilde{f}_{\alpha_p}(z_1, \mathbf{X}_{z_1})^2 | \mathcal{A}_\varrho \right) \right]^{\frac{1}{2}} \\
& \quad \times \left[E\left((\tilde{I}_\alpha[g(\cdot)]_{\varrho,z_1})^2 | \mathcal{A}_\varrho \right) \right]^{\frac{1}{2}} dz_1 \dots dz_p, \tag{4.5.38}
\end{aligned}$$

where α_p is defined as the multi-index with length p and all zeros, which means $l(\alpha_p) = n(\alpha_p) = p$. Therefore, $\tilde{f}_{\alpha_p}(t, \mathbf{X}_t)$ denotes the stochastic process resulting from applying p times the operator L^0 , see (4.3.7), to $h(t, \mathbf{X}_t)$. Note that C denotes here a different constant from that in (4.5.35).

Finally, by applying the result (4.5.24) of Lemma 4.5.3, with $q = 1$, to the last term in (4.5.38) and considering that one has the estimates

$E(\tilde{f}_{\alpha_p}(t, \mathbf{X}_t)^2 | \mathcal{A}_\varrho) \leq K < \infty$ and $E(g(t, v)^2 | \mathcal{A}_\varrho) \leq K(v)$ a.s. for $t \in [\varrho, \tau]$, with $K(v)$ being $\varphi(dv)$ -integrable, we obtain

$$\begin{aligned}
& \left| E\left(Z_\tau \tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau} | \mathcal{A}_\varrho \right) \right| \leq C (\tau - \varrho)^{n+1} \\
& \quad + K \int_\varrho^\tau \dots \int_\varrho^{z_2} (\tau - \varrho)^{\frac{l(\alpha)+n(\alpha)}{2}} dz_1 \dots dz_p \\
& \leq C (\tau - \varrho)^{l(\alpha)},
\end{aligned}$$

which proves the case of $\alpha = (j_1, \dots, j_{l+1})$ with $j_{l+1} = -1$.

2. If $\alpha = (j_1, \dots, j_{l+1})$ with $j_{l+1} = 0$, we can write

$$\tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau} = \int_\varrho^\tau \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,z} dz.$$

The product rule for stochastic integrals yields

$$\begin{aligned}
Z_\tau \tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau} &= \int_\varrho^\tau \left\{ (L^0 h(z, \mathbf{X}_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho,z} + h(z, \mathbf{X}_z) \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,z} \right\} dz \\
&\quad + \sum_{i=1}^m \int_\varrho^\tau (L^i h(z, \mathbf{X}_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho,z} dW_z^i \\
&\quad + \int_\varrho^\tau \int_{\mathcal{E}} \left\{ (L_v^{-1} h(z, \mathbf{X}_{z-})) \tilde{I}_\alpha[g(\cdot)]_{\varrho,z-} \right\} p_\varphi(dv, dz).
\end{aligned}$$

Therefore, similar to (4.5.35), by the properties of Itô's integral and the induction hypothesis one obtains

$$\begin{aligned} \left| E\left(Z_\tau \tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau} | \mathcal{A}_\varrho\right) \right| &\leq C (\tau - \varrho)^{n+1} \\ &+ \int_\varrho^\tau \left| E\left((L^0 h(z, \mathbf{X}_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho,z} | \mathcal{A}_\varrho\right) \right| dz. \end{aligned}$$

Then, by applying again this procedure $p - 1$ times, and by using the Cauchy-Schwarz inequality and the same estimates as before, we have

$$\begin{aligned} \left| E\left(Z_\tau \tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau} | \mathcal{A}_\varrho\right) \right| &\leq C (\tau - \varrho)^{n+1} + \int_\varrho^\tau \dots \int_\varrho^{z_2} \left[E\left(\tilde{f}_{\bar{\alpha}_p}(z_1, \mathbf{X}_{z_1})^2 | \mathcal{A}_\varrho\right) \right]^{\frac{1}{2}} \\ &\times \left[E\left((\tilde{I}_\alpha[g(\cdot)]_{\varrho,z_1})^2 | \mathcal{A}_\varrho\right) \right]^{\frac{1}{2}} dz_1 \dots dz_p \\ &\leq C (\tau - \varrho)^{l(\alpha)}, \end{aligned}$$

where $\tilde{f}_{\bar{\alpha}_p}(t, \mathbf{X}_t)$ denotes the stochastic process resulting from applying p times the operator L^0 , see (4.3.4), to $h(t, \mathbf{X}_t)$.

3. Let us finally consider the case $\alpha = (j_1, \dots, j_{l+1})$ with $j_{l+1} = (j)$ and $j \in \{1, 2, \dots, m\}$. Here, we have

$$\tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau} = \int_\varrho^\tau \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,z} dW_z^j.$$

Therefore, by applying the product rule for stochastic integrals we obtain

$$\begin{aligned} Z_\tau \tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau} &= \int_\varrho^\tau \left\{ (L^0 h(z, \mathbf{X}_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho,z} + (L^j h(z, \mathbf{X}_z)) \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,z} \right\} dz \\ &+ \sum_{i=1}^m \int_\varrho^\tau (L^i h(z, \mathbf{X}_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho,z} dW_z^i \\ &+ \int_\varrho^\tau h(z, \mathbf{X}_z) \tilde{I}_{\alpha-}[g(\cdot)]_{\varrho,z} dW_z^j \\ &+ \int_\varrho^\tau \int_{\mathcal{E}} \left\{ (L_v^{-1} h(z, \mathbf{X}_{z-})) \tilde{I}_\alpha[g(\cdot)]_{\varrho,z-} \right\} p_\varphi(dv, dz). \end{aligned}$$

Again, by the properties of the Itô integral and the inductive hypothesis we have

$$\begin{aligned} \left| E\left(Z_\tau \tilde{I}_\alpha[g(\cdot)]_{\varrho,\tau} | \mathcal{A}_\varrho\right) \right| &\leq C (\tau - \varrho)^{n+1} \\ &+ \int_\varrho^\tau \left| E\left((L^0 h(z, \mathbf{X}_z)) \tilde{I}_\alpha[g(\cdot)]_{\varrho,z} | \mathcal{A}_\varrho\right) \right| dz, \end{aligned}$$

as in the previous case of $\alpha = (j_1, \dots, j_{l+1})$ with $j_{l+1} = 0$. Therefore, in the same way, we can obtain the assertion (4.5.33).

To prove the estimate (4.5.34) we need only to check the case of $l(\alpha) = n+1$ with $\alpha = (j_1, \dots, j_{n+1})$ and $j_{n+1} = -1$.

By using the product rule of stochastic calculus one obtains

$$\begin{aligned} Z_\tau I_\alpha[g(\cdot)]_{\varrho, \tau} &= \int_\varrho^\tau (L^0 h(z, \mathbf{X}_z)) I_\alpha[g(\cdot)]_{\varrho, z} dz \\ &\quad + \sum_{i=1}^m \int_\varrho^\tau (L^i h(z, \mathbf{X}_z)) I_\alpha[g(\cdot)]_{\varrho, z} dW_z^i \\ &\quad + \int_\varrho^\tau \int_{\mathcal{E}} \left\{ h(z, \mathbf{X}_{z-}) I_{\alpha-}[g(\cdot)]_{\varrho, z-} + (L_v^{-1} h(z, \mathbf{X}_{z-})) I_\alpha[g(\cdot)]_{\varrho, z-} \right. \\ &\quad \left. + (L_v^{-1} h(z, \mathbf{X}_{z-})) I_{\alpha-}[g(\cdot)]_{\varrho, z-} \right\} p_\varphi(dv, dz). \end{aligned}$$

By the properties of Itô integrals and the induction hypothesis we obtain

$$\begin{aligned} \left| E\left(Z_\tau I_\alpha[g(\cdot)]_{\varrho, \tau} \mid \mathcal{A}_\varrho \right) \right| &= \left| \int_\varrho^\tau \int_{\mathcal{E}} E\left(\{h(z, \mathbf{X}_z) + L_v^{-1} h(z, \mathbf{X}_z)\} \right. \right. \\ &\quad \times I_{\alpha-}[g(\cdot)]_{\varrho, z} \mid \mathcal{A}_\varrho \Big) \varphi(dv) dz \\ &\quad + \int_\varrho^\tau E\left((L^0 h(z, \mathbf{X}_z)) I_\alpha[g(\cdot)]_{\varrho, z} \mid \mathcal{A}_\varrho \right) dz \\ &\quad + \int_\varrho^\tau \int_{\mathcal{E}} E\left((L_v^{-1} h(z, \mathbf{X}_z)) I_\alpha[g(\cdot)]_{\varrho, z} \mid \mathcal{A}_\varrho \right) \varphi(dv) dz \Big| \\ &\leq C (\tau - \varrho)^{n+1} \\ &\quad + \int_\varrho^\tau \left| E\left((\tilde{L}^0 h(z, \mathbf{X}_z)) I_\alpha[g(\cdot)]_{\varrho, z} \mid \mathcal{A}_\varrho \right) \right| dz, \quad (4.5.39) \end{aligned}$$

where we have used again the equality (4.3.7).

From this point we can proceed in the same way as in the proof of assertion (4.5.33). This completes the proof of Lemma 4.5.4. \square

Expectations of Products with Powers

We also derive the following lemma, similar to another result in Liu & Li (2000).

Lemma 4.5.5 *Let $\alpha \in \mathcal{M}_m$, ϱ and τ denote two stopping times with τ being \mathcal{A}_ϱ -measurable and $0 \leq \varrho \leq \tau \leq T$ almost surely. Moreover, let $h = \{h(t), t \in [\varrho, \tau]\}$ be an adapted process such that $E(h(t)^2 \mid \mathcal{A}_\varrho) \leq K < \infty$ for $t \in [\varrho, \tau]$. Additionally, consider an adapted process $g(\cdot) \in \mathcal{H}_\alpha$, where*

$g = g(\cdot, v)$ with $v \in \mathcal{E}^{s(\alpha)}$. If for a given $q \in \{1, 2, \dots\}$ there exists a positive, $\varphi(dv)$ -integrable function $K(v)$ such that $E(g(t, v)^{2^{s(\alpha)+3}q} | \mathcal{A}_\varrho) < K(v)$ a.s. for $t \in [\varrho, \tau]$, then we obtain

$$\tilde{F}_\tau^\alpha = E \left(h(\tau) \left| \tilde{I}_\alpha[g(\cdot)]_{\varrho, \tau} \right|^{2q} \middle| \mathcal{A}_\varrho \right) \leq C_1 (\tau - \varrho)^{q(l(\alpha) + n(\alpha) - s(\alpha)) + s(\alpha)}, \quad (4.5.40)$$

and

$$F_\tau^\alpha = E \left(h(\tau) \left| I_\alpha[g(\cdot)]_{\varrho, \tau} \right|^{2q} \middle| \mathcal{A}_\varrho \right) \leq C_2 (\tau - \varrho)^{q(l(\alpha) + n(\alpha) - s(\alpha)) + s(\alpha)}, \quad (4.5.41)$$

where the positive constants C_1 and C_2 do not depend on $(\tau - \varrho)$.

Proof: We first prove the estimate (4.5.40) by induction with respect to $s(\alpha)$.

1. If $s(\alpha) = 0$, then by the Cauchy-Schwarz inequality and Lemma 4.5.3 we obtain

$$\begin{aligned} \tilde{F}_\tau^\alpha &= E \left(h(\tau) \left| \tilde{I}_\alpha[g(\cdot)]_{\varrho, \tau} \right|^{2q} \middle| \mathcal{A}_\varrho \right) \\ &\leq \left[E \left(h(\tau)^2 \middle| \mathcal{A}_\varrho \right) \right]^{\frac{1}{2}} \left[E \left(\left| \tilde{I}_\alpha[g(\cdot)]_{\varrho, \tau} \right|^{4q} \middle| \mathcal{A}_\varrho \right) \right]^{\frac{1}{2}} \\ &\leq C (\tau - \varrho)^{q(l(\alpha) + n(\alpha))}. \end{aligned}$$

Note that here and in the following we denote again by C any positive constant that does not depend on $(\tau - \varrho)$.

2. Let us consider the case $s(\alpha) = s \in \{1, 2, \dots\}$.

By the relationship (4.2.12) and some straightforward estimates we obtain

$$\tilde{F}_\tau^\alpha \leq C \left\{ E \left(h(\tau) \left| I_\alpha[g(\cdot)]_{\varrho, \tau} \right|^{2q} \middle| \mathcal{A}_\varrho \right) + \sum_{i=1}^{2^{s(\alpha)}-1} E \left(h(\tau) \left| H_{\alpha,i} \right|^{2q} \middle| \mathcal{A}_\varrho \right) \right\}, \quad (4.5.42)$$

where terms $H_{\alpha,i}$ are described in Remark 4.2.1.

Since $s(\alpha) = s \in \{1, 2, \dots\}$, we can write $\alpha = \alpha_1 \star (-1) \star \alpha_2 \star (-1) \star \dots \star (-1) \star \alpha_{s+1}$, where $s(\alpha_j) = 0$ for $j \in \{1, 2, \dots, s+1\}$ and \star denotes the concatenation operation on multi-indices defined in (4.2.7).

Thus,

$$I_{\alpha_1 \star (-1)}[g(\cdot)]_{\varrho, \tau} = \sum_{i=p_\varphi(\varrho)+1}^{p_\varphi(\tau)} I_{\alpha_1}[g(\cdot)]_{\varrho, \tau_i},$$

where τ_i , with $i \in \{1, 2, \dots, p_\varphi(\tau)\}$, are the jump times generated by the Poisson measure. Similarly, one can show that

$$I_\alpha[g(\cdot)]_{\varrho, \tau} = \sum_{i=p_\varphi(\varrho)+1}^{p_\varphi(\tau)} I_{\alpha_1}[g(\cdot)]_{\varrho, \tau_i} I_{\alpha_2; \tau_i, \tau_{i+1}} \dots I_{\alpha_{s+1}; \tau_{i+s-1}, \tau},$$

where $I_{\alpha_j; \tau_i, \tau_{i+1}} = I_{\alpha_j}[1]_{\tau_i, \tau_{i+1}}$ denotes the multiple stochastic integral for the multi-index α_j over the time interval $[\tau_i, \tau_{i+1}]$. Let us note that $I_\alpha[g(\cdot)]_{\varrho, \tau} = 0$ if $p_\varphi([\varrho, \tau)) < s$. Therefore, by the Cauchy-Schwarz inequality and Lemma 4.5.3, we obtain

$$\begin{aligned} & E \left(h(\tau) |I_\alpha[g(\cdot)]_{\varrho, \tau}|^{2q} \middle| \mathcal{A}_\varrho \right) \\ &= \sum_{n \geq s} E \left(h(\tau) \left| \sum_{i=1}^{n-s+1} I_{\alpha_1}[g(\cdot)]_{\varrho, \tau_i} I_{\alpha_2; \tau_i, \tau_{i+1}} \dots I_{\alpha_{s+1}; \tau_{i+s-1}, \tau} \right|^{2q} \middle| \mathcal{A}_\varrho \right) \\ & \quad \times P(p_\varphi([\varrho, \tau)) = n) \\ & \leq e^{-\lambda(\tau-\varrho)} \sum_{n \geq s} (n-s+1)^{2q-1} \frac{(\lambda(\tau-\varrho))^n}{n!} \sum_{i=1}^{n-s+1} \left[E \left(h(\tau)^2 \middle| \mathcal{A}_\varrho \right) \right]^{\frac{1}{2}} \\ & \quad \times \left[E \left(|I_{\alpha_1}[g(\cdot)]_{\varrho, \tau_i}|^{8q} \middle| \mathcal{A}_\varrho \right) \right]^{\frac{1}{4}} \left[E \left(|I_{\alpha_2; \tau_i, \tau_{i+1}}|^{16q} \middle| \mathcal{A}_\varrho \right) \right]^{\frac{1}{8}} \\ & \quad \times \dots \times \left[E \left(|I_{\alpha_{s+1}; \tau_{i+s-1}, \tau}|^{2^{s+3}q} \middle| \mathcal{A}_\varrho \right) \right]^{\frac{1}{2^{s+2}}} \\ & \leq C e^{-\lambda(\tau-\varrho)} \sum_{n \geq s} (n-s+1)^{2q-1} \frac{(\lambda(\tau-\varrho))^n}{n!} \\ & \quad \times \sum_{i=1}^{n-s+1} (\tau-\varrho)^{q \sum_{i=1}^{s+1} (l(\alpha_i) + n(\alpha_i))} \\ & \leq C (\tau-\varrho)^{q(l(\alpha)+n(\alpha)-s(\alpha))} e^{-\lambda(\tau-\varrho)} (\tau-\varrho)^s \\ & \quad \times \sum_{n \geq s} (n-s+1)^{2q} \frac{\lambda^n (\tau-\varrho)^{n-s}}{n!} \\ & \leq C (\tau-\varrho)^{q(l(\alpha)+n(\alpha)-s(\alpha))+s(\alpha)}. \end{aligned} \tag{4.5.43}$$

Note that the last line of (4.5.43) follows by

$$\begin{aligned}
& \sum_{n \geq s} (n-s+1)^{2q} \frac{\lambda^n (\tau - \varrho)^{n-s}}{n!} \\
& \leq (\lambda)^s \sum_{j=0}^{\infty} (j+1)^{2q} (j+s) \dots (j+1) \frac{(\lambda(\tau - \varrho))^j}{j!} \\
& \leq K \sum_{j=0}^{\infty} (j+s)^{2q+s} \frac{(\lambda(\tau - \varrho))^j}{j!} \\
& \leq K \left\{ \sum_{j=0}^{\infty} \frac{(\lambda(\tau - \varrho))^j}{j!} + \sum_{j=0}^{\infty} j^{2q+s} \frac{(\lambda(\tau - \varrho))^j}{j!} \right\} \\
& = K e^{\lambda(\tau - \varrho)} (1 + B_{2q+s}(\lambda(\tau - \varrho))) \\
& \leq K e^{\lambda(\tau - \varrho)},
\end{aligned}$$

where we denote by $B_n(x)$ the Bell polynomial of degree n , see Bell (1934). To complete the proof we have to show that the bound (4.5.43) also holds for the other terms in (4.5.42). Note that, as shown in Remark 4.2.1, the terms $H_{\alpha,i}$, for $i \in \{1, \dots, 2^{s(\alpha)} - 1\}$, involve multiple stochastic integrals with the same multiplicity $l(\alpha)$ of the multiple stochastic integrals I_α and replace some of the integrations with respect to the Poisson jump measure p_φ with integrations with respect to time and to the intensity measure φ . Therefore, by following the steps above, one can establish the bound (4.5.43) for all terms in (4.5.42). This completes the proof of Lemma 4.5.5.

□

4.6 Exercises

4.1. Use the Wagner-Platen expansion with time increment $h > 0$ and Wiener process increment $W_{t_0+h} - W_{t_0}$ in the expansion part to expand the increment $X_{t_0+h} - X_{t_0}$ of a geometric Brownian motion at time t_0 , where

$$dX_t = a X_t dt + b X_t dW_t.$$

4.2. Expand the geometric Brownian motion from Exercise 4.1 such that all double integrals appear in the expansion part.

4.3. For multi-indices $\alpha = (0, 0, 0, 0), (1, 0, 2), (0, 2, 0, 1)$ determine $-\alpha$, $\alpha-$, $\ell(\alpha)$ and $n(\alpha)$.

4.4. Write out in full the multiple Itô stochastic integrals $I_{(0,0),t}, I_{(1,0),t}, I_{(1,1),t}$ and $I_{(1,2),t}$.

4.5. Express the multiple Itô integral $I_{(0,1)}$ in terms of $I_{(1)}$, $I_{(0)}$ and $I_{(1,0)}$.

4.6. Verify that $I_{(1,0),\Delta}$ is Gaussian distributed with

$$E(I_{(1,0),\Delta}) = 0, \quad E((I_{(1,0),\Delta})^2) = \frac{\Delta^3}{3},$$

$$E(I_{(1,0),\Delta} I_{(1),\Delta}) = \frac{\Delta^2}{2}.$$

4.7. For the case $d = m = 1$ determine the Itô coefficient functions $f_{(1,0)}$ and $f_{(1,1,1)}$.

4.8. Which of the following sets of multi-indices are not hierarchical sets:
 \emptyset , $\{(1)\}$, $\{v, (1)\}$, $\{v, (0), (0, 1)\}$, $\{v, (0), (1), (0, 1)\}$?

4.9. Determine the remainder sets that correspond to the hierarchical sets $\{v, (1)\}$ and $\{v, (0), (1), (0, 1)\}$.

4.10. Determine the truncated Wagner-Platen expansion at time $t = 0$ for the solution of the Itô SDE

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

using the hierarchical set $\mathcal{A} = \{v, (0), (1), (1, 1)\}$.

4.11. In the notation of Sect. 4.2 where the component -1 denotes in a multi-index α a jump term, determine for $\alpha = (1, 0, -1)$ its length $\ell(\alpha)$, its number of zeros and its number of time integrations.

4.12. For the multi-index $\alpha = (1, 0, -1)$ derive in the notation of Sect. 4.2 $-\alpha$.

Introduction to Scenario Simulation

In this chapter we introduce scenario simulation methods for SDEs. We consider pathwise converging simulation methods that use discrete-time approximations. This chapter follows classical ideas, as described in [Kloeden & Platen \(1999\)](#), and examines different numerical schemes. It also considers some examples of typical problems that can be handled by the simulation of approximating discrete-time trajectories.

5.1 Approximating Solutions of Ordinary Differential Equations

In this section we summarize some basic concepts of the numerical analysis of deterministic *ordinary differential equations* (ODEs). The methods are presented in such a way as to facilitate later their generalization to the stochastic setting. This also allows us to highlight the differences in the stochastic case.

Initial Value Problem for ODEs

In general, one does not have in practice an explicit solution $X_t = X(t; t_0, x_0)$ at time t of an initial value problem for an ODE of the form

$$dX_t = a(X_t) dt \quad (5.1.1)$$

for $t \in [t_0, T]$ with $X(t_0) = x_0 \in \Re$. Deterministic ordinary differential equations occur frequently in financial and economic models. Even when such a solution can be found, it may be too complicated to be calculated exactly. It can often be more convenient or computationally efficient to evaluate a solution numerically by using a discrete-time approximation. Practical necessity has lead to the development of numerical methods for calculating such discrete-time approximations to the solutions of initial value problems

for ODEs. The most widely applicable and commonly used are the methods in which the continuous time differential equation is replaced by a discrete-time difference equation, generating values $Y_1, Y_2, \dots, Y_n, \dots$ to approximate $X(t_1; t_0, x_0), X(t_2; t_0, x_0), \dots, X(t_n; t_0, x_0), \dots$ at given discrete times

$$t_0 < t_1 < t_2 < \dots < t_n < \dots$$

These approximations should become accurate if the time increments $\Delta_n = t_{n+1} - t_n$ for $n \in \{0, 1, \dots\}$ are made sufficiently small.

As background for the construction of discretization methods we shall review in this section basic *difference methods* used for ODEs and discuss their convergence and numerical stability.

Euler Method

The simplest difference method for the approximate solution of the ODE (5.1.1) is the *Euler method*

$$Y_{n+1} = Y_n + a(Y_n) \Delta \quad (5.1.2)$$

for a given time discretization with equidistant increments $\Delta = t_{n+1} - t_n$, where $n \in \{0, 1, \dots\}$. First one specifies the initial value $Y_0 = x_0$. Then one computes recursively the values Y_1, Y_2, \dots . The difference

$$l_{n+1} = X(t_{n+1}; t_n, Y_n) - Y_{n+1} \quad (5.1.3)$$

is generally not zero for $n \in \{0, 1, \dots\}$. This difference is called the *local discretization error* for the n th time step. The *global discretization error* at time t_{n+1} is defined as

$$e_{n+1} = X(t_{n+1}; t_0, x_0) - Y_{n+1}. \quad (5.1.4)$$

This is the error obtained with respect to the exact solution $X(t_{n+1}; t_0, x_0)$ of the ODE (5.1.1). Theoretically the errors given in (5.1.3) and (5.1.4) assume that we can perform all arithmetic calculations exactly. In practice, however, a computer is restricted to a finite number of decimal places when performing calculations. It needs to round off all excess decimal places, thus, introducing a *roundoff error*, which we shall denote by r_{n+1} for the n th time step.

Estimating the Local Discretization Error

The key to estimating the size of the discretization errors is the deterministic Taylor formula. For a twice continuously differentiable function X_t it follows that

$$X_{t_{n+1}} = X_{t_n} + \frac{dX_{t_n}}{dt} \Delta + \frac{1}{2!} \frac{d^2X_{\theta_n}}{dt^2} \Delta^2 \quad (5.1.5)$$

with some θ_n satisfying $t_n < \theta_n < t_{n+1}$. For $X_t = X(t; t_n, Y_n)$, the solution of the ODE with $X_{t_n} = Y_n$ has then, according to (5.1.1), the representation

$$X_{t_{n+1}} = X_{t_n} + a(X_{t_n}) \Delta + \frac{1}{2!} \frac{d^2 X_{\theta_n}}{dt^2} \Delta^2. \quad (5.1.6)$$

Setting $X_{t_n} = Y_n$, and subtracting (5.1.2) from (5.1.6) we find that the local discretization error (5.1.3) has the form

$$l_{n+1} = \frac{1}{2!} \frac{d^2 X_{\theta_n}}{dt^2} \Delta^2.$$

If we knew that $|\frac{d^2 X_{\theta_n}}{dt^2}| < M$ for all t in the given interval $[t_0, T]$, then we would have the estimate

$$|l_{n+1}| \leq \frac{1}{2!} M \Delta^2 \quad (5.1.7)$$

for any discretization time interval $[t_n, t_{n+1}]$ with $t_0 \leq t_n < t_{n+1} \leq T$. One can obtain a bound on $|\frac{d^2 X_{\theta_n}}{dt^2}|$ by using the fact that

$$\frac{d^2 X_t}{dt^2} = \frac{d}{dt} \frac{dX_t}{dt} = \frac{da(X_t)}{dt} = \frac{\partial a(X_t)}{\partial x} \frac{dX_t}{dt} = \frac{\partial a(X_t)}{\partial x} a(X_t).$$

If a and $\frac{\partial a}{\partial x}$ are continuous and all solutions X_t under consideration remain in some closed and bounded set \mathcal{C} , then we can use the bound

$$M = \max \left| a(x) \frac{\partial a(x)}{\partial X} \right|,$$

where the maximum is taken over $x \in \mathcal{C}$. The above bound M will usually be a substantial overestimate. Still, from (5.1.7) we can see that the local discretization error for the Euler method (5.1.2) is of order Δ^2 .

Estimating the Global Discretization Error

To estimate the global discretization error we assume, for simplicity, that the drift function $a(x)$ satisfies a uniform Lipschitz condition

$$|a(x) - a(y)| \leq K |x - y|$$

with some constant $K < \infty$ for all $x, y \in \mathfrak{N}$ and that the time discretization is equidistant with step size $\Delta > 0$. Applying the Taylor formula (5.1.5) to the solution $X_t = X(t; t_0, x_0)$ we have the representation (5.1.6) with $X_{t_n} \neq Y_n$, in general. Subtracting (5.1.2) from (5.1.6) we get

$$e_{n+1} = e_n + (a(X_{t_n}) - a(Y_n)) \Delta + \frac{1}{2} \frac{d^2 X_{t_{\theta_n}}}{dt^2} \Delta^2. \quad (5.1.8)$$

By using the Lipschitz condition on a and a bound M on $|\frac{d^2 X_t}{dt^2}|$ we obtain

$$|e_{n+1}| \leq |e_n| + K |e_n| \Delta + \frac{1}{2} M \Delta^2$$

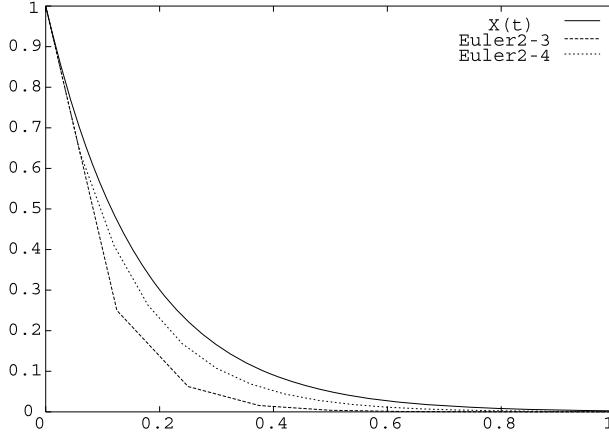


Fig. 5.1.1. Euler approximations and X_t

for $n \in \{0, 1, \dots\}$. By induction it follows from the inequality

$$|e_{n+1}| \leq (1 + K \Delta) |e_n| + \frac{1}{2} M \Delta^2 \quad (5.1.9)$$

with $e_0 = x_0 - Y_0 = 0$ that

$$|e_{n+1}| \leq \frac{1}{2} \left(\frac{(1 + K \Delta)^n - 1}{(1 + K \Delta) - 1} \right) M \Delta^2.$$

Since

$$(1 + K \Delta)^n \leq e^{nK\Delta}$$

we obtain

$$|e_{n+1}| \leq \frac{1}{2} (e^{nK\Delta} - 1) \frac{M}{K} \Delta.$$

Consequently, the global discretization error for the Euler method (5.1.2) satisfies the estimate

$$|e_{n+1}| \leq \frac{1}{2} (e^{K(T-t_0)} - 1) \frac{M}{K} \Delta \quad (5.1.10)$$

for discretization times $t_n = t_0 + n \Delta \leq T$. Obviously, the global discretization error has one power of Δ less than the local discretization error.

To give an example, let us apply the Euler method (5.1.2) to the ODE

$$dX_t = -\eta X_t dt$$

for $t \in [0, 1]$ with $X_0 = 1$ using an equidistant time discretization with time step sizes $\Delta = 2^{-3}$ and 2^{-4} . In Fig. 5.1.1 we plot the results together with the exact solution $X_t = \exp\{-\eta t\}$ against $t \in [0, 1]$ for $\eta = 6$.

In Fig. 5.1.1 the upper curve provides the exact solution, and the lower and middle curves the Euler approximation with step sizes $\Delta = 2^{-3}$ and

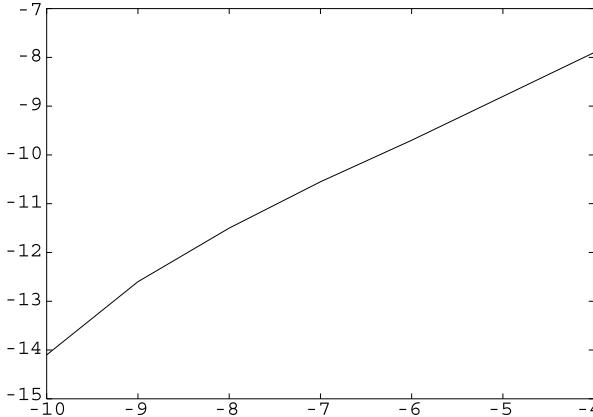


Fig. 5.1.2. $\log_2(e_{\frac{1}{\Delta}})$ versus $\log_2(\Delta)$ for Euler method

2^{-4} , respectively. Obviously, the global discretization error is smaller for the smaller time step size.

Let us now look more closely at the dependence of the global discretization error on the step size. For this we recall that a function

$$e(\Delta) = A \Delta^\gamma$$

becomes linear when considering logarithms. This means, we obtain

$$\log_a(e(\Delta)) = \log_a(A) + \gamma \log_a(\Delta)$$

for logarithms to the base $a \neq 1$. Throughout the following comparative studies we choose time steps of the form $\Delta = 2^{-n}$ for $n \in \{1, 2, \dots\}$. This is because we usually halve the time step successively, in which case logarithms to the base $a = 2$ are appropriate. In Fig. 5.1.2 we plot the values of $\log_2(e_{\frac{1}{\Delta}})$ against $\log_2(\Delta)$ for the Euler approximations of the above example for $\eta = 5$ and using the time step sizes $\Delta = 2^{-4}, 2^{-5}, \dots, 2^{-10}$. We note from Fig. 5.1.2 that the logarithm of the global error dependent on the log of the step size is linear as expected from (5.1.10). The graph is almost a straight line with slope one.

We repeat this numerical experiment on our computer for the wider range of time step sizes $2^0, 2^{-1}, \dots, 2^{-14}$. From Fig. 5.1.3 we see that the calculated discretization error e_{n+1} for the Euler method is only proportional to the step size Δ in the range from 2^{-3} to 2^{-11} . For a step size Δ that is too large, the numerical approximation fails. Here the Taylor expansion has too larger errors that cannot be easily neglected. For $\Delta \leq 2^{-11}$ the error e_{n+1} begins to increase as Δ is further decreased. This does not contradict the global error estimate (5.1.10), but occurs because e_{n+1} also includes a roundoff error that arises on our computer due to its finite arithmetic precision. To estimate e_{n+1} we must add the roundoff error r_{n+1} to the right hand side of (5.1.9). For $|r_{n+1}| \leq R$, and each n we then obtain the estimate

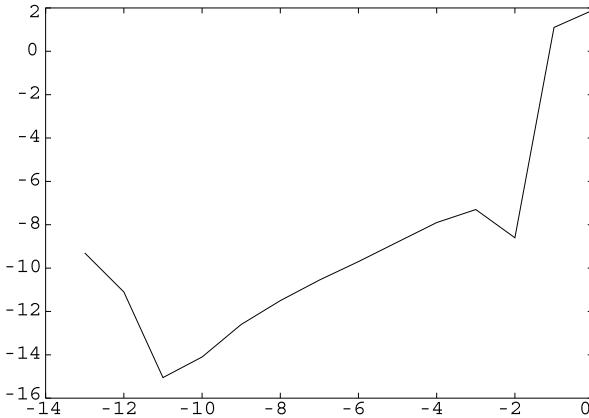


Fig. 5.1.3. Log-log plot for global error of Euler method

$$|e_{n+1}| \leq \frac{1}{2} \left(e^{K(T-t_0)} - 1 \right) \left(\frac{2R}{K} \frac{1}{\Delta} + \frac{M}{K} \Delta \right)$$

instead of (5.1.10). This illustrates that for very small time step size Δ the term with the reciprocal $\frac{1}{\Delta}$ dominates the estimation term. The above global error estimate represents in some sense the worst case scenario. However, this bound is rather indicative of the cumulative effect of the roundoff error, since for smaller Δ more calculations are required to reach the given end time.

The presence of roundoff errors means, in practice, that there is a minimum step size Δ_{\min} for each ODE approximation, below which one cannot improve the accuracy of the approximations calculated by way of the simple Euler method.

Higher Order and Implicit Methods

For introductory purposes let us say that a discrete-time approximation Y with time step size $\Delta > 0$ has the *order of convergence* $\gamma > 0$ for a given ODE if there exists a constant $C < \infty$ such that

$$e_{\frac{T}{\Delta}} \leq C \Delta^\gamma. \quad (5.1.11)$$

The Euler method has for an ODE the order of convergence $\gamma = 1$. To obtain a more accurate discrete-time approximation we need to use other methods with higher order of convergence.

For the Euler method we simply froze the right hand side of the ODE at the value $a(Y_n)$ of the last approximation point for each discretization subinterval $t_n \leq t < t_{n+1}$. One should obtain a more accurate approximation if one could use the average of the values at both end points of the discretization interval. In this case we obtain the *trapezoidal method*

$$Y_{n+1} = Y_n + \frac{1}{2} (a(Y_n) + a(Y_{n+1})) \Delta. \quad (5.1.12)$$

This is called an *implicit* scheme because the unknown quantity Y_{n+1} appears on both sides of equation (5.1.12). More generally, we have the *family of implicit Euler schemes*

$$Y_{n+1} = Y_n + (\alpha a(Y_{n+1}) + (1 - \alpha) a(Y_n)) \Delta, \quad (5.1.13)$$

where $\alpha \in [0, 1]$ is the *degree of implicitness*. In the case $\alpha = 1$, we have the *fully implicit Euler scheme*. Unfortunately, in general, the value Y_{n+1} cannot be obtained algebraically. To overcome this difficulty we could use the Euler method (5.1.2) to approximate the Y_{n+1} term on the right hand side of the scheme (5.1.12). This then leads to the *modified trapezoidal method*

$$Y_{n+1} = Y_n + \frac{1}{2}(a(Y_n) + a(\bar{Y}_{n+1})) \Delta,$$

with supporting value

$$\bar{Y}_{n+1} = Y_n + a(Y_n) \Delta.$$

This scheme can then be written as

$$Y_{n+1} = Y_n + \frac{1}{2}\left(a(Y_n) + a(Y_n + a(Y_n) \Delta)\right) \Delta, \quad (5.1.14)$$

which is also known as the *Heun method*. It is a simple example of a *predictor-corrector method* with the predictor \bar{Y}_{n+1} inserted into the corrector equation to give the next iterate Y_{n+1} .

Both the trapezoidal and the modified trapezoidal methods have local discretization errors of third order. This can be verified by using the classical Taylor formula with a third order remainder term for the expansion at $a(Y_{n+1})$ or $a(\bar{Y}_{n+1})$ about Y_n . The global discretization error for both methods (5.1.12) and (5.1.14) is of second order, which is again one order less than the local discretization error.

With the standard arithmetic of a computer we repeat the previous numerical experiment for the modified trapezoidal method (5.1.14). Fig. 5.1.4 indicates that the resulting global error for the modified trapezoidal method is proportional to Δ^2 for $\Delta \leq 2^{-3}$, whereas that of the Euler method in Fig. 5.1.3 was proportional to Δ .

Further higher order difference methods can be systematically derived by using more accurate approximations of the right hand side of the ODE (5.1.1) over each discretization subinterval $t_n \leq t \leq t_{n+1}$. The resulting schemes are called *one-step methods* as they involve only the values Y_n and Y_{n+1} in addition to t_n and Δ . Explicit one-step methods are usually written in the form

$$Y_{n+1} = Y_n + \Psi(Y_n, \Delta) \Delta, \quad (5.1.15)$$

for some function $\Psi(x, \Delta)$, which is called the *increment function*. It is a p th order method if its global discretization error is bounded by some constant times the p th power of Δ . We will show below that if the functions a and Ψ

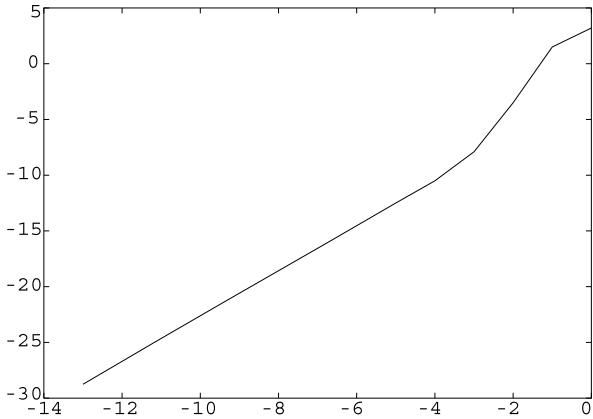


Fig. 5.1.4. Log-log plot for global error for modified trapezoidal method

are sufficiently smooth, then a $(p+1)$ th order local discretization error implies a p th order global discretization error. For example, the Euler method (5.1.2) is a first order one-step scheme with $\Psi(x, \Delta) = a(x)$ and the Heun method (5.1.14) is a first order one-step scheme with

$$\Psi(x, \Delta) = \frac{1}{2} \left(a(x) + a(x + a(x) \Delta) \right).$$

The increment function Ψ cannot be chosen completely arbitrarily. For instance, it should be *consistent* with the differential equation, that is, it should satisfy the limit

$$\lim_{\Delta \rightarrow 0} \Psi(x, \Delta) = a(x),$$

if the values calculated from (5.1.15) are to converge to the solution of the ODE. The Euler method is obviously consistent and the Heun method is consistent when $a(x)$ is continuous.

Extrapolation

One can sometimes obtain higher order accuracy from a one-step scheme by the method of *extrapolation*. For example, suppose we use the Euler scheme (5.1.2) with N equal time steps $\Delta = \frac{T}{N}$ on the interval $t \in [0, T]$. If X_T is the true value at time T and $Y_N(\Delta)$ the corresponding value from the Euler scheme, then we have

$$Y_N(\Delta) = X_T + e_T \Delta + O(\Delta^2), \quad (5.1.16)$$

where we have written the global truncation error as $e_T \Delta + O(\Delta^2)$. If, instead, we use the Euler scheme with $2N$ time steps of equal length $\frac{\Delta}{2}$, then we have

$$Y_{2N} \left(\frac{1}{2} \Delta \right) = X_T + \frac{1}{2} e_T \Delta + O(\Delta^2). \quad (5.1.17)$$

We can eliminate the leading error term e_T from (5.1.16) and (5.1.17) by forming the linear combination

$$X_T = 2Y_{2N} \left(\frac{1}{2} \Delta \right) - Y_N(\Delta) + O(\Delta^2). \quad (5.1.18)$$

Thus we have a second order approximation

$$Z_N(\Delta) = 2Y_{2N} \left(\frac{1}{2} \Delta \right) - Y_N(\Delta) \quad (5.1.19)$$

for X_T using only the first order Euler scheme. Of course, this requires repeating the Euler scheme calculations using half the original time step size. However, for a complicated ODE it may involve fewer and simpler calculations than a second order one-step scheme would require. This extrapolation method is known as *Richardson extrapolation*. It can also be applied to more general one-step schemes.

Truncated Taylor Methods

A one-step difference method with local discretization error of order $p+1$ is readily suggested by the classical Taylor formula with $(p+1)$ th order remainder term

$$X_{t_{n+1}} = X_{t_n} + \frac{dX_{t_n}}{dt} \Delta + \dots + \frac{1}{p!} \frac{d^p X_{t_n}}{dt^p} \Delta^p + \frac{1}{(p+1)!} \frac{d^{p+1} X_{\theta_n}}{dt^{p+1}} \Delta^{p+1}, \quad (5.1.20)$$

where $t_n < \theta_n < t_{n+1}$ and $\Delta = t_{n+1} - t_n$, for $p \in \mathcal{N}$. One can apply this Taylor expansion to a solution X_t of the ODE (5.1.1) if the drift function $a(x)$ and its derivatives of orders up to and including p are continuous, as this assures that X_t is $(p+1)$ times continuously differentiable. Indeed, from (5.1.1) and the chain rule one obtains by repeatedly differentiating $a(X_t)$ the derivatives $\frac{dX_t}{dt} = a'(X_t)$, $\frac{d^2X_t}{dt^2} = a''(X_t)$, Evaluating these terms at Y_n and omitting the remainder term in (5.1.20) we obtain a one-step method for Y_{n+1} , which we shall call the *p th order truncated Taylor method*. This method, obviously, has a local discretization error of order $p+1$. Under appropriate conditions it can be shown to have global discretization error of order p . The first order truncated Taylor method is then just the Euler method (5.1.2), the *second order truncated Taylor method* is

$$Y_{n+1} = Y_n + a(Y_n) \Delta + \frac{1}{2!} (a'(Y_n) a(Y_n)) \Delta^2 \quad (5.1.21)$$

and the *third order truncated Taylor method* has the form

$$\begin{aligned} Y_{n+1} = Y_n + a(Y_n) \Delta + \frac{1}{2!} (a'(Y_n) a(Y_n)) \Delta^2 \\ + \frac{1}{3!} (a''(Y_n) a^2(Y_n) + (a'(Y_n))^2 a(Y_n)) \Delta^3. \end{aligned} \quad (5.1.22)$$

One Step Methods

The coefficients in higher order truncated Taylor methods soon become complex. Moreover, considerable computational effort is needed for the evaluation of these coefficients. For this reason these methods are not particularly convenient. They are almost never used in practice, except to provide a reference point for the development and analysis of other, more efficient higher order schemes.

One way of simplifying the coefficients in a truncated Taylor method is to replace the derivatives by their difference quotients, for example, replacing $a'(Y_n)$ by

$$\frac{a(Y_{n+1}) - a(Y_n)}{Y_{n+1} - Y_n}.$$

This leads to an implicit scheme because Y_{n+1} appears on both sides of the corresponding recursion formula and, generally, cannot be solved algebraically. As in the trapezoidal method (5.1.12) we could use the Euler method (5.1.2) to predict a value of Y_{n+1} to be used on the right hand side of a formula, thus obtaining an explicit method. For the 2nd order truncated Taylor method this results first in the trapezoidal method (5.1.12) and then in the Heun method (5.1.14). The higher order coefficients will usually be considerably more complicated, but are at least derivative-free.

The standard procedure with most one-step methods

$$Y_{n+1} = Y_n + \Psi(Y_n, \Delta) \Delta \quad (5.1.23)$$

is to first derive the function $\Psi(x, \Delta)$ by a heuristic argument, then compare the method with a truncated Taylor method to determine the order of its discretization error. The, so called, *Runge-Kutta methods* are typical of this approach. For what will turn out to be the family of second order methods of this type, Ψ is chosen in the form

$$\Psi(x, \Delta) = \alpha a(x) + \beta a(x + \delta a(x) \Delta), \quad (5.1.24)$$

for certain constants α , β and δ , which represent a weighted averaging of the right hand side of the ODE (5.1.1) over two points. Expanding the second term about x , we obtain

$$\Psi(x, \Delta) \approx (\alpha + \beta) a(x) + \delta \beta a'(x) a(x) \Delta + \frac{1}{2} \delta^2 \beta a''(x) (a(x))^2 \Delta^2$$

neglecting higher order terms. Hence, subtracting (5.1.23) with this expansion for Ψ evaluated at (Y_n, Δ) from the third order truncated Taylor method (5.1.22) we obtain the expression

$$\begin{aligned} \Psi(Y_n, \Delta) &\approx (1 - \alpha - \beta) a(Y_n) \Delta + \left(\frac{1}{2!} - \delta \beta \right) a'(Y_n) a(Y_n) \Delta^2 \\ &+ \frac{1}{2} \left(\frac{1}{3} - \delta^2 \beta \right) a''(Y_n) (a(Y_n))^2 \Delta^3 + \frac{1}{6} (a'(Y_n))^2 a(Y_n) \Delta^3, \end{aligned}$$

where everything is now evaluated at Y_n and we neglect higher order terms. The first two terms here drop out if we choose the weighting parameters α , β and δ such that

$$\alpha + \beta = 1 \quad \text{and} \quad \delta \beta = \frac{1}{2}. \quad (5.1.25)$$

In general, it will not be possible to eliminate both Δ^3 terms by the choice of parameters β and δ as the second of these terms may not vanish identically. The parameter constraints (5.1.25) assure that a difference method with Ψ given by (5.1.24) will have local discretization error of order three and hence global discretization error of order two. Since one of the parameters in (5.1.25) can be chosen arbitrarily, this gives an infinite number of 2nd order difference schemes. Note that the first constraint in (5.1.25) assures that all of these methods are consistent, as previously defined. The choice $\alpha = \beta = \frac{1}{2}$, $\delta = 1$ gives the Heun method (5.1.14), which is also called the 2nd order Runge-Kutta method.

Global Order of Convergence

Let $X_t = x(t; t_0, x_0)$ be the solution of the ODE (5.1.1) and let Y_n be generated by

$$Y_{n+1} = Y_n + \Psi(Y_n, \Delta) \Delta \quad (5.1.26)$$

with $Y_0 = X_0$. Then the global discretization error (5.1.4) satisfies the relation

$$\begin{aligned} e_{n+1} &= Y_{n+1} - X_{t_{n+1}} \\ &= e_n + \Psi(Y_n, \Delta) \Delta + X_{t_n} - X_{t_{n+1}} \\ &= e_n + (\Psi(Y_n, \Delta) - \Psi(X_{t_n}, \Delta)) \Delta + (\Psi(X_{t_n}, \Delta) \Delta + X_{t_n} - X_{t_{n+1}}), \end{aligned}$$

where the very last term is the local discretization error. From the global Lipschitz condition

$$|\Psi(x', \Delta') - \Psi(x, \Delta)| \leq K (|x' - x| + |\Delta' - \Delta|) \quad (5.1.27)$$

and using the assumption that the local discretization error is of order $p + 1$ we obtain

$$\begin{aligned} |e_{n+1}| &= |e_n| + K |e_n| \Delta + D \Delta^{p+1} \\ &= (1 + K \Delta) |e_n| + D \Delta^{p+1}. \end{aligned}$$

Here D is some positive constant, and it follows that

$$|e_n| \leq \frac{D}{K} \left(e^{K(T-t_0)} - 1 \right) \Delta^p$$

on the interval $[t_0, T]$. Under appropriate conditions, the global discretization error is thus of order p if the local discretization error is of order $p + 1$.

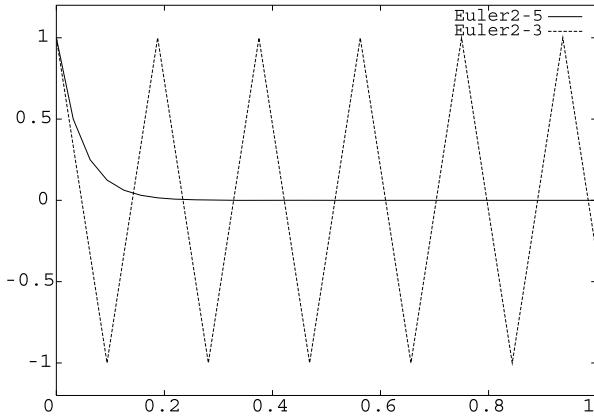


Fig. 5.1.5. Log-log plot for global error the Euler method

Numerical Stability

We may still encounter difficulties when trying to implement a difference method which is known to be convergent. For example, the differential equation

$$\frac{dX_t}{dt} = -\eta X_t \quad (5.1.28)$$

has exact solution $X_t = x_0 e^{-\eta t}$, which converges rapidly to zero. For this differential equation the Euler method with constant time step size Δ ,

$$Y_{n+1} = (1 - \eta \Delta) Y_n,$$

has exact iterates

$$Y_n = (1 - \eta \Delta)^n Y_0.$$

If we choose $\Delta > 2^{-4}$ for $\eta = 33$ these iterates oscillate, as shown in Fig. 5.1.5, instead of converging to zero like the exact solutions of (5.1.28). This is a simple example of a *numerical instability*, which, in this particular case, can be overcome by taking a time step $\Delta < 2^{-4}$, see Fig. 5.1.5. For some other methods or ODEs the numerical instabilities may persist no matter how small one makes Δ . The structure of some methods can make them intrinsically unstable, causing small errors such as roundoff errors to grow rapidly and ultimately rendering the computation useless.

The idea of *numerical stability* of a one-step method is that errors will remain bounded with respect to an initial error for any ODE of type (5.1.1) with right hand side $a(x)$ satisfying a Lipschitz condition. To be specific, let Y_n and \tilde{Y}_n denote the approximations obtained by the method (5.1.26) when starting in Y_0 and \tilde{Y}_0 , respectively. We say that a discrete-time approximation is *numerically stable* if there exist finite positive constants Δ_θ and M such that

$$|Y_n - \tilde{Y}_n| \leq M |Y_0 - \tilde{Y}_0| \quad (5.1.29)$$

for $n \in \{0, 1, \dots, n_T\}$ and any two solutions Y_n, \tilde{Y}_n of (5.1.26) corresponding to a time discretization with $\Delta < \Delta_\theta$. The constants Δ_0 and M here, may depend on the particular time interval $t \in [t_0, T]$ in addition to the ODE under consideration. The inequality (5.1.29) is analogous to the continuity in initial conditions of the solutions of the ODE (5.1.1). The following result is then straightforward to prove:

The one-step method (5.1.26) is numerically stable if the increment function Ψ satisfies a global Lipschitz condition of the form (5.1.27).

To see this, let $E_n = Y_n - \tilde{Y}_n$, then

$$\begin{aligned} |E_{n+1}| &= \left| E_n + \Delta \left(\Psi(Y_n, \Delta) - \Psi(\tilde{Y}_n, \Delta) \right) \right| \\ &\leq |E_n| + \Delta |\Psi(Y_n, \Delta) - \Psi(\tilde{Y}_n, \Delta)| \\ &\leq |E_n| + \Delta K |Y_n - \tilde{Y}_n| \leq (1 + K \Delta) |E_n| \end{aligned}$$

Hence with $M = (1 + K \Delta)^{n_T}$

$$|E_n| \leq (1 + K \Delta)^n |E_0| \leq (1 + K \Delta)^{n_T} |E_0| \leq M |E_0|$$

for $n \in \{0, 1, \dots, n_T\}$.

Above we tried to ensure that the global error does not grow over an increasing time horizon. This leads to the idea of asymptotic numerical stability. We shall say that a one-step method (5.1.26) is *asymptotically numerically stable* for a given ODE if there exist positive constants Δ_a and M such that

$$\lim_{n \rightarrow \infty} |Y_n - \tilde{Y}_n| \leq M |Y_0 - \tilde{Y}_0| \quad (5.1.30)$$

for any two solutions y and \tilde{y} of (5.1.26) corresponding to any time discretization with $\Delta < \Delta_a$.

It is easy to see that the Euler method is asymptotically numerically stable for the ODE (5.1.28) with $\Delta_a \leq 2^{-4}$ and $\eta = 32$. For larger step sizes the method may fail to be asymptotically numerically stable. The numerical stability of a method has to have priority over potential higher order convergence properties. We will see that the numerical problems we discussed for solving ODEs arise in a similar fashion for SDEs.

5.2 Scenario Simulation

Essentially, only in the case of linear SDEs and its transformations, see Sect. 1.7 and Chap. 2, do we have explicit solutions for SDEs that we can exactly or almost exactly simulate. In finance and other areas of application it is often helpful, and in some cases essential, to be able to simulate at least approximate trajectories of solutions of SDEs that model the quantities under consideration. These scenarios of solutions of SDEs are typically generated using pseudo-random number generators as discussed in Kloeden et al. (2003).

However, one can use also natural or physical random numbers, see [Pivitt \(1999\)](#), and even observed random returns, log-returns or increments from observed asset prices. The latter allows us to perform *historical simulations*, that is, to generate reasonably realistic paths of securities and portfolios. If such simulation is focussed on extreme scenarios, then it is called *stress testing*. In insurance, the area of dynamic financial analysis is attracting increasing attention, see [Kaufmann, Gadmer & Klett \(2001\)](#). The combination of the benchmark approach with scenario simulation of entire market models appears to be rather promising. In the following we will give an introduction to *scenario simulation*, which aims to approximate the path of a solution of a given SDE by the path of a discrete-time approximation.

Discrete-Time Approximation

Consider a given discretization $0 = t_0 < t_1 < \dots < t_n < \dots < t_N = T < \infty$ of the time interval $[0, T]$. We shall approximate a process $X = \{X_t, t \in [0, T]\}$ satisfying the one-dimensional SDE

$$dX_t = a(t, X_t) dt + b(t, X_t) dW_t \quad (5.2.1)$$

on $t \in [0, T]$ with initial value $X_0 \in \Re$. For simplicity, we focus in this introduction on the case of SDEs without jumps.

Let us first introduce one of the simplest, reasonable discrete-time approximations, which is the *Euler scheme* or *Euler-Maruyama scheme*, see [Maruyama \(1955\)](#). An Euler scheme is a continuous time stochastic process $Y = \{Y_t, t \in [0, T]\}$ satisfying the recursive algorithm

$$Y_{n+1} = Y_n + a(t_n, Y_n) (t_{n+1} - t_n) + b(t_n, Y_n) (W_{t_{n+1}} - W_{t_n}), \quad (5.2.2)$$

for $n \in \{0, 1, \dots, N-1\}$ with initial value

$$Y_0 = X_0, \quad (5.2.3)$$

where we have written

$$Y_n = Y_{t_n} \quad (5.2.4)$$

for the value of the approximation at the discretization time t_n . We shall also write

$$\Delta_n = t_{n+1} - t_n \quad (5.2.5)$$

for the n th increment of the time discretization and call

$$\Delta = \max_{n \in \{0, 1, \dots, N-1\}} \Delta_n \quad (5.2.6)$$

the *maximum step size*. For much of this book we shall consider, for simplicity, *equidistant time discretizations* with

$$t_n = n \Delta \quad (5.2.7)$$

for $\Delta_n = \Delta = \frac{T}{N}$ and some integer N large enough so that $\Delta \in (0, 1)$.

The sequence $(Y_n)_{n \in \{0, 1, \dots, N\}}$ of values of the Euler approximation (5.2.2) at the instants of the time discretization t_0, t_1, \dots, t_N can be recursively computed. We need to generate the random increments

$$\Delta W_n = W_{t_{n+1}} - W_{t_n}, \quad (5.2.8)$$

for $n \in \{0, 1, \dots, N-1\}$ of the Wiener process $W = \{W_t, t \in [0, T]\}$. From (1.1.16) we know that these increments are independent Gaussian distributed random variables with mean

$$E(\Delta W_n) = 0 \quad (5.2.9)$$

and variance

$$E((\Delta W_n)^2) = \Delta_n. \quad (5.2.10)$$

For the increments (5.2.8) of the Wiener process we shall use a sequence of independent Gaussian random numbers.

For describing a numerical scheme we will typically use the abbreviation

$$f = f(t_n, Y_n) \quad (5.2.11)$$

for a function f defined on $[0, T] \times \mathbb{R}^d$ and $n \in \{0, 1, \dots, N-1\}$, when no misunderstanding is possible. This makes the notation more efficient. We can then rewrite the Euler scheme (5.2.2) in the form

$$Y_{n+1} = Y_n + a \Delta_n + b \Delta W_n, \quad (5.2.12)$$

for $n \in \{0, 1, \dots, N-1\}$. Usually, we do not mention the initial condition, however, we typically set

$$Y_0 = X_0. \quad (5.2.13)$$

The recursive structure of the Euler scheme, which generates approximate values of the diffusion process X at the discretization instants only, is the key to its successful implementation on a computer. We shall focus on *discrete-time approximations* that are generated by recursive algorithms. In general, we will use the term *scheme* or *approximation* to denote a recursive algorithm, which provides the values of a discrete-time approximation at given discretization times.

Interpolation

We emphasize that we shall consider a discrete-time approximation to be a stochastic process defined on the entire interval $[0, T]$, although it will be often sufficient to consider its values at the discretization times. If required, values at the intermediate instants can be determined by an appropriate interpolation method. The simplest one is *piecewise constant interpolation* with

$$Y_t = Y_{n_t} \quad (5.2.14)$$

for $t \in [0, T]$. Here n_t is the integer defined by

$$n_t = \max\{n \in \{0, 1, \dots, N\} : t_n \leq t\}, \quad (5.2.15)$$

which is the largest integer n for which t_n does not exceed t .

Furthermore, the *linear interpolation*

$$Y_t = Y_{n_t} + \frac{t - t_{n_t}}{t_{n_t+1} - t_{n_t}} (Y_{n_t+1} - Y_{n_t}) \quad (5.2.16)$$

for $t \in [0, T]$ is often used because it provides continuous sample paths in a convenient manner.

In general, the sample paths of a diffusion process inherit the irregularities of the sample paths of its driving Wiener process, in particular, its nondifferentiability. It is impossible to fully reproduce the microstructure of such paths in a scenario simulation on a digital computer. Here we shall concentrate on simulating values of a discrete-time approximation at given discretization times.

Simulating Geometric Brownian Motion

To illustrate various aspects of a *scenario simulation* via a discrete-time approximation of a diffusion process we shall examine a simple but important example in some detail. In finance one observes growth processes, which are under the standard market model interpreted as geometric Brownian motions, see (1.6.3) and (1.7.13). Let us consider the geometric Brownian motion process $X = \{X_t, t \in [0, T]\}$ satisfying the linear SDE

$$dX_t = a X_t dt + b X_t dW_t \quad (5.2.17)$$

for $t \in [0, T]$ with initial value $X_0 > 0$. This is also called the Black-Scholes SDE. The geometric Brownian motion X has *drift coefficient*

$$a(t, x) = a x \quad (5.2.18)$$

and *diffusion coefficient*

$$b(t, x) = b x. \quad (5.2.19)$$

Here a denotes the *appreciation rate* and $b \neq 0$ the *volatility*. We know from (1.7.14) that the SDE (5.2.17) has the explicit solution

$$X_t = X_0 \exp \left\{ \left(a - \frac{1}{2} b^2 \right) t + b W_t \right\} \quad (5.2.20)$$

for $t \in [0, T]$ and given Wiener process $W = \{W_t, t \in [0, T]\}$. The availability of an explicit solution makes the simulation an easy task, as shown in Sect. 2.1, because the solution of the SDE (5.2.17) is a direct function (5.2.20) of the Wiener process value. This makes geometric Brownian motion a good example for studying the performance of schemes since an explicit exact solution is available for comparison.

Scenario Simulation

Knowing the solution (5.2.20) of the SDE (5.2.17) explicitly, gives us the possibility of comparing the Euler approximation Y , see (5.2.12), with the exact solution X and hence calculate the appropriate errors.

To simulate a trajectory of the Euler approximation for a given time discretization we start from the initial value $Y_0 = X_0$, and proceed recursively by generating the next value

$$Y_{n+1} = Y_n + a Y_n \Delta_n + b Y_n \Delta W_n \quad (5.2.21)$$

for $n \in \{0, 1, \dots, N-1\}$ according to the Euler scheme (5.2.12) with drift coefficient (5.2.18) and diffusion coefficient (5.2.19). Here

$$\Delta W_n = W_{t_{n+1}} - W_{t_n}, \quad (5.2.22)$$

see (5.2.8), is the $N(0, \Delta_n)$ distributed Gaussian increment of the Wiener process W over the subinterval $[t_n, t_{n+1}]$, which we generate by a random number generator.

For comparison, we can use (5.2.20) for the given Black-Scholes dynamics to determine the corresponding exact values of the solution. This uses the same sample path of the Wiener process in (5.2.20), as for the Euler scheme (5.2.21) one obtains as an explicit solution at time t_n the value

$$X_{t_n} = X_0 \exp \left\{ \left(a - \frac{1}{2} b^2 \right) t_n + b \sum_{i=1}^n \Delta W_{i-1} \right\} \quad (5.2.23)$$

for $n \in \{0, 1, \dots, N-1\}$.

For illustration, we generate a piecewise constant interpolated Euler approximation on the time interval $[0, 1]$ with equal step size $\Delta = 0.25$ for the geometric Brownian motion X_t satisfying (5.2.17) with $X_0 = 1.0$, $a = 1.0$ and $b = 1.0$. We plot in Fig. 5.2.1 the piecewise constant Euler approximations for the step sizes $\Delta = 0.25$ and $\Delta = 0.0625$ together with the exact solution for the same sample path of the Wiener process. At the discretization times the Euler approximations differ considerably from the exact solution of the SDE (5.2.17) for the given rather large step size. However, we obtained a closer resemblance when we use the smaller step size $\Delta = 0.0625$, as shown in Fig. 5.2.1.

We need to be careful when using a discrete-time scheme, such as the Euler scheme (5.2.21), to generate an approximate solution of an SDE. The resulting discrete-time approximation is mathematically an object which is different from the exact solution of the given SDE. We have to make sure that the resulting expressions in the computation are always meaningful. For instance, inconsistencies may arise because the increments in the noise part of the Euler scheme can take extremely large values of either sign. Even though this can occur only with small probability, large negative stochastic increments

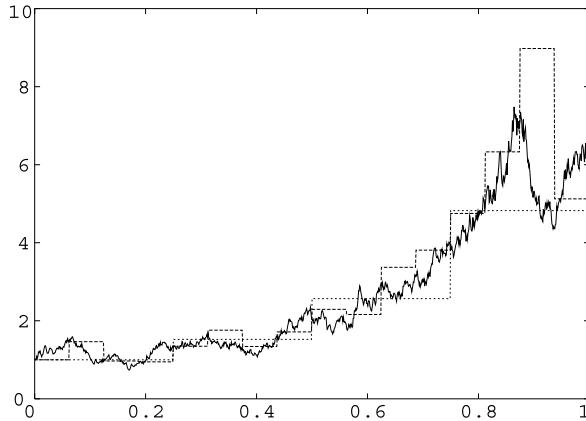


Fig. 5.2.1. Euler approximations for $\Delta = 0.25$ and $\Delta = 0.0625$ and exact solution for Black-Scholes SDE

can, for instance, make the Euler approximation (5.2.21) negative. This is not consistent with our aim of modeling positive asset prices by a Black-Scholes model. According to (5.2.23) these prices have to stay nonnegative. The possibility of generating negative asset prices is a quite serious defect in the Euler method when directly applied to the Black-Scholes dynamics. For extremely small step sizes this phenomenon becomes highly unlikely, however, it is never excluded and one may have to consider alternative ways of avoiding such a situation. One way is demonstrated in Sect. 2.1. However, for other SDEs this way is not available.

For Black-Scholes dynamics the potential negativity can be easily avoided if one simulates first the logarithm of the geometric Brownian motion by a corresponding Euler scheme and then derives the approximate value of X by taking the exponential of the approximated logarithm. One notes that it is worth first carefully analyzing the underlying model dynamics and only afterwards choosing an appropriate way of simulating the desired scenarios. By simulating transforms of a model, as discussed in Chap. 2 when using simple transformed Wiener processes, it is possible to make the simulation highly accurate. Furthermore, such an approach helps to respect conservation laws, such as the nonnegativity of asset prices.

We emphasize that one should be careful in interpreting the simulation of an approximate discrete-time scenario as being sufficiently accurate. It may in certain cases significantly differ from the exact solution due to potential numerical instabilities, as will be discussed in Chap. 14. Later we will address some problems that result from numerical instabilities which can make a simulation useless. Small errors always occur naturally. However, if these are uncontrolled they can propagate and may result in prohibitive errors.

Types of Approximation

So far we have not specified a criterion for the classification of the overall accuracy of a discrete-time approximation with respect to vanishing time step size. Such a criterion should reflect the main goal of a simulation. It turns out that there are *two* basic types of tasks that are typical for the simulation of solutions of SDEs.

The first type, which we have discussed previously, is called *scenario simulation* and applies in situations, where a *pathwise approximation* is required. For instance, scenario simulation is needed for judging the overall behavior of modeled quantities, for hedge simulations for dynamic financial analysis and for the testing of calibration methods and statistical estimators, see Chap. 9.

The second type is called Monte Carlo simulation and focuses on *approximating probabilities* and thus *expectations* of functionals of the approximated process. For instance, via Monte Carlo methods one simulates probabilities, moments, prices of contingent claims, expected utilities and risk measures. The second type is highly relevant for many quantitative methods in finance and insurance. It is generally used when quantities cannot be analytically determined. The first type of tasks is relevant for verifying quantitative methods and extracting information from models and the observed data, as is the case in filtering, see Chap. 10.

In this and the following chapters we concentrate on the first type of tasks, the scenario simulation, which we relate to the *strong approximation* of solutions of SDEs. Later we will separately study in Chap. 11 Monte Carlo methods, which we will then relate to the *weak approximation* of SDEs.

Order of Strong Convergence

In general, one does not know much about the solution of a given SDE. However, one can try to use a scenario simulation to discover some of its properties. Theoretically one can estimate and in some cases also practically calculate the error of an approximation using the following *absolute error criterion*. For a given maximum step size Δ it is defined as the expectation of the absolute difference between the discrete-time approximation Y_T^Δ and the solution X_T of the SDE at the time horizon T , that is

$$\varepsilon(\Delta) = E(|X_T - Y_T^\Delta|). \quad (5.2.24)$$

This provides some measure for the pathwise closeness between approximation and exact value at the end of the time interval $[0, T]$ and its dependence on the maximum step size Δ of the time discretization.

While the Euler approximation is the simplest, useful discrete-time approximation, it is, generally, not very efficient numerically. We shall therefore derive and investigate other discrete-time approximations in the following sections. In order to assess and classify different discrete-time approximations, we introduce their *order of strong convergence*.

Definition 5.2.1. We shall say that a discrete-time approximation Y^Δ converges strongly with order $\gamma > 0$ at time T if there exists a positive constant C , which does not depend on Δ , and a $\delta_0 > 0$ such that

$$\varepsilon(\Delta) = E(|X_T - Y_T^\Delta|) \leq C \Delta^\gamma \quad (5.2.25)$$

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

We emphasize that this criterion has been constructed for the *classification* of strong approximations. There exist in the literature more general results, see, for instance, Kloeden & Platen (1999), which provide uniform estimates. For instance, the supremum of the squared difference between X_t and Y_t^Δ for $t \in [0, T]$ can also be estimated by an expression $C\Delta^\gamma$. These estimates typically provide the same strong order as classified by (5.2.25) but are more difficult to verify or need stronger assumptions. For this reason, the classification criterion (5.2.25) for the order of strong convergence appears to be natural and is sufficient for categorizing strong approximations.

5.3 Strong Taylor Schemes

In this section we use the Wagner-Platen expansion (4.4.7) for the case without jumps to derive discrete-time approximations with respect to the strong convergence criterion (5.2.25). By appropriate truncation of this stochastic Taylor expansion for each discretization step we obtain corresponding strong Taylor schemes of given strong order of convergence.

Operators and Multiple Stochastic Integrals

For compact notation in this and in the following sections we shall use the operators:

$$L^0 = \frac{\partial}{\partial t} + \sum_{k=1}^d a^k \frac{\partial}{\partial x^k} + \frac{1}{2} \sum_{k,\ell=1}^d \sum_{j=1}^m b^{k,j} b^{\ell,j} \frac{\partial}{\partial x^k \partial x^\ell}, \quad (5.3.1)$$

$$\underline{L}^0 = \frac{\partial}{\partial t} + \sum_{k=1}^d \underline{a}^k \frac{\partial}{\partial x^k} \quad (5.3.2)$$

and

$$L^j = \underline{L}^j = \sum_{k=1}^d b^{k,j} \frac{\partial}{\partial x^k} \quad (5.3.3)$$

for $j \in \{1, 2, \dots, m\}$, where

$$\underline{a}^k = a^k - \frac{1}{2} \sum_{j=1}^m \underline{L}^j b^{k,j} \quad (5.3.4)$$

for $k \in \{1, 2, \dots, d\}$. Here the functions a^k and $b^{k,j}$ depend on the time and the state variable. In addition, similar as to Chap. 4, we shall abbreviate, for the n th discretization interval, multiple Itô integrals by

$$I_{(j_1, \dots, j_\ell)} = \int_{t_n}^{t_{n+1}} \dots \int_{t_n}^{s_2} dW_{s_1}^{j_1} \dots dW_{s_\ell}^{j_\ell}, \quad (5.3.5)$$

see (4.2.9), and multiple Stratonovich integrals analogously by

$$J_{(j_1, \dots, j_\ell)} = \int_{t_n}^{t_{n+1}} \dots \int_{t_n}^{s_2} \circ dW_{s_1}^{j_1} \dots \circ dW_{s_\ell}^{j_\ell}, \quad (5.3.6)$$

where we use Stratonovich integration instead of Itô integration, see Kloeden & Platen (1999), for $j_1, \dots, j_\ell \in \{0, 1, \dots, m\}$, $\ell \in \{1, 2, \dots\}$ and $n \in \{0, 1, \dots\}$ with the convention that

$$W_t^0 = t \quad (5.3.7)$$

for all $t \in [0, T]$. In (5.3.5) and (5.3.6) we suppress the dependence on t_n and Δ . Similarly, as introduced in (5.2.11), we will use in the following schemes the abbreviation $f = f(t_n, Y_n)$, for $n \in \{0, 1, \dots\}$ and any function f defined on $[0, T] \times \mathbb{R}^d$. Furthermore, we will usually not explicitly mention the initial value $\mathbf{Y}_0 = \mathbf{X}_0$. These conventions simplify the formulae for the various schemes that we will discuss later on.

If we refer in this section to the order of convergence of a scheme, then we mean the order of strong convergence of the corresponding discrete-time approximation. Finally, we suppose that the vector process \mathbf{X} under consideration satisfies the Itô SDE

$$d\mathbf{X}_t = \mathbf{a}(t, \mathbf{X}_t) dt + \sum_{j=1}^m \mathbf{b}^j(t, \mathbf{X}_t) dW_t^j, \quad (5.3.8)$$

which is equivalent to the Stratonovich SDE

$$d\mathbf{X}_t = \underline{\mathbf{a}}(t, \mathbf{X}_t) dt + \sum_{j=1}^m \mathbf{b}^j(t, \mathbf{X}_t) \circ dW_t^j \quad (5.3.9)$$

for $t \in [0, T]$ with $\mathbf{X}_0 \in \mathbb{R}^d$ and adjusted drift $\underline{\mathbf{a}}$, see (4.1.11) and Kloeden & Platen (1999).

Euler Scheme

Let us begin with the Euler scheme (5.2.21), which has been already mentioned in the previous section. It represents the simplest strong Taylor approximation and, as we shall see, attains the order of strong convergence $\gamma = 0.5$.

In the one-dimensional case, $d = m = 1$, the *Euler scheme* has the form

$$Y_{n+1} = Y_n + a \Delta + b \Delta W, \quad (5.3.10)$$

where

$$\Delta = t_{n+1} - t_n \quad (5.3.11)$$

is the length of the time discretization subinterval $[t_n, t_{n+1}]$ and we write in the following

$$\Delta W = \Delta W_n = W_{t_{n+1}} - W_{t_n} \quad (5.3.12)$$

for the $N(0, \Delta)$ independent Gaussian distributed increment of the Wiener process W on $[t_n, t_{n+1}]$. Note that we use here and in the following our abbreviation, where we suppress in the coefficient functions the dependence on time t_n and approximate value Y_n of the last discretization point. We also suppress the dependence of the random Wiener process increments on n .

In the multi-dimensional case with scalar noise $m = 1$ and $d \in \{1, 2, \dots\}$, the k th component of the *Euler scheme* is given by

$$Y_{n+1}^k = Y_n^k + a^k \Delta + b^k \Delta W, \quad (5.3.13)$$

for $k \in \{1, 2, \dots, d\}$, where the drift and diffusion coefficients are d -dimensional vectors $\mathbf{a} = (a^1, \dots, a^d)^\top$ and $\mathbf{b} = (b^1, \dots, b^d)^\top$.

For the general multi-dimensional case with $d, m \in \{1, 2, \dots\}$ the k th component of the *Euler scheme* has the form

$$Y_{n+1}^k = Y_n^k + a^k \Delta + \sum_{j=1}^m b^{k,j} \Delta W^j. \quad (5.3.14)$$

Here

$$\Delta W^j = W_{t_{n+1}}^j - W_{t_n}^j \quad (5.3.15)$$

is the $N(0, \Delta)$ independent Gaussian distributed increment of the j th component of the m -dimensional standard Wiener process W on $[t_n, t_{n+1}]$, and ΔW^{j_1} and ΔW^{j_2} are independent for $j_1 \neq j_2$. The diffusion coefficient $\mathbf{b} = [b^{k,j}]_{k,j=1}^{d,m}$ is here a $d \times m$ -matrix.

The Euler scheme (5.3.10) corresponds to a truncated Wagner-Platen expansion, see (4.4.2) or (4.4.7), containing only the time and Wiener integrals of multiplicity one. Assuming Lipschitz and linear growth conditions on the coefficients \mathbf{a} and \mathbf{b} , we will see that the Euler approximation is of strong order $\gamma = 0.5$.

In special cases, this means for specific SDEs, the Euler scheme may actually sometimes achieve a higher order of strong convergence. For example, when the noise is *additive*, that is when the diffusion coefficient is a differentiable, deterministic function of time and has the form

$$\mathbf{b}(t, \mathbf{x}) = \mathbf{b}_t \quad (5.3.16)$$

for all $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d$, then the Euler scheme has typically the order of strong convergence $\gamma = 1.0$.

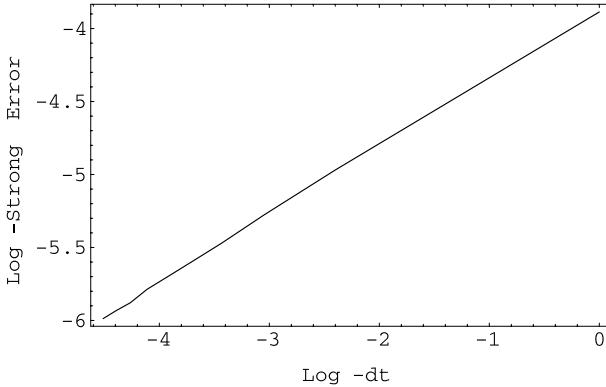


Fig. 5.3.1. Log-log plot of the absolute error $\varepsilon(\Delta)$ for an Euler scheme against $\ln(\Delta)$

The Euler scheme gives reasonable numerical results when the drift and diffusion coefficients are nearly constant and the time step size is sufficiently small. In general however, and if possible, the use of more sophisticated higher order schemes is recommended, which we will introduce below.

A Simulation Example

Let us check experimentally the order $\gamma = 0.5$ of strong convergence for the Euler scheme in our standard example. We suppose that X satisfies the Black-Scholes SDE

$$dX_t = \mu X_t dt + \sigma X_t dW_t \quad (5.3.17)$$

for $t \in [0, T]$ with initial value $X_0 > 0$, appreciation rate $\mu > 0$ and volatility $\sigma > 0$.

In this case X is a geometric Brownian motion and we compare the Euler approximation with the exact solution

$$X_T = X_0 \exp \left\{ \left(\mu - \frac{\sigma^2}{2} \right) T + \sigma W_T \right\}, \quad (5.3.18)$$

see (5.2.20). In the corresponding log-log plot we show in Fig. 5.3.1 for the Euler scheme the logarithm of the absolute error $\ln(\varepsilon(\Delta))$ against $\ln(\Delta)$. We observe in this figure how the absolute error $\varepsilon(\Delta) = E(|X_T - Y_N^\Delta|)$ behaves for increasing time step size. We have used the following default parameters: $X_0 = 1$, $\mu = 0.06$, $\sigma = 0.2$ and $T = 1$. Furthermore, we generated 5000 simulations, where the time step size Δ ranges from 0.01 to 1. One observes in Fig. 5.3.1 an almost perfect line. We omit any confidence intervals since they are quite small and can be further reduced by increasing the number of simulations.

Regressing the log of the absolute error on the log of the size of the time step size provides the following fitted line

$$\ln(\varepsilon(\Delta)) = -3.86933 + 0.46739 \ln(\Delta).$$

The empirical order γ of strong convergence is with about 0.47 close to the theoretical strong order of $\gamma = 0.5$ for the Euler scheme.

Milstein Scheme

We shall now introduce a scheme, the *Milstein scheme*, which turns out to be of strong order $\gamma = 1.0$. It is obtained by including one additional term from the Wagner-Platen expansion (4.4.7). This scheme was first suggested in Milstein (1974).

The Milstein scheme has in the one-dimensional case with $d = m = 1$ the form

$$Y_{n+1} = Y_n + a \Delta + b \Delta W + \frac{1}{2} b b' \{ (\Delta W)^2 - \Delta \}. \quad (5.3.19)$$

Recall that we suppress here in the coefficient functions the dependence on t_n and Y_n . The additional term in the Milstein scheme, when compared with the Euler scheme, marks the point of divergence of stochastic numerical analysis from deterministic numerics. In some sense one can regard the Milstein scheme as the proper stochastic generalization of the deterministic Euler scheme for the strong convergence criterion. As we will show later, it gives the same order $\gamma = 1.0$ of strong convergence as the Euler scheme does in the deterministic case.

In the multi-dimensional case with $m = 1$ and $d \in \{1, 2, \dots\}$ the k th component of the Milstein scheme is given by

$$Y_{n+1}^k = Y_n^k + a^k \Delta + b^k \Delta W + \left(\sum_{\ell=1}^d b^\ell \frac{\partial b^k}{\partial x^\ell} \right) \frac{1}{2} \{ (\Delta W)^2 - \Delta \}. \quad (5.3.20)$$

In the general multi-dimensional case with $d, m \in \{1, 2, \dots\}$ the k th component of the Milstein scheme has the form

$$Y_{n+1}^k = Y_n^k + a^k \Delta + \sum_{j=1}^m b^{k,j} \Delta W^j + \sum_{j_1, j_2=1}^m L^{j_1} b^{k, j_2} I_{(j_1, j_2)} \quad (5.3.21)$$

involving terms with some multiple Itô integrals $I_{(j_1, j_2)}$. Alternatively, by using multiple Stratonovich integrals $J_{(j_1, j_2)}$ this scheme can be written by (5.3.4) in the form

$$Y_{n+1}^k = Y_n^k + \underline{a}^k \Delta + \sum_{j=1}^m b^{k,j} \Delta W^j + \sum_{j_1, j_2=1}^m \underline{L}^{j_1} b^{k, j_2} J_{(j_1, j_2)}. \quad (5.3.22)$$

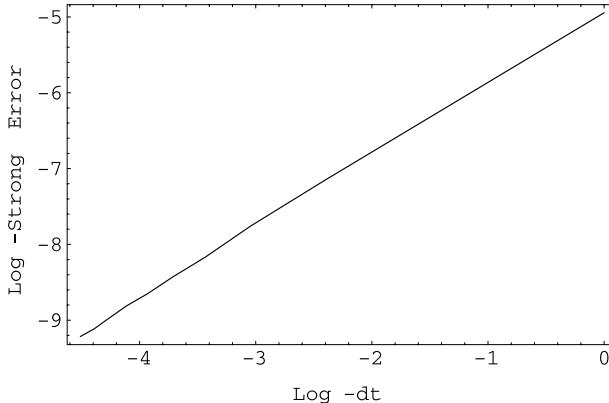


Fig. 5.3.2. Log-log plot of the absolute error against log-step size for Milstein scheme

We remark that for $j_1 \neq j_2$ with $j_1, j_2 \in \{1, 2, \dots, m\}$ the multiple stochastic integrals

$$J_{(j_1, j_2)} = I_{(j_1, j_2)} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} dW_s^{j_1} dW_s^{j_2}, \quad (5.3.23)$$

appearing above, are equal. They cannot be simply expressed in terms of the increments ΔW^{j_1} and ΔW^{j_2} of the components of the Wiener process and some effort has to be made to generate or approximate these, as we will discuss later. However, in the case $j_1 = j_2$ we have

$$I_{(j_1, j_1)} = \frac{1}{2} \{(\Delta W^{j_1})^2 - \Delta\} \quad \text{and} \quad J_{(j_1, j_1)} = \frac{1}{2} (\Delta W^{j_1})^2, \quad (5.3.24)$$

for $j_1 \in \{1, 2, \dots, m\}$, see (4.2.16).

Let us generate for the Milstein scheme the corresponding log-log plot of the logarithm of the absolute error against the logarithm of the step size for the previous example (5.3.17). We can see in Fig. 5.3.2 the log-absolute error for a geometric Brownian motion when approximated with a Milstein scheme by using the same default parameters as in Fig. 5.3.1.

Regressing the log of the absolute error on the log of the time step size we obtain the following fitted line

$$\ln(\varepsilon(\Delta)) = -4.91021 + 0.95 \ln(\Delta).$$

The experimentally determined estimate for the order of strong convergence is about 0.95 and is thus close to the theoretical value of $\gamma = 1.0$.

Diffusion Commutativity Condition

For some important dynamics in financial modeling the corresponding SDEs have special structural properties, which allow the Milstein scheme to be sim-

plified. This avoids in some cases the use of double integrals of the type (5.3.23) that involve different Wiener processes. For instance, in the case with *additive noise*, that is when the diffusion coefficients depend at most on time t and not on the variable x , the Milstein scheme (5.3.19) reduces to the Euler scheme (5.3.10).

Another important special case is that of commutative noise for which the diffusion matrix satisfies the *diffusion commutativity condition*

$$L^{j_1} b^{k,j_2}(t, \mathbf{x}) = L^{j_2} b^{k,j_1}(t, \mathbf{x}) \quad (5.3.25)$$

for all $j_1, j_2 \in \{1, 2, \dots, m\}$, $k \in \{1, 2, \dots, d\}$ and $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d$. For instance, the diffusion commutativity condition is for asset prices satisfied if one simulates systems of Black-Scholes SDEs, see (1.7.25) and (1.7.30). Systems of SDEs with additive noise also satisfy the diffusion commutativity condition (5.3.25). Obviously, the diffusion commutativity condition is satisfied if the given SDE is driven by a single Wiener process only.

It turns out that the *Milstein scheme* under the diffusion commutativity condition can be written as

$$Y_{n+1}^k = Y_n^k + \underline{a}^k \Delta + \sum_{j=1}^m b^{k,j} \Delta W^j + \frac{1}{2} \sum_{j_1, j_2=1}^m \underline{L}^{j_1} b^{k,j_2} \Delta W^{j_1} \Delta W^{j_2} \quad (5.3.26)$$

for $k \in \{1, 2, \dots, d\}$. This scheme does not use any double integrals as given in (5.3.23).

A Black-Scholes Example

A typical example is given by the simulation of correlated Black-Scholes dynamics with $d = m = 2$, where the first risky security satisfies the SDE

$$\begin{aligned} dX_t^1 &= X_t^1 [r dt + \theta^1(\theta^1 dt + dW_t^1) + \theta^2(\theta^2 dt + dW_t^2)] \\ &= X_t^1 \left[\left(r + \frac{1}{2} ((\theta^1)^2 + (\theta^2)^2) \right) dt + \theta^1 \circ dW_t^1 + \theta^2 \circ dW_t^2 \right] \end{aligned}$$

and the second risky security is given by the SDE

$$\begin{aligned} dX_t^2 &= X_t^2 [r dt + (\theta^1 - \sigma^{1,1})(\theta^1 dt + dW_t^1)] \\ &= X_t^2 \left[\left(r + (\theta^1 - \sigma^{1,1}) \left(\frac{1}{2} \theta^1 + \frac{1}{2} \sigma^{1,1} \right) \right) dt + (\theta^1 - \sigma^{1,1}) \circ dW_t^1 \right] \end{aligned}$$

for $t \in [0, T]$. Here W^1 and W^2 are independent Wiener processes and $r, \theta^1, \theta^2, \sigma^{1,1} > 0$ are assumed to be strictly positive constants. Obviously, we obtain

$$L^1 b^{1,2} = \sum_{k=1}^2 b^{k,1} \frac{\partial}{\partial x^k} b^{1,2} = \theta^1 X_t^1 \theta^2 = \sum_{k=1}^2 b^{k,2} \frac{\partial}{\partial x^k} b^{1,1} = L^2 b^{1,1}$$

and

$$L^1 b^{2,2} = \sum_{k=1}^2 b^{k,1} \frac{\partial}{\partial x^k} b^{2,2} = 0 = \sum_{k=1}^2 b^{k,2} \frac{\partial}{\partial x^k} b^{2,1} = L^2 b^{2,1}$$

such that the diffusion commutativity condition (5.3.25) is satisfied. The corresponding Milstein scheme is then of the form

$$\begin{aligned} Y_{n+1}^1 &= Y_n^1 + Y_n^1 \left(r + \frac{1}{2} ((\theta^1)^2 + (\theta^2)^2) \Delta + \theta^1 \Delta W^1 + \theta^2 \Delta W^2 \right. \\ &\quad \left. + \frac{1}{2} (\theta^1)^2 (\Delta W^1)^2 + \frac{1}{2} (\theta^2)^2 (\Delta W^2)^2 + \theta^1 \theta^2 \Delta W^1 \Delta W^2 \right) \end{aligned} \quad (5.3.27)$$

$$\begin{aligned} Y_{n+1}^2 &= Y_n^2 + Y_n^2 \left(\left(r + \frac{1}{2} (\theta^1 - \sigma^{1,1}) (\theta^1 + \sigma^{1,1}) \right) \Delta + (\theta^1 - \sigma^{1,1}) \Delta W^1 \right. \\ &\quad \left. + \frac{1}{2} (\theta^1 - \sigma^{1,1})^2 (\Delta W^1)^2 \right). \end{aligned} \quad (5.3.28)$$

Note that if the diffusion commutativity condition (5.3.25) were not satisfied, then multiple stochastic integrals with respect to the two Wiener processes would appear. This kind of situation will be discussed later. We remark that the scheme (5.3.27)–(5.3.28) may produce negative trajectories, which could create severe problems for certain applications. As shown in Sect. 2.1, there exist alternative methods that keep the approximate security prices under the Black-Scholes model positive.

A Square Root Process Example

As another example let us consider the simulation of a real valued square root process X of dimension ν , see (2.1.2), with SDE

$$dX_t = \frac{\nu}{4} \eta \left(\frac{1}{\eta} - X_t \right) dt + \sqrt{X_t} dW_t \quad (5.3.29)$$

for $t \in [0, T]$ and $X_0 = \frac{1}{\eta}$ for $\nu > 2$. Since there is only one driving Wiener process the noise is commutative. The Milstein scheme here has according to (5.3.19), (5.3.20) or (5.3.26) the form

$$Y_{n+1} = Y_n + \frac{\nu}{4} \eta \left(\frac{1}{\eta} - Y_n \right) \Delta + \sqrt{Y_n} \Delta W + \frac{1}{4} (\Delta W^2 - \Delta). \quad (5.3.30)$$

This scheme generates, for the parameters $\nu = 4$, $\eta = 0.048$, $T = 100$ and step size $\Delta = 0.05$ the trajectory exhibited in Fig. 5.3.3. Note that the above scheme (5.3.30) can generate a negative trajectory, which may cause problems in applications. Alternative methods that preserve nonnegativity have been described in Chap. 2. Furthermore, we remark that the strong convergence results, as given in Kloeden & Platen (1999), require Lipschitz continuity of the diffusion coefficient in (5.3.29), which is not given. Still the scheme performs quite well when the trajectory does not come too close to zero.

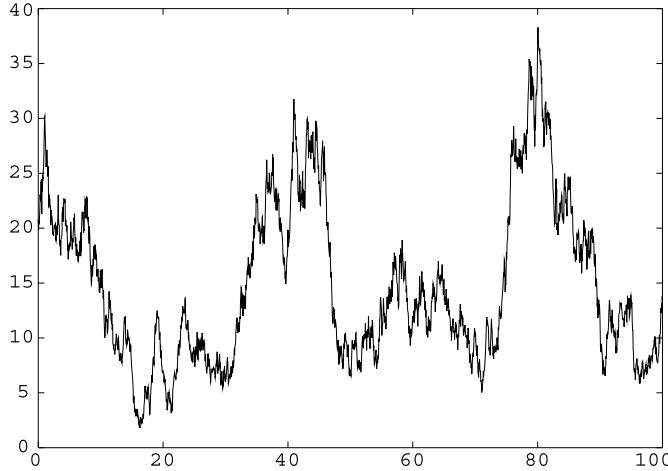


Fig. 5.3.3. Square root process simulated by the Milstein scheme

Double Wiener Integrals and Lévy Area

In the multi-dimensional case application of the Milstein scheme is practicable with the following approximation of the multiple Wiener integral that appears in (5.3.23). For $j_1 \neq j_2$ with $j_1, j_2 \in \{1, 2, \dots, m\}$ one can approximate the double stochastic integral $J_{(j_1, j_2)} = I_{(j_1, j_2)}$ by

$$\begin{aligned} J_{(j_1, j_2)}^p &= \Delta \left(\frac{1}{2} \xi_{j_1} \xi_{j_2} + \sqrt{\varrho_p} (\mu_{j_1, p} \xi_{j_2} - \mu_{j_2, p} \xi_{j_1}) \right) \\ &+ \frac{\Delta}{2\pi} \sum_{r=1}^p \frac{1}{r} \left(\zeta_{j_1, r} (\sqrt{2} \xi_{j_2} + \eta_{j_2, r}) - \zeta_{j_2, r} (\sqrt{2} \xi_{j_1} + \eta_{j_1, r}) \right), \end{aligned} \quad (5.3.31)$$

where

$$\varrho_p = \frac{1}{12} - \frac{1}{2\pi^2} \sum_{r=1}^p \frac{1}{r^2} \quad (5.3.32)$$

and $\xi_j, \mu_{j,p}, \eta_{j,r}$ and $\zeta_{j,r}$ are independent $N(0, 1)$ Gaussian distributed random variables with

$$\xi_j = \frac{1}{\sqrt{\Delta}} \Delta W^j \quad (5.3.33)$$

for $j \in \{1, 2, \dots, m\}$, $r \in \{1, \dots, p\}$ and $p \in \{1, 2, \dots\}$. The number of terms p in the expansion (5.3.31) influences the difference between $J_{(j_1, j_2)}^p$ and $J_{(j_1, j_2)}$. It has been shown in Kloeden & Platen (1999) that if one chooses

$$p = p(\Delta) \geq \frac{K}{\Delta} \quad (5.3.34)$$

for some positive constant K , then one obtains the order of strong convergence $\gamma = 1.0$ for the Milstein scheme with the above approximation of the double stochastic integral (5.3.23).

An alternative way to approximate these multiple stochastic integrals is the following:

We apply the Euler scheme to the SDE which describes the dynamics of the required multiple stochastic integrals with a time step size δ , smaller than the original time step size Δ . Since the Euler scheme has strong order of convergence $\gamma = 0.5$, by choosing $\delta = \Delta^2$, the strong order $\gamma = 1.0$ of the scheme is preserved, see [Milstein \(1995a\)](#) and [Kloeden \(2002\)](#).

Another way to obtain the desired double Wiener integrals is the technique of [Gaines & Lyons \(1994\)](#) who suggest sampling the distribution of the Lévy area $I_{(j_1,j_2)} - I_{(j_2,j_1)}$, $j_1 \neq j_2$, conditional on ΔW^{j_1} and ΔW^{j_2} , which is known. Unfortunately, this technique does not apply for $m > 2$.

Strong Order 1.5 Taylor Scheme

There are certain simulation tasks that require more accurate schemes than the Milstein scheme. For instance, if one wants to study the distribution of extreme log-returns one needs higher order schemes to capture reasonably well the behavior of the trajectory in the tails of the log-return distribution. In general, we obtain more accurate strong Taylor schemes by including further multiple stochastic integrals from the Wagner-Platen expansion into the scheme. Each of these multiple stochastic integrals contains additional information about the sample path of the driving Wiener process. The necessity of the inclusion of further multiple stochastic integrals to achieve higher order convergence is a fundamental feature of the numerical analysis of SDEs compared to that of deterministic ODEs. For ODEs it was enough to have sufficient smoothness in the drift coefficient to obtain higher order schemes.

By adding more terms from the expansion (4.4.7) to the Milstein scheme one obtains, as in [Platen & Wagner \(1982\)](#), in the autonomous one-dimensional case $d = m = 1$ the *strong order 1.5 Taylor scheme*

$$\begin{aligned} Y_{n+1} = & Y_n + a \Delta + b \Delta W + \frac{1}{2} b b' \{ (\Delta W)^2 - \Delta \} + a' b \Delta Z \\ & + \frac{1}{2} \left(a a' + \frac{1}{2} b^2 a'' \right) \Delta^2 + \left(a b' + \frac{1}{2} b^2 b'' \right) \{ \Delta W \Delta - \Delta Z \} \\ & + \frac{1}{2} b (b b'' + (b')^2) \left\{ \frac{1}{3} (\Delta W)^2 - \Delta \right\} \Delta W. \end{aligned} \quad (5.3.35)$$

Here the additional random variable ΔZ is required, which represents the double integral

$$\Delta Z = I_{(1,0)} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} dW_{s_1} ds_2. \quad (5.3.36)$$

One can show that ΔZ is Gaussian distributed with mean $E(\Delta Z) = 0$, variance $E((\Delta Z)^2) = \frac{1}{3}\Delta^3$ and covariance $E(\Delta Z \Delta W) = \frac{1}{2}\Delta^2$.

Note that a pair of appropriately correlated normally distributed random variables $(\Delta W, \Delta Z)$ can be constructed from two independent $N(0, 1)$ standard Gaussian distributed random variables U_1 and U_2 by means of the linear transformation

$$\Delta W = U_1 \sqrt{\Delta}, \quad \Delta Z = \frac{1}{2} \Delta^{\frac{3}{2}} \left(U_1 + \frac{1}{\sqrt{3}} U_2 \right). \quad (5.3.37)$$

All multiple stochastic integrals that appear in (4.4.7) and which are also built into (5.3.35) can be expressed in terms of Δ , ΔW and ΔZ . In particular, the last term in (5.3.35) contains the triple Wiener integral

$$I_{(1,1,1)} = \frac{1}{2} \left\{ \frac{1}{3} (\Delta W^1)^2 - \Delta \right\} \Delta W^1, \quad (5.3.38)$$

as described by the corresponding Hermite polynomial given in (4.2.17).

In the multi-dimensional case with $d \in \{1, 2, \dots\}$ and $m = 1$ the k th component of the strong order 1.5 Taylor scheme is given by

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + a^k \Delta + b^k \Delta W \\ &\quad + \frac{1}{2} L^1 b^k \{(\Delta W)^2 - \Delta\} + L^1 a^k \Delta Z \\ &\quad + L^0 b^k \{\Delta W \Delta - \Delta Z\} + \frac{1}{2} L^0 a^k \Delta^2 \\ &\quad + \frac{1}{2} L^1 L^1 b^k \left\{ \frac{1}{3} (\Delta W)^2 - \Delta \right\} \Delta W. \end{aligned} \quad (5.3.39)$$

In the general multi-dimensional case with $d, m \in \{1, 2, \dots\}$ the k th component of the strong order 1.5 Taylor scheme takes the form

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + a^k \Delta + \frac{1}{2} L^0 a^k \Delta^2 \\ &\quad + \sum_{j=1}^m (b^{k,j} \Delta W^j + L^0 b^{k,j} I_{(0,j)} + L^j a^k I_{(j,0)}) \\ &\quad + \sum_{j_1, j_2=1}^m L^{j_1} b^{k,j_2} I_{(j_1, j_2)} + \sum_{j_1, j_2, j_3=1}^m L^{j_1} L^{j_2} b^{k,j_3} I_{(j_1, j_2, j_3)}. \end{aligned} \quad (5.3.40)$$

In important practical situations, for instance in the case of additive noise, the strong Taylor scheme (5.3.40) simplifies considerably. It will be shown, see also Kloeden & Platen (1999), that the above strong schemes have, under appropriate assumptions on the drift and diffusion coefficients, the strong order $\gamma = 1.5$.

Approximate Multiple Itô Integrals

As with the general multi-dimensional version of the Milstein scheme (5.3.21), we also have to approximate some multiple stochastic integrals involving different components of the Wiener process. Otherwise, it is difficult to implement the above schemes in practice for $m > 1$, because there are no explicit expressions for the multiple stochastic integrals directly available. We continue the approach taken in (5.3.31), which is based on a Brownian bridge type expansion of the Wiener process, see Kloeden & Platen (1999).

Let $\xi_j, \zeta_{j,1}, \dots, \zeta_{j,p}, \eta_{j,1}, \dots, \eta_{j,p}, \mu_{j,p}$ and $\phi_{j,p}$ be independent standard Gaussian random variables. Then for $j, j_1, j_2, j_3 \in \{1, 2, \dots, m\}$ and some $p \in \{1, 2, \dots\}$ we set

$$I_{(j)} = \Delta W^j = \sqrt{\Delta} \xi_j, \quad I_{(j,0)} = \frac{1}{2} \Delta \left(\sqrt{\Delta} \xi_j + a_{j,0} \right) \quad (5.3.41)$$

with

$$a_{j,0} = -\frac{\sqrt{2\Delta}}{\pi} \sum_{r=1}^p \frac{1}{r} \zeta_{j,r} - 2\sqrt{\Delta} \varrho_p \mu_{j,p},$$

where

$$\varrho_p = \frac{1}{12} - \frac{1}{2\pi^2} \sum_{r=1}^p \frac{1}{r^2},$$

$$I_{(0,j)} = \Delta W^j \Delta - I_{(j,0)}, \quad I_{(j,j)} = \frac{1}{2} \{ (\Delta W^j)^2 - \Delta \},$$

$$I_{(j,j,j)} = \frac{1}{2} \left\{ \frac{1}{3} (\Delta W^j)^2 - \Delta \right\} \Delta W^j,$$

$$I_{(j_1,j_2)}^p = \frac{1}{2} \Delta \xi_{j_1} \xi_{j_2} - \frac{1}{2} \sqrt{\Delta} (\xi_{j_1} a_{j_2,0} - \xi_{j_2} a_{j_1,0}) + A_{j_1,j_2}^p \Delta$$

for $j_1 \neq j_2$, where

$$A_{j_1,j_2}^p = \frac{1}{2\pi} \sum_{r=1}^p \frac{1}{r} (\zeta_{j_1,r} \eta_{j_2,r} - \eta_{j_1,r} \zeta_{j_2,r}),$$

and

$$I_{(j_1,j_2,j_3)}^p = J_{(j_1,j_2,j_3)}^p - \frac{1}{2} (\mathbf{1}_{\{j_1=j_2\}} I_{(0,j_3)} + \mathbf{1}_{\{j_2=j_3\}} I_{(j_1,0)})$$

with

$$\begin{aligned} J_{(j_1,j_2,j_3)}^p &= \frac{1}{\sqrt{\Delta}} \xi_{j_1} J_{(0,j_2,j_3)}^p + \frac{1}{2} a_{j_1,0} J_{(j_2,j_3)}^p + \frac{1}{2\pi} \Delta b_{j_1} \xi_{j_2} \xi_{j_3} \\ &\quad - \Delta^{\frac{3}{2}} \xi_{j_2} B_{j_1,j_3}^p + \Delta^{\frac{3}{2}} \xi_{j_3} \left(\frac{1}{2} A_{j_1,j_2}^p - C_{j_2,j_1}^p \right) + \Delta^{\frac{3}{2}} D_{j_1,j_2,j_3}^p \end{aligned}$$

with

$$\begin{aligned}
D_{j_1, j_2, j_3}^p = & -\frac{1}{\pi^2 2^{\frac{5}{2}}} \sum_{r, \ell=1}^p \frac{1}{l(\ell+r)} [\zeta_{j_2, \ell} (\zeta_{j_3, \ell+r} \eta_{j_1, r} - \zeta_{j_1, r} \eta_{j_3, \ell+r}) \\
& + \eta_{j_2, \ell} (\zeta_{j_1, r} \zeta_{j_3, \ell+r} + \eta_{j_1, r} \eta_{j_3, \ell+r})] \\
& + \frac{1}{\pi^2 2^{\frac{5}{2}}} \sum_{\ell=1}^p \sum_{r=1}^{\ell-1} \frac{1}{r(\ell-r)} [\zeta_{j_2, \ell} (\zeta_{j_1, r} \eta_{j_3, \ell-r} + \zeta_{j_3, \ell-r} \eta_{j_1, r}) \\
& - \eta_{j_2, \ell} (\zeta_{j_1, r} \zeta_{j_3, \ell-r} - \eta_{j_1, r} \eta_{j_3, \ell-r})] \\
& + \frac{1}{\pi^2 2^{\frac{5}{2}}} \sum_{\ell=1}^p \sum_{r=\ell+1}^{2p} \frac{1}{r(r-\ell)} [\zeta_{j_2, \ell} (\zeta_{j_3, r-\ell} \eta_{j_1, r} - \zeta_{j_1, r} \eta_{j_3, r-\ell}) \\
& + \eta_{j_2, \ell} (\zeta_{j_1, r} \zeta_{j_3, r-\ell} + \eta_{j_1, r} \eta_{j_3, r-\ell})],
\end{aligned}$$

where, for $r > p$ we set $\eta_{j,r} = 0$ and $\zeta_{j,r} = 0$ for all $j \in \{1, 2, \dots, m\}$. Here $\mathbf{1}_{\{A\}}$ denotes the indicator function for the event A .

One can show, see [Kloeden & Platen \(1999\)](#), that if one chooses p so that

$$p = p(\Delta) \geq \frac{K}{\Delta^2} \quad (5.3.42)$$

for an appropriate positive constant K , then the scheme (5.3.40) with the above approximations of multiple stochastic integrals converges with strong order $\gamma = 1.5$.

We remark, as mentioned earlier, that one can alternatively take a very fine time discretization of each time interval $[t_n, t_{n+1}]$ and simply approximate the multiple Itô integrals as suggested by the definition of the Itô integral.

Strong Order 2.0 Taylor Scheme

We now use the Stratonovich-Taylor expansion from [Kloeden & Platen \(1999\)](#), and retain those terms which are needed to obtain a strong order 2.0 scheme.

In the one-dimensional case $d = m = 1$ with time independent a and b the *strong order 2.0 Taylor scheme* has the form

$$\begin{aligned}
Y_{n+1} = & Y_n + \underline{a} \Delta + b \Delta W + \frac{1}{2!} b b' (\Delta W)^2 + b \underline{a}' \Delta Z \\
& + \frac{1}{2} \underline{a} \underline{a}' \Delta^2 + \underline{a} b' \{ \Delta W \Delta - \Delta Z \} \\
& + \frac{1}{3!} b (b b')' (\Delta W)^3 + \frac{1}{4!} b \left(b (b b')' \right)' (\Delta W)^4 \\
& + \underline{a} (b b')' J_{(0,1,1)} + b (\underline{a} b')' J_{(1,0,1)} + b (b \underline{a}')' J_{(1,1,0)}. \quad (5.3.43)
\end{aligned}$$

The Gaussian random variables ΔW and ΔZ here are the same as given in (5.3.37), where ΔZ is defined in (5.3.36).

Approximate Multiple Stratonovich Integrals

Also for this scheme we prepare a set of approximate multiple stochastic integrals. According to [Kloeden & Platen \(1999\)](#), the multiple Stratonovich integrals of multiplicity three appearing in the above scheme can be approximated by the random variables

$$\Delta W = J_{(1)}^p = \sqrt{\Delta} \zeta_1, \quad \Delta Z = J_{(1,0)}^p = \frac{1}{2} \Delta \left(\sqrt{\Delta} \zeta_1 + a_{1,0} \right),$$

$$J_{(1,0,1)}^p = \frac{1}{3!} \Delta^2 \zeta_1^2 - \frac{1}{4} \Delta a_{1,0}^2 + \frac{1}{\pi} \Delta^{\frac{3}{2}} \zeta_1 b_1 - \Delta^2 B_{1,1}^p,$$

$$J_{(0,1,1)}^p = \frac{1}{3!} \Delta^2 \zeta_1^2 - \frac{1}{2\pi} \Delta^{\frac{3}{2}} \zeta_1 b_1 + \Delta^2 B_{1,1}^p - \frac{1}{4} \Delta^{\frac{3}{2}} a_{1,0} \zeta_1 + \Delta^2 C_{1,1}^p,$$

$$J_{(1,1,0)}^p = \frac{1}{3!} \Delta^2 \zeta_1^2 + \frac{1}{4} \Delta a_{1,0}^2 - \frac{1}{2\pi} \Delta^{\frac{3}{2}} \zeta_1 b_1 + \frac{1}{4} \Delta^{\frac{3}{2}} a_{1,0} \zeta_1 - \Delta^2 C_{1,1}^p \quad (5.3.44)$$

with

$$a_{1,0} = -\frac{1}{\pi} \sqrt{2\Delta} \sum_{r=1}^p \frac{1}{r} \xi_{1,r} - 2\sqrt{\Delta \varrho_p} \mu_{1,p},$$

$$\varrho_p = \frac{1}{12} - \frac{1}{2\pi^2} \sum_{r=1}^p \frac{1}{r^2},$$

$$b_1 = \sqrt{\frac{\Delta}{2}} \sum_{r=1}^p \frac{1}{r^2} \eta_{1,r} + \sqrt{\Delta \alpha_p} \phi_{1,p},$$

$$\alpha_p = \frac{\pi^2}{180} - \frac{1}{2\pi^2} \sum_{r=1}^p \frac{1}{r^4},$$

$$B_{1,1}^p = \frac{1}{4\pi^2} \sum_{r=1}^p \frac{1}{r^2} (\xi_{1,r}^2 + \eta_{1,r}^2)$$

and

$$C_{1,1}^p = -\frac{1}{2\pi^2} \sum_{\substack{r,l=1 \\ r \neq l}}^p \frac{r}{r^2 - l^2} \left(\frac{1}{l} \xi_{1,r} \xi_{1,\ell} - \frac{l}{r} \eta_{1,r} \eta_{1,\ell} \right).$$

Here $\zeta_1, \xi_{1,r}, \eta_{1,r}, \mu_{1,p}$ and $\phi_{1,p}$ denote for $r \in \{1, \dots, p\}$ and $p \in \{1, 2, \dots\}$ independent standard Gaussian random variables.

Multi-dimensional Strong Order 2.0 Taylor Scheme

For the multi-dimensional case $d \in \{1, 2, \dots\}$ but with scalar noise, that is for $m = 1$, the k th component of the strong order 2.0 Taylor scheme is given by

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + \underline{a}^k \Delta + b^k \Delta W + \frac{1}{2!} \underline{L}^1 b^k (\Delta W)^2 + \underline{L}^1 \underline{a}^k \Delta Z \\ &\quad + \frac{1}{2} \underline{L}^0 \underline{a}^k \Delta^2 + \underline{L}^0 b^k \{\Delta W \Delta - \Delta Z\} \\ &\quad + \frac{1}{3!} \underline{L}^1 \underline{L}^1 b^k (\Delta W)^3 + \frac{1}{4!} \underline{L}^1 \underline{L}^1 \underline{L}^1 b^k (\Delta W)^4 \\ &\quad + \underline{L}^0 \underline{L}^1 b^k J_{(0,1,1)} + \underline{L}^1 \underline{L}^0 b^k J_{(1,0,1)} + \underline{L}^1 \underline{L}^1 \underline{a}^k J_{(1,1,0)}. \end{aligned} \quad (5.3.45)$$

In the general multi-dimensional case with $d, m \in \{1, 2, \dots\}$ the k th component of the *strong order 2.0 Taylor scheme* has the form

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + \underline{a}^k \Delta + \frac{1}{2} \underline{L}^0 \underline{a}^k \Delta^2 + \sum_{j=1}^m (b^{k,j} \Delta W^j + \underline{L}^0 b^{k,j} J_{(0,j)} + \underline{L}^j \underline{a}^k J_{(j,0)}) \\ &\quad + \sum_{j_1, j_2=1}^m \left(\underline{L}^{j_1} b^{k,j_2} J_{(j_1, j_2)} + \underline{L}^0 \underline{L}^{j_1} b^{k,j_2} J_{(0, j_1, j_2)} \right. \\ &\quad \left. + \underline{L}^{j_1} \underline{L}^0 b^{k,j_2} J_{(j_1, 0, j_2)} + \underline{L}^{j_1} \underline{L}^{j_2} \underline{a}^k J_{(j_1, j_2, 0)} \right) \\ &\quad + \sum_{j_1, j_2, j_3=1}^m \underline{L}^{j_1} \underline{L}^{j_2} b^{k,j_3} J_{(j_1, j_2, j_3)} + \sum_{j_1, j_2, j_3, j_4=1}^m \underline{L}^{j_1} \underline{L}^{j_2} \underline{L}^{j_3} b^{k,j_4} J_{(j_1, j_2, j_3, j_4)}. \end{aligned} \quad (5.3.46)$$

Conditions under which the above strong approximations converge with a certain strong order are given in [Kloeden & Platen \(1999\)](#), but follow also from the results of the next chapter. We emphasize that it is always worth checking whether a substantial simplification of a strong scheme is possible when the given SDE has some special structure. In the case of commutative noise, see (5.3.25), such simplifications naturally emerge.

The multiple Stratonovich integrals appearing in (5.3.46) can be generated as follows by using a very fine discretization of the time interval $[t, t_{n+1}]$. First, one simulates all the Wiener paths involved for the short period in the fine time discretization. Second, one uses the definition (1.4.35) of the Stratonovich integral and approximates the multiple Stratonovich integrals by the corresponding multiple sums.

5.4 Derivative-Free Strong Schemes

In this section we consider strong schemes which avoid the use of derivatives in the same way as Runge-Kutta schemes do for ODEs in a deterministic setting.

The resulting schemes are similar to Runge-Kutta schemes for ODEs, however, it must be emphasized that they cannot simply be constructed as heuristic generalizations of deterministic Runge-Kutta schemes. The extra terms in the Itô formula that occur additionally to the typical terms in the deterministic chain rule make it necessary to develop an appropriate class of *derivative-free schemes*.

Explicit Strong Order 1.0 Schemes

In a strong Taylor approximation derivatives of various orders of the drift and diffusion coefficients must be evaluated at each step. This makes the construction and implementation of such schemes complicated.

Various first order derivative-free schemes can be obtained from the Milstein scheme (5.3.21) simply by replacing the derivatives in the Milstein scheme by the corresponding differences. The inclusion of these differences requires the computation of supporting values of the coefficients at additional supporting points. An example is given by the following scheme, which is due to [Platen \(1984\)](#) and called the *Platen scheme*. In the one-dimensional case with $d = m = 1$ it takes the form

$$Y_{n+1} = Y_n + a \Delta + b \Delta W + \frac{1}{2\sqrt{\Delta}} \{b(\tau_n, \bar{Y}_n) - b\} \{(\Delta W)^2 - \Delta\} \quad (5.4.1)$$

with supporting value

$$\bar{Y}_n = Y_n + a \Delta + b \sqrt{\Delta}. \quad (5.4.2)$$

In the multi-dimensional case $d \in \{1, 2, \dots\}$ with scalar noise, that is with $m = 1$, the k th component of the Platen scheme has the form

$$Y_{n+1}^k = Y_n^k + a^k \Delta + b^k \Delta W + \frac{1}{2\sqrt{\Delta}} \{b^k(\tau_n, \bar{\mathbf{Y}}_n) - b^k\} \{(\Delta W)^2 - \Delta\} \quad (5.4.3)$$

with the vector supporting value

$$\bar{\mathbf{Y}}_n = \mathbf{Y}_n + \mathbf{a} \Delta + \mathbf{b} \sqrt{\Delta}. \quad (5.4.4)$$

In the general multi-dimensional case $d, m \in \{1, 2, \dots\}$ the k th component of this explicit strong order 1.0 scheme is

$$Y_{n+1}^k = Y_n^k + a^k \Delta + \sum_{j=1}^m b^{k,j} \Delta W^j + \frac{1}{\sqrt{\Delta}} \sum_{j_1, j_2=1}^m \left\{ b^{k, j_2} (\tau_n, \bar{\mathbf{Y}}_n^{j_1}) - b^{k, j_2} \right\} I_{(j_1, j_2)} \quad (5.4.5)$$

with the vector supporting values

$$\bar{\mathbf{Y}}_n^j = \mathbf{Y}_n + \mathbf{a} \Delta + \mathbf{b}^j \sqrt{\Delta} \quad (5.4.6)$$

for $j \in \{1, 2, \dots\}$. If the noise is commutative, as described in (5.3.25), then the scheme (5.4.5)–(5.4.6) has the Stratonovich analogue

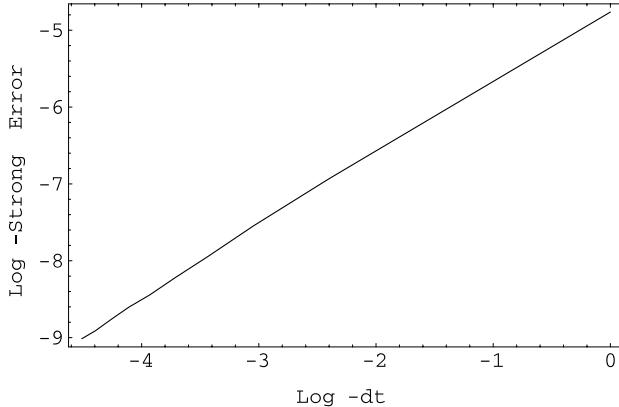


Fig. 5.4.1. Log-log plot of the absolute error of a Platen scheme

$$Y_{n+1}^k = Y_n^k + \underline{a}^k \Delta + \frac{1}{2} \sum_{j=1}^m \{ b^{k,j} (\tau_n, \bar{\mathbf{Y}}_n) + b^{k,j} \} \Delta W^j \quad (5.4.7)$$

with vector supporting values

$$\bar{\mathbf{Y}}_n = \mathbf{Y}_n + \underline{a} \Delta + \sum_{j=1}^m \mathbf{b}^j \Delta W^j. \quad (5.4.8)$$

Here \underline{a} denotes again the adjusted Stratonovich drift (5.3.4). In practice, the above schemes turn out to be convenient since they provide the strong order $\gamma = 1.0$ that is achieved by the Milstein scheme and avoid the computation of derivatives.

In the log-log plot shown in Fig. 5.4.1 we display the absolute error for a geometric Brownian motion approximated by the first order Platen scheme using the same default parameters from the example of the previous section for the Euler and Milstein schemes in Figs. 5.3.1 and 5.3.2.

Regressing the log of the absolute error on the log of the time step size results in the fit:

$$\ln(\varepsilon(\Delta)) = -4.71638 + 0.946112 \ln(\Delta).$$

The order of strong convergence is close to the theoretical value of $\gamma = 1$.

In Burrage (1998) a two stage Runge-Kutta method of strong order $\gamma = 1.0$ has been proposed that for $m = d = 1$ has in the autonomous case the form

$$Y_{n+1} = Y_n + (\underline{a}(Y_n) + 3\underline{a}(\bar{Y}_n)) \frac{\Delta}{4} + \frac{1}{4} (b(Y_n) + 3b(\bar{Y}_n)) \Delta W \quad (5.4.9)$$

with

$$\bar{Y}_n = Y_n + \frac{2}{3} (a(Y_n) \Delta + b(Y_n) \Delta W). \quad (5.4.10)$$

This method minimizes the leading error term within a class of first order Runge-Kutta methods and is, therefore, of particular interest.

Explicit Strong Order 1.5 Schemes

One can also construct derivative-free schemes of order $\gamma = 1.5$ by replacing the derivatives in the strong order 1.5 Taylor scheme by corresponding finite differences.

In the autonomous, one-dimensional case $d = m = 1$ an *explicit strong order 1.5 scheme* has been proposed in [Platen \(1984\)](#) in the form

$$\begin{aligned} Y_{n+1} = & Y_n + b \Delta W + \frac{1}{2\sqrt{\Delta}} \{a(\bar{Y}_+) - a(\bar{Y}_-)\} \Delta Z \\ & + \frac{1}{4} \{a(\bar{Y}_+) + 2a + a(\bar{Y}_-)\} \Delta \\ & + \frac{1}{4\sqrt{\Delta}} \{b(\bar{Y}_+) - b(\bar{Y}_-)\} \{(\Delta W)^2 - \Delta\} \\ & + \frac{1}{2\Delta} \{b(\bar{Y}_+) - 2b + b(\bar{Y}_-)\} \{\Delta W \Delta - \Delta Z\} \\ & + \frac{1}{4\Delta} \left[b(\bar{\Phi}_+) - b(\bar{\Phi}_-) - b(\bar{Y}_+) + b(\bar{Y}_-) \right] \left\{ \frac{1}{3}(\Delta W)^2 - \Delta \right\} \Delta W \end{aligned} \quad (5.4.11)$$

with

$$\bar{Y}_{\pm} = Y_n + a \Delta \pm b \sqrt{\Delta} \quad (5.4.12)$$

and

$$\bar{\Phi}_{\pm} = \bar{Y}_+ \pm b(\bar{Y}_+) \sqrt{\Delta}. \quad (5.4.13)$$

Here ΔZ is the multiple Itô integral $I_{(1,0)}$ defined in [\(5.3.36\)](#).

In the general multi-dimensional autonomous case with $d, m \in \{1, 2, \dots\}$ the k th component of the explicit strong order 1.5 scheme satisfies the relation

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + a^k \Delta + \sum_{j=1}^m b^{k,j} \Delta W^j \\
&+ \frac{1}{2\sqrt{\Delta}} \sum_{j_2=0}^m \sum_{j_1=1}^m \left\{ b^{k,j_2} \left(\bar{\mathbf{r}}_+^{j_1} \right) - b^{k,j_2} \left(\bar{\mathbf{r}}_-^{j_1} \right) \right\} I_{(j_1,j_2)} \\
&+ \frac{1}{2\Delta} \sum_{j_2=0}^m \sum_{j_1=1}^m \left\{ b^{k,j_2} \left(\bar{\mathbf{r}}_+^{j_1} \right) - 2b^{k,j_2} + b^{k,j_2} \left(\bar{\mathbf{r}}_-^{j_1} \right) \right\} I_{(0,j_2)} \\
&+ \frac{1}{2\Delta} \sum_{j_1,j_2,j_3=1}^m \left[b^{k,j_3} \left(\bar{\Phi}_+^{j_1,j_2} \right) - b^{k,j_3} \left(\bar{\Phi}_-^{j_1,j_2} \right) \right. \\
&\quad \left. - b^{k,j_3} \left(\bar{\mathbf{r}}_+^{j_1} \right) + b^{k,j_3} \left(\bar{\mathbf{r}}_-^{j_1} \right) \right] I_{(j_1,j_2,j_3)} \tag{5.4.14}
\end{aligned}$$

with

$$\bar{\mathbf{r}}_\pm^j = \mathbf{Y}_n + \frac{1}{m} \mathbf{a} \Delta \pm \mathbf{b}^j \sqrt{\Delta} \tag{5.4.15}$$

and

$$\bar{\Phi}_\pm^{j_1,j_2} = \bar{\mathbf{r}}_+^{j_1} \pm \mathbf{b}^{j_2} \left(\bar{\mathbf{r}}_+^{j_1} \right) \sqrt{\Delta}, \tag{5.4.16}$$

where we interpret $b^{k,0}$ as a^k .

Further higher order explicit methods and also multi-step schemes with corresponding strong order of convergence can be found in [Kloeden & Platen \(1999\)](#) and [Burrage \(1998\)](#).

A Simulation Study

Simulation studies similar to the previous ones for the strong Taylor schemes generate the expected slopes for the logarithm of the absolute errors. In Fig. 5.4.2 we summarize experimental results on strong approximation. Log-log plots of absolute errors for various strong schemes are displayed. We notice that the Euler method is less accurate and achieves a lower order of strong convergence than the Milstein and the Runge-Kutta or Platen method. Additionally we implement the strong order $\gamma = 1.5$ Taylor scheme as described in (5.3.35).

In practice, for a simulation it is also important to know how much computational time is involved when performing a scenario simulation. Usually it is of secondary concern since typically one wants to study only a few trajectories via scenario simulation. The CPU times needed to generate one path with 500,000 time steps by each scheme are as follows. The Euler method is with 3.5 seconds less computationally intensive than the other methods. Thus, one has to pay a price in computational time for accuracy. The Milstein and strong order 1.0 Platen scheme need 4.1 seconds. Even the strong

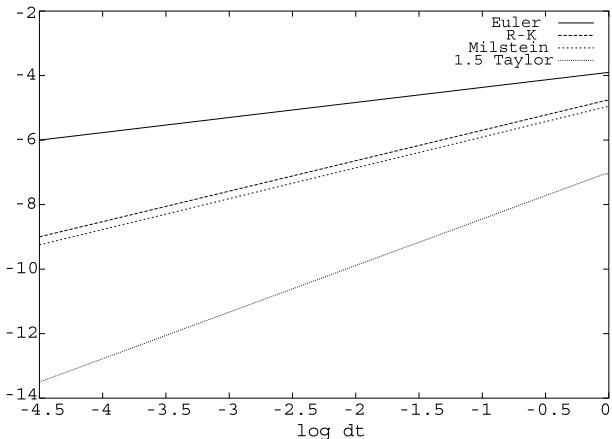


Fig. 5.4.2. Log-log plot of the absolute error for various strong schemes

order 1.5 Taylor scheme consumes only 4.4 seconds. As mentioned previously, the above schemes are adequate for scenario simulation, dynamic financial analysis, filtering, testing of estimators and for hedge simulation.

5.5 Exercises

5.1. Write down for the linear SDE

$$dX_t = (\mu X_t + \eta) dt + \gamma X_t dW_t$$

with $X_0 = 1$ the Euler scheme and the Milstein scheme.

5.2. Determine for the Vasicek short rate model

$$dr_t = \gamma(\bar{r} - r_t) dt + \beta dW_t$$

the Euler and Milstein schemes. What are the differences between the resulting two schemes?

5.3. Write down for the SDE in Exercise 5.1 the strong order 1.5 Taylor scheme.

5.4. Derive for the Black-Scholes SDE

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

with $X_0 = 1$ the explicit strong order 1.0 scheme.

Regular Strong Taylor Approximations with Jumps

In this chapter we start to go beyond the work described in [Kloeden & Platen \(1999\)](#) on the numerical solution of SDEs. We now allow the driving noise of the SDEs to have jumps. We present regular strong approximations obtained directly from a truncated Wagner-Platen expansion with jumps. The term regular refers to the time discretizations used to construct these approximations. These do not include the jump times of the Poisson random measure, as opposed to the jump-adapted strong approximations that will be presented later in Chap. 8. A convergence theorem for approximations of a given strong order of convergence will be presented at the end of this chapter. The reader who aims to simulate a solution of an SDE with low jump intensity is referred directly to Chap. 8 which describes jump-adapted schemes that are convenient to use.

6.1 Discrete-Time Approximation

SDE with Jumps

First we consider a one-dimensional SDE, $d = 1$, in the form

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t + \int_{\mathcal{E}} c(t, X_{t-}, v) p_\varphi(dv, dt), \quad (6.1.1)$$

for $t \in [0, T]$, with $X_0 \in \mathfrak{R}$, and $W = \{W_t, t \in [0, T]\}$ an $\underline{\mathcal{A}}$ -adapted one-dimensional Wiener process. We assume an $\underline{\mathcal{A}}$ -adapted Poisson measure $p_\varphi(dv, dt)$ with mark space $\mathcal{E} \subseteq \mathfrak{R} \setminus \{0\}$, and with intensity measure $\varphi(dv) dt = \lambda F(dv) dt$, where $F(\cdot)$ is a given probability distribution function for the realizations of the marks, see Sect. 1.8.

The SDE (6.1.1) can be written in integral form as

$$\begin{aligned} X_t &= X_0 + \int_0^t a(s, X_s) ds + \int_0^t b(s, X_s) dW_s + \int_0^t \int_{\mathcal{E}} c(s, X_{s-}, v) p_\varphi(dv, ds) \\ &= X_0 + \int_0^t a(s, X_s) ds + \int_0^t b(s, X_s) dW_s + \sum_{i=1}^{p_\varphi(t)} c(\tau_i, X_{\tau_i-}, \xi_i), \end{aligned} \quad (6.1.2)$$

where $\{(\tau_i, \xi_i), i \in \{1, 2, \dots, p_\varphi(t)\}\}$ is the double sequence of pairs of jump times and corresponding marks generated by the Poisson random measure. We express the i th mark at time τ_i by $\xi_i \in \mathcal{E}$. For simplicity, we have assumed a one-dimensional mark space $\mathcal{E} \subseteq \mathbb{R} \setminus \{0\}$. Multi-dimensional mark spaces can be similarly considered.

The case of a mark-independent jump size, which means $c(t, x, v) = c(t, x)$, is of particular interest as it simplifies the derivation and the implementation of numerical schemes. Therefore, we also consider a one-dimensional SDE with mark-independent jump size, which, in integral form, is of the form

$$X_t = X_0 + \int_0^t a(s, X_s) ds + \int_0^t b(s, X_s) dW_s + \sum_{i=1}^{p_\varphi(t)} c(\tau_i, X_{\tau_i-}). \quad (6.1.3)$$

Later, we will present strong Taylor approximations for a d -dimensional SDE, as introduced in Sect. 1.8, given by

$$d\mathbf{X}_t = \mathbf{a}(t, \mathbf{X}_t) dt + \mathbf{b}(t, \mathbf{X}_t) d\mathbf{W}_t + \int_{\mathcal{E}} \mathbf{c}(t, \mathbf{X}_{t-}, v) p_\varphi(dv, dt), \quad (6.1.4)$$

for $t \in [0, T]$, with $\mathbf{X}_0 \in \mathbb{R}^d$ and $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \in [0, T]\}$ an \mathcal{A} -adapted m -dimensional Wiener process. Moreover, p_φ is again an \mathcal{A} -adapted Poisson measure. Here $\mathbf{a}(t, \mathbf{x})$ and $\mathbf{c}(t, \mathbf{x}, v)$ are d -dimensional vectors of real valued functions on $[0, T] \times \mathbb{R}^d$ and on $[0, T] \times \mathbb{R}^d \times \mathcal{E}$, respectively. Furthermore, $\mathbf{b}(t, \mathbf{x})$ is a $d \times m$ -matrix of real valued functions on $[0, T] \times \mathbb{R}^d$. We recall that we use superscripts to denote vector components.

We will use again the following operators, defined earlier in (4.3.4)–(4.3.6),

$$\begin{aligned} L^0 f(t, \mathbf{x}, \mathbf{u}) &= \frac{\partial}{\partial t} f(t, \mathbf{x}, \mathbf{u}) + \sum_{i=1}^d a^i(t, \mathbf{x}) \frac{\partial}{\partial x^i} f(t, \mathbf{x}, \mathbf{u}) \\ &\quad + \frac{1}{2} \sum_{i,r=1}^d \sum_{j=1}^m b^{i,j}(t, \mathbf{x}) b^{r,j}(t, \mathbf{x}) \frac{\partial^2}{\partial x^i \partial x^j} f(t, \mathbf{x}, \mathbf{u}) \\ &\quad + \int_{\mathcal{E}} \left\{ f(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x}, v), \mathbf{u}) - f(t, \mathbf{x}, \mathbf{u}) \right\} \varphi(dv), \end{aligned} \quad (6.1.5)$$

$$L^k f(t, \mathbf{x}, \mathbf{u}) = \sum_{i=1}^d b^{i,k}(t, \mathbf{x}) \frac{\partial}{\partial x^i} f(t, \mathbf{x}, \mathbf{u}), \quad \text{for } k \in \{1, 2, \dots, m\} \quad (6.1.6)$$

and

$$L_v^{-1}f(t, \mathbf{x}, \mathbf{u}) = f(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x}, v), \mathbf{u}) - f(t, \mathbf{x}, \mathbf{u}), \quad (6.1.7)$$

for all $t \in [0, T]$, $\mathbf{x} \in \Re^d$ and $\mathbf{u} \in \mathcal{E}^{s(\alpha)}$.

Regular Time Discretization

Moreover, we consider a *regular time discretization* $0 = t_0 < t_1 < \dots < t_N = T$, on which we will construct a *discrete-time approximation* $Y^\Delta = \{Y_t^\Delta, t \in [0, T]\}$ of the solution of the SDE (6.1.4). For a given maximum time step size $\Delta \in (0, 1)$ we require the regular time discretization

$$(t)_\Delta = \{0 = t_0 < t_1 < \dots < t_N = T\}, \quad (6.1.8)$$

to satisfy the following conditions:

$$P(t_{n+1} - t_n \leq \Delta) = 1, \quad (6.1.9)$$

where

$$t_{n+1} \text{ is } \mathcal{A}_{t_n} - \text{measurable}, \quad (6.1.10)$$

for $n \in \{0, 1, \dots, N-1\}$ and

$$n_t = \max\{n \in \mathcal{N} : t_n \leq t\} < \infty \text{ a.s.}, \quad (6.1.11)$$

denoting the largest integer n such that $t_n \leq t$, for all $t \in [0, T]$. Such a time discretization is called *regular*, as opposed to the jump-adapted scheme presented later in Chap. 8. For instance, we could consider an equidistant time discretization with n th discretization time $t_n = n\Delta$, $n \in \{0, 1, \dots, N\}$, and time step size $\Delta = \frac{T}{N}$. However, the discretization times could also be random, as is needed if one wants to employ a step size control. Conditions (6.1.9), (6.1.10) and (6.1.11) pose some restrictions on the choice of the random discretization times. Condition (6.1.9) requires that the maximum step size in the time discretization cannot be larger than Δ . Condition (6.1.10) ensures that the length $\Delta_n = t_{n+1} - t_n$ of the next time step depends only on the information available at the last discretization time t_n . Condition (6.1.11) guarantees a finite number of discretization points in any bounded interval $[0, t]$.

For simplicity, when describing the schemes we will use the abbreviation

$$f = f(t_n, Y_n) \quad (6.1.12)$$

for a function f when no misunderstanding is possible. For the jump coefficient we may also write

$$c(v) = c(t_n, Y_n, v) \quad \text{and} \quad c(\xi_i) = c(t_n, Y_n, \xi_i), \quad (6.1.13)$$

if convenient, where ξ_i is the i th mark of the Poisson measure. Similarly, we write

$$c'(v) = c'(t_n, Y_n, v) \quad \text{and} \quad c'(\xi_i) = c'(t_n, Y_n, \xi_i). \quad (6.1.14)$$

Note that here and in the sequel, the prime ' in (6.1.14) denotes the derivative with respect to the spatial variable, which is the second argument. Moreover as in the previous chapter, we will omit mentioning the initial value Y_0 and the step numbers $n \in \{0, 1, \dots, N\}$ as this is self-explanatory.

Euler Scheme

The simplest scheme is again the well-known *Euler scheme*, which, in the one-dimensional case $d = m = 1$, is given by the algorithm

$$\begin{aligned} Y_{n+1} &= Y_n + a\Delta_n + b\Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) p_\varphi(dv, dz) \\ &= Y_n + a\Delta_n + b\Delta W_n + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c(\xi_i), \end{aligned} \quad (6.1.15)$$

for $n \in \{0, 1, \dots, N - 1\}$ with initial value $Y_0 = X_0$. Note that $a = a(t_n, Y_n)$, $b = b(t_n, Y_n)$, $c(v) = c(t_n, Y_n, v)$ and $c(\xi_i) = c(t_n, Y_n, \xi_i)$, according to the abbreviation defined in (6.1.12)–(6.1.14). Here

$$\Delta_n = t_{n+1} - t_n = I_{0,n} \quad (6.1.16)$$

is the length of the time interval $[t_n, t_{n+1}]$ and

$$\Delta W_n = W_{t_{n+1}} - W_{t_n} \quad (6.1.17)$$

is the n th Gaussian $N(0, \Delta_n)$ distributed increment of the Wiener process W , $n \in \{0, 1, \dots, N - 1\}$. Furthermore,

$$p_\varphi(t) = p_\varphi(\mathcal{E}, [0, t]) \quad (6.1.18)$$

represents the total number of jumps of the Poisson random measure up to time t , which is Poisson distributed with mean λt . Finally,

$$\xi_i \in \mathcal{E} \quad (6.1.19)$$

is the i th mark of the Poisson random measure p_φ at the i th jump time τ_i , with distribution function $F(\cdot)$. The Euler scheme (6.1.15) generally achieves a strong order of convergence $\gamma = 0.5$, as will be shown at the end of this chapter.

When we have a mark-independent jump size, which means $c(t, x, v) = c(t, x)$, the Euler scheme reduces to

$$Y_{n+1} = Y_n + a\Delta_n + b\Delta W_n + c\Delta p_n, \quad (6.1.20)$$

where

$$\Delta p_n = p_\varphi(t_{n+1}) - p_\varphi(t_n) \quad (6.1.21)$$

follows a Poisson distribution with mean $\lambda \Delta_n$.

In the multi-dimensional case with scalar driving Wiener process, which means $m = 1$, and mark-dependent jump size, the k th component of the *Euler scheme* is given by

$$Y_{n+1}^k = Y_n^k + a^k \Delta_n + b^k \Delta W_n + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c^k(\xi_i), \quad (6.1.22)$$

for $k \in \{1, 2, \dots, d\}$, where a^k , b^k , and c^k are the k th components of the drift, diffusion and jump coefficients, respectively.

In the multi-dimensional case with scalar Wiener process, $m = 1$, and mark-independent jump size, the k th component of the *Euler scheme* is given by

$$Y_{n+1}^k = Y_n^k + a^k \Delta_n + b^k \Delta W_n + c^k \Delta p_n, \quad (6.1.23)$$

for $k \in \{1, 2, \dots, d\}$.

For the general multi-dimensional case with mark-dependent jump size the k th component of the *Euler scheme* is of the form

$$Y_{n+1}^k = Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c^k(\xi_i), \quad (6.1.24)$$

where a^k and c^k are the k th components of the drift and the jump coefficients, respectively, and $b^{k,j}$ is the component of the k th row and j th column of the diffusion matrix b , for $k \in \{1, 2, \dots, d\}$, and $j \in \{1, 2, \dots, m\}$. Moreover,

$$\Delta W_n^j = W_{t_{n+1}}^j - W_{t_n}^j \quad (6.1.25)$$

is the $N(0, \Delta_n)$ distributed n th increment of the j th Wiener process. We recall that

$$\xi_i \in \mathcal{E} \quad (6.1.26)$$

is the i th mark generated by the Poisson random measure.

In the multi-dimensional case with mark-independent jump size we obtain the k th component of the *Euler scheme*

$$Y_{n+1}^k = Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + c^k \Delta p_n, \quad (6.1.27)$$

for $k \in \{1, 2, \dots, d\}$.

The Euler scheme includes only the time integral and the stochastic integrals of multiplicity one from the Wagner-Platen expansions (4.4.4) and (4.4.5), which both give the same truncated expansion. As we will see later, it

can be interpreted as the strong order 0.5 Taylor scheme. In the following we will use the Wagner-Platen expansion (4.4.4) for the construction of strong Taylor schemes. Similarly, one can start from the compensated Wagner-Platen expansion (4.4.5) obtaining slightly different compensated Taylor schemes, as will be discussed later.

6.2 Strong Order 1.0 Taylor Scheme

When accuracy and efficiency are required, it is important to construct numerical methods with higher strong order of convergence. This can be achieved by adding more terms of the Wagner-Platen expansion (4.4.4) to the scheme.

One-dimensional Strong Order 1.0 Taylor Scheme

It is now possible to derive the *strong order 1.0 Taylor scheme*, which, in the one-dimensional case, $d = m = 1$, is given by

$$\begin{aligned}
 Y_{n+1} = & Y_n + a\Delta_n + b\Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) p_\varphi(dv, dz) \\
 & + bb' \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW(z_1) dW(z_2) \\
 & + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b c'(v) dW(z_1) p_\varphi(dv, dz_2) \\
 & + \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ b(t_n, Y_n + c(v)) - b \right\} p_\varphi(dv, dz_1) dW(z_2) \quad (6.2.1) \\
 & + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ c(t_n, Y_n + c(v_1), v_2) - c(v_2) \right\} p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2),
 \end{aligned}$$

where

$$b' = b'(t, x) = \frac{\partial b(t, x)}{\partial x} \quad \text{and} \quad c'(v) = c'(t, x, v) = \frac{\partial c(t, x, v)}{\partial x}. \quad (6.2.2)$$

For simplicity, the abbreviations (6.1.12)–(6.1.14) were used in (6.2.1). The scheme (6.2.1) achieves strong order $\gamma = 1.0$, as we will see later. It represents a generalization of the Milstein scheme, see Milstein (1974), to the case of jump diffusions.

In view of applications to scenario simulations, an issue is the generation of the multiple stochastic integrals appearing in (6.2.1). By application of Itô's formula for jump diffusion processes, see Ikeda & Watanabe (1989), and the integration by parts formula, we can simplify the four double stochastic integrals appearing in (6.2.1) and rewrite the strong order 1.0 Taylor scheme (6.2.1) as

$$\begin{aligned}
Y_{n+1} = & Y_n + a\Delta_n + b\Delta W_n + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c(\xi_i) + \frac{bb'}{2} \left((\Delta W_n)^2 - \Delta_n \right) \\
& + b \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c'(\xi_i) (W(\tau_i) - W(t_n)) \\
& + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} \left\{ b(Y_n + c(\xi_i)) - b \right\} (W(t_{n+1}) - W(\tau_i)) \\
& + \sum_{j=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(\tau_j)} \left\{ c(Y_n + c(\xi_i), \xi_j) - c(\xi_j) \right\}. \tag{6.2.3}
\end{aligned}$$

This scheme is readily applicable for scenario simulation.

In the special case of a mark-independent jump coefficient, $c(t, x, v) = c(t, x)$, the strong order 1.0 Taylor scheme reduces to

$$\begin{aligned}
Y_{n+1} = & Y_n + a\Delta_n + b\Delta W_n + c\Delta p_n + b b' I_{(1,1)} + b c' I_{(1,-1)} \\
& + \{b(t_n, Y_n + c) - b\} I_{(-1,1)} + \{c(t_n, Y_n + c) - c\} I_{(-1,-1)}, \tag{6.2.4}
\end{aligned}$$

with multiple stochastic integrals

$$\begin{aligned}
I_{(1,1)} &= \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} dW_{s_1} dW_{s_2} = \frac{1}{2} \left\{ (\Delta W_n)^2 - \Delta_n \right\}, \\
I_{(1,-1)} &= \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{s_2} dW_{s_1} p_\varphi(ds, dv, ds_2) = \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} W_{\tau_i} - \Delta p_n W_{t_n}, \\
I_{(-1,1)} &= \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} \int_{\mathcal{E}} p_\varphi(ds, dv, ds_1) dW_{s_2} = \Delta p_n \Delta W_n - I_{(1,-1)}, \tag{6.2.5} \\
I_{(-1,-1)} &= \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{s_2} \int_{\mathcal{E}} p_\varphi(ds_1, dv_1) p_\varphi(ds_2, dv_2, ds) = \frac{1}{2} \left\{ (\Delta p_n)^2 - \Delta p_n \right\}.
\end{aligned}$$

On Multiple Stochastic Integrals

The level of complexity of the scheme (6.2.1), even in the simpler case (6.2.4) of mark-independent jump size, is quite substantial when compared to the Euler scheme (6.1.20). Indeed, it requires not only function evaluations of the drift, diffusion and jump coefficients, but also evaluations of their derivatives. The calculation of derivatives can be avoided by constructing derivative-free schemes that will be presented in the next chapter.

We point out that the simulation of some of the multiple stochastic integrals is computationally demanding. The generation of $I_{(1,1)}$ and $I_{(-1,-1)}$ is

straightforward once we have generated the random variables ΔW_n and Δp_n . The generation of the mixed multiple stochastic integrals $I_{(1,-1)}$ and $I_{(-1,1)}$ is more complex since it requires keeping track of the jump times between discretization points for the evaluation of W_{τ_i} . Conditioned on the number of jump events realized on the time interval $(t_n, t_{n+1}]$, these jump times being independent and uniformly distributed. Therefore, once we have generated the number of jumps Δp_n , we can sample Δp_n independent outcomes from a uniform distribution on $(t_n, t_{n+1}]$ in order to obtain the exact location of the jump times. However, from the computational point of view, this demonstrates that the efficiency of the algorithm is heavily dependent on the level of the intensity of the Poisson measure.

For the special case of a mark-independent jump coefficient, the number of calculations involved in an algorithm, such as the Euler scheme (6.1.20), does not depend on the level of the intensity. Note that here, we neglect the additional time needed to sample from a Poisson distribution with higher intensity. On the other hand, even in this special case, for the scheme (6.2.4) the number of computations is directly related to the number of jumps because of the generation of the two double stochastic integrals $I_{(1,-1)}$ and $I_{(-1,1)}$. Therefore, this algorithm is inefficient for the simulation of jump diffusion SDEs driven by a Poisson measure with high intensity.

We mention here, as on previous occasions that one can use for each discretization interval, a very fine time discretization, then simulate on this time grid the driving Wiener processes and the Poisson measure, and approximate the multiple stochastic integrals by using the corresponding sums according to the Itô integral definition.

It is, in principle, possible to derive strong Taylor schemes of any given order, as will be demonstrated later in Sect. 6.4. However, the schemes become rather complex. Moreover, as explained above, for SDEs driven by high intensity Poisson measures, these schemes are computationally inefficient. For these reasons we do not present in this section any scheme with order of strong convergence higher than $\gamma = 1.0$. For the construction of higher order schemes, we refer to the reader Chap. 8, where we describe jump-adapted approximations that avoid multiple stochastic integrals involving the Poisson measure. This makes it much easier to derive and implement these schemes.

Let us illustrate the higher accuracy achieved by the strong order 1.0 Taylor scheme in a scenario simulation. We consider the SDE (1.8.5) describing the Merton model with constant jump coefficient $c(t, x) = \psi x$. Using the following parameters: $X_0 = 1$, $\mu = 0.05$, $\sigma = 0.2$, $\psi = -0.25$, $T = 10$ and $\lambda = 0.3$, in Fig. 6.2.1 we show the path $X = \{X_t, t \in [0, 10]\}$ of the Merton model with explicit solution (1.8.6).

In Fig. 6.2.2 we illustrate the accuracy of the Euler scheme against that of the strong order 1.0 Taylor scheme in approximating the true solution X on the same sample path shown in Fig. 6.2.1, using a time step size $\Delta = 0.5$. We plot in Fig. 6.2.2 the absolute error $|X_{t_n} - Y_{t_n}|$, for $n \in \{0, 1, \dots, n_T\}$,

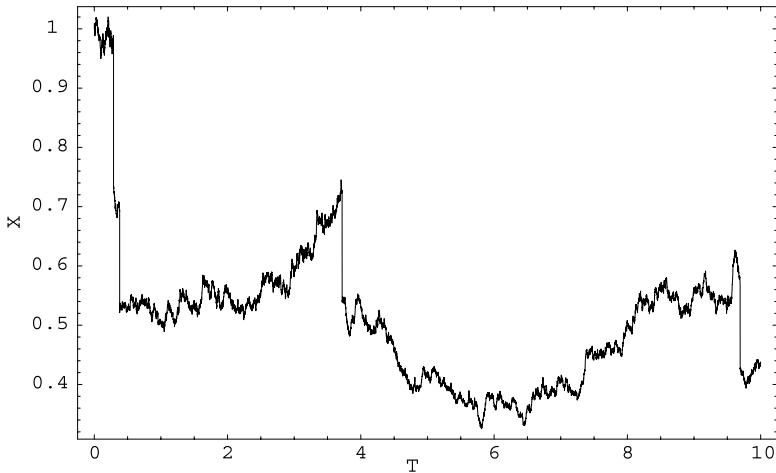


Fig. 6.2.1. A path of the Merton model

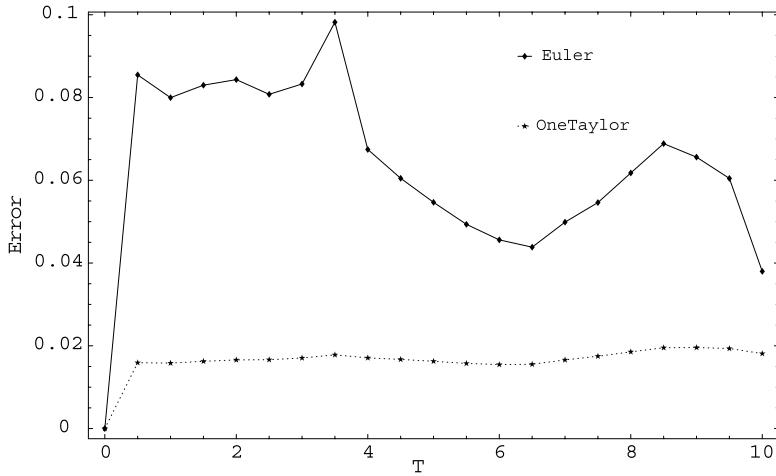


Fig. 6.2.2. Strong error for the Euler and the strong order 1.0 Taylor schemes

generated by these schemes. One clearly notices the higher accuracy of the strong order 1.0 Taylor scheme.

Multi-dimensional Strong Order 1.0 Taylor Scheme

In the multi-dimensional case with scalar Wiener process, $m = 1$, and mark-dependent jump size, the k th component of the *strong order 1.0 Taylor scheme* is given by the algorithm

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + a^k \Delta_n + b^k \Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c^k(v) p_\varphi(dv, dz) \\
&\quad + \sum_{l=1}^d \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} b^l \frac{\partial b^k}{\partial x^l} dW(z_1) dW(z_2) \\
&\quad + \sum_{l=1}^d \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b^l \frac{\partial c^k(v)}{\partial x^l} dW(z_1) p_\varphi(dv, dz_2) \\
&\quad + \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ b^k(t_n, \mathbf{Y}_n + \mathbf{c}(v)) - b^k \right\} p_\varphi(dv, dz_1) dW(z_2) \\
&\quad + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ c^k(t_n, \mathbf{Y}_n + \mathbf{c}(v_1), v_2) - c^k(v_2) \right\} \\
&\quad \times p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2), \tag{6.2.6}
\end{aligned}$$

where a^k , b^k , and c^k are the k th components of the drift, diffusion and jump coefficients, respectively, for $k \in \{1, 2, \dots, d\}$. Similar to (6.2.3) we can rewrite this scheme in a form that is readily applicable for scenario simulation.

For the multi-dimensional case with one driving Wiener process and mark-independent jump size the k th component of the *strong order 1.0 Taylor scheme* simplifies to

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + a^k \Delta_n + b^k \Delta W_n + c^k \Delta p_n + \sum_{l=1}^d b^l \frac{\partial b^k}{\partial x^l} I_{(1,1)} \\
&\quad + \sum_{l=1}^d b^l \frac{\partial c^k}{\partial x^l} I_{(1,-1)} + \left\{ b^k(t_n, \mathbf{Y}_n + \mathbf{c}) - b^k \right\} I_{(-1,1)} \\
&\quad + \left\{ c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k \right\} I_{(-1,-1)}, \tag{6.2.7}
\end{aligned}$$

for $k \in \{1, 2, \dots, d\}$. Here the four double stochastic integrals involved can be generated as described in (6.2.5).

In the general multi-dimensional case the k th component of the *strong order 1.0 Taylor scheme* is given by

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c^k(v) p_\varphi(dv, dz) \\
&\quad + \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW^{j_1}(z_1) dW^{j_2}(z_2) \\
&\quad + \sum_{j_1=1}^m \sum_{i=1}^d \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b^{i, j_1} \frac{\partial c^k(v)}{\partial x^i} dW^{j_1}(z_1) p_\varphi(dv, dz_2)
\end{aligned}$$

$$\begin{aligned}
& + \sum_{j_1=1}^m \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \left\{ b^{k,j_1}(t_n, \mathbf{Y}_n + \mathbf{c}(v)) - b^{k,j_1} \right\} p_\varphi(dv, dz_2) dW^{j_1}(z_2) \\
& + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ c^k(t_n, \mathbf{Y}_n + \mathbf{c}(v_1), v_2) - c^k(v_2) \right\} \\
& \times p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2),
\end{aligned} \tag{6.2.8}$$

for $k \in \{1, 2, \dots, d\}$.

In the multi-dimensional case with mark-independent jump size, the k th component of the *strong order 1.0 Taylor scheme* is given by

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + c^k \Delta p_n \\
& + \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{i,j_1} \frac{\partial b^{k,j_2}}{\partial x^i} I_{(j_1, j_2)} + \sum_{j_1=1}^m \sum_{i=1}^d b^{i,j_1} \frac{\partial c^k}{\partial x^i} I_{(j_1, -1)} \\
& + \sum_{j_1=1}^m \left\{ b^{k,j_1}(t_n, \mathbf{Y}_n + \mathbf{c}) - b^{k,j_1} \right\} I_{(-1, j_1)} + \left\{ c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k \right\} I_{(-1, -1)},
\end{aligned} \tag{6.2.9}$$

for $k \in \{1, 2, \dots, d\}$.

Mixed Multiple Stochastic Integrals

The considerations on the generation of the mixed multiple stochastic integrals involving Wiener processes and the Poisson random measure, presented for the one-dimensional case, apply in a similar way to the implementation of the scheme (6.2.9). Indeed,

$$\begin{aligned}
I_{(j_1, -1)} &= \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{s_2} dW_{s_1}^{j_1} p_\varphi(dv, ds_2) = \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} W_{\tau_i}^{j_1} - \Delta p_n W_{t_n}^{j_1}, \\
I_{(-1, j_1)} &= \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} \int_{\mathcal{E}} p_\varphi(dv, ds_1) dW_{s_2}^{j_1} = \Delta p_n \Delta W_n^{j_1} - I_{(j_1, -1)},
\end{aligned}$$

for $j_1 \in \{1, 2, \dots, m\}$. Moreover, in (6.2.9) we also require the multiple stochastic integrals

$$I_{(j_1, j_2)} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW^{j_1}(z_1) dW^{j_2}(z_2), \tag{6.2.10}$$

for $j_1, j_2 \in \{1, 2, \dots, m\}$. When $j_1 = j_2$, the corresponding double Wiener integral is given by

$$I_{(j_1, j_1)} = \frac{1}{2} \{ (\Delta W_n^{j_1})^2 - \Delta_n \}. \tag{6.2.11}$$

However, when $j_1 \neq j_2$, this integral cannot be easily expressed in terms of the increments $\Delta W_n^{j_1}$ and $\Delta W_n^{j_2}$ of the corresponding Wiener processes. Nonetheless, employing the expansion, presented in Sect. 5.3, it is possible to generate approximations of the required multiple stochastic integrals with the desired accuracy.

Strong Order 1.0 Compensated Taylor Scheme

The compensated strong Taylor schemes that arise from the compensated Wagner-Platen expansion (4.4.5) are generally different from those resulting from the Wagner-Platen expansion (4.4.4) used in this section. As an illustration, we describe below the *strong order 1.0 compensated Taylor scheme* in the special case $d = m = 1$. This scheme is given by the algorithm

$$\begin{aligned} Y_{n+1} = & Y_n + \tilde{a} \Delta_n + b \Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) \tilde{p}_\varphi(dv, dz) \\ & + bb' \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW(z_1) dW(z_2) \\ & + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b c'(v) dW(z_1) \tilde{p}_\varphi(dv, dz_2) \\ & + \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ b(t_n, Y_n + c(v)) - b \right\} p_\varphi(dv, dz_1) dW(z_2) \quad (6.2.12) \\ & + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ c(t_n, Y_n + c(v_1), v_2) - c(v_2) \right\} \tilde{p}_\varphi(dv_1, dz_1) \tilde{p}_\varphi(dv_2, dz_2). \end{aligned}$$

For application in scenario simulation, the generation of a stochastic integral involving the compensated Poisson measure \tilde{p}_φ can be implemented as follows: one could first generate the corresponding multiple stochastic integral with jump integrations with respect to the Poisson jump measure p_φ and then subtract its mean. This is given by the corresponding multiple stochastic integral with the integrations with respect to the Poisson jump measure p_φ replaced by integrations with respect to time and the intensity measure φ , see (4.2.1). For example, to generate the stochastic integral

$$\int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) \tilde{p}_\varphi(dv, dz),$$

we can use the representation

$$\int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) \tilde{p}_\varphi(dv, dz) = \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) p_\varphi(dv, dz) - \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) \varphi(dv) dz. \quad (6.2.13)$$

We recall, that, according to (6.1.13), we have used the abbreviation $c(v) = c(t_n, Y_n, v)$. Therefore, the last term on the right-hand side of (6.2.13) is \mathcal{A}_{t_n} -measurable, but its implementation generally requires a numerical integration at each time step.

To compare the computational effort required by the strong order 1.0 compensated Taylor scheme (6.2.12) with that of the strong order 1.0 Taylor scheme (6.2.1), let us rewrite the former using integrations with respect to the Poisson jump measure p_φ . In this case the strong order 1.0 compensated Taylor scheme is given by

$$\begin{aligned}
Y_{n+1} = & Y_n + a\Delta_n + b\Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) p_\varphi(dv, dz) \\
& + b b' \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW(z_1) dW(z_2) \\
& + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b c'(v) dW(z_1) p_\varphi(dv, dz_2) \\
& + \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{b(t_n, Y_n + c(v)) - b\} p_\varphi(dv, dz_1) dW(z_2) \\
& + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{c(t_n, Y_n + c(v_1), v_2) - c(v_2)\} p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2) \\
& - \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b c'(v) dW(z_1) \varphi(dv) dz_2 \\
& - \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{c(t_n, Y_n + c(v_1), v_2) - c(v_2)\} \varphi(dv_1) dz_1 p_\varphi(dv_2, dz_2) \\
& - \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{c(t_n, Y_n + c(v_1), v_2) - c(v_2)\} p_\varphi(dv_1, dz_1) \varphi(dv_2) dz_2 \\
& + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{c(t_n, Y_n + c(v_1), v_2) - c(v_2)\} \varphi(dv_1) dz_1 \varphi(dv_2) dz_2.
\end{aligned}$$

Therefore, the implementation of the strong order 1.0 compensated Taylor scheme requires the same computational effort as the strong order 1.0 Taylor scheme plus that required for the generation of the last four terms in the above scheme. Note that both schemes have the same order of strong convergence $\gamma = 1.0$. Nonetheless, the compensated strong Taylor scheme might be more accurate when dealing with high intensity jump diffusions.

6.3 Commutativity Conditions

Jump Commutativity Condition

As previously discussed, higher order Taylor schemes, even with mark-independent jump size, become computationally inefficient when the intensity of the Poisson measure is high. In this case the number of operations involved is almost proportional to the intensity level. Also the jump-adapted schemes to be presented in Chap. 8 show that their efficiency is strongly dependent on the intensity of the jumps.

By analyzing the multiple stochastic integrals required for the scheme (6.2.4), we observe that the dependence on the jump times affects only the mixed multiple stochastic integrals $I_{(1,-1)}$ and $I_{(-1,1)}$. However, since by (6.2.5) we have

$$I_{(-1,1)} = \Delta p_n \Delta W_n - I_{(1,-1)},$$

the sum of these integrals is obtained as

$$I_{(1,-1)} + I_{(-1,1)} = \Delta p_n \Delta W_n, \quad (6.3.1)$$

which is independent of the particular jump times. Let us consider a one-dimensional SDE with mark-independent jump size, $c(t, x, v) = c(t, x)$, satisfying the *jump commutativity condition*

$$b(t, x) \frac{\partial c(t, x)}{\partial x} = b(t, x + c(t, x)) - b(t, x), \quad (6.3.2)$$

for all $t \in [0, T]$ and $x \in \mathfrak{R}$. In this case the strong order 1.0 Taylor scheme (6.2.4) depends only on the sum $I_{(1,-1)} + I_{(-1,1)}$ expressed in (6.3.1). One does not need to keep track of the exact location of the jump times. Hence, its computational complexity is independent of the intensity level.

This is an important observation from a practical point of view. If a given SDE satisfies the jump commutativity condition (6.3.2), then considerable savings in computational time can be achieved. In practice, one can take advantage of this insight and save valuable computational time.

When we have a linear diffusion coefficient of the form

$$b(t, x) = b_1(t) + b_2(t) x, \quad (6.3.3)$$

with $b(t, x) > 0$, as it frequently occurs in finance, the jump commutativity condition (6.3.2) implies the following ODE for the jump coefficient:

$$\frac{\partial c(t, x)}{\partial x} = \frac{b_2(t) c(t, x)}{b_1(t) + b_2(t) x}, \quad (6.3.4)$$

for all $t \in [0, T]$ and $x \in \mathfrak{R}$. Therefore, for linear diffusion coefficients of the form (6.3.3) the class of SDEs satisfying the jump commutativity condition (6.3.2) is identified by mark-independent jump coefficients of the form

$$c(t, x) = e^{K(t)} \left(b_1(t) + b_2(t) x \right), \quad (6.3.5)$$

where $K(t)$ is an arbitrary function of time.

For instance, the SDE (1.8.5) with mark-independent, multiplicative jump size $c(t, x, v) = x \beta$, for $\beta \geq -1$, satisfies the jump commutativity condition (6.3.2). The corresponding strong order 1.0 Taylor scheme is given by

$$\begin{aligned} Y_{n+1} &= Y_n + \mu Y_n \Delta_n + \sigma Y_n \Delta W_n + \beta Y_n \Delta p_n + \frac{1}{2} \sigma^2 Y_n \{(\Delta W_n)^2 - \Delta_n\} \\ &\quad + \sigma \beta Y_n \Delta p_n \Delta W_n + \frac{1}{2} \beta^2 Y_n \{(\Delta p_n)^2 - \Delta p_n\}. \end{aligned} \quad (6.3.6)$$

Multi-dimensional Jump Commutativity

In the multi-dimensional case with scalar Wiener process and mark-independent jump size we obtain the *jump commutativity condition*

$$\sum_{l=1}^d b^l(t, \mathbf{x}) \frac{\partial c^k(t, \mathbf{x})}{\partial x^l} = b^k(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x})) - b^k(t, \mathbf{x}) \quad (6.3.7)$$

for $k \in \{1, 2, \dots, d\}$, $t \in [0, T]$ and $\mathbf{x} \in \mathbb{R}^d$.

For the general multi-dimensional case, one can show that the sum of two multiple stochastic integrals with respect to the j_1 th component of the Wiener process and the Poisson measure, given by

$$I_{(j_1, -1)} + I_{(-1, j_1)} = \Delta p_n \Delta W_n^{j_1}, \quad (6.3.8)$$

is independent of the particular jump times. Therefore, for a general multi-dimensional SDE with mark-independent jump size, we obtain the *jump commutativity condition*

$$\sum_{l=1}^d b^{l, j_1}(t, \mathbf{x}) \frac{\partial c^k(t, \mathbf{x})}{\partial x^l} = b^{k, j_1}(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x})) - b^{k, j_1}(t, \mathbf{x}) \quad (6.3.9)$$

for $j_1 \in \{1, 2, \dots, m\}$, $k \in \{1, 2, \dots, d\}$, $t \in [0, T]$ and $\mathbf{x} \in \mathbb{R}^d$. Thus, once the diffusion coefficients are specified, a solution of the first order semilinear $d \times m$ -dimensional system of PDEs (6.3.9) provides a d -dimensional commutative jump coefficient. Note that the system (6.3.9) has $d \times m$ equations and only d unknown functions. Therefore, even for simple diffusion coefficients, there may not exist any jump coefficient satisfying (6.3.9).

Consider, for instance, the multi-dimensional case with *additive diffusion coefficient* $b(t, x) = b(t)$. The jump commutativity condition (6.3.9) reduces to the $m \times d$ system of first order homogeneous linear PDEs

$$\sum_{l=1}^d b^{l, j}(t) \frac{\partial c^k(t, \mathbf{x})}{\partial x^l} = 0 \quad (6.3.10)$$

for $j \in \{1, 2, \dots, m\}$, $k \in \{1, 2, \dots, d\}$, $t \in [0, T]$ and $\mathbf{x} = (x^1, x^2, \dots, x^d)^\top \in \Re^d$. An *additive jump coefficient* $\mathbf{c}(t, \mathbf{x}) = \mathbf{c}(t)$ satisfies the condition (6.3.10). In the scalar case, $m = 1$, or in the trivial case where $b^{i,j_1}(t) = b^{i,j_2}(t)$ for all $i \in \{1, 2, \dots, d\}$, $j_1, j_2 \in \{1, 2, \dots, m\}$ and $t \in [0, T]$, we obtain the solution

$$c^k(t, \mathbf{x}) = f(t, \mathbf{y}) \quad (6.3.11)$$

for $k \in \{1, 2, \dots, d\}$. Here $\mathbf{y} = (y^1, \dots, y^{d-1})^\top \in \Re^{d-1}$ has components

$$y^i = \frac{-b^{i+1}(t)x^1 + b^1(t)x^{i+1}}{b^1(t)}$$

and f is an arbitrary function, differentiable with respect to the second argument \mathbf{y} .

Let us consider the multi-dimensional, *multiplicative diffusion coefficient* $\mathbf{b}(t, \mathbf{x})$, where the element in the i th row and j th column is given by $b^{i,j}(t, \mathbf{x}) = \sigma^{i,j}(t)x^i$, with $\sigma^{i,j}(t) \in \Re$ for $i \in \{1, 2, \dots, d\}$ and $j \in \{1, 2, \dots, m\}$. In this case the jump commutativity condition (6.3.9) reduces to the $d \times m$ system of first order linear PDEs

$$\sum_{l=1}^d \sigma^{l,j}(t)x^l \frac{\partial c^k(t, \mathbf{x})}{\partial x^l} = \sigma^{k,j}(t)c^k(t, \mathbf{x}) \quad (6.3.12)$$

for $j \in \{1, 2, \dots, m\}$, $k \in \{1, 2, \dots, d\}$, $t \in [0, T]$ and $\mathbf{x} = (x^1, x^2, \dots, x^d)^\top \in \Re^d$. In the scalar case, with $m = 1$, or in the trivial case where $\sigma^{i,j_1}(t) = \sigma^{i,j_2}(t)$ for $i \in \{1, 2, \dots, d\}$, $j_1, j_2 \in \{1, 2, \dots, m\}$ and $t \in [0, T]$, we obtain the solution

$$c^k(t, \mathbf{x}) = (x^1)^{\frac{\sigma^k(t)}{\sigma^1(t)}} f(t, \mathbf{y}) \quad (6.3.13)$$

for $k \in \{1, 2, \dots, d\}$. Here $\mathbf{y} = (y^1, \dots, y^{d-1})^\top \in \Re^{d-1}$ has components

$$y^i = (x^1)^{\frac{\sigma^{i+1}(t)}{\sigma^1(t)}} x^{i+1}$$

and f is an arbitrary function, differentiable with respect to the components of the second argument \mathbf{y} .

Combined Commutativity Conditions

In the context with jumps we now discuss some commutativity conditions involving only the diffusion coefficients. These conditions were discussed already in Sect. 5.3, in the context of the approximation of pure diffusion SDEs. As noticed in Sect. 6.2, when we have a multi-dimensional driving Wiener process, the corresponding strong order 1.0 Taylor scheme requires multiple stochastic integrals with respect to the different components of the Wiener process. In general, these can be generated only by resorting to approximations, such as described in Sect. 5.3. However, in the special case of a diffusion commutativity condition, where

$$L^{j_1} b^{k,j_2}(t, \mathbf{x}) = L^{j_2} b^{k,j_1}(t, \mathbf{x}) \quad (6.3.14)$$

for $j_1, j_2 \in \{1, 2, \dots, m\}$, $k \in \{1, 2, \dots, d\}$, $t \in [0, T]$ and $\mathbf{x} \in \Re^d$, see (5.3.25), it is possible to express all the double Wiener integrals in terms of the increments $\Delta W_n^{j_1}$ and $\Delta W_n^{j_2}$ of the Wiener processes. Therefore, for a multi-dimensional SDE satisfying the diffusion commutativity condition (6.3.14), the jump commutativity condition (6.3.9) and having a mark-independent jump size, we obtain a computationally efficient strong order 1.0 Taylor scheme, whose k th component is given by

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + c^k \Delta p_n \\ &+ \frac{1}{2} \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \left\{ \Delta W_n^{j_1} \Delta W_n^{j_2} - \Delta_n \right\} \\ &+ \sum_{j_1=1}^m \{b^{k, j_1}(t_n, \mathbf{Y}_n + \mathbf{c}) - b^{k, j_1}\} (\Delta p_n \Delta W_n^{j_1}) \\ &+ \frac{1}{2} \{c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k\} ((\Delta p_n)^2 - \Delta p_n), \end{aligned} \quad (6.3.15)$$

for $k \in \{1, 2, \dots, d\}$. For instance, the special case of additive diffusion and jump coefficients, which means $\mathbf{b}(t, \mathbf{x}) = \mathbf{b}(t)$ and $\mathbf{c}(t, \mathbf{x}) = \mathbf{c}(t)$, satisfies all the required commutativity conditions and, therefore, leads to an efficient strong order 1.0 Taylor scheme. Note that in this particular case the last two lines of (6.3.15) vanish and, thus, only increments of Wiener processes and the Poisson measure are needed.

From this analysis it becomes clear that when selecting a suitable numerical scheme for a specific model it is important to check for particular commutativity properties of the SDE under consideration to potentially save complexity and computational time. For the modeler it also suggests to focus on models that satisfy certain commutativity conditions. The resulting model may be more tractable than others with an *ad hoc* structure of diffusion and jumps coefficients.

6.4 Convergence Results

Higher Order Strong Taylor Schemes

In this section we introduce strong Taylor approximations and compensated strong Taylor approximations of given strong order $\gamma \in \{0.5, 1, 1.5, 2, \dots\}$. The key results underpinning the construction of these strong approximations are the Wagner-Platen expansions (4.4.4) and (4.4.5) presented in Chap. 4. By

including in a scheme enough terms from these expansions, we can obtain approximations with the desired order of strong convergence. More precisely, for a strong order $\gamma \in \{0.5, 1, 1.5, 2, \dots\}$ Taylor scheme we need to use the hierarchical set

$$\mathcal{A}_\gamma = \left\{ \alpha \in \mathcal{M} : l(\alpha) + n(\alpha) \leq 2\gamma \quad \text{or} \quad l(\alpha) = n(\alpha) = \gamma + \frac{1}{2} \right\}, \quad (6.4.1)$$

where $l(\alpha)$ denotes again the length of the multi-index α , and $n(\alpha)$ the number of its components equal to zero. Note that in Chap. 4 we have derived two types of Wagner-Platen expansions. The first uses the Poisson measure as the integrator of jump type and the second employs the compensated Poisson measure as integrator involving jumps. Therefore, we obtain two different types of strong approximations: the strong Taylor approximations and the compensated strong Taylor approximations.

For a regular time discretization with maximum step size $\Delta \in (0, 1)$, we define the *strong order γ Taylor scheme* by the vector equation

$$\mathbf{Y}_{n+1}^\Delta = \mathbf{Y}_n^\Delta + \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} I_\alpha \left[f_\alpha(t_n, \mathbf{Y}_n^\Delta) \right]_{t_n, t_{n+1}} = \sum_{\alpha \in \mathcal{A}_\gamma} I_\alpha \left[f_\alpha(t_n, \mathbf{Y}_n^\Delta) \right]_{t_n, t_{n+1}} \quad (6.4.2)$$

for $n \in \{0, 1, \dots, n_T - 1\}$. Similarly, we define the *strong order γ compensated Taylor scheme* by the vector equation

$$\mathbf{Y}_{n+1}^\Delta = \mathbf{Y}_n^\Delta + \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \tilde{I}_\alpha \left[\tilde{f}_\alpha(t_n, \mathbf{Y}_n^\Delta) \right]_{t_n, t_{n+1}} = \sum_{\alpha \in \mathcal{A}_\gamma} \tilde{I}_\alpha \left[\tilde{f}_\alpha(t_n, \mathbf{Y}_n^\Delta) \right]_{t_n, t_{n+1}} \quad (6.4.3)$$

for $n \in \{0, 1, \dots, n_T - 1\}$.

Equations (6.4.2) and (6.4.3) provide recursive numerical routines generating approximate values of the solution of the SDE (6.1.1) at the time discretization points. In order to assess the strong order of convergence of these schemes, we define, through a specific interpolation, the *strong order γ Taylor approximation* $\mathbf{Y}^\Delta = \{\mathbf{Y}_t^\Delta, t \in [0, T]\}$, by

$$\mathbf{Y}_t^\Delta = \sum_{\alpha \in \mathcal{A}_\gamma} I_\alpha [f_\alpha(t_{n_t}, \mathbf{Y}_{t_{n_t}}^\Delta)]_{t_{n_t}, t} \quad (6.4.4)$$

and the *strong order γ compensated Taylor approximation* $\mathbf{Y}^\Delta = \{\mathbf{Y}_t^\Delta, t \in [0, T]\}$, by

$$\mathbf{Y}_t^\Delta = \sum_{\alpha \in \mathcal{A}_\gamma} \tilde{I}_\alpha [\tilde{f}_\alpha(t_{n_t}, \mathbf{Y}_{t_{n_t}}^\Delta)]_{t_{n_t}, t} \quad (6.4.5)$$

for $t \in [0, T]$, starting from a given \mathcal{A}_0 -measurable random variable \mathbf{Y}_0 , where n_t was defined in (6.1.1).

These two approximations define stochastic processes $\mathbf{Y}^\Delta = \{\mathbf{Y}_t^\Delta, t \in [0, T]\}$, whose values coincide with those of the *strong order γ Taylor scheme*

(6.4.2) and of the *strong order γ compensated Taylor scheme* (6.4.3), respectively, at the time discretization points. Between the discretization points the multiple stochastic integrals have constant coefficient functions but evolve randomly as a stochastic process.

The strong Taylor schemes, given by (6.4.3), are constructed with jump integrals with respect to the compensated Poisson measure \tilde{p}_φ , as follows from the definition of the multiple stochastic integrals (4.2.9). Note that the Euler scheme and the strong order 1.0 Taylor scheme presented in the previous sections were expressed in terms of the Poisson measure p_φ . When rewritten in terms of the compensated Poisson measure \tilde{p}_φ , one recovers the strong Taylor schemes (6.4.5), with $\gamma = 0.5$ and $\gamma = 1.0$, respectively.

Strong Convergence Theorem

The strong order of convergence of the compensated strong Taylor schemes presented above can be derived from the following theorem. This convergence theorem enables us to construct a compensated strong Taylor approximation $\mathbf{Y}^\Delta = \{\mathbf{Y}_t^\Delta, t \in [0, T]\}$ of any given strong order $\gamma \in \{0.5, 1, 1.5, 2, \dots\}$.

Theorem 6.4.1. *For given $\gamma \in \{0.5, 1, 1.5, 2, \dots\}$, let $\mathbf{Y}^\Delta = \{\mathbf{Y}_t^\Delta, t \in [0, T]\}$ be the strong order γ compensated Taylor approximation defined in (6.4.5), corresponding to a time discretization with maximum step size $\Delta \in (0, 1)$. We assume that*

$$E(|\mathbf{X}_0|^2) < \infty \quad \text{and} \quad E(|\mathbf{X}_0 - \mathbf{Y}_0^\Delta|^2) \leq K_1 \Delta^{2\gamma}. \quad (6.4.6)$$

Moreover, suppose that the coefficient functions \tilde{f}_α satisfy the following conditions:

For $\alpha \in \mathcal{A}_\gamma$, $t \in [0, T]$, $\mathbf{u} \in \mathcal{E}^{s(\alpha)}$ and $\mathbf{x}, \mathbf{y} \in \Re^d$ the coefficient function \tilde{f}_α satisfies the Lipschitz type condition

$$\left| \tilde{f}_\alpha(t, \mathbf{x}, \mathbf{u}) - \tilde{f}_\alpha(t, \mathbf{y}, \mathbf{u}) \right| \leq K_1(\mathbf{u}) |\mathbf{x} - \mathbf{y}|, \quad (6.4.7)$$

where $(K_1(\mathbf{u}))^2$ is $\varphi(d\mathbf{u}^1) \times \dots \times \varphi(d\mathbf{u}^{s(\alpha)})$ -integrable.

For all $\alpha \in \mathcal{A}_\gamma \cup \mathcal{B}(\mathcal{A}_\gamma)$ we assume

$$\tilde{f}_{-\alpha} \in \mathcal{C}^{1,2} \quad \text{and} \quad \tilde{f}_\alpha \in \mathcal{H}_\alpha, \quad (6.4.8)$$

there exists a set G with $\mathcal{A}_{\gamma-1} \subseteq G \subseteq \mathcal{A}_\gamma$, where for all $\alpha \in (G \cup \mathcal{B}(G)) \setminus \{v\}$: $f_{-\alpha} \in \cap_{k=-1}^m \mathcal{L}^k$, for all $\alpha \in G$: $f_\alpha(\tau_n, \mathbf{x}_{\tau_n}) \in \mathcal{M}_{s(\alpha)}$, $n \in \{0, \dots, n_T + 1\}$ and for all $\alpha \in \mathcal{B}(G)$: $\tilde{f}_\alpha(\cdot, \mathbf{x}_\cdot) \in \mathcal{M}_{s(\alpha)}$, and for $\alpha \in \mathcal{A}_\gamma \cup \mathcal{B}(\mathcal{A}_\gamma)$, $t \in [0, T]$, $\mathbf{u} \in \mathcal{E}^{s(\alpha)}$ and $\mathbf{x} \in \Re^d$, we require

$$\left| \tilde{f}_\alpha(t, \mathbf{x}, \mathbf{u}) \right|^2 \leq K_2(\mathbf{u})(1 + |\mathbf{x}|^2), \quad (6.4.9)$$

where $K_2(\mathbf{u})$ is $\varphi(du^1) \times \dots \times \varphi(du^{s(\alpha)})$ -integrable.

Then the estimate

$$\sqrt{E\left(\sup_{0 \leq z \leq T} |\mathbf{X}_z - \mathbf{Y}_z^\Delta|^2 | \mathcal{A}_0\right)} \leq K_3 \Delta^\gamma \quad (6.4.10)$$

holds, where the constant K_3 does not depend on Δ .

Remark 6.4.2. By using the definitions of the operators (4.3.4)–(4.3.7) and of the sets in (4.2.8), it is possible to obtain conditions on the coefficients \mathbf{a} , \mathbf{b} and \mathbf{c} of the SDE (4.4.1) which imply the conditions (6.4.7)–(6.4.9) on the coefficient functions f_α . For instance, if the drift, diffusion and jump coefficients of the SDE (1.8.2) have $2(\gamma + 2)$ times continuously differentiable components $a^k, b^{k,j}, c^k$, for all $k \in \{1, 2, \dots, d\}$ and $j \in \{1, 2, \dots, m\}$ that are uniformly bounded with uniformly bounded derivatives, then conditions (6.4.7)–(6.4.9) are fulfilled.

Theorem 6.4.1 generalizes a similar result for pure diffusions described in Kloeden & Platen (1999). The proof will be given in Sect. 6.6. A related result, under slightly different conditions, was mentioned without proof in Platen (1982a).

Theorem 6.4.1 establishes the order of strong convergence of the strong order γ compensated Taylor approximation defined in (6.4.5). The following corollary permits us to obtain the order of strong convergence of the strong Taylor schemes presented in this chapter.

Corollary 6.4.3 Let $\mathbf{Y}^\Delta = \{\mathbf{Y}_t^\Delta, t \in [0, T]\}$ be the strong order γ Taylor approximation (6.4.4). Assume that the Itô coefficient functions f_α satisfy the conditions of Theorem 6.4.1. Then, if the conditions on the initial data (6.4.6) also hold, we obtain the estimate

$$\sqrt{E\left(\sup_{0 \leq z \leq T} |\mathbf{X}_z - \mathbf{Y}_z^\Delta|^2\right)} \leq K \Delta^\gamma, \quad (6.4.11)$$

where K is a finite positive constant independent of Δ .

A similar result, limited to SDEs driven by Wiener processes and homogeneous Poisson processes, can be found in Gardoñ (2004).

6.5 Lemma on Multiple Itô Integrals

We derive here a lemma on multiple stochastic integrals that we will need in the proof of Theorem 6.4.1. This lemma can be also used for other approximations.

Lemma 6.5.1 For a given multi-index $\alpha \in \mathcal{M}_m \setminus \{v\}$, a time discretization $(t)_\Delta$ with $\Delta \in (0, 1)$ and $g \in \mathcal{H}_\alpha$ let

$$\begin{aligned} V_{t_0, u, s(\alpha)} &= \int_{\mathcal{E}} \cdots \int_{\mathcal{E}} E \left(\sup_{t_0 \leq z \leq u} |g(z, v^1, \dots, v^{s(\alpha)})|^2 \middle| \mathcal{A}_{t_0} \right) \varphi(dv^1) \cdots \varphi(dv^{s(\alpha)}) \\ &< \infty, \end{aligned} \quad (6.5.1)$$

$$\tilde{F}_t^\alpha = E \left(\sup_{t_0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[g(\cdot)]_{t_{n_z}, z} \right|^2 \middle| \mathcal{A}_{t_0} \right) \quad (6.5.2)$$

and

$$F_t^\alpha = E \left(\sup_{t_0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} I_\alpha[g(\cdot)]_{t_n, t_{n+1}} + I_\alpha[g(\cdot)]_{t_{n_z}, z} \right|^2 \middle| \mathcal{A}_{t_0} \right). \quad (6.5.3)$$

Then

$$\tilde{F}_t^\alpha \leq \begin{cases} (t - t_0) \Delta^{2(l(\alpha)-1)} \int_{t_0}^t V_{t_0, u, s(\alpha)} du & \text{when : } l(\alpha) = n(\alpha) \\ 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0, u, s(\alpha)} du & \text{when : } l(\alpha) \neq n(\alpha) \end{cases}$$

and

$$F_t^\alpha \leq \begin{cases} (t - t_0) \Delta^{2(l(\alpha)-1)} \int_{t_0}^t V_{t_0, u, s(\alpha)} du & \text{when : } l(\alpha) = n(\alpha) \\ 4^{l(\alpha)-n(\alpha)+2} \hat{C}^{s(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0, u, s(\alpha)} du & \text{when : } l(\alpha) \neq n(\alpha) \end{cases}$$

almost surely, for every $t \in [t_0, T]$. Here $\hat{C} = 4 + \lambda(T - t_0)$.

Proof: Let us first prove the assertion of the lemma for \tilde{F}_t^α .

1. By definition (6.1.11) of n_z we get, for $z \in [t_n, t_{n+1})$, the relation $t_{n_z} = t_n$. Then, for a multi-index $\alpha = (j_1, \dots, j_n)$ with $j_n = 0$, we obtain

$$\begin{aligned} \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[g(\cdot)]_{t_{n_z}, z} \\ = \sum_{n=0}^{n_z-1} \int_{t_n}^{t_{n+1}} \tilde{I}_{\alpha-}[g(\cdot)]_{t_n, s} ds + \int_{t_{n_z}}^z \tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_z}, s} ds \\ = \sum_{n=0}^{n_z-1} \int_{t_n}^{t_{n+1}} \tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_s}, s} ds + \int_{t_{n_z}}^z \tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_s}, s} ds \\ = \int_{t_0}^z \tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_s}, s} ds. \end{aligned} \quad (6.5.4)$$

The same type of equality holds analogously for every $j_n \in \{-1, 0, \dots, m\}$.

2. Let us first consider the case with $l(\alpha) = n(\alpha)$. By the Cauchy-Schwarz inequality we have

$$\begin{aligned}
\tilde{F}_t^\alpha &= E \left(\sup_{t_0 \leq z \leq t} \left| \int_{t_0}^z \tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u}, u} du \right|^2 \middle| \mathcal{A}_{t_0} \right) \\
&\leq E \left(\sup_{t_0 \leq z \leq t} (z - t_0) \int_{t_0}^z |\tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u}, u}|^2 du \middle| \mathcal{A}_{t_0} \right) \\
&\leq (t - t_0) E \left(\int_{t_0}^t |\tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u}, u}|^2 du \middle| \mathcal{A}_{t_0} \right) \\
&= (t - t_0) \int_{t_0}^t E \left(|\tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u}, u}|^2 \middle| \mathcal{A}_{t_0} \right) du \\
&\leq (t - t_0) \int_{t_0}^t E \left(\sup_{t_{n_u} \leq z \leq u} |\tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u}, z}|^2 \middle| \mathcal{A}_{t_0} \right) du \\
&= (t - t_0) \int_{t_0}^t E \left(E \left(\sup_{t_{n_u} \leq z \leq u} |\tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u}, z}|^2 \middle| \mathcal{A}_{t_{n_u}} \right) \middle| \mathcal{A}_{t_0} \right) du, \quad (6.5.5)
\end{aligned}$$

where the last line holds because $t_0 \leq t_{n_u}$ a.s. and then $\mathcal{A}_{t_0} \subseteq \mathcal{A}_{t_{n_u}}$ for $u \in [t_0, t]$. Therefore, applying Lemma 4.5.2 to

$$E \left(\sup_{t_{n_u} \leq z \leq u} |\tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u}, z}|^2 \middle| \mathcal{A}_{t_{n_u}} \right) \quad (6.5.6)$$

yields

$$\begin{aligned}
\tilde{F}_t^\alpha &\leq (t - t_0) 4^{l(\alpha-) - n(\alpha-)} \\
&\times \int_{t_0}^t E \left((u - t_{n_u})^{l(\alpha-) + n(\alpha-) - 1} \int_{t_{n_u}}^u V_{t_{n_u}, z, s(\alpha-)} dz \middle| \mathcal{A}_{t_0} \right) du \\
&\leq (t - t_0) 4^{l(\alpha-) - n(\alpha-)} \int_{t_0}^t E \left((u - t_{n_u})^{l(\alpha-) + n(\alpha-)} V_{t_{n_u}, u, s(\alpha-)} \middle| \mathcal{A}_{t_0} \right) du \\
&\leq (t - t_0) 4^{l(\alpha-) - n(\alpha-)} \Delta^{l(\alpha-) + n(\alpha-)} \int_{t_0}^t E \left(V_{t_{n_u}, u, s(\alpha-)} \middle| \mathcal{A}_{t_0} \right) du, \quad (6.5.7)
\end{aligned}$$

where the last line holds as $(u - t_{n_u}) \leq \Delta$ for $u \in [t_0, t]$ and $t \in [t_0, T]$. Since $\mathcal{A}_{t_0} \subseteq \mathcal{A}_{t_{n_u}}$, we notice that for $u \in [t_0, t]$

$$\begin{aligned}
& E(V_{t_{n_u}, u, s(\alpha-)} | \mathcal{A}_{t_0}) \\
&= E\left(\int_{\mathcal{E}} \dots \int_{\mathcal{E}} E\left(\sup_{t_{n_u} \leq z \leq u} |g(z, v^1, \dots, v^{s(\alpha-)})|^2 \mid \mathcal{A}_{t_{n_u}}\right) \right. \\
&\quad \times \varphi(dv^1) \dots \varphi(dv^{s(\alpha-)}) \Big| \mathcal{A}_{t_0}\Big) \\
&= \int_{\mathcal{E}} \dots \int_{\mathcal{E}} E\left(E\left(\sup_{t_{n_u} \leq z \leq u} |g(z, v^1, \dots, v^{s(\alpha-)})|^2 \mid \mathcal{A}_{t_{n_u}}\right) \mid \mathcal{A}_{t_0}\right) \\
&\quad \times \varphi(dv^1) \dots \varphi(dv^{s(\alpha-)}) \\
&= \int_{\mathcal{E}} \dots \int_{\mathcal{E}} E\left(\sup_{t_{n_u} \leq z \leq u} |g(z, v^1, \dots, v^{s(\alpha-)})|^2 \mid \mathcal{A}_{t_0}\right) \varphi(dv^1) \dots \varphi(dv^{s(\alpha-)}) \\
&\leq \int_{\mathcal{E}} \dots \int_{\mathcal{E}} E\left(\sup_{t_0 \leq z \leq u} |g(z, v^1, \dots, v^{s(\alpha-)})|^2 \mid \mathcal{A}_{t_0}\right) \varphi(dv^1) \dots \varphi(dv^{s(\alpha-)}) \\
&= V_{t_0, u, s(\alpha-)}. \tag{6.5.8}
\end{aligned}$$

It then follows

$$\begin{aligned}
\tilde{F}_t^\alpha &\leq (t - t_0) 4^{l(\alpha-) - n(\alpha-)} \Delta^{l(\alpha-) + n(\alpha-)} \int_{t_0}^t V_{t_0, u, s(\alpha-)} du \\
&= (t - t_0) \Delta^{2(l(\alpha)-1)} \int_{t_0}^t V_{t_0, u, s(\alpha)} du, \tag{6.5.9}
\end{aligned}$$

since $l(\alpha-) = n(\alpha-)$, $s(\alpha) = s(\alpha-)$ and this completes the proof for the case $l(\alpha) = n(\alpha)$.

3. Let us now consider the case with a multi-index $\alpha = (j_1, \dots, j_l)$ with $l(\alpha) \neq n(\alpha)$ and $j_l \in \{1, 2, \dots, m\}$. In this case the multiple stochastic integral is a martingale. Hence, by Doob's inequality, Itô's isometry and Lemma 4.5.2 we obtain

$$\begin{aligned}
\tilde{F}_t^\alpha &= E\left(\sup_{t_0 \leq z \leq t} \left| \int_{t_0}^z \tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u}, u} dW_u^{j_l} \right|^2 \mid \mathcal{A}_{t_0}\right) \\
&\leq 4 E\left(\left| \int_{t_0}^t \tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u}, u} dW_u^{j_l} \right|^2 \mid \mathcal{A}_{t_0}\right)
\end{aligned}$$

$$\begin{aligned}
&\leq 4 \int_{t_0}^t E \left(|\tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u},u}|^2 \middle| \mathcal{A}_{t_0} \right) du \\
&= 4 \int_{t_0}^t E \left(E \left(|\tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u},u}|^2 \middle| \mathcal{A}_{t_{n_u}} \right) \middle| \mathcal{A}_{t_0} \right) du \\
&\leq 4 \int_{t_0}^t E \left(E \left(\sup_{t_{n_u} \leq z \leq u} |\tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u},z}|^2 \middle| \mathcal{A}_{t_{n_u}} \right) \middle| \mathcal{A}_{t_0} \right) du \\
&\leq 4^{l(\alpha)-n(\alpha)} \\
&\quad \times \int_{t_0}^t E \left((u - t_{n_u})^{l(\alpha)+n(\alpha)-1} \int_{t_{n_u}}^u V_{t_{n_u},z,s(\alpha-)} dz \middle| \mathcal{A}_{t_0} \right) du \\
&\leq 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0,u,s(\alpha-)} du \\
&= 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0,u,s(\alpha-)} du \\
&\leq 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0,u,s(\alpha)} du,
\end{aligned} \tag{6.5.10}$$

where the last passage holds since $s(\alpha) = s(\alpha-)$. This completes the proof in this case.

4. Let us now consider the case with a multi-index $\alpha = (j_1, \dots, j_l)$ with $l(\alpha) \neq n(\alpha)$ and $j_l = -1$. The multiple stochastic integral is again a martingale. Therefore, by Doob's inequality, Lemma 4.5.2 and steps similar to the previous case we obtain

$$\begin{aligned}
\tilde{F}_t^\alpha &= E \left(\sup_{t_0 \leq z \leq t} \left| \int_{t_0}^z \int_{\mathcal{E}} \tilde{I}_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{t_{n_u},u-} \tilde{p}_\varphi(dv^{s(\alpha)}, du) \right|^2 \middle| \mathcal{A}_{t_0} \right) \\
&\leq 4E \left(\left| \int_{t_0}^t \int_{\mathcal{E}} \tilde{I}_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{t_{n_u},u-} \tilde{p}_\varphi(dv^{s(\alpha)}, du) \right|^2 \middle| \mathcal{A}_{t_0} \right) \\
&= 4 \int_{t_0}^t \int_{\mathcal{E}} E \left(|\tilde{I}_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{t_{n_u},u}|^2 \middle| \mathcal{A}_{t_0} \right) \varphi(dv^{s(\alpha)}) du \\
&= 4 \int_{t_0}^t \int_{\mathcal{E}} E \left(E \left(|\tilde{I}_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{t_{n_u},u}|^2 \middle| \mathcal{A}_{t_{n_u}} \right) \middle| \mathcal{A}_{t_0} \right) \varphi(dv^{s(\alpha)}) du \\
&\leq 4 \int_{t_0}^t \int_{\mathcal{E}} E \left(E \left(\sup_{t_{n_u} \leq z \leq u} |\tilde{I}_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{t_{n_u},z}|^2 \middle| \mathcal{A}_{t_{n_u}} \right) \middle| \mathcal{A}_{t_0} \right) \varphi(dv^{s(\alpha)}) du
\end{aligned}$$

$$\begin{aligned}
&\leq 4^{l(\alpha)-n(\alpha)+1} \\
&\quad \times \int_{t_0}^t \int_{\mathcal{E}} E \left((u - t_{n_u})^{l(\alpha)-n(\alpha)-1} \int_{t_{n_u}}^u V_{t_{n_u}, z, s(\alpha-)} dz \middle| \mathcal{A}_{t_0} \right) \varphi(dv^{s(\alpha)}) du \\
&\leq 4^{l(\alpha)-n(\alpha)+1} \Delta^{l(\alpha)-n(\alpha)} \int_{t_0}^t \int_{\mathcal{E}} E \left(V_{t_{n_u}, u, s(\alpha-)} \middle| \mathcal{A}_{t_0} \right) \varphi(dv^{s(\alpha)}) du.
\end{aligned} \tag{6.5.11}$$

Hence, using (6.5.8) we have

$$\begin{aligned}
\tilde{F}_t^\alpha &\leq 4^{l(\alpha)-n(\alpha)+1} \Delta^{l(\alpha)-n(\alpha)} \int_{t_0}^t \int_{\mathcal{E}} V_{t_0, u, s(\alpha-)} \varphi(dv^{s(\alpha)}) du \\
&= 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0, u, s(\alpha)} du \\
&\leq 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0, u, s(\alpha)} du
\end{aligned} \tag{6.5.12}$$

since $l(\alpha) = l(\alpha-) + 1$, $n(\alpha) = n(\alpha-)$, $s(\alpha) = s(\alpha-) + 1$ and this completes the proof in this case.

5. Finally, we assume that $\alpha = (j_1, \dots, j_l)$ with $l(\alpha) \neq n(\alpha)$ and $j_l = 0$. It can be shown that the discrete-time process

$$\left\{ \sum_{n=0}^k \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}}, k \in \{0, 1, \dots, n_T - 1\} \right\} \tag{6.5.13}$$

is a discrete-time martingale.

Using Cauchy-Schwarz inequality we obtain

$$\begin{aligned}
\tilde{F}_t^\alpha &= E \left(\sup_{t_0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[g(\cdot)]_{t_{n_z}, z} \right|^2 \middle| \mathcal{A}_{t_0} \right) \\
&\leq 2 E \left(\sup_{t_0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} \right|^2 \middle| \mathcal{A}_{t_0} \right) \\
&\quad + 2 E \left(\sup_{t_0 \leq z \leq t} \left| \tilde{I}_\alpha[g(\cdot)]_{t_{n_z}, z} \right|^2 \middle| \mathcal{A}_{t_0} \right).
\end{aligned} \tag{6.5.14}$$

Applying Doob's inequality to the first term of the equation (6.5.14) we have

$$\begin{aligned}
& E \left(\sup_{t_0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} \right|^2 \middle| \mathcal{A}_{t_0} \right) \\
& \leq 4 E \left(\left| \sum_{n=0}^{n_t-1} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} \right|^2 \middle| \mathcal{A}_{t_0} \right) \\
& \leq 4 E \left(\left[\left| \sum_{n=0}^{n_t-2} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} \right|^2 \right. \right. \\
& \quad \left. \left. + 2 \left| \sum_{n=0}^{n_t-2} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} \right| E(|\tilde{I}_\alpha[g(\cdot)]_{t_{n_t-1}, t_{n_t}}| \mid \mathcal{A}_{t_{n_t-1}}) \right. \right. \\
& \quad \left. \left. + E(|\tilde{I}_\alpha[g(\cdot)]_{t_{n_t-1}, t_{n_t}}|^2 \mid \mathcal{A}_{t_{n_t-1}}) \right] \middle| \mathcal{A}_{t_0} \right) \\
& \leq 4 E \left(\left[\left| \sum_{n=0}^{n_t-2} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} \right|^2 \right. \right. \\
& \quad \left. \left. + E(|\tilde{I}_\alpha[g(\cdot)]_{t_{n_t-1}, t_{n_t}}|^2 \mid \mathcal{A}_{t_{n_t-1}}) \right] \middle| \mathcal{A}_{t_0} \right). \tag{6.5.15}
\end{aligned}$$

Here the last line holds by the discrete-time martingale property of the involved stochastic integrals, which is $E(|\tilde{I}_\alpha[g(\cdot)]_{t_{n_t-1}, t_{n_t}}| \mid \mathcal{A}_{t_{n_t-1}}) = 0$. Then by applying Lemma 4.5.2 we obtain

$$\begin{aligned}
& E \left(\sup_{t_0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} \right|^2 \middle| \mathcal{A}_{t_0} \right) \\
& \leq 4 E \left(\left[\left| \sum_{n=0}^{n_t-2} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} \right|^2 \right. \right. \\
& \quad \left. \left. + 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_{n_t-1}}^{t_{n_t}} V_{t_{n_t-1}, u, s(\alpha)} du \right] \middle| \mathcal{A}_{t_0} \right) \\
& \leq 4 E \left(\left[\left| \sum_{n=0}^{n_t-3} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} \right|^2 \right. \right. \\
& \quad \left. \left. + 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_{n_t-2}}^{t_{n_t-1}} V_{t_{n_t-2}, u, s(\alpha)} du \right] \middle| \mathcal{A}_{t_0} \right)
\end{aligned}$$

$$\begin{aligned}
& + 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_{n_t-1}}^{t_{n_t}} V_{t_{n_t-1}, u, s(\alpha)} du \Big| \mathcal{A}_{t_0} \Big) \\
& \leq 4 E \left(\left| \sum_{n=0}^{n_t-3} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} \right|^2 \right. \\
& \quad \left. + 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_{n_t-2}}^{t_{n_t-1}} V_{t_{n_t-2}, u, s(\alpha)} du \right. \\
& \quad \left. + 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_{n_t-1}}^{t_{n_t}} V_{t_{n_t-2}, u, s(\alpha)} du \right| \mathcal{A}_{t_0} \right), \quad (6.5.16)
\end{aligned}$$

where the last passage holds since $V_{t_{n_t-1}, u, s(\alpha)} \leq V_{t_{n_t-2}, u, s(\alpha)}$. Applying this procedure repetitively and using (6.5.8) we finally obtain

$$\begin{aligned}
& E \left(\sup_{t_0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[g(\cdot)]_{t_n, t_{n+1}} \right|^2 \Big| \mathcal{A}_{t_0} \right) \\
& \leq 4^{l(\alpha)-n(\alpha)+1} \Delta^{l(\alpha)+n(\alpha)-1} E \left(\int_{t_0}^t V_{t_0, u, s(\alpha)} du \Big| \mathcal{A}_{t_0} \right) \\
& = 4^{l(\alpha)-n(\alpha)+1} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0, u, s(\alpha)} du. \quad (6.5.17)
\end{aligned}$$

For the second term of equation (6.5.14), by applying the Cauchy-Schwarz inequality, similar steps as those used previously and Lemma 4.5.2, we obtain

$$\begin{aligned}
& E \left(\sup_{t_0 \leq z \leq t} \left| \tilde{I}_\alpha[g(\cdot)]_{t_{n_z}, z} \right|^2 \Big| \mathcal{A}_{t_0} \right) \\
& = E \left(\sup_{t_0 \leq z \leq t} \left| \int_{t_{n_z}}^z \tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_z}, u} du \right|^2 \Big| \mathcal{A}_{t_0} \right) \\
& \leq E \left(\sup_{t_0 \leq z \leq t} (z - t_{n_z}) \int_{t_{n_z}}^z |\tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_z}, u}|^2 du \Big| \mathcal{A}_{t_0} \right) \\
& \leq \Delta \int_{t_0}^t E \left(E \left(\sup_{t_{n_u} \leq z \leq u} |\tilde{I}_{\alpha-}[g(\cdot)]_{t_{n_u}, z}|^2 \Big| \mathcal{A}_{t_{n_u}} \right) \Big| \mathcal{A}_{t_0} \right) du \\
& \leq \Delta 4^{l(\alpha-)-n(\alpha-)} \Delta^{l(\alpha-)+n(\alpha-)-1} \int_{t_0}^t E \left(\int_{t_{n_u}}^u V_{t_{n_u}, z, s(\alpha-)} dz \Big| \mathcal{A}_{t_0} \right) du
\end{aligned}$$

$$\begin{aligned} &\leq \Delta^{4^{l(\alpha)-n(\alpha)}} \Delta^{l(\alpha)+n(\alpha)-1} \Delta \int_{t_0}^t V_{t_0,u,s(\alpha)} du \\ &= 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0,u,s(\alpha)} du, \end{aligned} \quad (6.5.18)$$

where the last passage holds since $l(\alpha) = l(\alpha-) + 1$, $n(\alpha) = n(\alpha-) + 1$ and $s(\alpha) = s(\alpha-)$.

Therefore, combining equations (6.5.17) and (6.5.18) we finally obtain

$$\begin{aligned} \tilde{F}_t^\alpha &\leq 2 \left(4^{l(\alpha)-n(\alpha)+1} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0,u,s(\alpha)} du \right. \\ &\quad \left. + 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0,u,s(\alpha)} du \right) \\ &\leq 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \int_{t_0}^t V_{t_0,u,s(\alpha)} du, \end{aligned} \quad (6.5.19)$$

which completes the proof of the lemma for \tilde{F}_τ^α . \square

Let us now prove the assertion for F_τ^α . The case of $l(\alpha) = n(\alpha)$ has been already proved since $F_\tau^\alpha = \tilde{F}_\tau^\alpha$.

Consider now the case of $l(\alpha) > n(\alpha) + s(\alpha)$ with $\alpha = (j_1, \dots, j_l)$. Note that in this case at least one of the elements of the multi-index α belongs to $\{1, 2, \dots, m\}$ and, thus, the process

$$\left\{ \sum_{n=0}^k I_\alpha[g(\cdot)]_{t_n, t_{n+1}}, k \in \{0, 1, \dots, n_T - 1\} \right\} \quad (6.5.20)$$

is a discrete-time martingale.

If $j_l = 0$, then using similar steps as in (6.5.14)–(6.5.19) we obtain

$$\begin{aligned} F_t^\alpha &\leq 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \hat{K}^{s(\alpha)} \int_{t_0}^t V_{t_0,u,s(\alpha)} du \\ &\leq 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \hat{C}^{s(\alpha)} \int_{t_0}^t V_{t_0,u,s(\alpha)} du \end{aligned} \quad (6.5.21)$$

where $\hat{K} = \frac{1}{2}(4 + \lambda(T - t_0))$, see Lemma 4.5.2, and $\hat{C} = 4 + \lambda(T - t_0)$.

If $j_l \in \{1, 2, \dots, m\}$, then by similar steps as in (6.5.10) we obtain

$$\begin{aligned} F_t^\alpha &\leq 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \hat{K}^{s(\alpha)} \int_{t_0}^t V_{t_0,u,s(\alpha)} du \\ &\leq 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \hat{C}^{s(\alpha)} \int_{t_0}^t V_{t_0,u,s(\alpha)} du. \end{aligned} \quad (6.5.22)$$

If $j_l = -1$, then by the decomposition (4.5.1) and the Cauchy-Schwarz inequality we obtain

$$\begin{aligned} F_t^\alpha &\leq 2E\left(\sup_{t_0 \leq z \leq t} \left| \int_{t_0}^z \int_{\mathcal{E}} I_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{t_{n_u}, u} \tilde{p}_\varphi(dv^{s(\alpha)}, du) \right|^2 \middle| \mathcal{A}_{t_0}\right) \\ &\quad + 2E\left(\sup_{t_0 \leq z \leq t} \left| \int_{t_0}^z \int_{\mathcal{E}} I_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{t_{n_u}, u} \varphi(dv^{s(\alpha)}) du \right|^2 \middle| \mathcal{A}_{t_0}\right). \end{aligned} \quad (6.5.23)$$

For the first term on the right-hand side of (6.5.23) by similar steps as those used in (6.5.11) and (6.5.12), we obtain

$$\begin{aligned} E\left(\sup_{t_0 \leq z \leq t} \left| \int_{t_0}^z \int_{\mathcal{E}} I_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{t_{n_u}, u} \tilde{p}_\varphi(dv^{s(\alpha)}, du) \right|^2 \middle| \mathcal{A}_{t_0}\right) \\ \leq 4^{l(\alpha)-n(\alpha)} \Delta^{l(\alpha)+n(\alpha)-1} \hat{K}^{s(\alpha)-1} \int_{t_0}^t V_{t_0, u, s(\alpha)} du. \end{aligned} \quad (6.5.24)$$

For the second term on the right-hand side of (6.5.23) by similar steps as those used in (6.5.5)–(6.5.9), we obtain

$$\begin{aligned} E\left(\sup_{t_0 \leq z \leq t} \left| \int_{t_0}^z \int_{\mathcal{E}} I_{\alpha-}[g(\cdot, v^{s(\alpha)})]_{t_{n_u}, u} \varphi(dv^{s(\alpha)}) du \right|^2 \middle| \mathcal{A}_{t_0}\right) \\ \leq \lambda(t-t_0) 4^{l(\alpha)-n(\alpha)-1} \Delta^{l(\alpha)+n(\alpha)-1} \hat{K}^{s(\alpha)-1} \int_{t_0}^t V_{t_0, u, s(\alpha)} du. \end{aligned} \quad (6.5.25)$$

By combining (6.5.23), (6.5.24) and (6.5.25), we obtain

$$\begin{aligned} F_t^\alpha &\leq \frac{2}{4^3} (4 + \lambda(T-t_0)) 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \hat{K}^{s(\alpha)-1} \int_{t_0}^t V_{t_0, u, s(\alpha)} du \\ &\leq 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \hat{K}^{s(\alpha)} \int_{t_0}^t V_{t_0, u, s(\alpha)} du \\ &\leq 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \hat{C}^{s(\alpha)} \int_{t_0}^t V_{t_0, u, s(\alpha)} du. \end{aligned} \quad (6.5.26)$$

Let us finally consider the case of $l(\alpha) = n(\alpha) + s(\alpha)$ with $s(\alpha) \geq 1$. By use of relationship (4.2.13) one can rewrite the multiple stochastic integrals $I_\alpha[g(\cdot)]_{t_n, t_{n+1}}$ appearing in F_t^α as the sum of $2^{s(\alpha)}$ multiple stochastic integrals involving integrations with respect to the compensated Poisson measure p_φ , the product of time and the intensity measure $\varphi(\cdot)$. Therefore, by applying the Cauchy-Schwarz inequality and similar steps as used before, we obtain

$$\begin{aligned}
F_t^\alpha &\leq 2^{s(\alpha)} 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \hat{K}^{s(\alpha)} \int_{t_0}^t V_{t_0, u, s(\alpha)} du \\
&= 4^{l(\alpha)-n(\alpha)+2} \Delta^{l(\alpha)+n(\alpha)-1} \hat{C}^{s(\alpha)} \int_{t_0}^t V_{t_0, u, s(\alpha)} du. \tag{6.5.27}
\end{aligned}$$

This completes the proof of Lemma 6.5.1. \square

6.6 Proof of the Convergence Theorem

Before proceeding to the proof of Theorem 6.4.1, we present a lemma on the second moment estimate of the strong order γ compensated Taylor approximation (6.4.5)

Lemma 6.6.1 *Under the conditions of Theorem 6.4.1, we obtain*

$$E\left(\sup_{0 \leq z \leq T} |\mathbf{Y}_z^\Delta|^2 \mid \mathcal{A}_0\right) \leq C(1 + |\mathbf{Y}_0^\Delta|^2), \tag{6.6.1}$$

where \mathbf{Y}^Δ is the strong order γ compensated Taylor approximation (6.4.5).

Proof: Note that the strong order γ compensated Taylor approximation \mathbf{Y}^Δ at time $t \in [0, T]$ is given by

$$\mathbf{Y}_t^\Delta = \mathbf{Y}_0^\Delta + \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \left\{ \sum_{n=0}^{n_t-1} \tilde{I}_\alpha[\tilde{f}_\alpha(t_n, \mathbf{Y}_{t_n}^\Delta)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[\tilde{f}_\alpha(t_{n_t}, \mathbf{Y}_{t_{n_t}}^\Delta)]_{t_{n_t}, t} \right\}.$$

Therefore, we have

$$\begin{aligned}
E\left(\sup_{0 \leq z \leq T} |\mathbf{Y}_z^\Delta|^2 \mid \mathcal{A}_0\right) &\leq E\left(\sup_{0 \leq z \leq T} (1 + |\mathbf{Y}_z^\Delta|^2) \mid \mathcal{A}_0\right) \\
&\leq E\left(\sup_{0 \leq z \leq T} \left(1 + \left|\mathbf{Y}_0^\Delta + \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \left\{ \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[\tilde{f}_\alpha(t_n, \mathbf{Y}_{t_n}^\Delta)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[\tilde{f}_\alpha(t_{n_z}, \mathbf{Y}_{t_{n_z}}^\Delta)]_{t_{n_z}, z} \right\}\right|^2\right) \mid \mathcal{A}_0\right) \\
&\leq E\left(\sup_{0 \leq z \leq T} \left(1 + 2|\mathbf{Y}_0^\Delta|^2 + 2 \left| \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \left\{ \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[\tilde{f}_\alpha(t_n, \mathbf{Y}_{t_n}^\Delta)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[\tilde{f}_\alpha(t_{n_z}, \mathbf{Y}_{t_{n_z}}^\Delta)]_{t_{n_z}, z} \right\}\right|^2\right) \mid \mathcal{A}_0\right)
\end{aligned}$$

$$\leq C_1 \left(1 + |\mathbf{Y}_0^\Delta|^2 \right) + 2K \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} E \left(\sup_{0 \leq z \leq T} \left| \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[\tilde{f}_\alpha(t_n, \mathbf{Y}_{t_n}^\Delta)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[\tilde{f}_\alpha(t_{n_z}, \mathbf{Y}_{t_{n_z}}^\Delta)]_{t_{n_z}, z} \right|^2 \middle| \mathcal{A}_0 \right), \quad (6.6.2)$$

where K is a positive constant depending only on the strong order γ of the approximation. By Lemma 6.5.1 and the linear growth condition (6.4.9) we obtain

$$\begin{aligned} & E \left(\sup_{0 \leq z \leq T} |\mathbf{Y}_z^\Delta|^2 \middle| \mathcal{A}_0 \right) \\ & \leq C_1 (1 + |\mathbf{Y}_0^\Delta|^2) + 2K_1 \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \left\{ \int_0^T \int_{\mathcal{E}} \dots \int_{\mathcal{E}} \right. \\ & \quad \times E \left(\sup_{0 \leq z \leq u} |\tilde{f}_\alpha(z, \mathbf{Y}_z^\Delta)|^2 \middle| \mathcal{A}_0 \right) \varphi(dv^1) \dots \varphi(dv^{s(\alpha)}) du \left. \right\} \\ & \leq C_1 (1 + |\mathbf{Y}_0^\Delta|^2) + 2K_1 \\ & \quad \times \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \left\{ \int_{\mathcal{E}} \dots \int_{\mathcal{E}} K_2(v^1, \dots, v^{s(\alpha)}) \varphi(dv^1) \dots \varphi(dv^{s(\alpha)}) \right. \\ & \quad \times \left. \int_0^T E \left(\sup_{0 \leq z \leq u} (1 + |\mathbf{Y}_z^\Delta|^2) \middle| \mathcal{A}_0 \right) du \right\} \\ & \leq C_1 (1 + |\mathbf{Y}_0^\Delta|^2) + C_2 \int_0^T E \left(\sup_{0 \leq z \leq u} (1 + |\mathbf{Y}_z^\Delta|^2) \middle| \mathcal{A}_0 \right) du. \end{aligned} \quad (6.6.3)$$

We then have

$$\begin{aligned} E \left(\sup_{0 \leq s \leq T} (1 + |\mathbf{Y}_s^\Delta|^2) \middle| \mathcal{A}_0 \right) & \leq C_1 (1 + |\mathbf{Y}_0^\Delta|^2) \\ & \quad + 2K \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} D(\alpha, \Delta, \gamma) \int_0^T \int_{\mathcal{E}} \dots \int_{\mathcal{E}} \\ & \quad E \left(\sup_{0 \leq s \leq u} |f_\alpha(s, \mathbf{Y}_s^\Delta)|^2 \middle| \mathcal{A}_0 \right) \varphi(dv_1) \dots \varphi(dv_{s(\alpha)}) du \end{aligned}$$

$$\begin{aligned}
&\leq C_1(1 + |\mathbf{Y}_0^\Delta|^2) + 2K(\gamma) \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} K_\alpha(\Delta) \\
&\quad \times \int_{\mathcal{E}} \dots \int_{\mathcal{E}} K_2(v_1, \dots, v_{s(\alpha)}) \varphi(dv_1) \dots \varphi(dv_{s(\alpha)}) \\
&\quad \times \int_0^T E\left(\sup_{0 \leq s \leq u} (1 + |\mathbf{Y}_s^\Delta|^2) | \mathcal{A}_0\right) du \\
&\leq C_1(1 + |\mathbf{Y}_0^\Delta|^2) \\
&\quad + K(\gamma, \lambda, \Delta) \int_0^T E\left(\sup_{0 \leq s \leq u} (1 + |\mathbf{Y}_s^\Delta|^2) | \mathcal{A}_0\right) du. \tag{6.6.4}
\end{aligned}$$

Then by applying the Gronwall inequality, we obtain

$$\begin{aligned}
E\left(\sup_{0 \leq s \leq T} |\mathbf{Y}_s^\Delta|^2 | \mathcal{A}_0\right) &\leq E\left(\sup_{0 \leq s \leq T} |1 + \mathbf{Y}_s^\Delta|^2 | \mathcal{A}_0\right) \\
&\leq C_1(1 + |\mathbf{Y}_0^\Delta|^2) \\
&\quad + K(\gamma, \lambda, \Delta) \int_0^T e^{K(\gamma, \lambda, \Delta)(T-u)} C_1(1 + |\mathbf{Y}_0^\Delta|^2) du \\
&\leq C_1(1 + |\mathbf{Y}_0^\Delta|^2) \\
&\quad + C_1(1 + |\mathbf{Y}_0^\Delta|^2)(e^{K(\gamma, \lambda, \Delta)T} - 1) \\
&\leq C_2(\gamma, \lambda, \Delta)(1 + |\mathbf{Y}_0^\Delta|^2), \tag{6.6.5}
\end{aligned}$$

where C is a positive finite constant independent of Δ . \square

Now by using Lemma 6.5.1 and Lemma 6.6.1, we can finally prove Theorem 6.4.1.

Proof of Theorem 6.4.1

- With the Wagner-Platen expansion (4.4.4) we can represent the solution of the SDE (1.8.2) as

$$\mathbf{X}_\tau = \sum_{\alpha \in \mathcal{A}_\gamma} \tilde{I}_\alpha[\tilde{f}_\alpha(\rho, \mathbf{X}_\rho)]_{\rho, \tau} + \sum_{\alpha \in \mathcal{B}(\mathcal{A}_\gamma)} \tilde{I}_\alpha[\tilde{f}_\alpha(\cdot, \mathbf{X}_\cdot)]_{\rho, \tau}, \tag{6.6.6}$$

for any two stopping times ρ and τ with $0 \leq \rho \leq \tau \leq T$ a.s. Therefore, we can express the solution of the SDE (1.8.2) at time $t \in [0, T]$ as

$$\begin{aligned}
\mathbf{X}_t &= \mathbf{X}_0 + \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \left\{ \sum_{n=0}^{n_t-1} \tilde{I}_\alpha[\tilde{f}_\alpha(t_n, \mathbf{X}_{t_n})]_{t_n, t_{n+1}} + \tilde{I}_\alpha[\tilde{f}_\alpha(t_{n_t}, \mathbf{X}_{t_{n_t}})]_{t_{n_t}, t} \right\} \\
&\quad + \sum_{\alpha \in \mathcal{B}(\mathcal{A}_\gamma)} \left\{ \sum_{n=0}^{n_t-1} \tilde{I}_\alpha[\tilde{f}_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[\tilde{f}_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_{n_t}, t} \right\}, \tag{6.6.7}
\end{aligned}$$

where n_t is defined as in equation (6.1.11).

We recall from (6.4.3) that the strong order γ compensated Taylor approximation \mathbf{Y}^Δ at time $t \in [0, T]$ is given by

$$\mathbf{Y}_t^\Delta = \mathbf{Y}_0^\Delta + \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \left\{ \sum_{n=0}^{n_t-1} \tilde{I}_\alpha[\tilde{f}_\alpha(t_n, \mathbf{Y}_{t_n}^\Delta)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[\tilde{f}_\alpha(t_{n_t}, \mathbf{Y}_{t_{n_t}}^\Delta)]_{t_{n_t}, t} \right\}. \quad (6.6.8)$$

From the moment estimate (1.9.5) provided by Theorem 1.9.3 we have

$$E \left(\sup_{0 \leq z \leq T} |\mathbf{X}_z|^2 \mid \mathcal{A}_0 \right) \leq C (1 + E(|\mathbf{X}_0|^2)). \quad (6.6.9)$$

By Lemma 6.6.1, we obtain a similar uniform estimate for the second moment of the approximation \mathbf{Y}^Δ in the form

$$E \left(\sup_{0 \leq z \leq T} |\mathbf{Y}_z^\Delta|^2 \mid \mathcal{A}_0 \right) \leq C (1 + |\mathbf{Y}_0^\Delta|^2). \quad (6.6.10)$$

2. Let us now analyze the mean square error of the strong order γ compensated Taylor approximation \mathbf{Y}^Δ . By (6.6.7), (6.6.8) and Cauchy-Schwarz inequality we obtain

$$\begin{aligned} Z(t) &= E \left(\sup_{0 \leq z \leq t} |\mathbf{X}_z - \mathbf{Y}_z^\Delta|^2 \mid \mathcal{A}_0 \right) \\ &= E \left(\sup_{0 \leq z \leq t} \left| \mathbf{X}_0 - \mathbf{Y}_0^\Delta \right. \right. \\ &\quad \left. \left. + \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \left\{ \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[\tilde{f}_\alpha(t_n, \mathbf{X}_{t_n}) - \tilde{f}_\alpha(t_n, \mathbf{Y}_{t_n}^\Delta)]_{t_n, t_{n+1}} \right. \right. \right. \\ &\quad \left. \left. \left. + \tilde{I}_\alpha[\tilde{f}_\alpha(t_{n_z}, \mathbf{X}_{t_{n_z}}) - \tilde{f}_\alpha(t_{n_z}, \mathbf{Y}_{t_{n_z}}^\Delta)]_{t_{n_z}, z} \right\} \right. \right. \\ &\quad \left. \left. + \sum_{\alpha \in \mathcal{B}(\mathcal{A}_\gamma)} \left\{ \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[\tilde{f}_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[\tilde{f}_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_{n_z}, z} \right\} \right|^2 \mid \mathcal{A}_0 \right) \\ &\leq C_3 \left\{ \left| \mathbf{X}_0 - \mathbf{Y}_0^\Delta \right|^2 + \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} S_t^\alpha + \sum_{\alpha \in \mathcal{B}(\mathcal{A}_\gamma)} U_t^\alpha \right\} \end{aligned} \quad (6.6.11)$$

for all $t \in [0, T]$, where S_t^α and U_t^α are defined as

$$\begin{aligned} S_t^\alpha &= E \left(\sup_{0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} \tilde{I}_\alpha [\tilde{f}_\alpha(t_n, \mathbf{X}_{t_n}) - \tilde{f}_\alpha(t_n, \mathbf{Y}_{t_n}^\Delta)]_{t_n, t_{n+1}} \right. \right. \\ &\quad \left. \left. + \tilde{I}_\alpha [\tilde{f}_\alpha(t_{n_z}, \mathbf{X}_{t_{n_z}}) - \tilde{f}_\alpha(t_{n_z}, \mathbf{Y}_{t_{n_z}}^\Delta)]_{t_{n_z}, z} \right|^2 \mid \mathcal{A}_0 \right), \quad (6.6.12) \end{aligned}$$

$$U_t^\alpha = E \left(\sup_{0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[\tilde{f}_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[\tilde{f}_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_{n_z}, z} \right|^2 \middle| \mathcal{A}_0 \right). \quad (6.6.13)$$

3. By using again Lemma 6.5.1 and the Lipschitz condition (6.4.7) we obtain

$$\begin{aligned} S_t^\alpha &= E \left(\sup_{0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} \tilde{I}_\alpha \left[\tilde{f}_\alpha(t_n, \mathbf{X}_{t_n}) - \tilde{f}_\alpha(t_n, \mathbf{Y}_{t_n}^\Delta) \right]_{t_n, t_{n+1}} \right. \right. \\ &\quad \left. \left. + \tilde{I}_\alpha \left[\tilde{f}_\alpha(t_{n_z}, \mathbf{X}_{t_{n_z}}) - \tilde{f}_\alpha(t_{n_z}, \mathbf{Y}_{t_{n_z}}^\Delta) \right]_{t_{n_z}, z} \right|^2 \middle| \mathcal{A}_0 \right) \\ &\leq C_4 \int_0^t \int_{\mathcal{E}} \dots \int_{\mathcal{E}} E \left(\sup_{0 \leq z \leq u} |\tilde{f}_\alpha(t_{n_z}, \mathbf{X}_{t_{n_z}}) - \tilde{f}_\alpha(t_{n_z}, \mathbf{Y}_{t_{n_z}}^\Delta)|^2 \middle| \mathcal{A}_0 \right) \\ &\quad \times \varphi(dv^1) \dots \varphi(dv^{s(\alpha)}) du \\ &\leq C_4 \int_{\mathcal{E}} \dots \int_{\mathcal{E}} (K_1(v^1, \dots, v^{s(\alpha)}))^2 \varphi(dv^1) \dots \varphi(dv^{s(\alpha)}) \\ &\quad \times \int_0^t E \left(\sup_{0 \leq z \leq u} |\mathbf{X}_{t_{n_z}} - \mathbf{Y}_{t_{n_z}}^\Delta|^2 \middle| \mathcal{A}_0 \right) du \\ &\leq C_5 \int_0^t Z(u) du. \end{aligned} \quad (6.6.14)$$

Applying again Lemma 6.5.1 and the linear growth condition (6.4.9) we obtain

$$\begin{aligned} U_t^\alpha &= E \left(\sup_{0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} \tilde{I}_\alpha[\tilde{f}_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_n, t_{n+1}} + \tilde{I}_\alpha[\tilde{f}_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_{n_z}, z} \right|^2 \middle| \mathcal{A}_0 \right) \\ &\leq C_5 \Delta^{\psi(\alpha)} \int_0^t \int_{\mathcal{E}} \dots \int_{\mathcal{E}} E \left(\sup_{0 \leq z \leq u} |\tilde{f}_\alpha(z, \mathbf{X}_z)|^2 \middle| \mathcal{A}_0 \right) \varphi(dv^1) \dots \varphi(dv^{s(\alpha)}) du \\ &\leq C_5 \Delta^{\psi(\alpha)} \int_{\mathcal{E}} \dots \int_{\mathcal{E}} K_2(v^1, \dots, v^{s(\alpha)}) \varphi(dv^1) \dots \varphi(dv^{s(\alpha)}) \\ &\quad \times \int_0^t E \left(\sup_{0 \leq z \leq u} (1 + |\mathbf{X}_z|^2) \middle| \mathcal{A}_0 \right) du \\ &\leq C_6 \Delta^{\psi(\alpha)} \left(t + \int_0^t E \left(\sup_{0 \leq z \leq u} |\mathbf{X}_z|^2 \middle| \mathcal{A}_0 \right) du \right), \end{aligned} \quad (6.6.15)$$

where

$$\psi(\alpha) = \begin{cases} 2l(\alpha) - 2 & : l(\alpha) = n(\alpha) \\ l(\alpha) + n(\alpha) - 1 & : l(\alpha) \neq n(\alpha). \end{cases}$$

Since we are now considering $\alpha \in \mathcal{B}(\mathcal{A}_\gamma)$, we have that $l(\alpha) \geq \gamma + 1$ when $l(\alpha) = n(\alpha)$ and $l(\alpha) + n(\alpha) \geq 2\gamma + 1$ when $l(\alpha) \neq n(\alpha)$, so that $\psi(\alpha) \geq 2\gamma$. Therefore, by applying the estimate (1.9.5) of Theorem 1.9.3 we obtain

$$\begin{aligned} U_t^\alpha &\leq C_6 \Delta^{2\gamma} \left(t + \int_0^t C_1 (1 + |\mathbf{X}_0|^2) du \right) \\ &\leq C_7 \Delta^{2\gamma} (1 + |\mathbf{X}_0|^2). \end{aligned} \quad (6.6.16)$$

4. Combining equations (6.6.11), (6.6.14) and (6.6.16) we obtain

$$Z(t) \leq C_8 \left\{ |\mathbf{X}_0 - \mathbf{Y}_0^\Delta|^2 + C_9 \Delta^{2\gamma} (1 + |\mathbf{X}_0|^2) + C_{10} \int_0^t Z(u) du \right\}. \quad (6.6.17)$$

By equations (6.6.9) and (6.6.10) $Z(t)$ is bounded. Therefore, by the Gronwall inequality we have

$$Z(T) \leq K_4 (1 + |\mathbf{X}_0|^2) \Delta^{2\gamma} + K_5 (|\mathbf{X}_0 - \mathbf{Y}_0^\Delta|^2). \quad (6.6.18)$$

Finally, by assumption (6.4.6), we obtain

$$\sqrt{E(\sup_{0 \leq z \leq T} |\mathbf{X}_z - \mathbf{Y}_z^\Delta|^2 | \mathcal{A}_0)} = \sqrt{Z(T)} \leq K_3 \Delta^\gamma, \quad (6.6.19)$$

which completes the proof of Theorem 6.4.1. \square

Remark 6.6.2. Note that the same result as that formulated in Theorem 6.4.1 holds also for the strong order γ Taylor approximation (6.4.4) as mentioned in Corollary 6.4.3. The proof uses analogous steps as those employed in the proof described above.

6.7 Exercises

6.1. Consider the two-dimensional SDE

$$\begin{aligned} dX_t^1 &= dW_t^1 \\ dX_t^2 &= X_t^1 dW_t^2 \end{aligned}$$

for $t \in [0, T]$, with $X_0^1 = X_0^2 = 1$. Here W^1 and W^2 are independent Wiener processes. Does the above SDE satisfy a diffusion commutativity condition?

6.2. Apply the Milstein scheme to the system of SDEs in Exercise 6.1.

6.3. Does the following diffusion process with jumps satisfy the jump commutativity condition if its SDE has the form

$$dX_t = a dt + 2 X_t dW_t + 2 X_{t-} p_\varphi(\mathcal{E}, dt)$$

for $t \geq 0$?

Regular Strong Itô Approximations

In this chapter we describe strong approximations on a regular time discretization that are more general than the regular strong Taylor approximations presented in the previous chapter. These approximations belong to the class of regular strong Itô schemes, which includes derivative-free, implicit and predictor-corrector schemes. More details on some of the results to be presented in this chapter can be found in Bruti-Liberati, Nikitopoulos-Sklibosios & Platen (2006) and Bruti-Liberati & Platen (2008).

7.1 Explicit Regular Strong Schemes

Introduction to Regular Strong Itô Schemes

The first types of schemes that we describe in this chapter are the so-called *derivative-free schemes*. Higher strong order Taylor schemes, as the strong order 1.0 Taylor scheme presented in Sect. 6.2, are rather complex as they involve the evaluation of derivatives of the drift, diffusion and jump coefficients at each time step. For the implementation of general numerical routines for the approximation of jump diffusion SDEs, without assuming a particular form for the coefficients, this constitutes a serious limitation. In principle, one is required to include a symbolic differentiation into a numerical algorithm. For these reasons, we present in this chapter derivative-free strong schemes that avoid the computation of derivatives.

In the second part of this chapter we present *implicit schemes*. As shown in Hofmann & Platen (1996) and Higham & Kloeden (2005, 2006), see also Chap. 14, in the presence of multiplicative noise explicit methods show narrow regions of numerical stability. We emphasize that SDEs with multiplicative noise are typically used when modeling asset prices in finance. They also arise in other important applications such as hidden Markov chain filtering, see Elliott et al. (1995) and Chap. 10. In order to construct approximate filters,

one needs a strong discrete-time approximation of an SDE with multiplicative noise, the Zakai equation. Moreover, in filtering problems for large systems it is not always possible to use small time step sizes, as the computations may not be performed fast enough to keep pace with the arrival of data. Therefore, for this kind of applications, higher strong order schemes with wide regions of numerical stability are crucial. To overcome some of these problems, we describe implicit schemes that have satisfactory numerical stability properties.

As we will see later in Chap. 14, explicit schemes have much narrower regions of numerical stability than corresponding implicit schemes. For this reason implicit schemes for SDEs with jumps have been proposed. Because of their improved numerical stability, implicit schemes can be used with much larger time step sizes than those required by explicit schemes. However, implicit schemes carry, in general, an additional computational burden since they usually require the solution of an algebraic equation at each time step. Therefore, in choosing between an explicit and an implicit scheme one faces a trade-off between computational efficiency and numerical stability. Additionally, as will be explained later, when designing an implicit scheme, it is not easy to introduce implicitness in its diffusion part. This is due to problems that arise with the presence of the reciprocal of Gaussian random variables for most SDEs when using *ad hoc* implicit schemes. For these reasons, we will present new *predictor-corrector schemes* that aim to combine good numerical stability properties and efficiency. A detailed investigation of the numerical stability of such predictor-corrector schemes will be given in Chap. 14.

Derivative-Free Strong Order 1.0 Scheme

By replacing the derivatives in the strong order 1.0 Taylor scheme, presented in Sect. 6.2, by the corresponding difference ratios, it is possible to obtain a scheme that does not require the evaluation of derivatives and achieves the same strong order of convergence. However, to construct the difference ratios we need supporting values of the coefficients at additional points.

In the one-dimensional case, $d = m = 1$, the *derivative-free strong order 1.0 scheme*, is given by

$$\begin{aligned}
 Y_{n+1} = & Y_n + a\Delta_n + b\Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) p_\varphi(dv, dz) \\
 & + \frac{(b(t_n, \bar{Y}_n) - b)}{\sqrt{\Delta_n}} \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1} dW_{z_2} \\
 & + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \frac{(c(t_n, \bar{Y}_n, v) - c(v))}{\sqrt{\Delta_n}} dW_{z_1} p_\varphi(dv, dz_2) \\
 & + \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ b(t_n, Y_n + c(v)) - b \right\} p_\varphi(dv, dz_1) dW_{z_2} \\
 & + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ c(t_n, Y_n + c(v_2), v_1) - c(v_1) \right\} p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2),
 \end{aligned} \tag{7.1.1}$$

with the supporting value

$$\bar{Y}_n = Y_n + b\sqrt{\Delta_n}. \quad (7.1.2)$$

The scheme (7.1.1)–(7.1.2) generally achieves a strong order $\gamma = 1.0$ and is a generalization of a corresponding scheme proposed in Platen (1984) for pure diffusions.

We can simplify the double stochastic integrals appearing in (7.1.1), as in Sect. 6.2, and rewrite the derivative-free strong order 1.0 Taylor scheme as

$$\begin{aligned} Y_{n+1} &= Y_n + a\Delta_n + b\Delta W_n + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c(\xi_i) + \frac{(b(t_n, \bar{Y}_n) - b)}{2\sqrt{\Delta_n}} ((\Delta W_n)^2 - \Delta_n) \\ &+ \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} \frac{(c(t_n, \bar{Y}_n, \xi_i) - c(\xi_i))}{\sqrt{\Delta_n}} (W_{\tau_i} - W_{t_n}) \\ &+ \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} \left\{ b(Y_n + c(\xi_i)) - b \right\} (W_{t_{n+1}} - W_{\tau_i}) \\ &+ \sum_{j=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(\tau_j)} \left\{ c(Y_n + c(\xi_i), \xi_j) - c(\xi_j) \right\}, \end{aligned} \quad (7.1.3)$$

with supporting value (7.1.2), which can be directly used in scenario simulation.

As discussed in Sect. 6.3, in the case of mark-independent jump size it is recommended to check the particular structure of the SDE under consideration. Indeed, we derived the jump commutativity condition (6.3.2) under which the strong order 1.0 Taylor scheme (6.2.4) exhibits a computational complexity that is independent of the intensity level of the Poisson random measure.

For the derivative-free strong order 1.0 scheme (7.1.1)–(7.1.2) with mark-independent jump size, the derivative-free coefficient of the multiple stochastic integral $I_{(1,-1)}$, which is of the form

$$\frac{c(t_n, \bar{Y}_n) - c(t_n, Y_n)}{\sqrt{\Delta_n}},$$

depends on the time step size Δ_n . On the other hand, the coefficient of the multiple stochastic integral $I_{(-1,1)}$,

$$b(t_n, Y_n + c(t_n, Y_n)) - b(t_n, Y_n),$$

is independent of Δ_n . Therefore, it is not possible to directly derive a commutativity condition similar to (6.3.2) that permits us to identify special classes

of SDEs for which the computational efficiency is independent of the jump intensity level.

For instance, for the SDE (1.8.5) with mark-independent jump size $c(t, x, v) = x\beta$, with $\beta \geq -1$, the derivative-free strong order 1.0 scheme is given by

$$\begin{aligned} Y_{n+1} &= Y_n + \mu Y_n \Delta_n + \sigma Y_n \Delta W_n + \beta Y_n \Delta p_n + \frac{\sigma}{\sqrt{\Delta}} \{\bar{Y}_n - Y_n\} I_{(1,1)} \\ &\quad + \frac{\beta}{\sqrt{\Delta}} \{\bar{Y}_n - Y_n\} I_{(1,-1)} + \sigma \beta Y_n I_{(-1,1)} + \beta^2 Y_n I_{(-1,-1)}, \end{aligned} \quad (7.1.4)$$

with the supporting value

$$\bar{Y}_n = Y_n + \sigma Y_n \sqrt{\Delta_n}. \quad (7.1.5)$$

Since the evaluation of the multiple stochastic integrals $I_{(1,-1)}$ and $I_{(-1,1)}$, as given in (6.2.5), depends on the number of jumps, the computational efficiency of the scheme (7.1.4)–(7.1.5) depends on the total intensity λ of the jump measure.

Let us consider the special class of one-dimensional SDEs satisfying the jump commutativity condition (6.3.2), which we recall here in the form

$$b(t, x) \frac{\partial c(t, x)}{\partial x} = b(t, x + c(t, x)) - b(t, x), \quad (7.1.6)$$

for all $t \in [0, T]$ and $x \in \mathfrak{R}$. Under this condition, we should first derive the strong order 1.0 Taylor scheme, using the relationship

$$I_{(1,-1)} + I_{(-1,1)} = \Delta p_n \Delta W_n, \quad (7.1.7)$$

obtaining

$$\begin{aligned} Y_{n+1} &= Y_n + a \Delta_n + b \Delta W_n + c \Delta p_n + \frac{b b'}{2} \{(\Delta W_n)^2 - \Delta_n\} \\ &\quad + \{b(t_n, Y_n + c) - b\} \Delta p_n \Delta W_n + \frac{\{c(t_n, Y_n + c) - c\}}{2} \{(\Delta p_n)^2 - \Delta p_n\}, \end{aligned} \quad (7.1.8)$$

where we have used again the abbreviation (6.1.12). Then, we replace the derivative b' by the corresponding difference ratio and obtain a *derivative-free strong order 1.0 scheme*

$$\begin{aligned} Y_{n+1} &= Y_n + a \Delta_n + b \Delta W_n + c \Delta p_n \\ &\quad + \frac{\{b(t_n, \bar{Y}_n) - b\}}{2\sqrt{\Delta_n}} \{(\Delta W_n)^2 - \Delta_n\} + \{b(t_n, Y_n + c) - b\} \Delta p_n \Delta W_n \\ &\quad + \frac{\{c(t_n, Y_n + c) - c\}}{2} \{(\Delta p_n)^2 - \Delta p_n\}, \end{aligned} \quad (7.1.9)$$

with supporting value given in (7.1.2). The computational efficiency is here independent of the intensity level.

Let us discuss an even more specific example. For the SDE (1.8.5) with $c(t, x, v) = x\beta$, for $\beta \geq -1$, we can derive the derivative-free strong order 1.0 scheme, which, due to the multiplicative form of the diffusion coefficient, is the same as the strong order 1.0 Taylor scheme (6.3.6).

Multi-dimensional Strong Order 1.0 Scheme

In the multi-dimensional case with scalar Wiener process, which means $m = 1$, and mark-dependent jump size, the k th component of the *derivative-free strong order 1.0 scheme* is given by

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + a^k \Delta_n + b^k \Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c^k(v) p_\varphi(dv, dz) \\ &\quad + \frac{(b^k(t_n, \bar{\mathbf{Y}}_n) - b^k)}{\sqrt{\Delta_n}} \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1} dW_{z_2} \\ &\quad + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \frac{(c^k(t_n, \bar{\mathbf{Y}}_n, v) - c^k(v))}{\sqrt{\Delta_n}} dW_{z_1} p_\varphi(dv, dz_2) \\ &\quad + \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{b^k(t_n, \mathbf{Y}_n + \mathbf{c}(v)) - b^k\} p_\varphi(dv, dz_1) dW_{z_2} \\ &\quad + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{c^k(t_n, \mathbf{Y}_n + \mathbf{c}(v_2), v_1) - c^k(v_1)\} \\ &\quad \times p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2), \end{aligned} \tag{7.1.10}$$

with the supporting value (7.1.2).

As noticed for the one-dimensional situation, even in the case of mark-independent jump size, it is not possible to derive a jump commutativity condition similar to (6.3.7), since the coefficient of the multiple stochastic integral $I_{(1,-1)}$ depends on the time step size Δ_n . However, as shown in Sect. 6.3, it makes sense to consider the special class of multi-dimensional SDEs with a scalar Wiener process and mark-independent jump size characterized by the jump commutativity condition

$$\sum_{l=1}^d b^l(t, \mathbf{x}) \frac{\partial c^k(t, \mathbf{x})}{\partial x^l} = b^k(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x})) - b^k(t, \mathbf{x}) \tag{7.1.11}$$

for $k \in \{1, 2, \dots, d\}$, $t \in [0, T]$ and $\mathbf{x} \in \mathbb{R}^d$. Using the relationship (7.1.7) one can derive in this case a strong order 1.0 Taylor scheme

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + a^k \Delta_n + b^k \Delta W_n + c^k \Delta p_n + \frac{1}{2} \sum_{l=1}^d b^l \frac{\partial b^k}{\partial x^l} \{(\Delta W_n)^2 - \Delta_n\} \\ &\quad + \{b^k(t_n, \mathbf{Y}_n + \mathbf{c}) - b^k\} \Delta p_n \Delta W_n + \frac{\{c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k\}}{2} \{(\Delta p_n)^2 - \Delta p_n\}, \end{aligned} \tag{7.1.12}$$

whose computational complexity is independent of the intensity λ . Replacing the coefficient

$$\sum_{l=1}^d b^l \frac{\partial b^k}{\partial x^l}$$

by the corresponding difference ratio, we obtain a derivative-free strong order 1.0 Taylor scheme of the form

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + a^k \Delta_n + b^k \Delta W_n + c^k \Delta p_n \\ &+ \frac{\{b^k(t_n, \bar{\mathbf{Y}}_n) - b^k\}}{2\sqrt{\Delta_n}} \{(\Delta W_n)^2 - \Delta_n\} + \{b^k(t_n, \mathbf{Y}_n + \mathbf{c}) - b^k\} \Delta p_n \Delta W_n \\ &+ \frac{\{c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k\}}{2} \{(\Delta p_n)^2 - \Delta p_n\}, \end{aligned} \quad (7.1.13)$$

with supporting value (7.1.2), for $k \in \{1, 2, \dots, d\}$. The computational complexity of the scheme (7.1.13) is independent of the jump intensity level.

In the general multi-dimensional case the k th component of the *derivative-free strong order 1.0 scheme* is given by

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c^k(v) p_\varphi(dv, dz) \\ &+ \frac{1}{\sqrt{\Delta_n}} \sum_{j_1, j_2=1}^m \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \{b^{k, j_1}(t_n, \bar{\mathbf{Y}}_n^{j_2}) - b^{k, j_1}\} dW_{z_1}^{j_1} dW_{z_2}^{j_2} \\ &+ \sum_{j_1=1}^m \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \{c^{k, j_1}(t_n, \bar{\mathbf{Y}}_n^{j_2}) - c^{k, j_1}\} dW_{z_1}^{j_1} p_\varphi(dv, dz_2) \\ &+ \sum_{j_1=1}^m \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \{b^{k, j_1}(t_n, \mathbf{Y}_n + \mathbf{c}(v)) - b^{k, j_1}\} p_\varphi(dv, dz_2) dW_{z_1}^{j_1} \\ &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{c^k(t_n, \mathbf{Y}_n + \mathbf{c}(v_1), v_2) - c^k(v_2)\} \\ &\times p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2), \end{aligned} \quad (7.1.14)$$

for $k \in \{1, 2, \dots, d\}$, with the vector supporting values

$$\bar{\mathbf{Y}}_n^j = \mathbf{Y}_n + \mathbf{b}^j \sqrt{\Delta_n}, \quad (7.1.15)$$

for $j \in \{1, 2, \dots, m\}$.

As shown in Sect. 6.3, for the special class of general multi-dimensional SDEs with mark-independent jump size, satisfying the jump commutativity condition

$$\sum_{l=1}^d b^{l,j_1}(t, \mathbf{x}) \frac{\partial c^k(t, \mathbf{x})}{\partial x^l} = b^{k,j_1}(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x})) - b^{k,j_1}(t, \mathbf{x}) \quad (7.1.16)$$

for $j_1 \in \{1, 2, \dots, m\}$, $k \in \{1, 2, \dots, d\}$, $t \in [0, T]$ and $\mathbf{x} \in \mathbb{R}^d$, it is possible to derive a strong order 1.0 Taylor scheme whose computational complexity is independent of the intensity level. Here one needs to use the relationship

$$I_{(j_1, -1)} + I_{(-1, j_1)} = \Delta p_n \Delta W_n^{j_1},$$

for $j_1 \in \{1, 2, \dots, m\}$. Then, by replacing the coefficients involving the derivatives, by the corresponding difference ratios, we obtain a derivative-free strong order 1.0 scheme that shows a computational complexity independent of the jump intensity level. This scheme is given by

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + c^k \Delta p_n \\ &+ \frac{1}{2\sqrt{\Delta_n}} \sum_{j_1, j_2=1}^m \{b^{k,j_1}(t_n, \bar{\mathbf{Y}}_n^{j_2}) - b^{k,j_1}(t_n, \mathbf{Y}_n)\} I_{(j_1, j_2)} \\ &+ \sum_{j=1}^m \{b^{k,j_1}(t_n, \mathbf{Y}_n + \mathbf{c}) - b^{k,j_1}\} \Delta p_n \Delta W_n^{j_1} \\ &+ \frac{\{c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k\}}{2} \{(\Delta p_n)^2 - \Delta p_n\}, \end{aligned} \quad (7.1.17)$$

for $k \in \{1, 2, \dots, d\}$, with the vector supporting values (7.1.15). For the generation of the multiple stochastic integral $I_{(j_1, j_2)}$, for $j_1, j_2 \in \{1, 2, \dots, d\}$, we refer to Sect. 6.3.

For the special case of a multi-dimensional SDE satisfying the jump commutativity condition (7.1.16), the diffusion commutativity condition

$$L^{j_1} b^{k,j_2}(t, \mathbf{x}) = L^{j_2} b^{k,j_1}(t, \mathbf{x}) \quad (7.1.18)$$

for $j_1, j_2 \in \{1, 2, \dots, m\}$, $k \in \{1, 2, \dots, d\}$, $t \in [0, T]$ and $\mathbf{x} \in \mathbb{R}^d$, and with mark-independent jump size, we obtain an efficiently implementable derivative-free strong order 1.0 scheme. Its k th component is given by

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + c^k \Delta p_n \\ &+ \frac{1}{2} \sum_{j_1, j_2=1}^m \{b^{k,j_1}(t_n, \bar{\mathbf{Y}}_n^{j_2}) - b^{k,j_1}(t_n, \mathbf{Y}_n)\} \{\Delta W_n^{j_1} \Delta W_n^{j_2} - \Delta_n\} \end{aligned}$$

$$\begin{aligned}
& + \sum_{j_1=1}^m \{b^{k,j_1}(t_n, \mathbf{Y}_n + \mathbf{c}) - b^{k,j_1}\} (\Delta p_n \Delta W_n^{j_1}) \\
& + \frac{1}{2} \{c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k\} ((\Delta p_n)^2 - \Delta p_n), \tag{7.1.19}
\end{aligned}$$

for $k \in \{1, 2, \dots, d\}$, with the vector supporting values (7.1.15).

7.2 Drift-Implicit Schemes

In general, given an explicit scheme of strong order γ it is usually possible to obtain a similar drift-implicit scheme of the same strong order. However, since the reciprocal of a Gaussian random variable does not have finite absolute moments, it is not easy to introduce implicitness in the diffusion part. As we will see later in Chap. 14, regions of numerical stability of drift-implicit schemes are typically wider than those of corresponding explicit schemes. Therefore, the former are often more suitable for a certain range of problems than corresponding explicit schemes. In this section we present drift-implicit strong schemes. In Higham & Kloeden (2005, 2006, 2007) a class of implicit methods of strong order $\gamma = 0.5$ for SDEs with jumps has been proposed. A detailed stability analysis in Chap. 14 will show that these schemes have good numerical stability properties. In the following we focus on drift-implicit schemes of strong order $\gamma = 0.5$ and $\gamma = 1.0$ for the SDE (6.1.4) with jumps.

Drift-Implicit Euler Scheme

In the one-dimensional case, $d = m = 1$, by introducing implicitness in the drift of the Euler scheme (6.1.15), we obtain the *drift-implicit Euler scheme*,

$$Y_{n+1} = Y_n + \{\theta a(t_{n+1}, Y_{n+1}) + (1 - \theta) a\} \Delta_n + b \Delta W_n + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c^k(\xi_i), \tag{7.2.1}$$

where the parameter $\theta \in [0, 1]$ characterizes the degree of implicitness and we have used the abbreviation defined in (6.1.12). For $\theta = 0$ we recover the Euler scheme (6.1.15), while for $\theta = 1$, we obtain a fully drift-implicit Euler scheme. The scheme (7.2.1) achieves a strong order of convergence $\gamma = 0.5$. This scheme was proposed and analyzed in Higham & Kloeden (2006) for SDEs driven by Wiener processes and homogeneous Poisson processes. It generalizes the drift-implicit Euler scheme for pure diffusions presented in Talay (1982b) and Milstein (1988a).

By comparing the drift-implicit Euler scheme (7.2.1) with the Euler scheme (6.1.15), one notices that the implementation of the former requires an additional computational effort. An algebraic equation has to be solved at

each time step, which can be performed, for instance, by a Newton-Raphson method. In special cases, however, the algebraic equation may admit an explicit solution. Note that the existence and uniqueness of the solution of this algebraic equation is guaranteed by Banach's fixed point theorem for every

$$\Delta \leq \frac{1}{\sqrt{K\theta}},$$

where K is the Lipschitz constant appearing in the Lipschitz condition (1.9.2) for the drift coefficient a , see for instance [Evans \(1999\)](#).

When we have a mark-independent jump size we obtain the *drift-implicit Euler scheme*

$$Y_{n+1} = Y_n + \{\theta a(t_{n+1}, Y_{n+1}) + (1 - \theta) a\} \Delta_n + b \Delta W_n + c \Delta p_n. \quad (7.2.2)$$

In the multi-dimensional case with scalar Wiener process, $m = 1$, and mark-dependent jump size, the k th component of the *drift-implicit Euler scheme* is given by

$$Y_{n+1}^k = Y_n^k + \{\theta a^k(t_{n+1}, Y_{n+1}) + (1 - \theta) a^k\} \Delta_n + b^k \Delta W_n + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c^k(\xi_i),$$

for $k \in \{1, 2, \dots, d\}$.

In the case of a mark-independent jump size the k -th component of the above *drift-implicit Euler scheme* reduces to

$$Y_{n+1}^k = Y_n^k + \{\theta a^k(t_{n+1}, Y_{n+1}) + (1 - \theta) a^k\} \Delta_n + b^k \Delta W_n + c^k \Delta p_n, \quad (7.2.3)$$

for $k \in \{1, 2, \dots, d\}$.

For the general multi-dimensional case with mark-dependent jump size the k th component of the *drift-implicit Euler scheme* is of the form

$$Y_{n+1}^k = Y_n^k + \{\theta a^k(t_{n+1}, \mathbf{Y}_{n+1}) + (1 - \theta) a^k\} \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c^k(\xi_i),$$

for $k \in \{1, 2, \dots, d\}$ and $j \in \{1, 2, \dots, m\}$.

Finally, in the multi-dimensional case with mark-independent jump size the k th component of the *drift-implicit Euler scheme* is given by

$$Y_{n+1}^k = Y_n^k + \{\theta a^k(t_{n+1}, \mathbf{Y}_{n+1}) + (1 - \theta) a^k\} \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + c^k \Delta p_n,$$

for $k \in \{1, 2, \dots, d\}$.

Drift-Implicit Strong Order 1.0 Scheme

In a similar way as for the Euler scheme, by introducing implicitness in the drift of the strong order 1.0 Taylor scheme presented in Sect. 6.2, we obtain the *drift-implicit strong order 1.0 scheme*.

In the one-dimensional case, $d = m = 1$, the *drift-implicit strong order 1.0 scheme* is given by

$$\begin{aligned}
 Y_{n+1} &= Y_n + \{\theta a(t_{n+1}, Y_{n+1}) + (1 - \theta) a\} \Delta_n + b \Delta W_n \\
 &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) p_\varphi(dv, dz) + bb' \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1} dW_{z_2} \\
 &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b c'(v) dW_{z_1} p_\varphi(dv, dz_2) \\
 &+ \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ b(t_n, Y_n + c(v)) - b \right\} p_\varphi(dv, dz_1) dW_{z_2} \\
 &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ c(t_n, Y_n + c(v_1), v_2) - c(v_2) \right\} p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2),
 \end{aligned} \tag{7.2.4}$$

where

$$b'(t, x) = \frac{\partial b(t, x)}{\partial x} \quad \text{and} \quad c'(t, x, v) = \frac{\partial c(t, x, v)}{\partial x}. \tag{7.2.5}$$

For simplicity, we have used the convention (6.1.12). Here the parameter $\theta \in [0, 1]$, characterizes again the degree of implicitness. This scheme achieves a strong order of convergence $\gamma = 1.0$. Note that the degree of implicitness θ can, in principle, be also chosen to be greater than one if this helps the numerically stability of the scheme. This scheme generalizes a first strong order drift-implicit scheme for pure diffusions presented in Talay (1982a) and Milstein (1988a).

One can simplify the double stochastic integrals appearing in the scheme (7.2.4) as shown for the strong order 1.0 Taylor scheme (6.2.3). This makes the resulting scheme more applicable to scenario simulation.

The jump commutativity condition (6.3.2), presented in Sect. 6.3, also applies to drift-implicit schemes. Therefore, for the class of SDEs identified by the jump commutativity condition (6.3.2) the computational efficiency of drift-implicit schemes of strong order $\gamma = 1.0$ is independent of the intensity level of the Poisson measure. For instance, for the SDE (1.8.5) with $c(t, x, v) = x\beta$ and $\beta \geq -1$, it is possible to derive a drift-implicit strong order 1.0 scheme, given by

$$\begin{aligned}
 Y_{n+1} &= \frac{Y_n}{1 - \mu\theta\Delta_n} \left\{ 1 + (1 - \theta) \mu\Delta_n + \sigma\Delta W_n + \beta\Delta p_n + \frac{1}{2} \sigma^2 \{(\Delta W_n)^2 - \Delta_n\} \right. \\
 &\quad \left. + \sigma\beta\Delta p_n\Delta W_n + \frac{1}{2} \beta^2 \{(\Delta p_n)^2 - \Delta p_n\} \right\},
 \end{aligned} \tag{7.2.6}$$

which is also efficient in the case of a high intensity jump measure. Note that θ should be chosen different from $(\mu\Delta_n)^{-1}$, which automatically arises for sufficiently small step sizes Δ_n .

In the multi-dimensional case with scalar Wiener process, $m = 1$, and mark-dependent jump size, the k th component of the *drift-implicit strong order 1.0 scheme* is given by

$$\begin{aligned}
 Y_{n+1}^k &= Y_n^k + \{\theta a^k(t_{n+1}, \mathbf{Y}_{n+1}) + (1 - \theta) a^k\} \Delta_n + b^k \Delta W_n \\
 &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c^k(v) p_\varphi(dv, dz) + \sum_{l=1}^d \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} b^l \frac{\partial b^k}{\partial x^l} dW_{z_1} dW_{z_2} \\
 &+ \sum_{l=1}^d \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b^l \frac{\partial c^k(v)}{\partial x^l} dW_{z_1} p_\varphi(dv, dz_2) \\
 &+ \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{b^k(t_n, \mathbf{Y}_n + \mathbf{c}(v)) - b^k\} p_\varphi(dv, dz_1) dW_{z_2} \\
 &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{c^k(t_n, \mathbf{Y}_n + \mathbf{c}(v_1), v_2) - c^k(v_2)\} \\
 &\quad \times p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2), \tag{7.2.7}
 \end{aligned}$$

for $k \in \{1, 2, \dots, d\}$.

For the special class of multi-dimensional SDEs with scalar Wiener process and mark-independent jump size, satisfying the jump commutativity condition (7.1.11), the k th component of the *drift-implicit strong order 1.0 scheme*, given by

$$\begin{aligned}
 Y_{n+1}^k &= Y_n^k + \{\theta a^k(t_{n+1}, \mathbf{Y}_{n+1}) + (1 - \theta) a^k\} \Delta_n + b^k \Delta W_n \\
 &+ c^k \Delta p_n + \frac{1}{2} \sum_{l=1}^d b^l \frac{\partial b^k}{\partial x^l} \{(\Delta W_n)^2 - \Delta_n\} \\
 &+ \{b^k(t_n, \mathbf{Y}_n + \mathbf{c}) - b^k\} \Delta p_n \Delta W_n + \frac{\{c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k\}}{2} \{(\Delta p_n)^2 - \Delta p_n\}, \tag{7.2.8}
 \end{aligned}$$

shows a computational efficiency independent of the jump intensity level.

In the general multi-dimensional case the k th component of the *drift-implicit strong order 1.0 scheme* is given by

$$\begin{aligned}
 Y_{n+1}^k &= Y_n^k + \{\theta a^k(t_{n+1}, \mathbf{Y}_{n+1}) + (1 - \theta) a^k\} \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j \\
 &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c^k(v) p_\varphi(dv, dz)
 \end{aligned}$$

$$\begin{aligned}
& + \sum_{j_1, j_2=1}^m \sum_{i=1}^d \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \left\{ b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \right\} dW_{z_1}^{j_1} dW_{z_2}^{j_2} \\
& + \sum_{j_1=1}^m \sum_{i=1}^d \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b^{i, j_1} \frac{\partial c^k(v)}{\partial x^i} dW_{z_1}^{j_1} p_\varphi(dv, dz_2) \\
& + \sum_{j_1=1}^m \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \left\{ b^{k, j_1}(t_n, \mathbf{Y}_n + \mathbf{c}(v)) - b^{k, j_1} \right\} \\
& \quad \times p_\varphi(dv, dz_2) dW_{z_2}^{j_1} \\
& + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ c^k(t_n, \mathbf{Y}_n + \mathbf{c}(v_1), v_2) - c^k(v_2) \right\} \\
& \quad \times p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2), \tag{7.2.9}
\end{aligned}$$

for $k \in \{1, 2, \dots, d\}$.

In the implementation of drift-implicit schemes for multi-dimensional SDEs it is important to exploit the specific structure of the SDE under consideration. For instance, for multi-dimensional SDEs satisfying the diffusion commutativity condition (7.1.18) as well as the jump commutativity condition (7.1.16) and with mark-independent jump size, we obtain an efficient, implementable *drift-implicit strong order 1.0 scheme*, whose k th component is given by

$$\begin{aligned}
Y_{n+1}^k & = Y_n^k + \left\{ \theta a^k(t_{n+1}, \mathbf{Y}_{n+1}) + (1 - \theta) a^k \right\} \Delta_n + \sum_{j=1}^m b^{k, j} \Delta W_n^j \\
& + c^k \Delta p_n + \frac{1}{2} \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \left\{ \Delta W_n^{j_1} \Delta W_n^{j_2} - \Delta_n \right\} \\
& + \sum_{j_1=1}^m \left\{ b^{k, j_1}(t_n, \mathbf{Y}_n + \mathbf{c}) - b^{k, j_1} \right\} (\Delta p_n \Delta W_n^{j_1}) \\
& + \frac{1}{2} \left\{ c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k \right\} ((\Delta p_n)^2 - \Delta p_n), \tag{7.2.10}
\end{aligned}$$

for $k \in \{1, 2, \dots, d\}$. The special case of additive diffusion and jump coefficient, which means $\mathbf{b}(t, \mathbf{x}) = \mathbf{b}(t)$ and $\mathbf{c}(t, \mathbf{x}) = \mathbf{c}(t)$, satisfies all the required commutativity conditions and, therefore, leads to an efficient drift-implicit strong order 1.0 scheme.

7.3 Balanced Implicit Methods

This section describes an important implicit method that enables one to overcome a range of numerical instabilities.

Implicit Approximation of Diffusion Terms

A major difficulty arises from the fact that all previously described strong schemes do not provide implicit expressions for the diffusion terms. Only drift terms are made implicit. This turns out to be an important limitation of the above strong approximation methods. One cannot simply introduce implicit diffusion terms as it was possible for the drift. As we will discuss below, in an *ad hoc* implicit, discrete-time approximation with implicit diffusion term that would typically result in terms which include the inverse of Gaussian random variables. Such terms naturally explode for certain values and, thus, do not permit the construction of reasonable schemes.

Drift implicit methods are well adapted for systems with small noise and additive noise. But for those cases in which the diffusion part plays an essential role in the dynamics, as with relatively large multiplicative noise processes, the application of fully implicit methods, involving also implicit diffusion terms, is unavoidable.

An illustration for such a situation is provided by the one-dimensional SDE

$$dX_t = \sigma X_t dW_t \quad (7.3.1)$$

for $t \in [0, T]$, starting at $X_0 = x_0$. Here $W = \{W_t, t \in [0, T]\}$ is a standard Wiener process. The volatility σ is, for simplicity, chosen to be constant. The SDE (7.3.1) describes important dynamics in finance. It characterizes for the standard Black-Scholes model the martingale dynamics of a discounted or benchmarked asset price under the respective probability measure. These are central dynamics that one has to deal with frequently when simulating in finance.

As we will show later on, explicit strong methods can generate large errors for large time step sizes. They may even lead to explosions resulting in computer overflow. On the other hand, using an extremely small time step size may require unrealistic computational time and can also lead to some kind of numerical instability, as will be demonstrated later in Chap. 14.

Obviously, one cannot apply drift-implicit schemes to improve the numerical stability of the approximate solution for the SDE (7.3.1) since it does not contain any drift term. Thus, we have to construct fully implicit methods, which involve implicitness in diffusion terms.

Simulation Example using the Balanced Method

The simplest useful simulation method for the SDE (7.3.1) is provided by the Euler scheme (5.3.10), which has for (7.3.1) the form

$$Y_{n+1} = Y_n + \sigma Y_n \Delta W_n. \quad (7.3.2)$$

Here ΔW is independent $N(0, \Delta)$ Gaussian distributed and $Y_0 = X_0$. The Euler scheme is an explicit method. In fact, there is no simple stochastic counterpart of the deterministic implicit Euler method since the *ad hoc* scheme

$$Y_{n+1} = Y_n + \sigma Y_{n+1} \Delta W_n, \quad (7.3.3)$$

that is

$$Y_{n+1} = \frac{Y_n}{1 - \sigma \Delta W_n}, \quad (7.3.4)$$

fails because it can create division by zero. Furthermore, the absolute moment of the inverse of a Gaussian random variable is infinite, that is

$$E|(1 - \sigma \Delta W_n)^{-1}| = +\infty. \quad (7.3.5)$$

This demonstrates that it is not trivial to construct a method that is implicit in its diffusion term.

The Balanced Implicit Method

In Milstein, Platen & Schurz (1998) a family of *balanced implicit methods* has been proposed that overcomes part of the problem. In the one-dimensional case for $m = d = 1$ a balanced implicit method can be written in the form

$$Y_{n+1} = Y_n + a \Delta + b \Delta W + (Y_n - Y_{n+1}) C_n, \quad (7.3.6)$$

where

$$C_n = c^0(Y_n) \Delta + c^1(Y_n) |\Delta W_n| \quad (7.3.7)$$

and c^0, c^1 represent positive, real valued uniformly bounded functions. The freedom of choosing c^0 and c^1 can be exploited to construct a numerically stable scheme, tailored for the dynamics of the given SDE. Note however, the balanced implicit method is only of strong order $\gamma = 0.5$ since it is, in principle, a variation of the Euler scheme. This low order strong convergence is the price that one has to pay for obtaining numerical stability. As mentioned already, any numerically stable scheme is better than an unstable one with higher order.

The balanced implicit method can be interpreted as a family of specific methods providing a kind of balance between approximating diffusion terms in the numerical scheme. One can expect that by an appropriate choice of the parameters involved in these schemes one is able to find an acceptable combination that is suitable for the approximate solution of a given SDE. Numerical experiments will demonstrate the better behavior of the balanced implicit method in comparison with the explicit Euler method. As will be discussed in Chap. 14, the method has a large working range of suitable time step sizes without numerical instabilities arising. This is in contrast to the

explicit Euler method, where one has to be careful when using relatively large time step sizes.

In a number of applications, in particular, those involving SDEs with multiplicative noise as is typical in finance and filtering, see [Fischer & Platen \(1999\)](#) and Sect. 10.4, balanced implicit methods show better numerical stability than most other methods.

The General Balanced Implicit Method

To formulate the balanced implicit method for more general systems without jumps, that is $m, d \in \{1, 2, \dots\}$, we suppose that the d -dimensional stochastic process $\mathbf{X} = \{\mathbf{X}_t, t \in [0, T]\}$ with $E(\mathbf{X}_0)^2 < \infty$ satisfies the d -dimensional SDE

$$d\mathbf{X}_t = \mathbf{a}(t, \mathbf{X}_t) dt + \sum_{j=1}^m \mathbf{b}^j(t, \mathbf{X}_t) dW_t^j, \quad (7.3.8)$$

where $\mathbf{a}, \mathbf{b}^1, \dots, \mathbf{b}^m$ are d -dimensional Lipschitz continuous vector-valued functions, which fulfill also a linear growth condition. The processes $W^j = \{W_t^j, t \in [0, T]\}$ for $j \in \{1, 2, \dots, m\}$ represent independent standard Wiener processes.

Now let us introduce a *family of balanced implicit methods*. A balanced implicit method applied to (7.3.8) can be written in the general form

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \mathbf{a}(\tau_n, \mathbf{Y}_n) \Delta + \sum_{j=1}^m \mathbf{b}^j(\tau_n, \mathbf{Y}_n) \Delta W_n^j + \mathbf{C}_n(\mathbf{Y}_n - \mathbf{Y}_{n+1}), \quad (7.3.9)$$

where

$$\mathbf{C}_n = \mathbf{c}^0(\tau_n, \mathbf{Y}_n) \Delta + \sum_{j=1}^m \mathbf{c}^j(\tau_n, \mathbf{Y}_n) |\Delta W_n^j| \quad (7.3.10)$$

with $\Delta W_n^j = W_{\tau_{n+1}}^j - W_{\tau_n}^j$, $\Delta = \tau_{n+1} - \tau_n$, $n \in \{0, 1, \dots, N-1\}$. Here $\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^m$ represent $d \times d$ -matrix-valued uniformly bounded functions.

We assume that for any sequence of real numbers (α_i) with $\alpha_0 \in [0, \bar{\alpha}]$, $\alpha_1 \geq 0, \dots, \alpha_m \geq 0$, where $\bar{\alpha} \geq \Delta$ for all time step sizes $\Delta \in (0, \Delta_0]$, $\Delta_0 < 1$, considered and $(t, x) \in [0, \infty) \times \mathbb{R}^d$, the matrix

$$\mathbf{M}(t, x) = \mathbf{I} + \alpha_0 \mathbf{c}^0(t, x) + \sum_{j=1}^m \alpha_j \mathbf{c}^j(t, x)$$

has an inverse and satisfies the condition

$$|(\mathbf{M}(t, x))^{-1}| \leq K < \infty. \quad (7.3.11)$$

Here \mathbf{I} is the unit matrix. Obviously, condition (7.3.11) can be easily fulfilled by keeping $\mathbf{c}^0, \mathbf{c}^1, \dots, \mathbf{c}^m$ all positive definite. Under these conditions one obtains directly the one-step increment $\mathbf{Y}_{n+1} - \mathbf{Y}_n$ of the balanced implicit method via the solution of a system of linear algebraic equations.

Let us emphasize, that due to its flexibility the method (7.3.9)–(7.3.10) turns out to be rather general. Obviously, in the deterministic case it covers the implicit Euler method. Now, let us state the following convergence theorem.

Theorem 7.3.1. (Milstein-Platen-Schurz) *Under the above assumptions the balanced implicit method (7.3.9)–(7.3.10) converges with strong order $\gamma = 0.5$. More precisely, for all $k \in \{0, 1, \dots, N\}$ and step sizes $\Delta = \frac{T}{N}$, $N \in \{1, 2, \dots\}$ one has*

$$E(|\mathbf{X}_{\tau_k} - \mathbf{Y}_k| \mid \mathcal{A}_0) \leq (E(|\mathbf{X}_{\tau_k} - \mathbf{Y}_k|^2 \mid \mathcal{A}_0))^{\frac{1}{2}} \leq K (1 + |\mathbf{X}_0|^2)^{\frac{1}{2}} \Delta^{\frac{1}{2}}, \quad (7.3.12)$$

where K does not depend on Δ .

Proof of Strong Convergence

For $0 \leq s \leq t < \infty$ and $\mathbf{y} \in \mathbb{R}^d$, let the variable $\mathbf{X}_t^{s,y}$ denote the value of a solution of (7.3.8) at time t which starts in $\mathbf{y} \in \mathbb{R}^d$ at time s . To prove Theorem 7.3.1 we mention the following useful theorem, given in Milstein (1995a), concerning the order of strong convergence of a scheme.

Theorem 7.3.2. (Milstein) *Assume for a one-step discrete-time approximation \mathbf{Y} that for all $N \in \{1, 2, \dots\}$ and $n \in \{0, 1, \dots, N-1\}$*

$$\left| E\left(\mathbf{X}_{\tau_{n+1}}^{\tau_n, Y_n} - \mathbf{Y}_{n+1} \mid \mathcal{A}_{\tau_n}\right) \right| \leq K (1 + |\mathbf{Y}_n|^2)^{\frac{1}{2}} \Delta^{p_1} \quad (7.3.13)$$

and

$$\left(E\left(|\mathbf{X}_{\tau_{n+1}}^{\tau_n, Y_n} - \mathbf{Y}_{n+1}|^2 \mid \mathcal{A}_{\tau_n}\right) \right)^{\frac{1}{2}} \leq K (1 + |\mathbf{Y}_n|^2)^{\frac{1}{2}} \Delta^{p_2} \quad (7.3.14)$$

with $p_2 \geq \frac{1}{2}$ and $p_1 \geq p_2 + \frac{1}{2}$. Then,

$$\left(E\left(|\mathbf{X}_{\tau_k}^{0, X_0} - \mathbf{Y}_k|^2 \mid \mathcal{A}_0\right) \right)^{\frac{1}{2}} \leq K (1 + |\mathbf{X}_0|^2)^{\frac{1}{2}} \Delta^{p_2 - \frac{1}{2}} \quad (7.3.15)$$

for each $k \in \{0, 1, \dots, N\}$.

Note that the various constants have been covered by the same letter K .

Proof of Theorem 7.3.1: At first, we show that the estimate (7.3.13) holds for the balanced implicit method (7.3.9) with $p_1 = \frac{3}{2}$. For this purpose, the local Euler approximation step

$$\mathbf{Y}_{k+1}^E = \mathbf{Y}_k + a(\tau_k, \mathbf{Y}_k) \Delta + \sum_{j=1}^m b^j(\tau_k, \mathbf{Y}_k) \Delta W_k^j, \quad (7.3.16)$$

$k \in \{0, 1, \dots, N-1\}$ is introduced and one can deduce from $n \in \{0, 1, \dots, N-1\}$ that

$$\begin{aligned}
H_1 &= \left| E \left(\mathbf{X}_{\tau_{n+1}}^{\tau_n, Y_n} - \mathbf{Y}_{n+1} \mid \mathcal{A}_{\tau_n} \right) \right| \\
&= \left| E \left(\mathbf{X}_{\tau_{n+1}}^{\tau_n, Y_n} - \mathbf{Y}_{n+1}^E \mid \mathcal{A}_{\tau_n} \right) + E \left(\mathbf{Y}_{n+1}^E - \mathbf{Y}_{n+1} \mid \mathcal{A}_{\tau_n} \right) \right| \\
&\leq K (1 + |\mathbf{Y}_n|^2)^{\frac{1}{2}} \Delta^2 + H_2
\end{aligned}$$

with

$$\begin{aligned}
H_2 &= \left| E \left(\mathbf{Y}_{n+1}^E - \mathbf{Y}_{n+1} \mid \mathcal{A}_{\tau_n} \right) \right| \\
&= \left| E \left((\mathbf{I} - (\mathbf{I} + C_n)^{-1}) \left(\mathbf{a}(\tau_n, \mathbf{Y}_n) \Delta + \sum_{j=1}^m \mathbf{b}^j(\tau_n, \mathbf{Y}_n) \Delta W_n^j \right) \mid \mathcal{A}_{\tau_n} \right) \right| \\
&= \left| E \left((\mathbf{I} + C_n)^{-1} \mathbf{C}_n \left(\mathbf{a}(\tau_n, \mathbf{Y}_n) \Delta + \sum_{j=1}^m \mathbf{b}^j(\tau_n, \mathbf{Y}_n) \Delta W_n^j \right) \mid \mathcal{A}_{\tau_n} \right) \right|.
\end{aligned}$$

Exploiting above the symmetry property of ΔW_n^j for $j \in \{1, 2, \dots, m\}$ in those expressions involving this zero-mean Gaussian random variable we obtain

$$H_2 = \left| E \left((\mathbf{I} + C_n)^{-1} \mathbf{C}_n \mathbf{a}(\tau_n, \mathbf{Y}_n) \Delta \mid \mathcal{A}_{\tau_n} \right) \right|$$

and it follows with (7.3.11) that

$$\begin{aligned}
H_2 &\leq K E \left(|\mathbf{C}_n \mathbf{a}(\tau_n, \mathbf{Y}_n) \Delta| \mid \mathcal{A}_{\tau_n} \right) \\
&\leq K (1 + |\mathbf{Y}_n|^2)^{\frac{1}{2}} \Delta^{\frac{3}{2}}.
\end{aligned}$$

Thus the assumption (7.3.13) with $p_1 = 1.5$ in Theorem 7.3.2 is satisfied for the balanced implicit method.

Similarly, we check assumption (7.3.14) for the local mean-square error of the balanced implicit method (7.3.9) and obtain for $n \in \{0, 1, \dots, N-1\}$ by standard arguments

$$\begin{aligned}
H_3 &= \left(E \left(\left| \mathbf{X}_{\tau_{n+1}}^{\tau_n, Y_n} - \mathbf{Y}_{n+1} \right|^2 \mid \mathcal{A}_{\tau_n} \right) \right)^{\frac{1}{2}} \\
&\leq \left(E \left(\left| \mathbf{X}_{\tau_{n+1}}^{\tau_n, Y_n} - \mathbf{Y}_{n+1}^E \right|^2 \mid \mathcal{A}_{\tau_n} \right) \right)^{\frac{1}{2}} + \left(E \left(\left| \mathbf{Y}_{n+1}^E - \mathbf{Y}_{n+1} \right|^2 \mid \mathcal{A}_{\tau_n} \right) \right)^{\frac{1}{2}} \\
&\leq K (1 + |\mathbf{Y}_n|^2)^{\frac{1}{2}} \Delta.
\end{aligned}$$

Thus we can choose in Theorem 7.3.2 the exponents $p_2 = 1.0$ together with $p_1 = 1.5$ and apply it to finally prove the strong order $\gamma = p_2 - \frac{1}{2} = 0.5$ of the balanced implicit method, as was claimed in Theorem 7.3.1. \square

First Strong Order Balanced Implicit Schemes

It is natural to try to construct a balanced implicit method for some higher strong order, say, strong order $\gamma = 1$. [Alcock & Burrage \(2006\)](#), as well as [Kahl & Schurz \(2006\)](#), suggest to take the Milstein scheme (5.3.19) and simply add an implicit higher strong order term that may improve the numerical stability of the resulting scheme. In the one-dimensional case for $m = d = 1$ a version of a *first strong order balanced implicit method* can be written in the form

$$\begin{aligned} Y_{n+1} = Y_n + a \Delta + b \Delta W + b \frac{\partial}{\partial x} b(\tau_n, Y_n) \frac{1}{2} (\Delta W^2 - \Delta) \\ + (Y_n - Y_{n+1}) D(Y_n) |\Delta W^2 - \Delta|. \end{aligned} \quad (7.3.17)$$

Here the function $D(Y_n)$ can be relatively freely chosen. This scheme is of strong order $\gamma = 1$ under appropriate conditions, see [Alcock & Burrage \(2006\)](#) and [Kahl & Schurz \(2006\)](#).

More generally, a multi-dimensional version of the first strong order balanced implicit method is of the form

$$\begin{aligned} \mathbf{Y}_{n+1} = \mathbf{Y}_n + \mathbf{a}(\tau_n, \mathbf{Y}_n) \Delta + \sum_{j=1}^m \mathbf{b}^j(\tau_n, \mathbf{Y}_n) \Delta W_n^j \\ + \sum_{j_1, j_2=1}^m \left\{ \sum_{i=1}^d b^{i, j_1} \frac{\partial \mathbf{b}^{j_2}}{\partial x^i} I_{j_1, j_2} \right\} + (\mathbf{Y}_n - \mathbf{Y}_{n+1}) \sum_{j_1, j_2=1}^m D_{j_1, j_2}(\mathbf{Y}_n) |I_{j_1, j_2}|. \end{aligned}$$

Conditions that guarantee first strong order convergence follow in a similar manner as for the original balanced implicit method.

7.4 Predictor-Corrector Schemes

The following strong predictor-corrector schemes, proposed in [Bruti-Liberati & Platen \(2008\)](#), are designed to retain some of the numerical stability properties of similar implicit schemes, while avoiding the additional computational effort required for solving an algebraic equation in each time step. This is achieved via the following procedure implemented at each time step: At first an explicit scheme is generated, the so-called *predictor*, and afterwards a *de facto* implicit scheme is used as *corrector*. The corrector is made explicit by using a predicted value \bar{Y}_{n+1} , instead of Y_{n+1} . Additionally, with this procedure one avoids the problem of the possible appearance of some reciprocal of a Gaussian random variable and can also introduce “implicitness” in the diffusion part, as will be shown below.

Another advantage of predictor-corrector methods is that the difference $Z_{n+1} = \bar{Y}_{n+1} - Y_{n+1}$ between the predicted and the corrected value provides an indication of the local error. This can be used to implement advanced schemes with step size control based on Z_{n+1} .

Predictor-Corrector Euler Scheme

In the one-dimensional case, $d = m = 1$, the *family of predictor-corrector Euler schemes*, is given by the corrector

$$\begin{aligned} Y_{n+1} &= Y_n + \{\theta \bar{a}_\eta(t_{n+1}, \bar{Y}_{n+1}) + (1 - \theta) \bar{a}_\eta\} \Delta_n \\ &\quad + \{\eta b(t_{n+1}, \bar{Y}_{n+1}) + (1 - \eta)b\} \Delta W_n + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c(\xi_i), \end{aligned} \quad (7.4.1)$$

where $\bar{a}_\eta = a - \eta b b'$, and the predictor

$$\bar{Y}_{n+1} = Y_n + a \Delta_n + b \Delta W_n + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c(\xi_i). \quad (7.4.2)$$

Here the parameters $\theta, \eta \in [0, 1]$ characterize the degree of implicitness in the drift and in the diffusion coefficients, respectively. We remark that with the choice of $\eta > 0$ one obtains a scheme with some degree of implicitness also in the diffusion coefficient. This was not achievable with the drift-implicit schemes presented in Sect. 7.2. The scheme (7.4.1)–(7.4.2) has a strong order of convergence $\gamma = 0.5$, as will be shown in the next section.

For the general multi-dimensional case, the k th component of the *family of predictor-corrector Euler schemes*, is given by the corrector

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + \{\theta \bar{a}_\eta^k(t_{n+1}, \bar{Y}_{n+1}) + (1 - \theta) \bar{a}_\eta^k\} \Delta_n \\ &\quad + \sum_{j=1}^m \{\eta b^{k,j}(t_{n+1}, \bar{Y}_{n+1}) + (1 - \eta)b^{k,j}\} \Delta W_n^j + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c^k(\xi_i), \end{aligned} \quad (7.4.3)$$

for $\theta, \eta \in [0, 1]$, where

$$\bar{a}_\eta = a - \eta \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{k, j_1} \frac{\partial b^{k, j_2}}{\partial x^i}, \quad (7.4.4)$$

and the predictor

$$\bar{Y}_{n+1}^k = Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c^k(\xi_i). \quad (7.4.5)$$

Predictor-Corrector Strong Order 1.0 Scheme

As explained above, it is challenging to design an efficient first strong order scheme with good numerical stability properties. To enhance the numerical

stability properties of the strong order 1.0 Taylor scheme (6.2.1) we have presented the drift-implicit strong order 1.0 Taylor scheme (7.2.4). However, this scheme is computationally expensive, since it generally requires the solution of an algebraic equation at each time step. In the following we propose the *predictor-corrector strong order 1.0 scheme* which combines good numerical stability properties and efficiency.

In the one-dimensional case, $d = m = 1$, the *predictor-corrector strong order 1.0 scheme* is given by the corrector

$$\begin{aligned}
 Y_{n+1} &= Y_n + \{\theta a(t_{n+1}, \bar{Y}_{n+1}) + (1 - \theta) a\} \Delta_n + b \Delta W_n \\
 &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) p_\varphi(dv, dz) + b b' \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1} dW_{z_2} \\
 &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b c'(v) dW_{z_1} p_\varphi(dv, dz_2) \\
 &+ \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{b(t_n, Y_n + c(v)) - b\} p_\varphi(dv, dz_1) dW_{z_2} \quad (7.4.6) \\
 &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{c(t_n, Y_n + c(v_1), v_2) - c(v_2)\} p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2),
 \end{aligned}$$

and the predictor

$$\begin{aligned}
 \bar{Y}_{n+1} &= Y_n + a \Delta_n + b \Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) p_\varphi(dv, dz) \\
 &+ b b' \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1} dW_{z_2} \\
 &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b c'(v) dW_{z_1} p_\varphi(dv, dz_2) \\
 &+ \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{b(t_n, Y_n + c(v)) - b\} p_\varphi(dv, dz_1) dW_{z_2} \quad (7.4.7) \\
 &+ \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{c(t_n, Y_n + c(v_1), v_2) - c(v_2)\} p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2).
 \end{aligned}$$

Here the parameter $\theta \in [0, 1]$ characterizes the degree of implicitness in the drift coefficient. This scheme attains strong order $\gamma = 1.0$, as will be shown in the next section. The generation of the multiple stochastic integrals involved can be simplified, see (6.2.3).

For the general multi-dimensional case, the k th component of the *predictor-corrector strong order 1.0 scheme* is given by the corrector

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + \{\theta a^k(t_{n+1}, \bar{\mathbf{Y}}_{n+1}) + (1 - \theta) a^k\} \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j \\
&\quad + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c^k(v) p_\varphi(dv, dz) \\
&\quad + \sum_{j_1, j_2=1}^m \sum_{i=1}^d \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \{b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i}\} dW_{z_1}^{j_1} dW_{z_2}^{j_2}(z_2) \\
&\quad + \sum_{j_1=1}^m \sum_{i=1}^d \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b^{i, j_1} \frac{\partial c^k(v)}{\partial x^i} dW_{z_1}^{j_1} p_\varphi(dv, dz_2) \\
&\quad + \sum_{j_1=1}^m \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \{b^{k, j_1}(t_n, \mathbf{Y}_n + \mathbf{c}(v)) - b^{k, j_1}\} \\
&\quad \times p_\varphi(dv, dz_2) dW_{z_2}^{j_1} \\
&\quad + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ c^k(t_n, \mathbf{Y}_n + \mathbf{c}(v_1), v_2) - c^k(v_2) \right\} \\
&\quad \times p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2), \tag{7.4.8}
\end{aligned}$$

and the predictor

$$\begin{aligned}
\bar{Y}_{n+1}^k &= Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c^k(v) p_\varphi(dv, dz) \\
&\quad + \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1}^{j_1} dW_{z_2}^{j_2} \\
&\quad + \sum_{j_1=1}^m \sum_{i=1}^d \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b^{i, j_1} \frac{\partial c^k(v)}{\partial x^i} dW_{z_1}^{j_1} p_\varphi(dv, dz_2) \\
&\quad + \sum_{j_1=1}^m \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \{b^{k, j_1}(t_n, \mathbf{Y}_n + \mathbf{c}(v)) - b^{k, j_1}\} p_\varphi(dv, dz_2) dW_{z_2}^{j_1} \\
&\quad + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ c^k(t_n, \mathbf{Y}_n + \mathbf{c}(v_1), v_2) - c^k(v_2) \right\} \\
&\quad \times p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2), \tag{7.4.9}
\end{aligned}$$

with $\theta \in [0, 1]$, for $k \in \{1, 2, \dots, d\}$.

The considerations discussed in relation to the generation of multiple stochastic integrals for the strong order 1.0 Taylor scheme (6.2.8) apply here

also. In particular, if the SDE under analysis satisfies the diffusion commutativity condition (7.1.18) together with the jump commutativity condition (7.1.16) and has mark-independent jump size, then we obtain an efficient, implementable *predictor-corrector strong order 1.0 scheme*. Its k th component is given by the corrector

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + \{\theta a^k(t_{n+1}, \bar{\mathbf{Y}}_{n+1}) + (1 - \theta) a^k\} \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j \\
&\quad + c^k \Delta p_n + \frac{1}{2} \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \{\Delta W_n^{j_1} \Delta W_n^{j_2} - \Delta_n\} \\
&\quad + \sum_{j_1=1}^m \{b^{k, j_1}(t_n, \mathbf{Y}_n + \mathbf{c}) - b^{k, j_1}\} (\Delta p_n \Delta W_n^{j_1}) \\
&\quad + \frac{1}{2} \{c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k\} ((\Delta p_n)^2 - \Delta p_n), \tag{7.4.10}
\end{aligned}$$

and the predictor

$$\begin{aligned}
\bar{Y}_{n+1}^k &= Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + c^k \Delta p_n \\
&\quad + \frac{1}{2} \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \{\Delta W_n^{j_1} \Delta W_n^{j_2} - \Delta_n\} \\
&\quad + \sum_{j_1=1}^m \{b^{k, j_1}(t_n, \mathbf{Y}_n + \mathbf{c}) - b^{k, j_1}\} (\Delta p_n \Delta W_n^{j_1}) \\
&\quad + \frac{1}{2} \{c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k\} ((\Delta p_n)^2 - \Delta p_n), \tag{7.4.11}
\end{aligned}$$

with $\theta \in [0, 1]$, for $k \in \{1, 2, \dots, d\}$. In the special case of additive diffusion and jump coefficient, $\mathbf{b}(t, x) = \mathbf{b}(t)$ and $\mathbf{c}(t, x) = \mathbf{c}(t)$, respectively, which satisfy the required commutativity conditions, we obtain an efficient predictor-corrector strong order 1.0 scheme.

We remark that as in the predictor-corrector Euler scheme as with the predictor-corrector strong order 1.0 scheme, one can introduce quasi-implicitness into the diffusion part. We emphasized that predictor-corrector schemes with higher strong order can be easily constructed, providing numerically stable, efficient and conveniently implementable discrete-time approximations of jump diffusion processes that are quite efficient in many cases, see also Chap. 14.

7.5 Convergence Results

Strong Itô Schemes

In this section we consider general strong schemes, that is, the strong Itô schemes, constructed with the same multiple stochastic integrals underlying the strong Taylor schemes (6.4.3) presented in Sect. 6.4, but with various approximations for the different coefficients. Under particular conditions on these coefficients, the strong Itô schemes converge to the solution X of the SDE (1.8.2) with the same strong order γ achieved by the corresponding strong Taylor schemes. In principle, we can construct more general strong approximations of any given order of those already presented in this chapter. In particular, we will show that derivative-free, drift-implicit, and predictor-corrector schemes are strong Itô schemes in the sense defined below. Again we will discuss two different types of schemes; those based on the Wagner-Platen expansion (4.4.4) and those based on the compensated expansion (4.4.5).

For a regular time discretization $(t)_\Delta$ with maximum step size $\Delta \in (0, 1)$, as that introduced in (6.1.8), we define the *strong order γ Itô scheme* by the vector equation

$$\mathbf{Y}_{n+1}^\Delta = \mathbf{Y}_n^\Delta + \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} I_\alpha [h_{\alpha,n}]_{t_n, t_{n+1}} + \mathbf{R}_n, \quad (7.5.1)$$

and the *strong order γ compensated Itô scheme* by

$$\mathbf{Y}_{n+1}^\Delta = \mathbf{Y}_n^\Delta + \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \tilde{I}_\alpha [\tilde{h}_{\alpha,n}]_{t_n, t_{n+1}} + \tilde{\mathbf{R}}_n, \quad (7.5.2)$$

with $n \in \{0, 1, \dots, n_T - 1\}$. We assume that the coefficients $h_{\alpha,n}$ and $\tilde{h}_{\alpha,n}$ are \mathcal{A}_{t_n} -measurable and satisfy the estimates

$$E \left(\max_{0 \leq n \leq n_T - 1} |h_{\alpha,n} - f_\alpha(t_n, \mathbf{Y}_n^\Delta)|^2 \right) \leq C(\mathbf{u}) \Delta^{2\gamma - \psi(\alpha)}, \quad (7.5.3)$$

and

$$E \left(\max_{0 \leq n \leq n_T - 1} |\tilde{h}_{\alpha,n} - \tilde{f}_\alpha(t_n, \mathbf{Y}_n^\Delta)|^2 \right) \leq C(\mathbf{u}) \Delta^{2\gamma - \psi(\alpha)}, \quad (7.5.4)$$

respectively, for all $\alpha \in \mathcal{A}_\gamma \setminus \{v\}$, where $C : \mathcal{E}^{s(a)} \rightarrow \mathfrak{R}$ is a $\varphi(du)$ -integrable function. Here

$$\psi(\alpha) = \begin{cases} 2l(\alpha) - 2 & \text{when } l(\alpha) = n(\alpha) \\ l(\alpha) + n(\alpha) - 1 & \text{when } l(\alpha) \neq n(\alpha). \end{cases}$$

Additionally, \mathbf{R}_n and $\tilde{\mathbf{R}}_n$ are assumed to satisfy

$$E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} \mathbf{R}_k \right|^2 \right) \leq K \Delta^{2\gamma}, \quad (7.5.5)$$

and

$$E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} \tilde{\mathbf{R}}_k \right|^2 \right) \leq K \Delta^{2\gamma} \quad (7.5.6)$$

where K is a finite positive constant independent of Δ .

The strong Itô schemes, given by (7.5.2), are constructed using jump integrals with respect to the compensated Poisson measure \tilde{p}_φ , which follows from the definition of the multiple stochastic integrals in Sect. 4.2. As discussed previously in Sect. 6.4, for notational convenience we presented in this and the previous chapter strong schemes with jump integrals with respect to the Poisson measure p_φ . One can show that all these schemes can be rewritten as strong Itô schemes.

Convergence of Compensated Itô Schemes

We now formulate a convergence theorem that enables us to construct strong Itô schemes of any given strong order, including derivative-free, drift-implicit and predictor-corrector schemes. Similar results are given in Bruti-Liberati & Platen (2008).

Theorem 7.5.1. *Let $\mathbf{Y}^\Delta = \{\mathbf{Y}_n^\Delta, n \in \{0, 1, \dots, n_T\}\}$ be a discrete-time approximation generated via the strong order γ compensated Itô scheme (7.5.2), for a given regular time discretization $(t)_\Delta$ with maximum time step size $\Delta \in (0, 1)$, and for $\gamma \in \{0.5, 1, 1.5, 2, \dots\}$. If the conditions of Theorem 6.4.1 are satisfied, then*

$$\sqrt{E \left(\max_{0 \leq n \leq n_T} \left| \mathbf{X}_{t_n} - \mathbf{Y}_n^\Delta \right|^2 \right)} \leq K \Delta^\gamma, \quad (7.5.7)$$

where K is a finite positive constant, independent of Δ .

Proof: Since we have already shown in Theorem 6.4.1 that the compensated strong Taylor approximation (6.4.3) converges with strong order γ , here it is sufficient to show that the compensated Itô scheme (7.5.2) converges with strong order γ to the corresponding compensated strong Taylor scheme.

By $\bar{\mathbf{Y}}^\Delta$ we denote here the compensated strong Taylor scheme (6.4.3). Let us also assume, for simplicity, that $\bar{\mathbf{Y}}_0 = \mathbf{Y}_0$. Then by application of Jensen's inequality together with the Cauchy-Schwarz inequality, we obtain for all $t \in [0, T]$ the estimate

$$\begin{aligned}
H_t &= E \left(\max_{1 \leq n \leq n_t} \left| \bar{\mathbf{Y}}_n^\Delta - \mathbf{Y}_n^\Delta \right|^2 \right) \\
&= E \left(\max_{1 \leq n \leq n_t} \left| \sum_{k=0}^{n-1} \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \tilde{I}_\alpha \left[\tilde{f}_\alpha(t_k, \bar{\mathbf{Y}}_k^\Delta) \right]_{t_k, t_{k+1}} \right. \right. \\
&\quad \left. \left. - \sum_{k=0}^{n-1} \left(\sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \tilde{I}_\alpha \left[\tilde{h}_{\alpha,k} \right]_{t_k, t_{k+1}} + \mathbf{R}_n \right) \right|^2 \right) \\
&\leq K_1 \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \left\{ E \left(\max_{1 \leq n \leq n_t} \left| \sum_{k=0}^{n-1} \tilde{I}_\alpha \left[\tilde{f}_\alpha(t_k, \bar{\mathbf{Y}}_k^\Delta) - \tilde{f}_\alpha(t_k, \mathbf{Y}_k^\Delta) \right]_{t_k, t_{k+1}} \right|^2 \right) \right. \\
&\quad \left. + E \left(\max_{1 \leq n \leq n_t} \left| \sum_{k=0}^{n-1} \tilde{I}_\alpha \left[\tilde{f}_\alpha(t_k, \mathbf{Y}_k^\Delta) - \tilde{h}_{\alpha,k} \right]_{t_k, t_{k+1}} \right|^2 \right) \right\} \\
&\quad + K_1 E \left(\max_{1 \leq n \leq n_t} \left| \sum_{k=0}^{n-1} \mathbf{R}_n \right|^2 \right). \tag{7.5.8}
\end{aligned}$$

By applying Lemma 6.5.1, condition (7.5.6), the Lipschitz condition (6.4.7), as well as condition (7.5.4), we obtain

$$\begin{aligned}
H_t &\leq K_2 \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \left\{ \int_0^t \int_{\mathcal{E}} \dots \int_{\mathcal{E}} \left(E \left(\max_{1 \leq n \leq n_u} |\tilde{f}_\alpha(t_k, \bar{\mathbf{Y}}_k^\Delta) - \tilde{f}_\alpha(t_k, \mathbf{Y}_k^\Delta)|^2 \right) \right. \right. \\
&\quad \left. \left. + E \left(\max_{1 \leq n \leq n_u} |\tilde{f}_\alpha(t_k, \mathbf{Y}_k^\Delta) - \tilde{h}_{\alpha,n}|^2 \right) \right) \varphi(dv^1) \dots \varphi(dv^{s(\alpha)}) du \right\} \Delta^{\psi(\alpha)} + K_3 \Delta^{2\gamma} \\
&\leq K_2 \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \left\{ \int_0^t E \left(\max_{1 \leq n \leq n_u} |\bar{\mathbf{Y}}_k^\Delta - \mathbf{Y}_k^\Delta|^2 \right) \int_{\mathcal{E}} \dots \int_{\mathcal{E}} K_4(v_1, \dots, v_{s(\alpha)}) \right. \\
&\quad \times \varphi(dv_1) \dots \varphi(dv_{s(\alpha)}) du \\
&\quad + \int_0^t \int_{\mathcal{E}} \dots \int_{\mathcal{E}} K_5(v_1, \dots, v_{s(\alpha)}) \Delta^{2\gamma - \psi(\alpha)} \varphi(dv_1) \dots \varphi(dv_{s(\alpha)}) du \Big\} \Delta^{\psi(\alpha)} \\
&\quad + K_3 \Delta^{2\gamma} \\
&\leq K_5 \int_0^t E \left(\max_{0 \leq n \leq n_u} |\bar{\mathbf{Y}}_k^\Delta - \mathbf{Y}_k^\Delta|^2 \right) du \sum_{\alpha \in \mathcal{A}_\gamma \setminus \{v\}} \Delta^{\psi(\alpha)} + K_6 \Delta^{2\gamma} \\
&\leq K_7 \int_0^t H_u du + K_6 \Delta^{2\gamma}. \tag{7.5.9}
\end{aligned}$$

From the second moment estimate (6.6.1) at the compensated strong Taylor scheme $\bar{\mathbf{Y}}^\Delta$ in Lemma 6.6.1 and a similar estimate of the compensated strong Itô scheme \mathbf{Y}^Δ , one can show that H_t is bounded. Therefore, by applying the Gronwall inequality to (7.5.9), we obtain

$$H_t \leq K_5 \Delta^{2\gamma} e^{K\tau t}. \quad (7.5.10)$$

Since we have assumed $\bar{\mathbf{Y}}_0^\Delta = \mathbf{Y}_0^\Delta$, we obtain

$$E(\max_{0 \leq n \leq n_T} |\bar{\mathbf{Y}}_n^\Delta - \mathbf{Y}_n^\Delta|^2) \leq K \Delta^{2\gamma}. \quad (7.5.11)$$

Finally, by the estimate of Theorem 6.4.1 we obtain

$$\begin{aligned} & \sqrt{E(\max_{0 \leq n \leq n_T} |\mathbf{Y}_n^\Delta - \mathbf{X}_{t_n}|^2)} \\ &= \sqrt{E\left(\max_{0 \leq n \leq n_T} |\mathbf{Y}_n^\Delta - \bar{\mathbf{Y}}_n^\Delta + \bar{\mathbf{Y}}_n^\Delta - \mathbf{X}_{t_n}|^2\right)} \\ &\leq \sqrt{2 \left\{ E\left(\max_{0 \leq n \leq n_T} |\mathbf{Y}_n^\Delta - \bar{\mathbf{Y}}_n^\Delta|^2\right) + E\left(\max_{0 \leq n \leq n_T} |\bar{\mathbf{Y}}_n^\Delta - \mathbf{X}_{t_n}|^2\right) \right\}} \\ &\leq K \Delta^\gamma, \end{aligned} \quad (7.5.12)$$

which completes the proof of Theorem 7.5.1. \square

By the same arguments used above, one can show the following result.

Corollary 7.5.2 *Let $\mathbf{Y}^\Delta = \{\mathbf{Y}_n^\Delta, n \in \{0, 1, \dots, n_T\}\}$ be a discrete-time approximation generated by the strong order γ Itô scheme (7.5.1). If the conditions of Corollary 6.4.3 are satisfied, then*

$$\sqrt{E\left(\max_{0 \leq n \leq n_T} |\mathbf{X}_{t_n} - \mathbf{Y}_n^\Delta|^2\right)} \leq K \Delta^\gamma, \quad (7.5.13)$$

where K is a finite positive constant independent of Δ .

Derivative-Free Schemes

The strong Itô schemes (7.5.2) and (7.5.1), and the Theorem 7.5.1 and Corollary 7.5.2 allow us to assess the strong order of convergence of general approximations. In this subsection we show how to rewrite derivative-free schemes, including those presented in Sect. 7.1, as strong Itô schemes.

We recall here that in the one-dimensional case the derivative-free strong order 1.0 scheme, presented in Sect. 7.1, is given as

$$\begin{aligned}
Y_{n+1} = & Y_n + a\Delta_n + b\Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) p_\varphi(dv, dz) \\
& + \frac{(b(t_n, \bar{Y}_n) - b)}{\sqrt{\Delta_n}} \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1} dW_{z_2} \\
& + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \frac{(c(t_n, \bar{Y}_n, v) - c(v))}{\sqrt{\Delta_n}} dW_{z_1} p_\varphi(dv, dz_2) \\
& + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{b(t_n, Y_n + c(v)) - b\} p_\varphi(dv, dz_1) dW_{z_2} \\
& + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \{c(t_n, Y_n + c(v_2), v_1) - c(v_1)\} p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2),
\end{aligned} \tag{7.5.14}$$

with the supporting value

$$\bar{Y}_n = Y_n + b\sqrt{\Delta_n}. \tag{7.5.15}$$

From the deterministic Taylor expansion, we obtain

$$\begin{aligned}
b(t_n, \bar{Y}_n) = & b(t_n, Y_n) + b'(t_n, Y_n) \{\bar{Y}_n - Y_n\} \\
& + \frac{b''(t_n, Y_n + \hat{\theta}(\bar{Y}_n - Y_n))}{2} \{\bar{Y}_n - Y_n\}^2,
\end{aligned} \tag{7.5.16}$$

with

$$b'(t, x) = \frac{\partial b(t, x)}{\partial x} \quad \text{and} \quad b''(t, x) = \frac{\partial^2 b(t, x)}{\partial x^2} \tag{7.5.17}$$

and

$$\begin{aligned}
c(t_n, \bar{Y}_n, v) = & c(t_n, Y_n, v) + c'(t_n, Y_n, v) \{\bar{Y}_n - Y_n\} \\
& + \frac{c''(t_n, Y_n + \hat{\theta}(\bar{Y}_n - Y_n), v)}{2} \{\bar{Y}_n - Y_n\}^2,
\end{aligned} \tag{7.5.18}$$

with

$$c'(t, x, v) = \frac{\partial c(t, x, v)}{\partial x} \quad \text{and} \quad c''(t, x, v) = \frac{\partial^2 c(t, x, v)}{\partial x^2} \tag{7.5.19}$$

for every $v \in \mathcal{E}$ and some corresponding $\hat{\theta} \in (0, 1)$.

Therefore, we can rewrite the scheme (7.5.14) as

$$\begin{aligned}
Y_{n+1} = & Y_n + I_{(0)}[h_{(0),n}]_{t_n, t_{n+1}} + I_{(1)}[h_{(1),n}]_{t_n, t_{n+1}} + I_{(-1)}[h_{(-1),n}]_{t_n, t_{n+1}} \\
& + I_{(1,1)}[h_{(1,1),n}]_{t_n, t_{n+1}} + I_{(1,-1)}[h_{(1,-1),n}]_{t_n, t_{n+1}} \\
& + I_{(-1,1),n}[h_{(-1,1),n}]_{t_n, t_{n+1}} + I_{(-1,-1)}[h_{(-1,-1),n}]_{t_n, t_{n+1}},
\end{aligned} \tag{7.5.20}$$

with

$$\begin{aligned}
h_{(0),n} &= a(t_n, Y_n), \quad h_{(1),n} = b(t_n, Y_n), \quad h_{(-1),n} = c(t_n, Y_n, v), \\
h_{(1,-1),n} &= \frac{1}{\sqrt{\Delta_n}} \{c(t_n, \bar{Y}_n, v) - c(t_n, Y_n, v)\}, \\
h_{(-1,1),n} &= b(t_n, Y_n + c(t_n, Y_n, v)) - b(t_n, Y_n), \\
h_{(-1,-1),n} &= c(t_n, Y_n + c(t_n, Y_n, v_2), v_1) - c(t_n, Y_n, v_1), \\
h_{(1,1),n} &= \frac{1}{\sqrt{\Delta_n}} \{b(t_n, \bar{Y}_n) - b(t_n, Y_n)\}.
\end{aligned} \tag{7.5.21}$$

Only for $\alpha = (1, 1)$ and $\alpha = (1, -1)$ the coefficients $h_{\alpha,n}$ are different from the coefficients $f_{\alpha,n}$ of the strong order 1.0 Taylor scheme (6.2.1). Therefore, to prove that the scheme (7.5.14)–(7.5.15) is a strong order 1.0 Itô scheme, it remains to check the condition (7.5.3) for these two coefficients.

By the linear growth condition (6.4.9) of Theorem 6.4.1, we have

$$\begin{aligned}
\left| b(t_n, Y_n)^2 b'' \left(t_n, Y_n + \hat{\theta} b(t_n, Y_n) \sqrt{\Delta_n} \right) \right|^2 &\leq K_1 (1 + |Y_n|^4) K_2 (1 + |Y_n|^2) \\
&= C_1 (1 + |Y_n|^2 + |Y_n|^4 + |Y_n|^6).
\end{aligned} \tag{7.5.22}$$

In a similar way we also obtain

$$\left| b(t_n, Y_n)^2 c'' \left(t_n, Y_n + \hat{\theta} b(t_n, Y_n) \sqrt{\Delta_n}, v \right) \right|^2 \leq C_2(v) (1 + |Y_n|^2 + |Y_n|^4 + |Y_n|^6), \tag{7.5.23}$$

where $C_2(v) : \mathcal{E} \rightarrow \mathbb{R}$ is a $\varphi(dv)$ -integrable function.

Following similar steps as those used in the first part of the proof of Theorem 6.4.1, one can show that

$$E \left(\max_{0 \leq n \leq n_T-1} |Y_n|^{2q} \right) \leq K (1 + E(|Y_0|^{2q})), \tag{7.5.24}$$

for $q \in \mathcal{N}$. Therefore, assuming $E(|Y_0|^6) < \infty$, by conditions (7.5.22), (7.5.23) and (7.5.24), we obtain

$$\begin{aligned}
E \left(\max_{0 \leq n \leq n_T-1} |h_{(1,1),n} - f_{(1,1)}(t_n, Y_n)|^2 \right) &\leq E \left(\max_{0 \leq n \leq n_T-1} \left| \frac{b(t_n, Y_n)^2 b''(t_n, Y_n)}{2} \sqrt{\Delta_n} \right|^2 \right) \\
&\leq K \Delta (1 + E(|Y_0|^6)) \\
&\leq K \Delta^{2\gamma - \psi(\alpha)}.
\end{aligned} \tag{7.5.25}$$

We also have

$$E \left(\max_{0 \leq n \leq n_T - 1} |h_{(1,-1),n} - f_{(1,-1)}(t_n, Y_n)|^2 \right) \leq C(v) \Delta^{2\gamma - \psi(\alpha)}, \quad (7.5.26)$$

where $C(v) : \mathcal{E} \rightarrow \mathbb{R}$ is a $\varphi(dv)$ - integrable function, which demonstrates that the scheme (7.5.14) is a strong Itô scheme of order $\gamma = 1.0$.

In a similar way one can show that some other higher order derivative-free strong schemes for general multi-dimensional SDEs can be expressed as strong Itô schemes and, therefore, achieve the corresponding strong order of convergence.

Drift-Implicit Strong Schemes

As explained in Sect. 7.2, for a strong Taylor scheme of order γ it is usually possible to obtain a corresponding drift-implicit strong scheme. Due to problems with the reciprocal of Gaussian random variables in a scheme, as discussed in Sect. 7.3, one often introduces implicitness only in the drift terms, see Higham & Kloeden (2005, 2006). Exceptions are the balanced implicit methods in Sect. 7.3, where the implicitness is introduced into the diffusion terms by a rather special construction.

Drift-implicit schemes of order γ can be derived by an application of a Wagner-Platen expansion to the drift terms of a correspondent strong Taylor scheme of order γ . If we apply the Wagner-Platen expansion (4.4.4) to the drift term a , then we can write

$$\begin{aligned} a(t, X_t) &= a(t + \Delta, X_{t+\Delta}) - L^0 a(t, X_t) \Delta - L^1 a(t, X_t) (W_{t+\Delta} - W_t) \\ &\quad - \int_t^{t+\Delta} \int_{\mathcal{E}} L_v^{-1} a(t, X_t) p_{\varphi}(dv, dz) - R_1(t), \end{aligned} \quad (7.5.27)$$

where

$$\begin{aligned} R_1(t) &= \int_t^{t+\Delta} \left\{ \int_t^s L^0 L^0 a(u, X_u) du + \int_t^s L^1 L^0 a(u, X_u) dW_u \right. \\ &\quad \left. + \int_t^s \int_{\mathcal{E}} L_{v_1}^{-1} L^0 a(u, X_{u-}) p_{\varphi}(dv_1, du) \right\} ds \\ &\quad + \int_t^{t+\Delta} \left\{ \int_t^s L^0 L^1 a(u, X_u) du + \int_t^s L^1 L^1 a(u, X_u) dW_u \right. \\ &\quad \left. + \int_t^s \int_{\mathcal{E}} L_{v_1}^{-1} L^1 a(u, X_{u-}) p_{\varphi}(dv_1, du) \right\} dW_s \\ &\quad + \int_t^{t+\Delta} \int_{\mathcal{E}} \left\{ \int_t^{s-} L^0 L_{v_2}^{-1} a(u, X_u) du + \int_t^{s-} L^1 L_{v_2}^{-1} a(u, X_u) dW_u \right. \\ &\quad \left. + \int_t^{s-} \int_{\mathcal{E}} L_{v_1}^{-1} L_{v_2}^{-1} a(u, X_{u-}) p_{\varphi}(dv_1, du) \right\} p_{\varphi}(dv_2, ds) \end{aligned} \quad (7.5.28)$$

and the operators L^0 , L^1 and L_v^{-1} are defined in (4.3.4), (4.3.5) and (4.3.6), respectively.

In the one-dimensional case, for any $\theta \in [0, 1]$, we can rewrite the Euler scheme (6.1.15) as

$$\begin{aligned} Y_{n+1} = & Y_n + \{\theta a(t_n, Y_n) + (1 - \theta) a(t_n, Y_n)\} \Delta_n + b \Delta W_n \\ & + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(t_n, Y_n, v) p_\varphi(dv, ds). \end{aligned} \quad (7.5.29)$$

By replacing the first drift coefficient $a(t_n, Y_n)$ with its implicit expansion (7.5.27), we obtain

$$\begin{aligned} Y_{n+1} = & Y_n + \{\theta a(t_{n+1}, Y_{n+1}) + (1 - \theta) a\} \Delta_n + b \Delta W_n \\ & + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(t_n, Y_n, v) p_\varphi(dv, ds) \\ & - \left\{ L^0 a \Delta_n + L^1 a \Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} L_v^{-1} a p_\varphi(dv, dz) + R_1(t) \right\} \theta \Delta_n. \end{aligned}$$

Here we have written $a = a(t_n, Y_n)$ and $b = b(t_n, Y_n)$ according to the abbreviation introduced in (6.1.12).

The terms in the last line of equation (7.5.30) are not necessary for a scheme with strong order $\gamma = 0.5$. Therefore, they can be discarded when deriving the implicit Euler scheme (7.2.1), which yields

$$\begin{aligned} Y_{n+1} = & Y_n + \{\theta a(t_{n+1}, Y_{n+1}) + (1 - \theta) a\} \Delta_n + b \Delta W_n \\ & + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(t_n, Y_n, v) p_\varphi(dv, ds). \end{aligned} \quad (7.5.30)$$

Note that the drift-implicit Euler scheme (7.5.30) is well defined for time step size

$$\Delta \leq \frac{1}{\sqrt{K\theta}},$$

where K is the Lipschitz constant appearing in the Lipschitz condition (1.9.2) for the drift coefficient a . As previously mentioned, Banach's fixed point theorem ensures existence and uniqueness of the solution Y_{n+1} of the algebraic equation in (7.5.30).

By applying the same arguments to every time integral appearing in a higher order strong Taylor scheme, even in the general multi-dimensional case, it is possible to derive multi-dimensional higher order implicit schemes as, for instance, the drift-implicit strong order 1.0 scheme (7.2.9). To prove the strong order of convergence of such drift-implicit schemes it is sufficient to show that one can rewrite these as strong Itô schemes. The drift-implicit Euler scheme

(7.5.30), for instance, can be written as a strong order 0.5 Itô scheme given by

$$Y_{n+1} = Y_n + I_{(0)}[h_{(0),n}]_{t_n, t_{n+1}} + I_{(1)}[h_{(1),n}]_{t_n, t_{n+1}} + I_{(-1)}[h_{(-1),n}]_{t_n, t_{n+1}} + R_n \quad (7.5.31)$$

with

$$h_{(0),n} = a(t_n, Y_n), \quad h_{(1),n} = b(t_n, Y_n), \quad h_{(-1),n} = c(t_n, Y_n, v), \quad (7.5.32)$$

and

$$R_n = \theta \Delta_n (a(t_{n+1}, Y_{n+1}) - a(t_n, Y_n)). \quad (7.5.33)$$

Since the coefficients $h_{\alpha,n}$ are the same as those of the Euler scheme (6.1.15), all that remains is to check condition (7.5.6) for the remainder term R_n . Following similar steps as those used in the proof of Lemma 6.6.1, one can derive the second moment estimate

$$E \left(\max_{0 \leq n \leq n_T - 1} |Y_n|^2 \right) \leq K (1 + E(|Y_0|^2)). \quad (7.5.34)$$

We refer to [Higham & Kloeden \(2006\)](#) for details of the proof of this estimate in the case of the drift-implicit Euler scheme for SDEs driven by Wiener processes and homogeneous Poisson processes.

By applying the Cauchy-Schwarz inequality, the linear growth condition (1.9.3) and the estimate (7.5.34), we obtain

$$\begin{aligned} & E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} R_k \right|^2 \right) \\ &= E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} \theta \Delta_k (a(t_{k+1}, Y_{k+1}) - a(t_k, Y_k)) \right|^2 \right) \\ &\leq E \left(\max_{1 \leq n \leq n_T} \left(\sum_{0 \leq k \leq n-1} |\theta \Delta_k|^2 \right) \left(\sum_{0 \leq k \leq n-1} |a(t_{k+1}, Y_{k+1}) - a(t_k, Y_k)|^2 \right) \right) \\ &\leq K \Delta E \left(\left(\sum_{0 \leq k \leq n_T - 1} \Delta_k \right) \left(\sum_{0 \leq k \leq n_T - 1} |a(t_{k+1}, Y_{k+1}) - a(t_k, Y_k)|^2 \right) \right) \\ &\leq K \Delta E \left(\sum_{0 \leq k \leq n_T - 1} |a(t_{k+1}, Y_{k+1}) - a(t_k, Y_k)|^2 \right) \\ &\leq K \Delta E \left(\sum_{0 \leq k \leq n_T - 1} |Y_{k+1} - Y_k|^2 \right). \end{aligned} \quad (7.5.35)$$

Since $n_T < \infty$ almost surely, see (6.1.11), we can write

$$\begin{aligned} & E \left(\sum_{0 \leq k \leq n_T - 1} |Y_{k+1} - Y_k|^2 \right) \\ &= \sum_{i=1}^{\infty} E \left(\sum_{0 \leq k \leq i-1} |Y_{k+1} - Y_k|^2 \right) \times P(n_T = i) \\ &= \sum_{i=1}^{\infty} E \left(\sum_{0 \leq k \leq i-1} E(|Y_{k+1} - Y_k|^2 | \mathcal{A}_{t_k}) \right) \times P(n_T = i), \quad (7.5.36) \end{aligned}$$

where we have used the properties of conditional expectations.

By the Cauchy-Schwarz inequality, Itô's isometry and the linear growth conditions, one obtains

$$\begin{aligned} & E \left(|Y_{k+1} - Y_k|^2 \mid \mathcal{A}_{t_k} \right) \\ &\leq K \left\{ E \left(\left| \int_{t_k}^{t_{k+1}} (\theta(a(t_{k+1}, Y_{k+1}) - a(t_k, Y_k)) + \tilde{a}(t_k, Y_k)) dz \right|^2 \mid \mathcal{A}_{t_k} \right) \right. \\ &\quad \left. + \sum_{\alpha \in \{-1, 1\}} E \left(\left| \tilde{I}_\alpha[f_\alpha(t_k, Y_k)]_{t_k, t_{k+1}} \right|^2 \mid \mathcal{A}_{t_k} \right) \right\} \\ &\leq K \left\{ (t_{k+1} - t_k) \int_{t_k}^{t_{k+1}} E(|a(t_{k+1}, Y_{k+1})|^2 + |a(t_k, Y_k)|^2 + |\tilde{a}(t_k, Y_k)|^2 | \mathcal{A}_{t_k}) dz \right. \\ &\quad \left. + \sum_{\alpha \in \{-1, 1\}} \int_{t_k}^{t_{k+1}} \int_{\mathcal{E}} E \left(\left| \tilde{f}_\alpha(t_k, Y_k) \right|^2 \mid \mathcal{A}_{t_k} \right) \varphi(dv^{s(\alpha)}) dz \right\} \\ &\leq K \left\{ \Delta \int_{t_k}^{t_{k+1}} E((1 + |Y_{k+1}|^2) + (1 + |Y_k|^2) | \mathcal{A}_{t_k}) dz \right. \\ &\quad \left. + \sum_{\alpha \in \{-1, 1\}} \int_{t_k}^{t_{k+1}} \int_{\mathcal{E}} E((1 + |Y_k|^2) | \mathcal{A}_{t_k}) \varphi(dv^{s(\alpha)}) dz \right\} \\ &\leq K \left(1 + \max_{0 \leq n \leq i} |Y_n|^2 \right) (t_{k+1} - t_k) \quad (7.5.37) \end{aligned}$$

for $k \in \{1, \dots, i\}$ with $i \in \mathcal{N}$.

Finally, by combining (7.5.35), (7.5.36) and (7.5.37) and by using the second moment estimate (7.5.34), we obtain

$$\begin{aligned}
& E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} R_k \right|^2 \right) \\
& \leq K \Delta \sum_{i=1}^{\infty} E \left(\left(1 + \max_{0 \leq n \leq i} |Y_n|^2 \right) \sum_{0 \leq k \leq i-1} (t_{k+1} - t_k) \right) \times P(n_T = i) \\
& \leq K \Delta (1 + E(|Y_0|^2)) T \sum_{i=1}^{\infty} P(n_T = i) \\
& \leq K \Delta = K \Delta^{2\gamma}, \tag{7.5.38}
\end{aligned}$$

for $\gamma = 0.5$.

Therefore, the convergence of the drift-implicit Euler scheme follows from Corollary 7.5.2 since we have shown that it can be rewritten as a strong Itô scheme of order $\gamma = 0.5$. In a similar way it is possible to show that the drift-implicit strong order 1.0 scheme (7.2.9) can be rewritten as a strong order $\gamma = 1.0$ Itô scheme.

Strong Predictor-Corrector Schemes

In this subsection we show how to rewrite strong predictor-corrector schemes as strong Itô schemes and how to derive the corresponding orders of strong convergence.

Let us consider the family of strong predictor-corrector Euler schemes (7.4.1)–(7.4.2). We recall that in the one-dimensional case, $d = m = 1$, this scheme is given by the corrector

$$\begin{aligned}
Y_{n+1} &= Y_n + \{ \theta \bar{a}_\eta(t_{n+1}, \bar{Y}_{n+1}) + (1 - \theta) \bar{a}_\eta \} \Delta_n \\
&+ \{ \eta b(t_{n+1}, \bar{Y}_{n+1}) + (1 - \eta) b \} \Delta W_n + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c(\xi_i), \tag{7.5.39}
\end{aligned}$$

where $\bar{a}_\eta = a - \eta b b'$, and the predictor

$$\bar{Y}_{n+1} = Y_n + a \Delta_n + b \Delta W_n + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c(\xi_i). \tag{7.5.40}$$

This scheme can be written as a strong order 0.5 Itô scheme given by

$$Y_{n+1} = Y_n + I_{(0)}[h_{(0),n}]_{t_n, t_{n+1}} + I_{(1)}[h_{(1),n}]_{t_n, t_{n+1}} + I_{(-1)}[h_{(-1),n}]_{t_n, t_{n+1}} + R_n \tag{7.5.41}$$

with

$$h_{(0),n} = a(t_n, Y_n), \quad h_{(1),n} = b(t_n, Y_n), \quad h_{(-1),n} = c(t_n, Y_n, v), \quad (7.5.42)$$

and

$$R_n = R_{1,n} + R_{2,n} + R_{3,n} + R_{4,n} \quad (7.5.43)$$

with

$$R_{1,n} = \theta \left\{ a(t_{n+1}, \bar{Y}_{n+1}) - a(t_n, Y_n) \right\} \Delta_n, \quad (7.5.44)$$

$$R_{2,n} = -\theta \eta \left\{ b(t_{n+1}, \bar{Y}_{n+1}) b'(t_{n+1}, \bar{Y}_{n+1}) - b(t_n, Y_n) b'(t_n, Y_n) \right\} \Delta_n, \quad (7.5.45)$$

$$R_{3,n} = -\eta b(t_n, Y_n) b'(t_n, Y_n) \Delta_n, \quad (7.5.46)$$

and

$$R_{4,n} = -\eta \left\{ b(t_{n+1}, \bar{Y}_{n+1}) - b(t_n, Y_n) \right\} \Delta W_n. \quad (7.5.47)$$

Since the coefficients $h_{\alpha,n}$ are the same as those of the Euler scheme, being the strong Taylor scheme of order $\gamma = 0.5$, one needs to show that the remainder term R_n satisfies condition (7.5.5).

We also need the following second moment estimate of the numerical approximation Y_n , where

$$E \left(\max_{0 \leq n \leq n_T - 1} |Y_n|^2 \right) \leq K (1 + E(|Y_0|^2)). \quad (7.5.48)$$

This estimate can be derived by similar steps as those used in the proof of Lemma 6.6.1.

By the Cauchy-Schwarz inequality, the Lipschitz condition on the drift coefficient and equation (7.5.40), we obtain

$$\begin{aligned} & E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} R_{1,k} \right|^2 \right) \\ &= E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} \theta \left\{ a(t_{k+1}, \bar{Y}_{k+1}) - a(t_k, Y_k) \right\} \Delta_k \right|^2 \right) \\ &\leq E \left(\max_{1 \leq n \leq n_T} \left(\sum_{0 \leq k \leq n-1} |\theta \Delta_k|^2 \right) \left(\sum_{0 \leq k \leq n-1} |a(t_{k+1}, \bar{Y}_{k+1}) - a(t_k, Y_k)|^2 \right) \right) \\ &\leq K \Delta E \left(\left(\sum_{0 \leq k \leq n_T - 1} \Delta_k \right) \left(\sum_{0 \leq k \leq n_T - 1} |a(t_{k+1}, \bar{Y}_{k+1}) - a(t_k, Y_k)|^2 \right) \right) \end{aligned}$$

$$\begin{aligned}
&\leq K \Delta E \left(\sum_{0 \leq k \leq n_T - 1} |a(t_{k+1}, \bar{Y}_{k+1}) - a(t_k, Y_k)|^2 \right) \\
&\leq K \Delta E \left(\sum_{0 \leq k \leq n_T - 1} |\bar{Y}_{k+1} - Y_k|^2 \right) \\
&\leq K \Delta E \left(\sum_{0 \leq k \leq n_T - 1} \left| \sum_{\alpha \in \mathcal{A}_{0.5} \setminus \{v\}} \tilde{I}_\alpha[\tilde{f}_\alpha(t_k, Y_k)]_{t_k, t_{k+1}} \right|^2 \right) \\
&\leq K \Delta \sum_{\alpha \in \mathcal{A}_{0.5} \setminus \{v\}} E \left(\sum_{0 \leq k \leq n_T - 1} \left| \tilde{I}_\alpha[\tilde{f}_\alpha(t_k, Y_k)]_{t_k, t_{k+1}} \right|^2 \right). \quad (7.5.49)
\end{aligned}$$

Since $n_T < \infty$ almost surely, see (6.1.11), we can write

$$\begin{aligned}
&E \left(\sum_{0 \leq k \leq n_T - 1} \left| \tilde{I}_\alpha[\tilde{f}_\alpha(t_k, Y_k)]_{t_k, t_{k+1}} \right|^2 \right) \\
&= \sum_{i=1}^{\infty} E \left(\sum_{0 \leq k \leq i-1} \left| \tilde{I}_\alpha[\tilde{f}_\alpha(t_k, Y_k)]_{t_k, t_{k+1}} \right|^2 \right) \times P(n_T = i) \quad (7.5.50) \\
&= \sum_{i=1}^{\infty} E \left(\sum_{0 \leq k \leq i-1} E \left(\left| \tilde{I}_\alpha[\tilde{f}_\alpha(t_k, Y_k)]_{t_k, t_{k+1}} \right|^2 \mid \mathcal{A}_{t_k} \right) \right) \times P(n_T = i)
\end{aligned}$$

for every $\alpha \in \mathcal{A}_{0.5} \setminus \{v\}$. In the last line of (7.5.50) we have applied standard properties of conditional expectations.

By using the Cauchy-Schwarz inequality for $\alpha = (0)$, the Itô isometry for $\alpha = (j)$ and $j \in \{-1, 1\}$, and the linear growth conditions (6.4.9), we obtain

$$\begin{aligned}
&E \left(\left| \tilde{I}_\alpha[\tilde{f}_\alpha(t_k, Y_{t_k})]_{t_k, t_{k+1}} \right|^2 \mid \mathcal{A}_{t_k} \right) \\
&\leq \Delta^{n(\alpha)} \int_{t_k}^{t_{k+1}} \int_{\mathcal{E}} E \left(|\tilde{f}_\alpha(t_k, Y_{t_k})|^2 \mid \mathcal{A}_{t_k} \right) \varphi(dv^{s(\alpha)}) dz \\
&\leq \lambda^{s(\alpha)} \Delta^{n(\alpha)} \int_{t_k}^{t_{k+1}} E \left(1 + |Y_{t_k}|^2 \mid \mathcal{A}_{t_k} \right) dz \quad (7.5.51)
\end{aligned}$$

for $k \in \{0, 1, \dots, i\}$ and $i \in \mathcal{N}$.

By (7.5.48), (7.5.49), (7.5.50) and (7.5.51), we obtain

$$\begin{aligned}
& E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} R_{1,k} \right|^2 \right) \\
& \leq K \Delta \sum_{\alpha \in \mathcal{A}_{0.5} \setminus \{v\}} \lambda^{s(\alpha)} \Delta^{n(\alpha)} \\
& \quad \times \sum_{i=1}^{\infty} E \left(\int_0^{t_i} E \left((1 + |Y_{n_z}|^2) |A_{t_{n_z}} \right) dz \right) \times P(n_T = i) \\
& \leq K \Delta \sum_{\alpha \in \mathcal{A}_{0.5} \setminus \{v\}} \lambda^{s(\alpha)} \Delta^{n(\alpha)} \sum_{i=1}^{\infty} \int_0^{t_i} E \left(1 + \max_{0 \leq n \leq i} |Y_n|^2 \right) dz \times P(n_T = i) \\
& \leq K \Delta \sum_{\alpha \in \mathcal{A}_{0.5} \setminus \{v\}} \lambda^{s(\alpha)} \Delta^{n(\alpha)} (1 + E(|Y_0|^2)) T \sum_{i=1}^{\infty} P(n_T = i) \\
& \leq K \Delta = K \Delta^{2\gamma}, \tag{7.5.52}
\end{aligned}$$

for $\gamma = 0.5$.

Note that in Theorem 6.4.1 and in Theorem 7.5.1 we assumed that the coefficient function f_α satisfies the Lipschitz condition (6.4.7) for every $\alpha \in \mathcal{A}_\gamma$. If instead we require this condition to hold for every $\alpha \in \mathcal{A}_{0.5} \cup \mathcal{B}(\mathcal{A}_{0.5})$, then by taking into consideration the coefficient function corresponding to the multi-index $(1, 1) \in \mathcal{B}(\mathcal{A}_{0.5})$, we see that the coefficient bb' satisfies the Lipschitz type condition (6.4.7). Therefore, with similar steps as in (7.5.49), (7.5.50) and (7.5.51), we obtain

$$E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} R_{2,k} \right|^2 \right) \leq K \Delta = K \Delta^{2\gamma}, \tag{7.5.53}$$

for $\gamma = 0.5$.

By the linear growth condition on the coefficient bb' , which follows from condition (6.4.9), and the second moment estimate (7.5.48), one can show that

$$E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} R_{3,k} \right|^2 \right) \leq K \Delta = K \Delta^{2\gamma}, \tag{7.5.54}$$

for $\gamma = 0.5$.

By Doob's inequality, Itô's isometry, the Lipschitz condition on the diffusion coefficient and (7.5.40), we obtain

$$\begin{aligned}
& E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} R_{4,k} \right|^2 \right) \\
&= E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} \eta \int_{t_k}^{t_{k+1}} (b(t_{k+1}, \bar{Y}_{k+1}) - b(t_k, Y_k)) dW_z \right|^2 \right) \\
&\leq 4E \left(\left| \int_0^T (b(t_{n_z+1}, \bar{Y}_{n_z+1}) - b(t_{n_z}, Y_{n_z})) dW_z \right|^2 \right) \\
&= 4 \int_0^T E \left(|b(t_{n_z+1}, \bar{Y}_{n_z+1}) - b(t_{n_z}, Y_{n_z})|^2 \right) dz \\
&\leq K \int_0^T E \left(|\bar{Y}_{n_z+1} - Y_{n_z}|^2 \right) dz \\
&\leq K \int_0^T E \left(\left| \tilde{I}_\alpha [\tilde{f}_\alpha(t_{n_z}, Y_{n_z})]_{t_{n_z}, t_{n_z+1}} \right|^2 \right) dz. \tag{7.5.55}
\end{aligned}$$

By the Lipschitz condition, the estimate (7.5.51) and the second moment estimate (7.5.48), we have

$$\begin{aligned}
& E \left(\max_{1 \leq n \leq n_T} \left| \sum_{0 \leq k \leq n-1} R_{4,k} \right|^2 \right) \\
&\leq K \sum_{\alpha \in \mathcal{A}_{0.5} \setminus \{v\}} \lambda^{s(\alpha)} \Delta^{n(\alpha)} \int_0^T E \left(\int_{t_{n_z}}^{t_{n_z+1}} E(1 + |Y_{t_{n_z}}|^2 | \mathcal{A}_{t_{n_z}}) ds \right) dz \\
&\leq K \sum_{\alpha \in \mathcal{A}_{0.5} \setminus \{v\}} \lambda^{s(\alpha)} \Delta^{n(\alpha)} \int_0^T E \left(\int_{t_{n_z}}^{t_{n_z+1}} E \left(1 + \max_{0 \leq n \leq n_T} |Y_{t_n}|^2 | \mathcal{A}_{t_{n_z}} \right) ds \right) dz \\
&\leq K \sum_{\alpha \in \mathcal{A}_{0.5} \setminus \{v\}} \lambda^{s(\alpha)} \Delta^{n(\alpha)+1} (1 + E(|Y_0|^2)) T \\
&\leq K \Delta = K \Delta^{2\gamma}, \tag{7.5.56}
\end{aligned}$$

for $\gamma = 0.5$.

Finally, by combining the estimates (7.5.52)–(7.5.53)–(7.5.54) and (7.5.56) we have shown that the family of strong predictor-corrector Euler schemes (7.4.1)–(7.4.2) can be rewritten as a strong order $\gamma = 0.5$ Itô scheme. Therefore, it achieves strong order $\gamma = 0.5$.

Also in the general multi-dimensional case one can analogously rewrite the family of predictor-corrector Euler schemes (7.4.3)–(7.4.5) and similarly the predictor-corrector strong order 1.0 scheme (7.4.10)–(7.4.11) as Itô scheme of strong order $\gamma = 0.5$ and $\gamma = 1.0$, respectively. Finally, strong predictor-corrector schemes of higher strong order can be derived by similar methods as the ones used above.

7.6 Exercises

7.1. Write down the drift implicit Euler scheme for the Black-Scholes model

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

with $X_0 = 1$.

7.2. Describe for the Vasicek interest rate model

$$dr_t = \gamma(\bar{r} - r_t) dt + \beta dW_t$$

the drift implicit strong order 1.0 Runge-Kutta method.

7.3. Construct for the Black-Scholes model in Exercise 7.1 a balanced implicit method. Make sure that for $\mu, \sigma > 0$ the resulting approximation stays always nonnegative.

7.4. For the Merton type SDE

$$dX_t = a X_t dt + b X_t dW_t + c X_{t-} p_\varphi(\mathcal{E}, dt)$$

write down a derivative-free order 1.0 scheme when $c > -1$.

7.5. For the Merton model in the previous exercise provide the fully drift implicit order 1.0 scheme.

Jump-Adapted Strong Approximations

This chapter describes jump-adapted strong schemes. The term *jump-adapted* refers to the time discretizations used to construct these schemes. These discretizations include all jump times generated by the Poisson jump measure. The form of the resulting schemes is much simpler than that of the regular schemes presented in Chaps. 6 and 7 which are based on regular time discretizations. The idea of jump-adapted time discretization goes back to Platen (1982a). It appeared later in various literature, for instance, Maghsoudi (1996). Jump-adapted schemes are not very efficient for SDEs driven by a Poisson measure with a high total intensity. In this case, regular schemes would usually be preferred. Some of the results of this chapter can be found in Bruti-Liberati et al. (2006) and in Bruti-Liberati & Platen (2007b). Results presented already in Kloeden & Platen (1999) and Chap. 5 are employed in the following when approximating the diffusion part of the solution of an SDE.

8.1 Introduction to Jump-Adapted Approximations

Jump-Adapted Time Discretization

In principle, by including enough terms from a Wagner-Platen expansion, it is possible to derive regular strong Taylor schemes of any given order of strong convergence, as shown by Theorems 6.4.1 and 7.5.1. However, as noticed in Sect. 6.2, even for a one-dimensional SDE, higher order regular schemes can become complex in that they involve multiple stochastic integrals jointly with respect to the Wiener process, the Poisson random measure and time. In particular, when we have a mark-dependent jump size the generation of the required multiple stochastic integrals involving the Poisson measure can be challenging.

As noticed before, there are applications, such as filtering, in which we are able to construct the multiple stochastic integrals directly from data. In

these cases the proposed regular strong schemes can be readily applied. However, for scenario simulation we need to generate the multiple stochastic integrals by using random number generators. To avoid the generation of multiple stochastic integrals with respect to the Poisson jump measure, Platen (1982a) proposed *jump-adapted approximations* that significantly reduce the complexity of higher order schemes. A jump-adapted time discretization makes these schemes easily implementable for scenario simulation, including the case of a mark-dependent jump size. Indeed, between jump times the evolution of the SDE (6.1.4) is that of a diffusion without jumps, which can be approximated by standard schemes for pure diffusions, as presented in Chap. 5 and Kloeden & Platen (1999). At jump times the prescribed jumps are performed. As we will show in this chapter, it is possible to develop tractable jump-adapted higher order strong schemes including the case of mark-dependent jump sizes. The associated multiple stochastic integrals involve only time and Wiener process integrations.

In this chapter we consider a *jump-adapted time discretization* $0 = t_0 < t_1 < \dots < t_{n_T} = T$, on which we construct a *jump-adapted discrete-time approximation* $Y^\Delta = \{Y_t^\Delta, t \in [0, T]\}$ of the solution of the SDE (6.1.4). Here

$$n_t = \max\{n \in \{0, 1, \dots\} : t_n \leq t\} < \infty \quad \text{a.s.} \quad (8.1.1)$$

denotes again the largest integer n such that $t_n \leq t$, for all $t \in [0, T]$. We require the jump-adapted time discretization to include all the jump times $\{\tau_1, \tau_2, \dots, \tau_{p_\varphi(T)}\}$ of the Poisson random measure p_φ . Moreover, as in Sect. 6.1, for a given maximum step size $\Delta \in (0, 1)$ we require the *jump-adapted time discretization*

$$(t)_\Delta = \{0 = t_0 < t_1 < \dots < t_{n_T} = T\}, \quad (8.1.2)$$

to satisfy the following conditions:

$$P(t_{n+1} - t_n \leq \Delta) = 1, \quad (8.1.3)$$

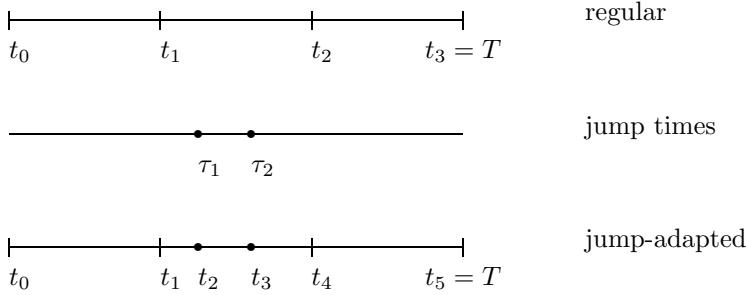
and

$$t_{n+1} \quad \text{is} \quad \mathcal{A}_{t_n} - \text{measurable}, \quad (8.1.4)$$

for $n \in \{0, 1, \dots, n_T - 1\}$, if it is not a jump time.

For instance, we could consider a jump-adapted time discretization $(t)_\Delta$ constructed by a superposition of the jump times $\{\tau_1, \tau_2, \dots, \tau_{p_\varphi(T)}\}$ of the Poisson random measure p_φ to a deterministic equidistant grid with n th discretization time $n\Delta$, $n \in \{0, 1, \dots, N\}$ and time step size $\Delta = \frac{T}{N}$. This means that we add all random jump times to an equidistant grid, as the one presented in Sect. 6.1. In this way the maximum time step size of the resulting jump-adapted discretization equals Δ .

In the graph below we show the composition of a jump-adapted time discretization.



Within this time grid we separate the diffusive part of the dynamics from the jumps, because the jumps can arise only at discretization times. Therefore, between discretization points we can approximate the diffusive part with a strong scheme for pure diffusion processes. We then add the effect of a jump to the evolution of the approximate solution when we encounter a jump time as a discretization time. We remark that with jump-adapted schemes a mark-dependent jump coefficient does not add any additional complexity to the numerical approximation. Therefore, in this chapter we consider the general case of jump diffusion SDEs with mark-dependent jump size as given in the SDE (6.1.4).

For convenience we set $\mathbf{Y}_{t_n} = \mathbf{Y}_n$ and define

$$\mathbf{Y}_{t_{n+1}-} = \lim_{s \uparrow t_{n+1}} \mathbf{Y}_s,$$

as the almost sure left-hand limit of \mathbf{Y} at time t_{n+1} . Moreover, to simplify the notation, we will use again the abbreviation

$$f = f(t_n, \mathbf{Y}_{t_n}) \quad (8.1.5)$$

for any coefficient function f when no misunderstanding is possible, see (6.1.12). Furthermore, we will again omit mentioning the initial value \mathbf{Y}_0 and the time step index $n \in \{0, 1, \dots, n_T\}$ if this is convenient. However, we aim to show the dependence on marks if it is relevant.

We will use again the operators

$$\begin{aligned} L^0 f(t, \mathbf{x}) &= \frac{\partial}{\partial t} f(t, \mathbf{x}) + \sum_{i=1}^d a^i(t, \mathbf{x}) \frac{\partial}{\partial x^i} f(t, \mathbf{x}) \\ &+ \frac{1}{2} \sum_{i,r=1}^d \sum_{j=1}^m b^{i,j}(t, \mathbf{x}) b^{r,j}(t, \mathbf{x}) \frac{\partial^2}{\partial x^i \partial x^j} f(t, \mathbf{x}) \end{aligned} \quad (8.1.6)$$

$$L^k f(t, \mathbf{x}) = \sum_{i=1}^d b^{i,k}(t, \mathbf{x}) \frac{\partial}{\partial x^i} f(t, \mathbf{x}), \quad \text{for } k \in \{1, 2, \dots, m\} \quad (8.1.7)$$

for all $t \in [0, T]$ and $\mathbf{x} \in \mathbb{R}^d$, similar to the pure diffusion case, see Chap. 5 and Kloeden & Platen (1999).

Essentially, throughout the remainder of this chapter we review strong discrete-time approximations for the diffusion part of the given SDE, as described in Chap. 5. Additionally, at each jump time of the driving Poisson measure we introduce a time discretization point into the jump-adapted discretization and approximate the jump as required.

8.2 Jump-Adapted Strong Taylor Schemes

In this section we present jump-adapted strong approximations whose diffusion part is given by the truncated Wagner-Platen expansion for pure diffusions.

Euler Scheme

For the one-dimensional case, being $d = m = 1$, we present the *jump-adapted Euler scheme* given by

$$Y_{t_{n+1}-} = Y_{t_n} + a\Delta_{t_n} + b\Delta W_{t_n} \quad (8.2.1)$$

and

$$Y_{t_{n+1}} = Y_{t_{n+1}-} + \int_{\mathcal{E}} c(t_{n+1}, Y_{t_{n+1}-}, v) p_\varphi(dv, \{t_{n+1}\}), \quad (8.2.2)$$

where

$$\Delta_{t_n} = t_{n+1} - t_n \quad (8.2.3)$$

is the length of the time step size $[t_n, t_{n+1}]$ and

$$\Delta W_{t_n} = W_{t_{n+1}} - W_{t_n} \quad (8.2.4)$$

is the n th Gaussian $N(0, \Delta_{t_n})$ distributed increment of the Wiener process W , $n \in \{0, 1, \dots, N-1\}$. Note that we write $a = a(t_n, Y_{t_n})$ and $b = b(t_n, Y_{t_n})$ according to the abbreviation (8.1.5). The impact of jumps is obtained by (8.2.2). If t_{n+1} is a jump time, then

$$\int_{\mathcal{E}} p_\varphi(dv, \{t_{n+1}\}) = 1 \quad (8.2.5)$$

and

$$\int_{\mathcal{E}} c(t_{n+1}, Y_{t_{n+1}-}, v) p_\varphi(dv, \{t_{n+1}\}) = c(t_{n+1}, Y_{t_{n+1}-}, \xi_{p_\varphi(t_{n+1})}). \quad (8.2.6)$$

If t_{n+1} is not a jump time one has

$$Y_{t_{n+1}} = Y_{t_{n+1}-}, \quad (8.2.7)$$

as

$$\int_{\mathcal{E}} p_{\varphi}(dv, \{t_{n+1}\}) = 0. \quad (8.2.8)$$

Therefore, the strong order of convergence of the jump-adapted Euler scheme is $\gamma = 0.5$, resulting from the strong order of the approximation (8.2.1) of the diffusive component.

For instance, for the SDE (1.8.5), the jump-adapted Euler scheme is given by

$$Y_{t_{n+1}-} = Y_{t_n} + \mu Y_{t_n} \Delta_{t_n} + \sigma Y_{t_n} \Delta W_{t_n} \quad (8.2.9)$$

and

$$Y_{t_{n+1}} = Y_{t_{n+1}-} + Y_{t_{n+1}-} \int_{\mathcal{E}} v p_{\varphi}(dv, \{t_{n+1}\}). \quad (8.2.10)$$

In the multi-dimensional case with scalar Wiener process, being $m = 1$, the k th component of the *jump-adapted Euler scheme* is given by

$$Y_{t_{n+1}-}^k = Y_{t_n}^k + a^k \Delta_{t_n} + b^k \Delta W_{t_n} \quad (8.2.11)$$

and

$$Y_{t_{n+1}}^k = Y_{t_{n+1}-}^k + \int_{\mathcal{E}} c^k(t_{n+1}, \mathbf{Y}_{t_{n+1}-}, v) p_{\varphi}(dv, \{t_{n+1}\}), \quad (8.2.12)$$

for $k \in \{1, 2, \dots, d\}$, where a^k , b^k , and c^k are the k th components of the drift, the diffusion and the jump coefficients, respectively.

For the general multi-dimensional case the k th component of the *jump-adapted Euler scheme* is of the form

$$Y_{t_{n+1}-}^k = Y_{t_n}^k + a^k \Delta_{t_n} + \sum_{j=1}^m b^{k,j} \Delta W_{t_n}^j \quad (8.2.13)$$

and

$$Y_{t_{n+1}}^k = Y_{t_{n+1}-}^k + \int_{\mathcal{E}} c^k(t_{n+1}, \mathbf{Y}_{t_{n+1}-}, v) p_{\varphi}(dv, \{t_{n+1}\}), \quad (8.2.14)$$

where a^k and c^k are the k th components of the drift and the jump coefficients, respectively. Furthermore, $b^{k,j}$ is the component of the k th row and j th column of the diffusion matrix \mathbf{b} , for $k \in \{1, 2, \dots, d\}$, and $j \in \{1, 2, \dots, m\}$. Additionally,

$$\Delta W_{t_n}^j = W_{t_{n+1}}^j - W_{t_n}^j \quad (8.2.15)$$

is the $N(0, \Delta_{t_n})$ distributed n th increment of the j th Wiener process.

Jump-Adapted Strong Order 1.0 Taylor Scheme

As the strong order of convergence of jump-adapted schemes is, in general, the one induced by the approximation of the diffusive part, by replacing the diffusive part of the jump-adapted Euler scheme with that of the strong order 1.0 Taylor scheme for diffusions, see Kloeden & Platen (1999), we obtain the jump-adapted strong order 1.0 Taylor scheme.

For a one-dimensional SDE, $d = m = 1$, the *jump-adapted strong order 1.0 Taylor scheme* is given by

$$Y_{t_{n+1}-} = Y_{t_n} + a \Delta_{t_n} + b \Delta W_{t_n} + \frac{b b'}{2} \left((\Delta W_{t_n})^2 - \Delta_{t_n} \right) \quad (8.2.16)$$

and (8.2.2), which achieves strong order $\gamma = 1.0$. This scheme can be interpreted as a jump-adapted version of the Milstein scheme for pure diffusions, see Milstein (1974).

The comparison of the jump-adapted strong order 1.0 scheme (8.2.16) with the regular strong order 1.0 Taylor scheme (6.2.1), shows that jump-adapted schemes are much simpler. Avoiding the problem of the generation of multiple stochastic integrals with respect to the Poisson measure.

For instance, for the SDE (1.8.5) of the Merton model, we have the jump-adapted strong order 1.0 Taylor scheme of the form

$$Y_{t_{n+1}-} = Y_{t_n} + \mu Y_{t_n} \Delta_{t_n} + \sigma Y_{t_n} \Delta W_{t_n} + \frac{\sigma^2 Y_{t_n}}{2} \left((\Delta W_{t_n})^2 - \Delta_{t_n} \right) \quad (8.2.17)$$

and (8.2.10).

For the multi-dimensional case with scalar Wiener process, being $m = 1$, the k th component of the *jump-adapted strong order 1.0 Taylor scheme* is given by

$$Y_{t_{n+1}-}^k = Y_{t_n}^k + a^k \Delta_{t_n} + b^k \Delta W_{t_n} + \frac{1}{2} \sum_{l=1}^d b^l \frac{\partial b^k}{\partial x^l} \left((\Delta W_{t_n})^2 - \Delta_{t_n} \right) \quad (8.2.18)$$

and (8.2.12), for $k \in \{1, 2, \dots, d\}$.

In the general multi-dimensional case the k th component of the *jump-adapted strong order 1.0 Taylor scheme* is given by

$$Y_{t_{n+1}-}^k = Y_{t_n}^k + a^k \Delta_{t_n} + \sum_{j=1}^m b^{k,j} \Delta W_{t_n}^j + \sum_{j_1, j_2=1}^m \sum_{i=1}^d \left(b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \right) I_{(j_1, j_2)} \quad (8.2.19)$$

and (8.2.14), for $k \in \{1, 2, \dots, d\}$. For the generation of the multiple stochastic integrals $I_{(j_1, j_2)}$ we refer to Sect. 5.3.

If we have a multi-dimensional SDE satisfying the diffusion commutativity condition, where

$$L^{j_1} b^{k, j_2}(t, \mathbf{x}) = L^{j_2} b^{k, j_1}(t, \mathbf{x}) \quad (8.2.20)$$

for $j_1, j_2 \in \{1, 2, \dots, m\}$, $k \in \{1, 2, \dots, d\}$, $t \in [0, T]$ and $\mathbf{x} \in \Re^d$, then it is possible to express all the multiple stochastic integrals in terms of the increments $\Delta W_{t_n}^{j_1}$ and $\Delta W_{t_n}^{j_2}$ of the Wiener process. Therefore, we obtain in this case an efficient *jump-adapted strong order 1.0 Taylor scheme*, whose k th component is given by

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + a^k \Delta_{t_n} + \sum_{j=1}^m b^{k,j} \Delta W_{t_n}^j \\ &\quad + \frac{1}{2} \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \{ \Delta W_{t_n}^{j_1} \Delta W_{t_n}^{j_2} - \Delta_{t_n} \} \end{aligned} \quad (8.2.21)$$

and (8.2.14), for $k \in \{1, 2, \dots, d\}$. The special case of additive diffusion noise, where $\mathbf{b}(t, \mathbf{x}) = \mathbf{b}(t)$, satisfies the required commutativity condition and, therefore, leads to an efficient jump-adapted strong order 1.0 Taylor scheme. Note that this holds for general jump coefficients, unlike the case of regular strong schemes in Chap. 6 and one could easily discuss further special cases of the above schemes.

Jump-Adapted Strong Order 1.5 Taylor Scheme

If we approximate the diffusive part of the SDE (6.1.4) with the strong order 1.5 Taylor scheme, see Kloeden & Platen (1999), then we obtain the *jump-adapted strong order 1.5 Taylor scheme*.

In the autonomous one-dimensional case, $d = m = 1$, the *jump-adapted strong order 1.5 Taylor scheme* is given by

$$\begin{aligned} Y_{t_{n+1}-} &= Y_{t_n} + a \Delta_{t_n} + b \Delta W_{t_n} + \frac{bb'}{2} ((\Delta W_{t_n})^2 - \Delta_{t_n}) + a' b \Delta Z_{t_n} \\ &\quad + \frac{1}{2} \left(a a' + \frac{1}{2} b^2 a'' \right) (\Delta_{t_n})^2 + \left(a b' + \frac{1}{2} b^2 b'' \right) (\Delta W_{t_n} \Delta_{t_n} - \Delta Z_{t_n}) \\ &\quad + \frac{1}{2} b (b b'' + b'^2) \left\{ \frac{1}{3} (\Delta W_{t_n})^2 - \Delta_{t_n} \right\} \Delta W_{t_n}, \end{aligned} \quad (8.2.22)$$

and equation (8.2.2), where we have used the abbreviation defined in (8.1.5). Here we need the multiple stochastic integral

$$\Delta Z_{t_n} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} dW_{s_1} ds_2. \quad (8.2.23)$$

Recall that ΔZ_{t_n} has a Gaussian distribution with mean $E(\Delta Z_{t_n}) = 0$, variance $E((\Delta Z_{t_n})^2) = \frac{1}{3} (\Delta_{t_n})^3$ and covariance $E(\Delta Z_{t_n} \Delta W_{t_n}) = \frac{1}{2} (\Delta_{t_n})^2$.

For example, for the SDE (1.8.5) of the Merton model the terms involving the random variable ΔZ_{t_n} cancel out, thus, yielding the rather simple jump-adapted strong order 1.5 Taylor scheme

$$\begin{aligned}
Y_{t_{n+1}-} &= Y_{t_n} + \mu Y_{t_n} \Delta_{t_n} + \sigma Y_{t_n} \Delta W_{t_n} + \frac{\sigma^2 Y_{t_n}}{2} \left((\Delta W_{t_n})^2 - \Delta_{t_n} \right) \\
&\quad + \frac{\mu^2 Y_{t_n}}{2} (\Delta_{t_n})^2 + \mu \sigma Y_{t_n} (\Delta W_{t_n} \Delta_{t_n}) \\
&\quad + \frac{1}{2} \sigma^3 Y_{t_n} \left\{ \frac{1}{3} (\Delta W_{t_n})^2 - \Delta_{t_n} \right\} \Delta W_{t_n}
\end{aligned} \tag{8.2.24}$$

and (8.2.10).

For the multi-dimensional case with scalar Wiener process, $m = 1$, the k th component of the *jump-adapted strong order 1.5 Taylor scheme* is given by

$$\begin{aligned}
Y_{t_{n+1}-}^k &= Y_{t_n}^k + a^k \Delta_{t_n} + b^k \Delta W_{t_n} + \frac{1}{2} L^1 b^k \left\{ (\Delta W_{t_n})^2 - \Delta_{t_n} \right\} \\
&\quad + L^1 a^k \Delta Z_{t_n} + L^0 b^k \{ \Delta W_{t_n} \Delta_{t_n} - \Delta Z_{t_n} \} + \frac{1}{2} L^0 a^k (\Delta_{t_n})^2 \\
&\quad + \frac{1}{2} L^1 L^1 b^k \left\{ \frac{1}{3} (\Delta W_{t_n})^2 - \Delta_{t_n} \right\} \Delta W_{t_n}
\end{aligned} \tag{8.2.25}$$

and (8.2.12), for $k \in \{1, 2, \dots, d\}$, where the differential operators L^0 and L^1 are defined in (8.1.6) and (8.1.7), respectively.

In the general multi-dimensional case the k th component of the *jump-adapted strong order 1.5 Taylor scheme* is given by

$$\begin{aligned}
Y_{t_{n+1}-}^k &= Y_{t_n}^k + a^k \Delta_{t_n} + \frac{1}{2} L^0 a^k (\Delta_{t_n})^2 \\
&\quad + \sum_{j=1}^m (b^{k,j} \Delta W_{t_n}^j + L^0 b^{k,j} I_{(0,j)} + L^j a^k I_{(j,0)}) \\
&\quad + \sum_{j_1, j_2=1}^m L^{j_1} b^{k, j_2} I_{(j_1, j_2)} + \sum_{j_1, j_2, j_3=1}^m L^{j_1} L^{j_2} b^{k, j_3} I_{(j_1, j_2, j_3)},
\end{aligned} \tag{8.2.26}$$

and (8.2.14), for $k \in \{1, 2, \dots, d\}$.

The double stochastic integrals appearing in the jump-adapted strong order 1.5 Taylor scheme can be generated as discussed in Sect. 5.3. We refer to Kloeden & Platen (1999) for the generation of the required triple stochastic integrals. The diffusion commutativity conditions can reduce the complexity of this scheme, see also Sect. 5.3.

Constructing strong schemes of higher order is, in principle, not difficult. However, since they involve multiple stochastic integrals of higher multiplicity, they can become quite complex. Therefore, we will not present here schemes of strong order higher than $\gamma = 1.5$. Instead we refer to the convergence theorem to be presented in Sect. 8.7 that provides the methodology for the construction of jump-adapted schemes of any given high strong order, which can be directly applied to a given SDE when required.

8.3 Jump-Adapted Derivative-Free Strong Schemes

As cited in Sect. 7.1, it is convenient to develop higher order numerical approximations that do not require the evaluation of derivatives of the coefficient functions. Within jump-adapted schemes, it is sufficient to replace the numerical scheme for the diffusive part by an equivalent derivative-free scheme. We refer to Kloeden & Platen (1999) and Chaps. 5 and 7 for derivative-free schemes for the diffusion part.

Derivative-Free Strong Order 1.0 Scheme

For a one-dimensional SDE, being $d = m = 1$, the *jump-adapted derivative-free strong order 1.0 scheme*, which achieves a strong order $\gamma = 1.0$, is given by

$$\begin{aligned} Y_{t_{n+1}-} &= Y_{t_n} + a\Delta_{t_n} + b\Delta W_{t_n} \\ &+ \frac{1}{2\sqrt{\Delta_{t_n}}}\{b(t_n, \bar{Y}_{t_n}) - b\}((\Delta W_{t_n})^2 - \Delta_{t_n}), \end{aligned} \quad (8.3.1)$$

and (8.2.2) with the supporting value

$$\bar{Y}_{t_n} = Y_{t_n} + b\sqrt{\Delta_{t_n}}. \quad (8.3.2)$$

In the multi-dimensional case with scalar Wiener process, being $m = 1$, the k th component of the *jump-adapted derivative-free strong order 1.0 scheme* is given by

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + a^k\Delta_{t_n} + b^k\Delta W_{t_n} \\ &+ \frac{1}{2\sqrt{\Delta_{t_n}}}\{b^k(t_n, \bar{\mathbf{Y}}_{t_n}) - b^k\}((\Delta W_{t_n})^2 - \Delta_{t_n}), \end{aligned} \quad (8.3.3)$$

and (8.2.12), with the vector supporting value

$$\bar{\mathbf{Y}}_{t_n} = \mathbf{Y}_{t_n} + \mathbf{b}\sqrt{\Delta_{t_n}}, \quad (8.3.4)$$

for $k \in \{1, 2, \dots, d\}$.

In the general multi-dimensional case the k th component of the *jump-adapted derivative-free strong order 1.0 scheme* is given by

$$\begin{aligned} Y_{t_{n+1}}^k &= Y_{t_n}^k + a^k\Delta_{t_n} + \sum_{j=1}^m b^{k,j}\Delta W_{t_n}^j \\ &+ \frac{1}{\sqrt{\Delta_{t_n}}}\sum_{j_1, j_2=1}^m \left\{b^{k, j_2}(t_n, \bar{\mathbf{Y}}_{t_n}^{j_1}) - b^{k, j_2}(t_n, \mathbf{Y}_{t_n})\right\} I_{(j_1, j_2)}, \end{aligned} \quad (8.3.5)$$

and (8.2.14), with the vector supporting value

$$\bar{\mathbf{Y}}_{t_n}^j = \mathbf{Y}_{t_n} + \mathbf{b}^j \sqrt{\Delta_{t_n}}, \quad (8.3.6)$$

for $k \in \{1, 2, \dots, d\}$ and for $j \in \{1, 2, \dots, m\}$. The multiple stochastic integrals can be generated, as described in Sect. 5.3. The diffusion commutativity condition presented in Sect. 8.2 may apply here also, and can therefore, lead to very efficient jump-adapted derivative-free schemes.

Derivative-Free Strong Order 1.5 Scheme

In the autonomous one-dimensional case, $d = m = 1$, the *jump-adapted derivative-free strong order 1.5 scheme* is given by

$$\begin{aligned} Y_{t_{n+1}-} &= Y_{t_n} + b\Delta W_{t_n} + \frac{1}{2\sqrt{\Delta_{t_n}}} \{a(\bar{Y}_{t_n}^+) - a(\bar{Y}_{t_n}^-)\} \Delta Z_{t_n} \\ &\quad + \frac{1}{4} \{a(\bar{Y}_{t_n}^+) + 2a + a(\bar{Y}_{t_n}^-)\} \Delta_{t_n} \\ &\quad + \frac{1}{4\sqrt{\Delta_{t_n}}} \{b(\bar{Y}_{t_n}^+) - b(\bar{Y}_{t_n}^-)\} ((\Delta W_{t_n})^2 - \Delta_{t_n}) \\ &\quad + \frac{1}{2\sqrt{\Delta_{t_n}}} \{b(\bar{Y}_{t_n}^+) + 2b + b(\bar{Y}_{t_n}^-)\} (\Delta W_{t_n} \Delta_{t_n} - \Delta Z_{t_n}) \\ &\quad + \frac{1}{4\sqrt{\Delta_{t_n}}} \{b(\bar{\Phi}_{t_n}^+) - b(\bar{\Phi}_{t_n}^-) - b(\bar{Y}_{t_n}^+) + b(\bar{Y}_{t_n}^-)\} \\ &\quad \times \left\{ \frac{1}{3} (\Delta W_{t_n})^2 - \Delta_{t_n} \right\} \Delta W_{t_n}, \end{aligned} \quad (8.3.7)$$

and (8.2.2), with

$$\bar{Y}_{t_n}^\pm = Y_{t_n} + a\Delta_{t_n} \pm b\Delta W_{t_n}, \quad (8.3.8)$$

and

$$\bar{\Phi}_{t_n}^\pm = \bar{Y}_{t_n}^\pm \pm b(\bar{Y}_{t_n}^+) \sqrt{\Delta_{t_n}}. \quad (8.3.9)$$

The multiple stochastic integral $\Delta Z_{t_n} = I_{(1,0)}$ can be generated as described in (5.3.37).

We could now continue to list multi-dimensional versions of this and other schemes, including those satisfying the diffusion commutativity condition. However, we refrain from doing so since the methodology is straightforward and already demonstrated above.

8.4 Jump-Adapted Drift-Implicit Schemes

As previously discussed, and as will be analyzed in detail in Chap. 16, for some applications it is crucial to construct higher order schemes with wide

regions of numerical stability. To achieve this, one needs to introduce implicitness into the schemes. For the derivation of jump-adapted drift-implicit schemes, it is sufficient to replace the explicit scheme for the diffusive part by a drift-implicit one. We refer to Chap. 7 and [Kloeden & Platen \(1999\)](#) for drift-implicit methods for SDEs driven by Wiener processes only.

Drift-Implicit Euler Scheme

For a one-dimensional SDE, $d = m = 1$, the *jump-adapted drift-implicit Euler scheme* is given by

$$Y_{t_{n+1}-} = Y_{t_n} + \{\theta a(t_{n+1}, Y_{t_{n+1}-}) + (1 - \theta) a\} \Delta_{t_n} + b \Delta W_{t_n}, \quad (8.4.1)$$

and (8.2.2), where the parameter $\theta \in [0, 1]$ characterizes the degree of implicitness.

In the multi-dimensional case with scalar Wiener noise, $m = 1$, the k th component of the *jump-adapted drift-implicit Euler scheme* is given by:

$$Y_{t_{n+1}-}^k = Y_{t_n}^k + \{\theta a^k(t_{n+1}, \mathbf{Y}_{t_{n+1}-}) + (1 - \theta) a^k\} \Delta_{t_n} + b^k \Delta W_{t_n}, \quad (8.4.2)$$

and (8.2.12), for $k \in \{1, 2, \dots, d\}$.

In the multi-dimensional case the k th component of the *jump-adapted drift-implicit Euler scheme* follows as

$$Y_{t_{n+1}-}^k = Y_{t_n}^k + \{\theta a^k(t_{n+1}, \mathbf{Y}_{t_{n+1}-}) + (1 - \theta) a^k\} \Delta_{t_n} + \sum_{j=1}^m b^{k,j} \Delta W_{t_n}^j,$$

and (8.2.14), for $k \in \{1, 2, \dots, d\}$.

Drift-Implicit Strong Order 1.0 Scheme

By using a drift-implicit strong order 1.0 scheme for the diffusive part, we obtain a jump-adapted drift-implicit strong order 1.0 scheme.

For a one-dimensional SDE, $d = m = 1$, the *jump-adapted drift-implicit strong order 1.0 scheme* is given by

$$\begin{aligned} Y_{t_{n+1}-} = Y_{t_n} &+ \{\theta a(t_{n+1}, Y_{t_{n+1}-}) + (1 - \theta) a\} \Delta_{t_n} + b \Delta W_{t_n} \\ &+ \frac{b b'}{2} ((\Delta W_{t_n})^2 - \Delta_{t_n}) \end{aligned} \quad (8.4.3)$$

and (8.2.2), which achieves strong order $\gamma = 1.0$.

In the multi-dimensional case with scalar Wiener process, $m = 1$, the k th component of the *jump-adapted drift-implicit strong order 1.0 scheme* is given by

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + \{\theta a^k(t_{n+1}, \mathbf{Y}_{t_{n+1}-}) + (1 - \theta) a^k\} \Delta_{t_n} + b^k \Delta W_{t_n} \\ &\quad + \frac{1}{2} \sum_{l=1}^d b^l \frac{\partial b^k}{\partial x^l} \left((\Delta W_{t_n})^2 - \Delta_{t_n} \right) \end{aligned} \quad (8.4.4)$$

and (8.2.12), for $k \in \{1, 2, \dots, d\}$.

In the general multi-dimensional case, the k th component of the *jump-adapted drift-implicit strong order 1.0 scheme* follows as

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + \{\theta a^k(t_{n+1}, \mathbf{Y}_{t_{n+1}-}) + (1 - \theta) a^k\} \Delta_{t_n} + \sum_{j=1}^m b^{k,j} \Delta W_{t_n}^j \\ &\quad + \sum_{j_1, j_2=1}^m \sum_{i=1}^d \left(b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \right) I_{(j_1, j_2)} \end{aligned} \quad (8.4.5)$$

and (8.2.14), for $k \in \{1, 2, \dots, d\}$.

Drift-Implicit Strong Order 1.5 Scheme

Finally, we present a jump-adapted drift-implicit strong order 1.5 Taylor type scheme.

In the autonomous one-dimensional case, $d = m = 1$, the *jump-adapted drift-implicit strong order 1.5 scheme* in its simplest form is given by

$$\begin{aligned} Y_{t_{n+1}-} &= Y_{t_n} + \frac{1}{2} \{a(Y_{t_{n+1}-}) + a\} \Delta_{t_n} + b \Delta W_{t_n} + \frac{b b'}{2} \left((\Delta W_{t_n})^2 - \Delta_{t_n} \right) \\ &\quad + \left(a b' + \frac{1}{2} b^2 b'' \right) (\Delta W_{t_n} \Delta_{t_n} - \Delta Z_{t_n}) + a' b \left\{ \Delta Z_{t_n} - \frac{1}{2} \Delta W_{t_n} \Delta_{t_n} \right\} \\ &\quad + \frac{1}{2} b \left(b b'' + (b')^2 \right) \left\{ \frac{1}{3} (\Delta W_{t_n})^2 - \Delta_{t_n} \right\} \Delta W_{t_n}, \end{aligned} \quad (8.4.6)$$

and (8.2.2), and achieves strong order $\gamma = 1.5$.

It is straightforward to formulate for the multi-dimensional case the generalization of this scheme. Under the diffusion commutativity condition the resulting schemes simplify further in an obvious manner.

We remark also that balanced implicit methods, see Sect. 7.3 and Milstein et al. (1998), Kahl & Schurz (2006) and Alcock & Burrage (2006), can be used to construct a numerically stable approximation of the diffusion part of the SDE. In general, when using balanced implicit methods, one still has to solve an algebraic equation at each time step. The following class of predictor-corrector methods avoids this extra computational effort.

8.5 Predictor-Corrector Strong Schemes

As previously discussed, predictor-corrector schemes combine good numerical stability properties with efficiency. In this section we present jump-adapted predictor-corrector schemes with strong order of convergence $\gamma \in \{0.5, 1\}$, obtained by a straightforward application of results in [Bruti-Liberati & Platen \(2008\)](#), see also Sect. 7.4.

Predictor-Corrector Euler Scheme

In the one-dimensional case, $d = m = 1$, the *family of jump-adapted predictor-corrector Euler schemes* is given by the corrector

$$\begin{aligned} Y_{t_{n+1}-} &= Y_n + \{\theta \bar{a}_\eta(t_{n+1}, \bar{Y}_{t_{n+1}-}) + (1 - \theta) \bar{a}_\eta\} \Delta_{t_n} \\ &\quad + \{\eta b(t_{n+1}, \bar{Y}_{t_{n+1}-}) + (1 - \eta)b\} \Delta W_{t_n}, \end{aligned} \quad (8.5.1)$$

where $\bar{a}_\eta = a - \eta b b'$, the predictor

$$\bar{Y}_{t_{n+1}-} = Y_{t_n} + a \Delta_{t_n} + b \Delta W_{t_n}, \quad (8.5.2)$$

and (8.2.2). The parameters $\theta, \eta \in [0, 1]$ characterize the degree of implicitness in the drift and in the diffusion coefficients, respectively. This scheme achieves a strong order of convergence $\gamma = 0.5$.

For the general multi-dimensional case, the k th component of the *family of jump-adapted predictor-corrector Euler schemes* is given by the corrector

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + \{\theta \bar{a}_\eta^k(t_{n+1}, \bar{Y}_{t_{n+1}-}) + (1 - \theta) \bar{a}_\eta^k\} \Delta_{t_n} \\ &\quad + \sum_{j=1}^m \{\eta b^{k,j}(t_{n+1}, \bar{Y}_{t_{n+1}}) + (1 - \eta)b^{k,j}\} \Delta W_{t_n}^j, \end{aligned} \quad (8.5.3)$$

for $\theta, \eta \in [0, 1]$, where

$$\bar{a}_\eta^k = a^k - \eta \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{k, j_1} \frac{\partial b^{k, j_2}}{\partial x^i}, \quad (8.5.4)$$

the predictor

$$\bar{Y}_{t_{n+1}-}^k = Y_{t_n}^k + a^k \Delta_{t_n} + \sum_{j=1}^m b^{k,j} \Delta W_{t_n}^j, \quad (8.5.5)$$

and (8.2.14).

Predictor-Corrector Strong Order 1.0 Scheme

Here we present the *jump-adapted predictor-corrector strong order 1.0 scheme*. For a one-dimensional SDE, $d = m = 1$, it is given by the corrector

$$\begin{aligned} Y_{t_{n+1}-} &= Y_{t_n} + \{\theta a(t_{n+1}, \bar{Y}_{t_{n+1}-}) + (1 - \theta) a\} \Delta_{t_n} \\ &\quad + b \Delta W_{t_n} + \frac{bb'}{2} ((\Delta W_{t_n})^2 - \Delta_{t_n}), \end{aligned} \quad (8.5.6)$$

the predictor

$$\bar{Y}_{t_{n+1}-} = Y_{t_n} + a \Delta_{t_n} + b \Delta W_{t_n} + \frac{bb'}{2} ((\Delta W_{t_n})^2 - \Delta_{t_n}) \quad (8.5.7)$$

and (8.2.2). The parameter $\theta \in [0, 1]$ characterizes again the degree of implicitness in the drift coefficient. This scheme attains a strong order $\gamma = 1.0$.

In the general multi-dimensional case, the k th component of the *jump-adapted predictor-corrector strong order 1.0 scheme* is given by the corrector

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + \{\theta a^k(t_{n+1}, \bar{Y}_{t_{n+1}-}) + (1 - \theta) a^k\} \Delta_{t_n} + \sum_{j=1}^m b^{k,j} \Delta W_{t_n}^j \\ &\quad + \sum_{j_1, j_2=1}^m \sum_{i=1}^d \left(b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \right) I_{(j_1, j_2)}, \end{aligned} \quad (8.5.8)$$

the predictor

$$\bar{Y}_{t_{n+1}-}^k = Y_{t_n}^k + a^k \Delta_{t_n} + \sum_{j=1}^m b^{k,j} \Delta W_{t_n}^j + \sum_{j_1, j_2=1}^m \sum_{i=1}^d \left(b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \right) I_{(j_1, j_2)} \quad (8.5.9)$$

and (8.2.14), for $k \in \{1, 2, \dots, d\}$. For the generation of the multiple stochastic integrals $I_{(j_1, j_2)}$ we refer again to Sect. 5.3.

For SDEs satisfying the diffusion commutativity condition (8.2.20), as in the case of an additive diffusion coefficient $\mathbf{b}(t, \mathbf{x}) = \mathbf{b}(t)$, we can express all of the multiple stochastic integrals in terms of the increments $\Delta W_{t_n}^{j_1}$ and $\Delta W_{t_n}^{j_2}$ of the Wiener process. This yields an efficiently implementable *jump-adapted predictor-corrector strong order 1.0 scheme* whose k th component is given by the corrector

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + \{\theta a^k(t_{n+1}, \bar{Y}_{t_{n+1}-}) + (1 - \theta) a^k\} \Delta_{t_n} \\ &\quad + \sum_{j=1}^m b^{k,j} \Delta W_{t_n}^j + \frac{1}{2} \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \{\Delta W_{t_n}^{j_1} \Delta W_{t_n}^{j_2} - \Delta_{t_n}\}, \end{aligned}$$

the predictor

$$\begin{aligned}\bar{Y}_{t_{n+1}-}^k &= Y_{t_n}^k + a^k \Delta_{t_n} + \sum_{j=1}^m b^{k,j} \Delta W_{t_n}^j \\ &\quad + \frac{1}{2} \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \{ \Delta W_{t_n}^{j_1} \Delta W_{t_n}^{j_2} - \Delta_{t_n} \} \quad (8.5.10)\end{aligned}$$

and (8.2.14), for $k \in \{1, 2, \dots, d\}$.

We remark that, as in Sect. 7.5, we can make the diffusion part quasi-implicit to obtain better numerical stability.

8.6 Jump-Adapted Exact Simulation

We presented in Chap. 2 the exact or almost exact simulation of solutions of SDEs.

In this section we extend this methodology and discuss the strong approximation of special classes of SDEs under which jump-adapted schemes have no or almost no discretization error. As will become clear later, in this case it is *not* necessary to specify a small maximum time step size Δ , see condition (8.1.3). Here the jump-adapted time discretization will be given by a superposition of the jump times generated by the Poisson measure and the times at which we are interested in sampling the simulated values of the solution X . For instance, if one needs the value of a jump diffusion X , only at the final time T , then the jump-adapted time discretization is given by $0 = t_0 < t_1 < \dots < t_{n_T} = T$, where n_T is defined in (6.1.11) and the sequence $t_1 < \dots < t_{n_T-1}$ equals that of the jump times $\tau_1 < \dots < \tau_{p_\varphi(T)}$ of the Poisson measure p_φ .

A Simple Example

Let us first present an illustrative example. Consider the jump diffusion SDE with multiplicative drift and diffusion coefficients and general jump coefficient $c(t, x, v)$, given by

$$dX_t = \mu X_t dt + \sigma X_t dW_t + \int_{\mathcal{E}} c(t, X_{t-}, v) p_\varphi(dv, dt). \quad (8.6.1)$$

Because of the general form of the jump coefficient c , the SDE (8.6.1) rarely admits an explicit solution. However, based on the above described jump-adapted time discretization, we can construct the following jump-adapted scheme given by

$$Y_{t_{n+1}-} = Y_{t_n} e^{(\mu - \frac{1}{2} \sigma^2) \Delta_{t_n} + \sigma \Delta W_{t_n}}, \quad (8.6.2)$$

and

$$Y_{t_{n+1}} = Y_{t_{n+1}-} + \int_{\mathcal{E}} c(t_{n+1}, Y_{t_{n+1}-}, v) p_\varphi(dv, \{t_{n+1}\}). \quad (8.6.3)$$

Since we are using the explicit solution of the diffusion part, see (1.8.6), no discretization error is introduced between jump times. Moreover, since by equation (8.6.3) the jump impact is added at the correct jump times, as such, even at the jump times we do not introduce any error. Therefore, we have described a method of simulating the solution of the jump diffusion SDE (8.6.1) that does not generate any discretization error.

Jump-Adapted Almost Exact Solutions

The above approach can be generalized for cases where we have an explicit or almost exact solution to the diffusion part of the SDE under consideration. Examples are described in Chap. 2. In the general case, we consider here the d -dimensional jump diffusion SDE

$$d\mathbf{X}_t = \mathbf{a}(t, \mathbf{X}_t) dt + \mathbf{b}(t, \mathbf{X}_t) d\mathbf{W}_t + \int_{\mathcal{E}} \mathbf{c}(t, \mathbf{X}_{t-}, v) p_\varphi(dv, dt), \quad (8.6.4)$$

that we aim to solve. One has then to check whether this SDE belongs to the special subclass of jump diffusion SDEs for which the corresponding diffusion SDE

$$d\mathbf{Z}_t = \mathbf{a}(t, \mathbf{Z}_t) dt + \mathbf{b}(t, \mathbf{Z}_t) d\mathbf{W}_t, \quad (8.6.5)$$

admits an explicit or almost exact solution in the sense of Chap. 2. If this is the case, then we can construct, as in (8.6.2)–(8.6.3), a jump-adapted scheme without discretization error. This means, by the above methodology it is possible to simulate exact solutions for a considerable range of SDEs with jumps.

As previously noticed, if the SDE under consideration is driven by a Poisson measure with high intensity, then a jump-adapted scheme may be computationally too expensive. In such a case, one may prefer to use a regular scheme that entails a discretization error, but permits the use of a coarser time discretization. There may be also cases where one can use the explicit solution of a Lévy process driven SDE when approximating a more complex SDE with a few low intensity major jumps.

8.7 Convergence Results

In this section we present a convergence theorem for jump-adapted approximations that allow us to assess the strong order of convergence of the methods presented in this chapter.

Sets of Multi-Indices

We consider here a jump-adapted time discretization $(t)_\Delta$, as defined in (8.1.2). The time discretization includes all jump times $\{\tau_1, \tau_2, \dots, \tau_{p_\varphi(T)}\}$ of the Poisson measure p_φ . As explained in Sect. 8.1, by construction, the jumps

arise only at discretization points. Therefore, between discretization points we can approximate the stochastic process X , for instance, with a strong Taylor scheme for diffusions. For this reason, here we slightly modified the notation used for the sets of multi-indices, introduced in Chap. 4, as outlined below.

For $m \in \mathcal{N}$ the set of all multi-indices α that do not include components equal to -1 is now denoted by

$$\widehat{\mathcal{M}}_m = \{(j_1, \dots, j_l) : j_i \in \{0, 1, 2, \dots, m\}, i \in \{1, 2, \dots, l\} \text{ for } l \in \mathcal{N}\} \cup \{v\},$$

where v is the multi-index of length zero.

We also recall the operators

$$\begin{aligned} L^0 f(t, \mathbf{x}) &= \frac{\partial}{\partial t} f(t, \mathbf{x}) + \sum_{i=1}^d a^i(t, \mathbf{x}) \frac{\partial}{\partial x^i} f(t, \mathbf{x}) \\ &+ \frac{1}{2} \sum_{i,r=1}^d \sum_{j=1}^m b^{i,j}(t, \mathbf{x}) b^{r,j}(t, \mathbf{x}) \frac{\partial^2}{\partial x^i \partial x^j} f(t, \mathbf{x}) \end{aligned} \quad (8.7.1)$$

and

$$L^k f(t, \mathbf{x}) = \sum_{i=1}^d b^{i,k}(t, \mathbf{x}) \frac{\partial}{\partial x^i} f(t, \mathbf{x}) \quad (8.7.2)$$

for $k \in \{1, 2, \dots, m\}$ and a function $f(t, \mathbf{x}) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ from $\mathcal{C}^{1,2}$.

For all $\alpha = (j_1, \dots, j_{l(\alpha)}) \in \widehat{\mathcal{M}}_m$ and a function $f : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, we define recursively the *Itô coefficient functions* f_α

$$f_\alpha(t, \mathbf{x}) = \begin{cases} f(t, \mathbf{x}) & \text{for } l(\alpha) = 0, \\ L^{j_1} f_{-\alpha}(t, \mathbf{x}) & \text{for } l(\alpha) \geq 1, \end{cases} \quad (8.7.3)$$

assuming that the coefficients of the SDE (1.8.2) are sufficiently smooth for the operators in (8.7.3) so as to be well defined.

Then given a set $\mathcal{A} \subset \widehat{\mathcal{M}}_m$, we also define the *remainder set* $\hat{\mathcal{B}}(\mathcal{A})$ of \mathcal{A} by

$$\hat{\mathcal{B}}(\mathcal{A}) = \{\alpha \in \widehat{\mathcal{M}}_m \setminus \mathcal{A} : -\alpha \in \mathcal{A}\}.$$

Moreover, for every $\gamma \in \{0.5, 1, 1.5, 2, \dots\}$ we define the hierarchical set

$$\hat{\mathcal{A}}_\gamma = \left\{ \alpha \in \widehat{\mathcal{M}} : l(\alpha) + n(\alpha) \leq 2\gamma \quad \text{or} \quad l(\alpha) = n(\alpha) = \gamma + \frac{1}{2} \right\}.$$

Jump-Adapted Strong Taylor Schemes

For a jump-adapted time discretization with maximum time step size $\Delta \in (0, 1)$ we define the *jump-adapted strong order γ Taylor scheme* by

$$\mathbf{Y}_{t_{n+1}-} = \mathbf{Y}_{t_n} + \sum_{\alpha \in \hat{\mathcal{A}}_\gamma \setminus \{v\}} f_\alpha(t_n, \mathbf{Y}_{t_n}) I_\alpha \quad (8.7.4)$$

and

$$\mathbf{Y}_{t_{n+1}} = \mathbf{Y}_{t_{n+1}-} + \int_{\mathcal{E}} \mathbf{c}(t_n, \mathbf{Y}_{t_{n+1}-}, v) p_\varphi(dv, \{t_{n+1}\}), \quad (8.7.5)$$

where I_α is the multiple stochastic integral of the multi-index α over the time period $(t_n, t_{n+1}]$ and for $n \in \{0, 1, \dots, n_T - 1\}$.

To assess the order of strong convergence of these schemes, we define through a specific interpolation the *jump-adapted strong order γ Taylor approximation* by

$$\mathbf{Y}_t = \sum_{\alpha \in \hat{\mathcal{A}}_\gamma \setminus \{v\}} I_\alpha[f_\alpha(t_{n_t}, \mathbf{Y}_{t_{n_t}})]_{t_{n_t}, t} \quad (8.7.6)$$

as there are no jumps between grid points.

Convergence Theorem

We can now formulate a convergence theorem for jump-adapted strong schemes similar to the result in [Platen \(1982a\)](#).

Theorem 8.7.1. *For a given $\gamma \in \{0.5, 1, 1.5, 2, \dots\}$, let $\mathbf{Y}^\Delta = \{\mathbf{Y}_t^\Delta, t \in [0, T]\}$ be the jump-adapted strong order γ Taylor approximation corresponding to a jump-adapted time discretization with maximum step size $\Delta \in (0, 1)$. We assume that*

$$E(|\mathbf{X}_0|^2) < \infty \quad \text{and} \quad E(|\mathbf{X}_0 - \mathbf{Y}_0^\Delta|^2) \leq C \Delta^{2\gamma}. \quad (8.7.7)$$

Moreover, suppose that the coefficient functions f_α satisfy the following conditions:

For $\alpha \in \hat{\mathcal{A}}_\gamma$, $t \in [0, T]$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, the coefficient f_α satisfies the Lipschitz type condition

$$|f_\alpha(t, \mathbf{x}) - f_\alpha(t, \mathbf{y})| \leq K_1 |\mathbf{x} - \mathbf{y}|. \quad (8.7.8)$$

For all $\alpha \in \hat{\mathcal{A}}_\gamma \cup \hat{\mathcal{B}}(\hat{\mathcal{A}}_\gamma)$ we assume

$$f_{-\alpha} \in \mathcal{C}^{1,2} \quad \text{and} \quad f_\alpha \in \mathcal{H}_\alpha, \quad (8.7.9)$$

and for $\alpha \in \hat{\mathcal{A}}_\gamma \cup \hat{\mathcal{B}}(\hat{\mathcal{A}}_\gamma)$, $t \in [0, T]$ and $\mathbf{x} \in \mathbb{R}^d$, we require

$$|f_\alpha(t, \mathbf{x})|^2 \leq K_2 (1 + |\mathbf{x}|^2). \quad (8.7.10)$$

Then the estimate

$$\sqrt{E \left(\sup_{0 \leq s \leq T} |\mathbf{X}_s - \mathbf{Y}_s^\Delta|^2 \mid \mathcal{A}_0 \right)} \leq K_3 \Delta^\gamma \quad (8.7.11)$$

holds, when the constant K_3 does not depend on Δ .

Remark 8.7.2. Instead of conditions (8.7.8)–(8.7.10) on the coefficients f_α , one can derive analogous conditions on the coefficients \mathbf{a} , \mathbf{b} and \mathbf{c} of the SDE (4.4.1), see also Remark 6.4.2.

Proof of Theorem 8.7.1 Since the jump-adapted time discretization contains all jump times of the solution \mathbf{X} of the SDE (1.8.2), then with the aid of the Wagner-Platen expansion for diffusion processes we can write

$$\begin{aligned} \mathbf{X}_t = \mathbf{X}_0 + & \sum_{\alpha \in \hat{\mathcal{A}}_\gamma \setminus \{v\}} \left\{ \sum_{n=0}^{n_t-1} I_\alpha[f_\alpha(t_n, \mathbf{X}_{t_n})]_{t_n, t_{n+1}} + I_\alpha[f_\alpha(t_{n_t}, \mathbf{X}_{t_{n_t}})]_{t_{n_t}, t} \right\} \\ & + \sum_{\alpha \in \hat{\mathcal{B}}(\hat{\mathcal{A}}_\gamma)} \left\{ \sum_{n=0}^{n_t-1} I_\alpha[f_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_n, t_{n+1}} + I_\alpha[f_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_{n_t}, t} \right\} \\ & + \int_0^t \int_{\mathcal{E}} \mathbf{c}(t_{n_z}, \mathbf{X}_{t_{n_z}-}, v) p_\varphi(dv, dz), \end{aligned} \quad (8.7.12)$$

for $t \in [0, T]$.

The jump-adapted strong order γ Taylor scheme can be written as

$$\begin{aligned} \mathbf{Y}_t = \mathbf{Y}_0 + & \sum_{\alpha \in \hat{\mathcal{A}}_\gamma \setminus \{v\}} \left\{ \sum_{n=0}^{n_t-1} I_\alpha[f_\alpha(t_n, \mathbf{Y}_{t_n})]_{t_n, t_{n+1}} + I_\alpha[f_\alpha(t_{n_t}, \mathbf{Y}_{t_{n_t}})]_{t_{n_t}, t} \right\} \\ & + \int_0^t \int_{\mathcal{E}} \mathbf{c}(t_{n_z}, \mathbf{Y}_{t_{n_z}-}, v) p_\varphi(dv, dz) \end{aligned} \quad (8.7.13)$$

for every $t \in [0, T]$.

From the estimate of Theorem 1.9.3 we have

$$E \left(\sup_{0 \leq z \leq T} |\mathbf{X}_z|^2 \mid \mathcal{A}_0 \right) \leq C (1 + E(|\mathbf{X}_0|^2)). \quad (8.7.14)$$

Moreover, with similar steps to those used in the proof of Lemma 6.6.1, we can show the estimate

$$E \left(\sup_{0 \leq z \leq T} |\mathbf{Y}_z^\Delta|^2 \mid \mathcal{A}_0 \right) \leq C (1 + E(|\mathbf{Y}_0^\Delta|^2)). \quad (8.7.15)$$

Now, the mean square error is given by

$$\begin{aligned}
Z(t) &= E \left(\sup_{0 \leq z \leq t} |\mathbf{X}_z - \mathbf{Y}_z^\Delta|^2 \middle| \mathcal{A}_0 \right) \\
&= E \left(\sup_{0 \leq z \leq t} \left| \mathbf{X}_0 - \mathbf{Y}_0^\Delta + \sum_{\alpha \in \hat{\mathcal{A}}_\gamma \setminus \{v\}} \left\{ \sum_{n=0}^{n_z-1} I_\alpha [f_\alpha(t_n, \mathbf{X}_{t_n}) - f_\alpha(t_n, \mathbf{Y}_{t_n}^\Delta)]_{t_n, t_{n+1}} \right. \right. \right. \\
&\quad \left. \left. \left. + I_\alpha [f_\alpha(t_{n_z}, \mathbf{X}_{t_{n_z}}) - f_\alpha(t_{n_z}, \mathbf{Y}_{t_{n_z}}^\Delta)]_{t_{n_z}, z} \right\} \right. \right. \\
&\quad \left. \left. + \sum_{\alpha \in \hat{\mathcal{B}}(\hat{\mathcal{A}}_\gamma)} \left\{ \sum_{n=0}^{n_z-1} I_\alpha [f_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_n, t_{n+1}} + I_\alpha [f_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_{n_z}, z} \right\} \right. \right. \\
&\quad \left. \left. + \int_0^z \int_{\mathcal{E}} \{ \mathbf{c}(t_{n_u}, \mathbf{X}_{t_{n_u}-}, v) - \mathbf{c}(t_{n_u}, \mathbf{Y}_{t_{n_u}-}, v) \} p_\varphi(dv, du) \right|^2 \middle| \mathcal{A}_0 \right) \\
&\leq C_3 \left\{ \left| \mathbf{X}_0 - \mathbf{Y}_0^\Delta \right|^2 + \sum_{\alpha \in \hat{\mathcal{A}}_\gamma \setminus \{v\}} S_t^\alpha + \sum_{\alpha \in \hat{\mathcal{B}}(\hat{\mathcal{A}}_\gamma)} U_t^\alpha + P_t \right\} \tag{8.7.16}
\end{aligned}$$

for all $t \in [0, T]$, where S_t^α , U_t^α and P_t are defined by

$$\begin{aligned}
S_t^\alpha &= E \left(\sup_{0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} I_\alpha [f_\alpha(t_n, \mathbf{X}_{t_n}) - f_\alpha(t_n, \mathbf{Y}_{t_n}^\Delta)]_{t_n, t_{n+1}} \right. \right. \\
&\quad \left. \left. + I_\alpha [f_\alpha(t_{n_z}, \mathbf{X}_{t_{n_z}}) - f_\alpha(t_{n_z}, \mathbf{Y}_{t_{n_z}}^\Delta)]_{t_{n_z}, z} \right|^2 \middle| \mathcal{A}_0 \right), \tag{8.7.17}
\end{aligned}$$

$$U_t^\alpha = E \left(\sup_{0 \leq z \leq t} \left| \sum_{n=0}^{n_z-1} I_\alpha [f_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_n, t_{n+1}} + I_\alpha [f_\alpha(\cdot, \mathbf{X}_\cdot)]_{t_{n_z}, z} \right|^2 \middle| \mathcal{A}_0 \right), \tag{8.7.18}$$

and

$$P_t = E \left(\sup_{0 \leq z \leq t} \left| \int_0^z \int_{\mathcal{E}} \{ \mathbf{c}(t_{n_u}, \mathbf{X}_{t_{n_u}-}, v) - \mathbf{c}(t_{n_u}, \mathbf{Y}_{t_{n_u}-}, v) \} p_\varphi(dv, du) \right|^2 \middle| \mathcal{A}_0 \right). \tag{8.7.19}$$

Therefore, the terms S_t^α and U_t^α can be estimated as in the proof of Theorem 6.4.1, while for P_t , by applying Jensen's and Doob's inequalities, Itô's isometry for jump processes, the Cauchy-Schwarz inequality and the Lipschitz condition (1.9.2), we obtain

$$\begin{aligned}
P_t &= E \left(\sup_{0 \leq z \leq t} \left| \int_0^z \int_{\mathcal{E}} \{ \mathbf{c}(t_{n_u}, \mathbf{X}_{t_{n_u}-}, v) - \mathbf{c}(t_{n_u}, \mathbf{Y}_{t_{n_u}-}, v) \} \tilde{p}_\varphi(dv, du) \right. \right. \\
&\quad \left. \left. + \int_0^z \int_{\mathcal{E}} \{ \mathbf{c}(t_{n_u}, \mathbf{X}_{t_{n_u}-}, v) - \mathbf{c}(t_{n_u}, \mathbf{Y}_{t_{n_u}-}, v) \} \varphi(dv) du \right|^2 \middle| \mathcal{A}_0 \right) \\
&\leq 8 E \left(\left| \int_0^t \int_{\mathcal{E}} \{ \mathbf{c}(t_{n_u}, \mathbf{X}_{t_{n_u}-}, v) - \mathbf{c}(t_{n_u}, \mathbf{Y}_{t_{n_u}-}, v) \} \tilde{p}_\varphi(dv, du) \right|^2 \middle| \mathcal{A}_0 \right) \\
&\quad + 2 E \left(\sup_{0 \leq z \leq t} \left| \int_0^z \int_{\mathcal{E}} \{ \mathbf{c}(t_{n_u}, \mathbf{X}_{t_{n_u}-}, v) - \mathbf{c}(t_{n_u}, \mathbf{Y}_{t_{n_u}-}, v) \} \varphi(dv) du \right|^2 \middle| \mathcal{A}_0 \right) \\
&\leq 8 E \left(\int_0^t \int_{\mathcal{E}} |\mathbf{c}(t_{n_u}, \mathbf{X}_{t_{n_u}-}, v) - \mathbf{c}(t_{n_u}, \mathbf{Y}_{t_{n_u}-}, v)|^2 \varphi(dv) du \middle| \mathcal{A}_0 \right) \\
&\quad + 2 \lambda t E \left(\int_0^t \int_{\mathcal{E}} |\mathbf{c}(t_{n_u}, \mathbf{X}_{t_{n_u}-}, v) - \mathbf{c}(t_{n_u}, \mathbf{Y}_{t_{n_u}-}, v)|^2 \varphi(dv) du \middle| \mathcal{A}_0 \right) \\
&\leq K E \left(\int_0^t |\mathbf{X}_{t_{n_u}-} - \mathbf{Y}_{t_{n_u}-}|^2 du \middle| \mathcal{A}_0 \right) \\
&\leq C \int_0^t Z(u) du. \tag{8.7.20}
\end{aligned}$$

Therefore, since by (8.7.14) and (8.7.15) the quantity $Z(t)$ is bounded, by application of the Gronwall inequality to (8.7.16) we complete the proof of Theorem 8.7.1. \square

Theorem 8.7.1 establishes the order of strong convergence of the jump-adapted strong Taylor schemes presented in Sect. 8.2. To prove the order of strong convergence of the jump-adapted schemes in the remainder of this chapter, one can define the *jump-adapted strong order γ Itô scheme*, with $\gamma \in \{0.5, 1, 1.5, \dots\}$ using the following procedure: Approximate the diffusion part by a strong order γ Itô scheme for pure diffusions, see Chap. 5 or Kloeden & Platen (1999), and generate the jump part as in (8.7.5).

One can prove the strong order of convergence of the jump-adapted strong order γ Itô scheme by first showing that this scheme converges with strong order γ to the jump-adapted strong order γ Taylor scheme. This can be done by using similar steps as to those described in the proof of Theorem 7.5.1. Thus, since Theorem 8.7.1 establishes the strong order of convergence of jump-adapted strong Taylor schemes, this also yields the strong order γ of the corresponding jump-adapted strong Itô schemes. Finally, the strong order of the jump-adapted derivative-free, implicit and predictor-corrector schemes presented in this chapter can be shown by rewriting these schemes as jump-adapted strong Itô schemes.

8.8 Numerical Results on Strong Schemes

This brief section provides some numerical results from the application of the strong schemes presented in Chaps. 6, 7 and 8. We investigate the accuracy of strong schemes, focussing on jump-adapted schemes.

Test Example

We study the strong approximation of the one-dimensional linear SDE

$$dX_t = X_{t-} \left(\mu dt + \sigma dW_t + \int_{\mathcal{E}} (v - 1) p_\varphi(dv, dt) \right) \quad (8.8.1)$$

for $t \in [0, T]$ and $X_0 > 0$, being the Merton model introduced in (1.8.5). We recall that this SDE admits the explicit solution

$$X_t = X_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t} \prod_{i=1}^{p_\varphi(t)} \xi_i, \quad (8.8.2)$$

where the marks ξ_i are distributed according to a given probability measure $F(dv) = \frac{\varphi(dv)}{\lambda}$ and $p_\varphi = \{p_\varphi(t), t \in [0, T]\}$ denotes a Poisson process with intensity $\lambda = \varphi(\mathcal{E}) < \infty$.

Strong Error

We consider the following strong order $\gamma = 0.5$ regular and jump-adapted drift-implicit and predictor-corrector Euler schemes. Moreover, we study the following strong order $\gamma = 1.0$ regular and jump-adapted Taylor and jump-adapted drift-implicit strong schemes. Additionally, we present the jump-adapted strong order 1.5 Taylor scheme, which attains strong order $\gamma = 1.5$. We report the strong error

$$\varepsilon(\Delta) = E(|X_T - Y_T^\Delta|), \quad (8.8.3)$$

as defined in (5.2.25), when comparing the results of these strong schemes to the closed form solution (8.8.2), we estimate the strong error $\varepsilon(\Delta)$ by running an extremely large number of simulations. The exact number depends on the scheme implemented, and is chosen such that the statistical error becomes negligible compared to the systematic time discretization error. In the corresponding plots we show the logarithm $\log_2(\varepsilon(\Delta))$ of the strong error versus the logarithm $\log_2(\Delta)$ of the time step size. The slopes of the estimated error in the log-log plots indicate the order of strong convergence attained by each scheme. Here we consider the case of an SDE with a driving jump process with small intensity $\lambda = 0.05$. Then later, to illustrate the impact of frequent jumps on the strong error, we consider a jump process with higher intensity.

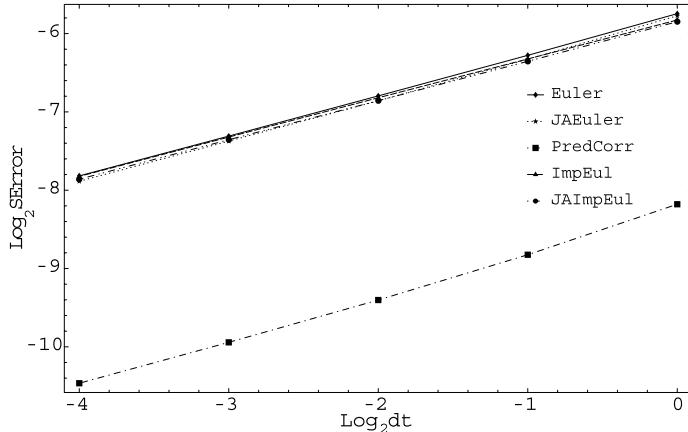


Fig. 8.8.1. Log-log plot of strong error versus time step size (constant marks)

The Case of Low Intensities

In this subsection we use the following default parameters: $\mu = 0.05$, $\sigma = 0.15$, $X_0 = 1$, $T = 1$ and $\lambda = 0.05$. First, we consider the case of the SDE (8.8.1) with degenerate marks, that is $\xi_i = \psi > 0$, with $\psi = 0.85$. This reduces the SDE (8.8.1) to an SDE with mark-independent jump coefficient $c(t, x, v) = x(\psi - 1) = -0.15x$.

In Fig. 8.8.1, we report the results obtained from the: regular and jump-adapted, regular and jump-adapted drift-implicit, and the regular predictor-corrector Euler schemes. We omit from Fig. 8.8.1 the results of the jump-adapted predictor-corrector Euler scheme as its strong error is indistinguishable from that of the jump-adapted Euler scheme. Note that, here and in the rest of this section, the implicitness parameters are set to $\theta = \eta = 0.5$. All schemes achieve an order of strong convergence of approximately $\gamma = 0.5$. This is consistent with the theoretical strong orders derived in previous chapters. Moreover all schemes, except for the regular predictor-corrector Euler scheme, are of similar accuracy. The regular predictor-corrector Euler scheme is significantly more accurate than the alternative schemes for all time step sizes considered, due to the corrector step's refinement of the estimate.

We now analyze the strong errors generated by the regular and jump-adapted strong order 1.0 Taylor, jump-adapted drift-implicit strong order 1.0, and the jump-adapted strong order 1.5 Taylor schemes. In Fig. 8.8.2 we present the results for these schemes, along with those of the regular predictor-corrector Euler scheme, already plotted in Fig. 8.8.1. We omit from Fig. 8.8.2 the result of the regular strong order 1.0 Taylor scheme, since its strong error is almost indistinguishable from that of the equivalent jump-adapted version. We notice that the jump-adapted strong order 1.0 Taylor and the jump-adapted drift-implicit strong order 1.0 schemes, achieve a theoretical strong order 1.0 in accordance with the convergence theorems given here, and in Chaps. 6 and 7.

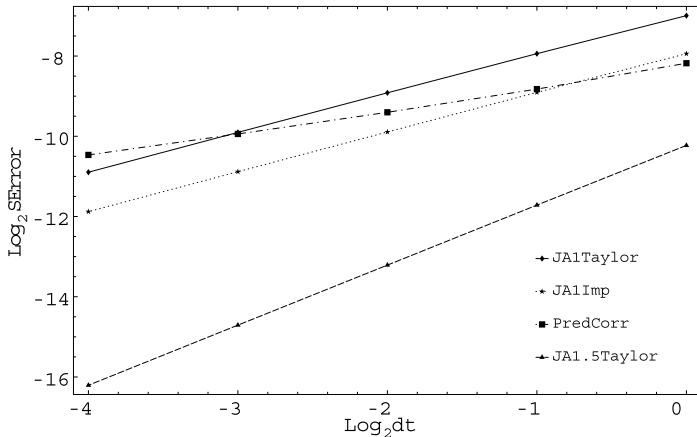


Fig. 8.8.2. Log-log plot of strong error versus time step size (constant marks)

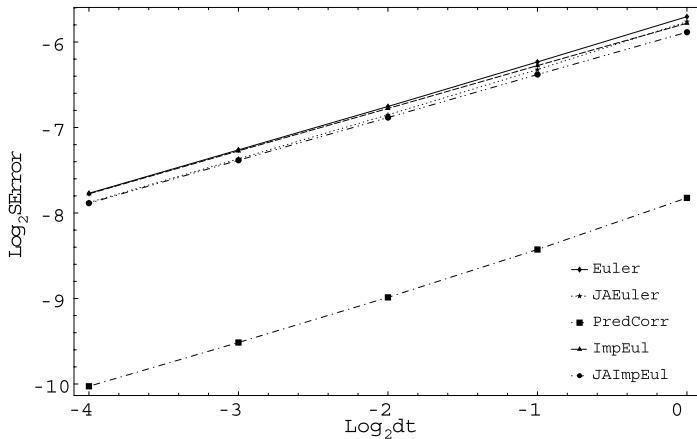


Fig. 8.8.3. Log-log plot of strong error versus time step size (lognormal marks)

The jump-adapted drift-implicit strong order 1.0 scheme is more accurate for small step sizes. In this plot we notice for the selected time step sizes that the accuracy of the regular predictor-corrector Euler scheme is of a magnitude equal to that of some first order schemes. Of course, since its strong order of convergence equals $\gamma = 0.5$, for smaller time step sizes it becomes less accurate than first order schemes. Finally, the most accurate scheme for all time step sizes considered is the jump-adapted strong order 1.5 Taylor scheme, which achieves an order of strong convergence of approximately $\gamma = 1.5$.

Let us now consider the case of lognormally distributed marks. Here the logarithm of mark $\zeta_i = \ln(\xi_i)$ is an independent Gaussian random variable, $\zeta_i \sim N(\varrho, \varsigma)$, with mean $\varrho = -0.1738$ and standard deviation $\sqrt{\varsigma} = 0.15$. These parameters imply that the expected value of the marks equals $E(\xi) = 0.85$.

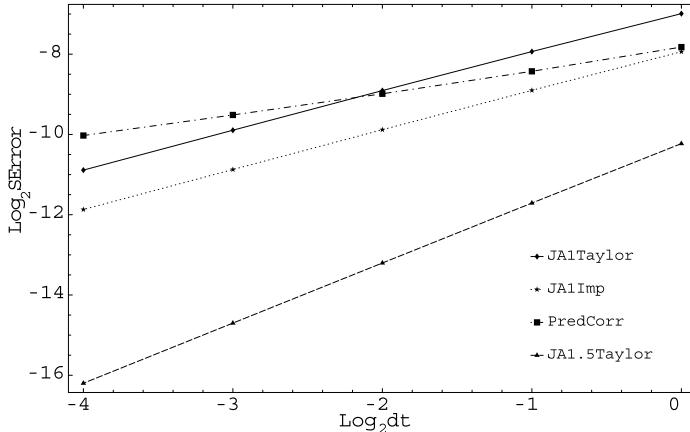


Fig. 8.8.4. Log-log plot of strong error versus time step size (lognormal marks)

In Fig. 8.8.3, we plot the results obtained from the regular and jump-adapted, regular and jump-adapted drift-implicit, and the regular predictor-corrector Euler schemes. Also in this case the results for the jump-adapted predictor-corrector Euler scheme are indistinguishable from those of the jump-adapted Euler scheme and, thus, omitted. We remark that these results are very similar to those obtained in Fig. 8.8.1 for the case of constant marks, confirming that the orders of strong convergence derived in the previous chapters hold also for the case of random marks. Again all schemes considered achieve a strong order of about $\gamma = 0.5$. Moreover, the regular predictor-corrector Euler scheme is the most accurate one. The remaining schemes have similar accuracy, with the regular Euler scheme the least accurate and the jump-adapted drift-implicit Euler scheme the most accurate. In Fig. 8.8.4 we report the results for the regular predictor-corrector Euler, jump-adapted strong order 1.0 Taylor, jump-adapted drift-implicit strong order 1.0 and the jump-adapted strong order 1.5 Taylor schemes. The results are again very similar to those obtained for the case of constant marks, reported in Fig. 8.8.2, with all schemes achieving the prescribed theoretical orders of strong convergence.

The Case of High Intensities

Let us now consider the strong errors generated by the strong schemes analyzed in the previous subsection, when using the larger intensity $\lambda = 2$. The remaining parameters of the SDE (8.8.1) are as in the previous subsection.

In Fig. 8.8.5 we show the results for the regular and jump-adapted, regular and jump-adapted predictor-corrector and jump-adapted drift-implicit Euler schemes. Note that the error generated by the regular drift-implicit scheme is very similar to that of the regular Euler scheme and, thus, is here omitted.

All schemes achieve their theoretical orders of strong convergence of approximately $\gamma = 0.5$. Here we can clearly see that jump-adapted schemes are

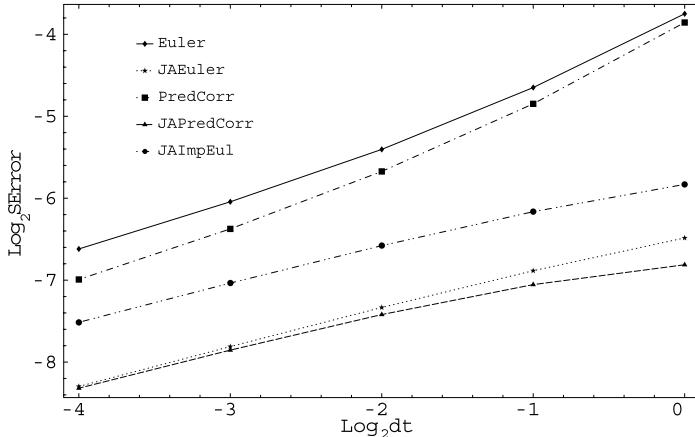


Fig. 8.8.5. Log-log plot of strong error versus time step size (constant marks)

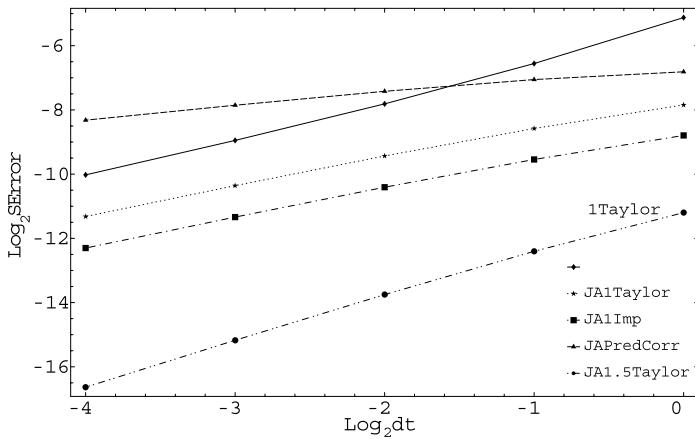


Fig. 8.8.6. Log-log plot of strong error versus time step size (constant marks)

more accurate than their regular counterparts. This is due to the simulation of the jump impact at the correct jump times in the jump-adapted scheme. Moreover, we report that of the jump-adapted schemes, the drift-implicit scheme is the least accurate, while the predictor-corrector scheme is the most accurate.

In Fig. 8.8.6 we show the results for the strong order 1.0 Taylor, jump-adapted strong order 1.0 Taylor, jump-adapted drift-implicit strong order 1.0, jump-adapted predictor-corrector Euler and the jump-adapted strong order 1.5 Taylor schemes. Also here all schemes achieve roughly the theoretical expected order of strong convergence. Note that while in the low intensity case the accuracy of the regular and jump-adapted versions of the strong order 1.0 Taylor scheme were very similar, in the high intensity case the jump-adapted versions are more accurate, and highlight that the jump-adapted strong order 1.5 Taylor scheme is the most accurate for all time step sizes considered.

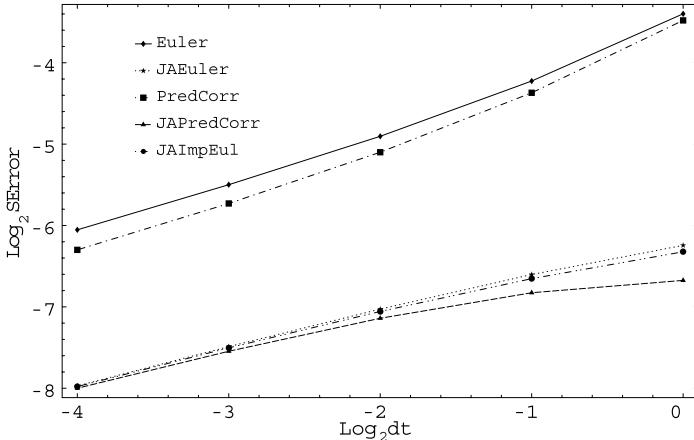


Fig. 8.8.7. Log-log plot of strong error versus time step size (lognormal marks)

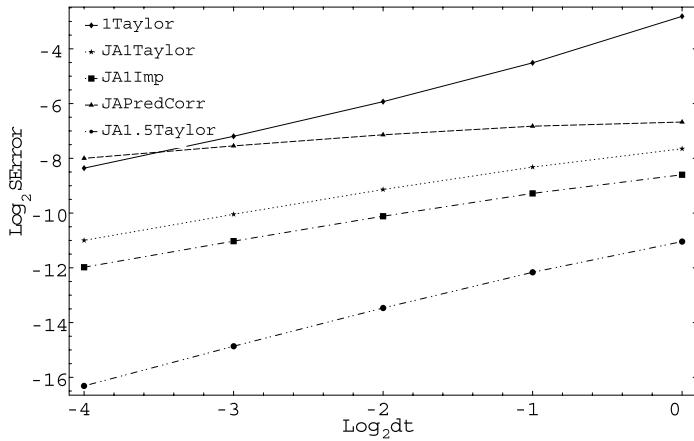


Fig. 8.8.8. Log-log plot of strong error versus time step size (lognormal marks)

Finally, in Figs. 8.8.7 and 8.8.8 we report the behavior of strong errors for all schemes analyzed previously in this chapter for the case of lognormal marks. The results are again very similar to those obtained in the case with constant marks. In particular, we report that all schemes achieve approximately their theoretical order of strong convergence.

In Fig. 8.8.9 we report the results obtained from the implicit Euler, implicit strong order 1.0 Taylor, jump-adapted implicit Euler and jump-adapted implicit order 1.0 Taylor schemes with $\theta = 1$. Also in this case the strong orders of convergence are consistent with those predicted by the convergence theorems. We notice again the higher accuracy of jump-adapted schemes due to the better simulation of the jump component.

We now consider the mark-dependent jump coefficient $c(t, x, v) = x(v-1)$, with marks drawn from a lognormal distribution with mean 1.1, and standard

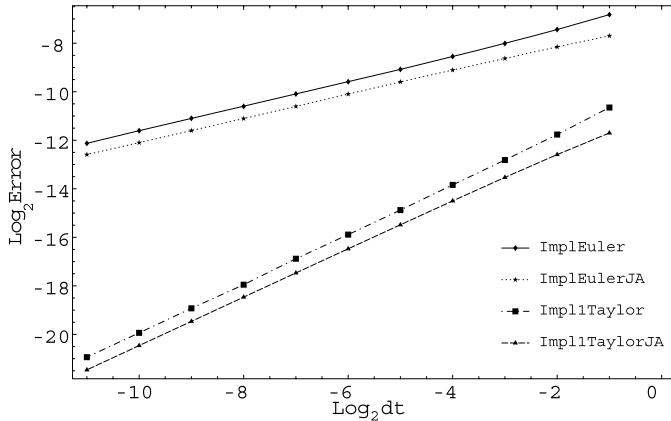


Fig. 8.8.9. Log-log plot of strong error versus time step size

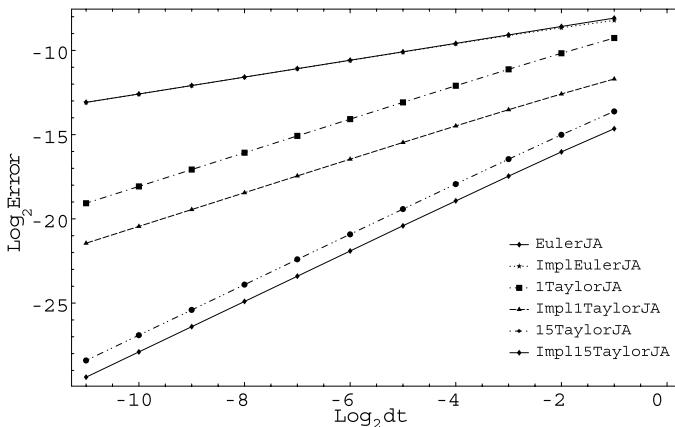


Fig. 8.8.10. Log-log plot of strong error versus time step size

deviation 0.02. As explained in Sect. 8.1, jump diffusion SDEs with a mark-dependent jump size can be handled efficiently by employing jump-adapted schemes. Therefore, in Fig. 8.8.10 we compare the following jump-adapted schemes: Euler, implicit Euler, strong order 1.0 Taylor, implicit strong order 1.0 Taylor, strong order 1.5 Taylor and implicit strong order 1.5 Taylor. For the implicit schemes the degree of implicitness θ , equals one. Again, the orders of strong convergence obtained from our numerical experiments are those predicted by the theory. Comparing explicit with implicit schemes, we report that for this choice of parameters the implicit schemes are more accurate. Since the jump impact is simulated without creating any additional error, these differences are due to the approximation of the diffusive part. We remark that implicit schemes, which offer wider regions of stability as we will see in Chap. 14, are more suitable for problems in which numerical stability is an

issue. This applies also for the areas of filtering and scenario simulation in finance, where SDEs with multiplicative noise naturally arise.

8.9 Approximation of Pure Jump Processes

In the context of modeling with continuous time Markov chains and related processes, pure jump SDEs naturally arise. In this section we consider strong discrete-time approximations of solutions of pure jump SDEs. The schemes to be presented are special cases of those considered here, and in Chaps. 6 and 7 when the drift and diffusion coefficients equal zero. The particular nature of the pure jump dynamics simplifies the implementation of the corresponding higher order strong schemes. Additionally, as we will see at the end of this section, strong orders of convergence are derived under weaker assumptions than those needed in the jump diffusion case. Most of the results in this chapter have been derived in [Bruti-Liberati & Platen \(2007b\)](#), and apply generally for a wide class of pure jump processes that do not need to be Markovian.

Pure Jump Process Approximations

We now present strong numerical approximations of solutions of pure jump SDEs. Such SDEs arise, for instance, when using a birth and death process or, more generally, a continuous time Markov chain. They play an important role in modeling credit ratings, bio-chemical reactions and piecewise constant quantities in other areas of application, see for instance [Turner et al. \(2004\)](#). The piecewise constant nature of pure jump dynamics simplifies the resulting numerical schemes. For instance, jump-adapted approximations, constructed on time discretizations including all jump times, produce no discretization error in this case. Therefore, in the case of low to medium jump intensities one can construct efficient schemes without discretization error. In the case of high intensity jump processes, jump-adapted schemes are often not feasible. However, we will demonstrate that one can derive higher order discrete-time approximations whose complexities turn out to be significantly less than those of numerical approximations of jump diffusions. In the case of SDEs driven by a Poisson process, the generation of the required multiple stochastic integrals for higher order approximations is straightforward and involves only one Poisson distributed random variable in each time step. Moreover, the simple structure of pure jump SDEs permits us to illustrate the use of a stochastic expansion in the derivation of higher order approximations. At the end of the section we will show that higher orders of strong convergence of discrete-time approximations for SDEs, driven purely by a Poisson random measure, can be derived under weaker conditions than those typically required for jump diffusions.

Pure Jump Model

Let us consider a counting process $N = \{N_t, t \in [0, T]\}$, which is right-continuous with left-hand limits and counts the arrival of certain events. Most of the following analysis applies for general counting processes. However, for illustration, we take N to be a *Poisson process* with constant intensity $\lambda \in (0, \infty)$ that starts at time $t = 0$ in $N_0 = 0$. It is defined on a filtered probability space $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ with $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0, T]}$ satisfying the usual conditions. Note that we can rewrite the Poisson process in terms of the Poisson random measure, introduced in Chap. 1, by setting

$$N_t = p_\varphi(\mathcal{E}, [0, t]), \quad (8.9.1)$$

for $t \in [0, T]$. An alternative representation of the Poisson process is provided by a Poisson random measure with mark space $\mathcal{E} = \{1\}$ and intensity measure $\varphi(\{1\}) = \lambda$. The Poisson process $N = \{N_t, t \in [0, T]\}$ generates an increasing sequence $(\tau_i)_{i \in \{1, 2, \dots, N_T\}}$ of jump times. Recall that for any right-continuous process $Z = \{Z_t, t \in [0, T]\}$ we define its *jump size* ΔZ_t at time t as the difference

$$\Delta Z_t = Z_t - Z_{t-}, \quad (8.9.2)$$

for $t \in [0, T]$, where Z_{t-} denotes again the left-hand limit of Z at time t . Thus, we can write

$$N_t = \sum_{s \in (0, t]} \Delta N_s$$

for $t \in [0, T]$.

For a pure jump process $X = \{X_t, t \in [0, T]\}$ that is driven by the Poisson process N we assume that its value X_t at time t satisfies the SDE

$$dX_t = c(t, X_{t-}) dN_t \quad (8.9.3)$$

for $t \in [0, T]$ with deterministic initial value $X_0 \in \mathfrak{R}$. This is a special case of the SDE (1.8.2), where the drift coefficient a and the diffusion coefficient b both equal zero and the jump coefficient c is mark-independent.

The *jump coefficient* $c : [0, T] \times \mathfrak{R} \rightarrow \mathfrak{R}$ is again assumed to be Borel measurable, Lipschitz continuous, such that

$$|c(t, x) - c(t, y)| \leq K |x - y|,$$

and to satisfy the growth condition

$$|c(t, x)|^2 \leq K (1 + |x|^2)$$

for $t \in [0, T]$ and $x, y \in \mathfrak{R}$ with some constant $K \in (0, \infty)$. These conditions can, in principle, be relaxed considerably and are here imposed for simpler presentation.

To provide for later illustration a simple, yet still interesting example, let us consider the linear SDE

$$dX_t = X_{t-} \psi dN_t \quad (8.9.4)$$

for $t \in [0, T]$ with $X_0 > 0$ and constant $\psi \in \mathbb{R}$. This is a degenerate case of the SDE (1.8.5), with drift coefficient $a(t, x) = 0$, diffusion coefficient $b(t, x) = 0$ and mark-independent jump coefficient $c(t, x) = x\psi$. By application of the Itô formula one can demonstrate that the solution $X = \{X_t, t \in [0, T]\}$ of the SDE (8.9.4) is a pure jump process with explicit representation

$$X_t = X_0 \exp\{N_t \ln(\psi + 1)\} = X_0 (\psi + 1)^{N_t} \quad (8.9.5)$$

for $t \in [0, T]$.

Jump-Adapted Schemes

We consider a *jump-adapted time discretization* $0 = t_0 < t_1 < \dots < t_{n_T} = T$, where n_T is as defined in (6.1.11) and the sequence $t_1 < \dots < t_{n_T-1}$ equals that of the jump times $\tau_1 < \dots < \tau_{n_T}$ of the Poisson process N . On this jump-adapted time grid we construct the *jump-adapted Euler scheme* by the algorithm

$$Y_{n+1} = Y_n + c \Delta N_n, \quad (8.9.6)$$

for $n \in \{0, 1, \dots, n_T - 1\}$, with initial value $Y_0 = X_0$, where $\Delta N_n = N_{t_{n+1}} - N_{t_n}$ is the n th increment of the Poisson process N . Between discretization times the right-continuous process Y is set to be piecewise constant. Note that here and in the sequel, when no misunderstanding is possible, we use the previously introduced abbreviation $c = c(t_n, Y_n)$.

Since the discretization points are constructed exactly at the jump times of N , and the simulation of the increments $N_{t_{i+1}} - N_{t_i} = 1$ is exact, the jump-adapted Euler scheme (8.9.6) produces *no discretization error*. Let us emphasize that this is a particularly attractive feature of jump-adapted schemes, when applied to pure jump SDEs. In the case of jump diffusion SDEs, the jump-adapted schemes typically produce a discretization error.

For the implementation of the scheme (8.9.6) one needs to compute the jump times τ_i , $i \in \{1, 2, \dots, n_T\}$, and then apply equation (8.9.6) recursively for every $i \in \{0, 1, 2, \dots, n_T - 1\}$. One can obtain the jump times via the corresponding waiting times between two consecutive jumps by sampling from an exponential distribution with parameter λ .

The computational demand of the algorithm (8.9.6) is heavily dependent on the intensity λ of the jump process. Indeed, the average number of steps, therefore the number of operations, is proportional to the intensity λ . Below we will introduce alternative methods suitable for large intensities, based on regular time discretizations.

Euler Scheme

In this subsection we develop discrete-time strong approximations whose computational complexity is independent of the jump intensity level. We consider an equidistant time discretization with time step size $\Delta \in (0, 1)$ as in Chap. 6. The simplest strong Taylor approximation $Y = \{Y_t, t \in [0, T]\}$ is the *Euler scheme*, given by

$$Y_{n+1} = Y_n + c \Delta N_n \quad (8.9.7)$$

for $n \in \{0, 1, \dots, n_T - 1\}$ with initial value $Y_0 = X_0$ and $\Delta N_n = N_{t_{n+1}} - N_{t_n}$. Between discretization times the right-continuous process Y is assumed to be piecewise constant.

By comparing the scheme (8.9.7) with the algorithm (8.9.6), we notice that the difference in the schemes occurs in the time discretization. We emphasize that the average number of operations and, thus, the computational complexity of the Euler scheme (8.9.7) is independent of the jump intensity. Therefore, a simulation based on the Euler scheme (8.9.7) is also feasible in the case of jump processes with high intensity. However, while the jump-adapted Euler scheme (8.9.6) produces no discretization error, the accuracy of the Euler scheme (8.9.7) depends on the size of the time step Δ and the nature of the jump coefficient.

For example, for the linear SDE (8.9.4), the Euler scheme (8.9.7) is of the form

$$Y_{n+1} = Y_n + Y_n \psi \Delta N_n = Y_n (1 + \psi \Delta N_n) \quad (8.9.8)$$

for $n \in \{0, 1, \dots, n_T - 1\}$ with $Y_0 = X_0$. Since the equidistant time discretization $t_n = \Delta n$ of this Euler scheme does not include the jump times of the underlying Poisson process, we have an approximation error.

Theorem 6.4.1 shows that the Euler approximation (8.9.7) achieves strong order of convergence $\gamma = 0.5$. This raises the question of how to construct higher order discrete-time approximations for the case of pure jump SDEs. The problem can be approached by a stochastic expansion for pure jump SDEs that we describe below. This expansion is a particular case of the Wagner-Platen expansion (4.4.4) for jump diffusions presented in Chap. 4. Therefore, the resulting strong approximations are particular cases of the strong schemes presented in Chap. 6.

Wagner-Platen Expansion

Since the use of the Wagner-Platen expansion for pure jump processes, see Sect. 4.1, may be unfamiliar to some readers, let us first illustrate the structure of this formula for a simple example. For any measurable function $f : \mathfrak{R} \rightarrow \mathfrak{R}$ and a given adapted counting process $N = \{N_t, t \in [0, T]\}$ we have the representation

$$f(N_t) = f(N_0) + \sum_{s \in (0, t]} \Delta f(N_s) \quad (8.9.9)$$

for all $t \in [0, T]$, where $\Delta f(N_s) = f(N_s) - f(N_{s-})$. We can formally write the equation (8.9.9) in the differential form of an SDE

$$df(N_t) = \Delta f(N_t) = (f(N_{t-} + 1) - f(N_{t-})) \Delta N_t$$

for $t \in [0, T]$. This equation can also be obtained from the Itô formula for semimartingales with jumps, see Sect. 1.5 or Protter (2005).

Obviously, the following difference expression $\tilde{\Delta}f(N_{s-})$ defines a measurable function, as long as,

$$\tilde{\Delta}f(N) = f(N + 1) - f(N) \quad (8.9.10)$$

is a measurable function of N . As in Sect. 4.1, by using this function we rewrite (8.9.9) in the form

$$f(N_t) = f(N_0) + \int_{(0,t]} \tilde{\Delta}f(N_{s-}) dN_s \quad (8.9.11)$$

for $t \in [0, T]$ and apply the formula (8.9.11) to $\tilde{\Delta}f(N_{s-})$ in (8.9.11), which yields

$$\begin{aligned} f(N_t) &= f(N_0) + \int_{(0,t]} \tilde{\Delta}f(N_0) dN_s + \int_{(0,t]} \int_{(0,s_2)} (\tilde{\Delta})^2 f(N_{s_1-}) dN_{s_1} dN_{s_2} \\ &= f(N_0) + \tilde{\Delta}f(N_0) \int_{(0,t]} dN_s + \int_{(0,t]} \int_{(0,s_2)} (\tilde{\Delta})^2 f(N_{s_1-}) dN_{s_1} dN_{s_2} \end{aligned} \quad (8.9.12)$$

for $t \in [0, T]$. Here $(\tilde{\Delta})^q$ denotes for an integer $q \in \{1, 2, \dots\}$ the q times consecutive application of the function $\tilde{\Delta}$ given in (8.9.10). Note that a double stochastic integral with respect to the counting process N naturally arises in (8.9.12). As such, one can now in (8.9.12) apply the formula (8.9.11) to the measurable function $(\tilde{\Delta})^2 f(N_{s_1-})$, which yields

$$f(N_t) = f(N_0) + \tilde{\Delta}f(N_0) \int_{(0,t]} dN_s + (\tilde{\Delta})^2 f(N_0) \int_{(0,t]} \int_{(0,s_2)} dN_{s_1} dN_{s_2} + \bar{R}_3(t) \quad (8.9.13)$$

with remainder term

$$\bar{R}_3(t) = \int_{(0,t]} \int_{(0,s_3)} \int_{(0,s_2)} (\tilde{\Delta})^3 f(N_{s_1-}) dN_{s_1} dN_{s_2} dN_{s_3}$$

for $t \in [0, T]$. In (8.9.13) we have obtained a double integral in the expansion part. Furthermore, we have a triple integral in the remainder term. We call (8.9.13) a Wagner-Platen expansion of the function $f(\cdot)$ with respect to the counting process N , see Sect. 4.1. The expansion part only depends on multiple stochastic integrals with respect to the counting process N . These multiple stochastic integrals are weighted by some constant coefficient functions with

values taken at the expansion point N_0 . It is obvious how to proceed to obtain higher order Wagner-Platen expansions by iterative application of formula (8.9.11).

Fortunately, the multiple stochastic integrals that arise can be easily computed. With (4.1.23) we can rewrite the Wagner-Platen expansion (8.9.13) in the form

$$f(N_t) = f(N_0) + \tilde{\Delta} f(N_0) \binom{N_t}{1} + (\tilde{\Delta})^2 f(N_0) \binom{N_t}{2} + \bar{R}_3(t),$$

where

$$\tilde{\Delta} f(N_0) = \tilde{\Delta} f(0) = f(1) - f(0),$$

$$(\tilde{\Delta})^2 f(N_0) = f(2) - 2f(1) + f(0).$$

In the given case this leads to the expansion (4.1.25). More generally, by induction it follows the Wagner-Platen expansion of the form

$$f(N_t) = \sum_{k=0}^l (\tilde{\Delta})^k f(N_0) \binom{N_t}{k} + \bar{R}_{l+1}(t) \quad (8.9.14)$$

with

$$\bar{R}_{l+1}(t) = \int_{(0,t]} \dots \int_{(0,s_2)} (\tilde{\Delta})^{l+1} f(N_{s_1-}) dN_{s_1} \dots dN_{s_{l+1}}$$

for $t \in [0, T]$ and $l \in \{0, 1, \dots\}$, where $(\tilde{\Delta})^0 f(N_0) = f(N_0)$. By neglecting the remainder term in (8.9.14) one neglects the potential occurrence of a higher number of jumps than l and obtains a useful truncated approximation of a measurable function f with respect to a counting process N .

Note that in (8.9.14) the truncated expansion is *exact* if no more than l jumps occur until time t in the realization of N . Consequently, if there is only a small probability that more than l jumps occur over the given time period, then the truncated Wagner-Platen expansion can be expected to be highly accurate under any reasonable criterion.

Expansion for Functions

Similar to (8.9.14) let us now derive a Wagner-Platen expansion for functions of solutions of the general pure jump SDE (8.9.3). We define similarly as above the measurable function $\tilde{\Delta}f(\cdot)$ such that

$$\tilde{\Delta} f(X_{t-}) = \Delta f(X_t) = f(X_t) - f(X_{t-}) \quad (8.9.15)$$

for all $t \in [0, T]$. In the same manner as previously shown, this leads to the expansion

$$\begin{aligned}
f(X_t) &= f(X_0) + \int_{(0,t]} \tilde{\Delta} f(X_{s-}) dN_s \\
&= f(X_0) + \int_{(0,t]} \left(\tilde{\Delta} f(X_0) + \int_{(0,s_2)} \left(\tilde{\Delta} \right)^2 f(X_{s_1-}) dN_{s_1} \right) dN_{s_2} \\
&= f(X_0) + \sum_{k=1}^l \left(\tilde{\Delta} \right)^k f(X_0) \int_{(0,t]} \cdots \int_{(0,s_2)} dN_{s_1} \cdots dN_{s_k} + \tilde{R}_{f,t}^{l+1} \\
&= f(X_0) + \sum_{k=1}^l \left(\tilde{\Delta} \right)^k f(X_0) \binom{N_t}{k} + \tilde{R}_{f,t}^{l+1} \tag{8.9.16}
\end{aligned}$$

with

$$\tilde{R}_{f,t}^{l+1} = \int_{(0,t]} \cdots \int_{(0,s_2)} \left(\tilde{\Delta} \right)^{l+1} f(X_{s_1-}) dN_{s_1} \cdots dN_{s_{l+1}}$$

for $t \in [0, T]$ and $l \in \{1, 2, \dots\}$. One notes that (8.9.16) generalizes (8.9.14) in an obvious fashion.

For the particular example of the linear SDE (8.9.4) we obtain for any measurable function f the function

$$\tilde{\Delta} f(X_{\tau-}) = f(X_{\tau-}(1 + \psi)) - f(X_{\tau-})$$

for the jump times $\tau \in [0, T]$ with $\Delta N_\tau = 1$. Therefore, in the case $l = 2$ we get from (4.1.23) and (8.9.16) the expression

$$\begin{aligned}
f(X_t) &= f(X_0) + (f(X_0(1 + \psi)) - f(X_0)) (N_t - N_0) \\
&\quad + \left(f(X_0(1 + \psi)^2) - 2f(X_0(1 + \psi)) + f(X_0) \right) \\
&\quad \times \frac{1}{2} (N_t - N_0) ((N_t - N_0) - 1) + \tilde{R}_{f,t}^3
\end{aligned}$$

for $t \in [0, T]$. By neglecting the remainder term $\tilde{R}_{f,t}^3$ we obtain in a systematic way, for this simple example a truncated Wagner-Platen expansion of $f(X_t)$ at the expansion point X_0 . We emphasize that in the derivation of the expansion (8.9.16), only measurability of the function f and the coefficients $(\tilde{\Delta})^k f(\cdot)$, for $k \in \{1, \dots, l\}$ is required. This contrasts with the case of diffusion and jump diffusion SDEs where sufficient differentiability conditions are needed to obtain a Wagner-Platen expansion. It also indicates that one can obtain useful approximations for a general counting process that may not even be Markovian. This is a remarkable fact since on the market micro structure level all observed quantities are, in principle, pure jump processes.

Strong Order 1.0 Taylor Scheme

The Euler scheme (8.9.7) can be interpreted as being derived from the expansion (8.9.16) applied to each time step by setting $f(x) = x$, choosing $l = 1$

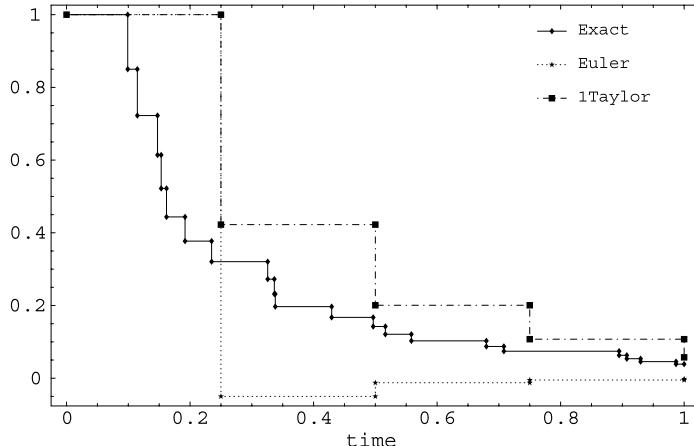


Fig. 8.9.1. Exact solution, Euler and order 1.0 Taylor approximations

and neglecting the remainder term. By choosing $l = 2$ in the corresponding truncated Wagner-Platen expansion, when applied to each time discretization interval $[t_n, t_{n+1}]$ with $f(x) = x$, we obtain the *strong order 1.0 Taylor approximation*

$$Y_{n+1} = Y_n + c \Delta N_n + (c(t_n, Y_n + c) - c) \frac{1}{2} (\Delta N_n) (\Delta N_n - 1) \quad (8.9.17)$$

for $n \in \{0, 1, \dots, n_T - 1\}$ with $Y_0 = X_0$ and $\Delta N_n = N_{t_{n+1}} - N_{t_n}$.

In the special case of the linear example (8.9.4), the strong order 1.0 Taylor approximation turns out to be of the form

$$Y_{n+1} = Y_n \left\{ 1 + \psi \Delta N_n + \frac{\psi^2}{2} \Delta N_n (\Delta N_n - 1) \right\} \quad (8.9.18)$$

for $n \in \{0, 1, \dots, n_T - 1\}$ with $Y_0 = X_0$.

For the linear SDE (8.9.4) and a given sample path of the counting process, here a Poisson process, we plot in Fig. 8.9.1 the exact solution (8.9.5), the Euler approximation (8.9.8) and the strong order 1.0 Taylor approximation (8.9.18). We selected a time step size $\Delta = 0.25$ and the following parameters: $X_0 = 1$, $T = 1$, $\psi = -0.15$ and $\lambda = 20$. Note in Fig. 8.9.1 that the strong order 1.0 Taylor approximation is at the terminal time $t = 1$ rather close to the exact solution. Visually it appears to better than the Euler approximation, which even becomes negative. Theorem 6.4.1, presented in Chap. 6, and Theorem 8.9.4, to be presented below, provide a firm basis for judging the convergence of such higher order schemes.

Strong Order 1.5 and 2.0 Taylor Schemes

If we use the truncated Wagner-Platen expansion (8.9.16) with $l = 3$, when applied to each time interval $[t_n, t_{n+1}]$ with $f(x) = x$, we obtain the *strong*

order 1.5 Taylor approximation

$$\begin{aligned} Y_{n+1} &= Y_n + c \Delta N_n + \left\{ c(t_n, Y_n + c) - c \right\} \binom{\Delta N_n}{2} \\ &\quad + \left\{ c(t_n, Y_n + c + c(t_n, Y_n + c)) - 2c(t_n, Y_n + c) + c \right\} \binom{\Delta N_n}{3} \end{aligned} \quad (8.9.19)$$

for $n \in \{0, 1, \dots, n_T - 1\}$ with $Y_0 = X_0$.

In the case of our particular example (8.9.4), the strong order 1.5 Taylor approximation is of the form

$$Y_{n+1} = Y_n \left\{ 1 + \psi \Delta N_n + \psi^2 \binom{\Delta N_n}{2} + \psi^3 \binom{\Delta N_n}{3} \right\}$$

for $n \in \{0, 1, \dots, n_T - 1\}$ with $Y_0 = X_0$.

To construct an approximation with second order strong convergence we need to choose $l = 4$ in the truncated expansion (8.9.16) with $f(x) = x$. In doing so we obtain the *strong order 2.0 Taylor approximation*

$$\begin{aligned} Y_{n+1} &= Y_n + c \Delta N_n + \left\{ c(Y_n + c(Y_n)) - c(Y_n) \right\} \binom{\Delta N_n}{2} \\ &\quad + \left\{ c(t_n, Y_n + c + c(t_n, Y_n + c)) - 2c(t_n, Y_n + c) + c \right\} \binom{\Delta N_n}{3} \\ &\quad + \left\{ c(t_n, Y_n + c + c(t_n, Y_n + c) + c(t_n, Y_n + c + c(t_n, Y_n + c))) \right. \\ &\quad \left. - 3c(t_n, Y_n + c + c(t_n, Y_n + c)) + 3c(t_n, Y_n + c) - c \right\} \binom{\Delta N_n}{4} \end{aligned} \quad (8.9.20)$$

for $n \in \{0, 1, \dots, n_T - 1\}$ with $Y_0 = X_0$.

For the linear SDE (8.9.4) the strong order 2.0 Taylor approximation is of the form

$$Y_{n+1} = Y_n \left\{ 1 + \psi \Delta N_n + \psi^2 \binom{\Delta N_n}{2} + \psi^3 \binom{\Delta N_n}{3} + \psi^4 \binom{\Delta N_n}{4} \right\}$$

for $n \in \{0, 1, \dots, n_T - 1\}$ with $Y_0 = X_0$. This is a rather convenient algorithm, which allows us to guess other higher order schemes for this SDE.

Convergence Results

It is desirable to be able to construct systematically, highly accurate discrete-time approximations for solutions of pure jump SDEs. For this purpose we use the Wagner-Platen expansion (8.9.16) to obtain the strong order γ Taylor scheme for pure jump processes, for $\gamma \in \{0.5, 1, 1.5, \dots\}$.

In this subsection we consider a pure jump process described by a more general SDE than the SDE (8.9.3) considered so far in this chapter. For the

pure jump SDE (8.9.3) driven by a single Poisson process, it is possible, as shown above, to derive higher strong order schemes that involve only one Poisson random variable in each time step. However, it is important to study also more general multi-dimensional pure jump processes, which allow the modeling of more complex quantities as, for instance, state-dependent intensities of defaults. For this reason, we consider the d -dimensional pure jump SDE

$$d\mathbf{X}_t = \int_{\mathcal{E}} \mathbf{c}(t, \mathbf{X}_{t-}, v) p_\varphi(dv, dt), \quad (8.9.21)$$

for $t \in [0, T]$, with $\mathbf{X}_0 \in \Re^d$. Here the jump coefficient \mathbf{c} and the Poisson random measure p_φ as defined in (1.8.2). Note that the mark space \mathcal{E} of the Poisson random measure can be made multi-dimensional or split into disjoint subsets and, thus, can generate several sources for modeling jumps. The case of a multi-dimensional SDE driven by several Poisson processes is a specific case of the SDE (8.9.21). The SDE (8.9.21) is equivalent to the jump diffusion SDE (1.8.2) when the drift coefficient a and the diffusion coefficient b equal zero. Note that birth and death processes and, more generally, continuous time Markov chains can be described by the SDE (8.9.21). It is sometimes not the most convenient or usual description of such dynamics. However, it is a very generic one as we can see.

Theorem 6.4.1, presented in Chap. 6, establishes the strong order of convergence of strong Taylor approximations for jump diffusion SDEs. When specifying the theorem mentioned to the case of SDEs driven by pure jump processes, it turns out that it is possible to weaken the assumptions on the coefficients of the Wagner-Platen expansion. As we will see below, the Lipschitz and growth conditions on the jump coefficient are sufficient enough to establish the convergence of strong Taylor schemes of any given strong order of convergence $\gamma \in \{0.5, 1, 1.5, 2, \dots\}$. Differentiability of the jump coefficient is not required. This is due to the structure of the increment operator L_v^{-1} , see (4.3.6), naturally appearing in the coefficient of the Wagner-Platen expansion for pure jump processes.

For a regular time discretization $(t)_\Delta$ with maximum step size $\Delta \in (0, 1)$ we define, the *strong order γ Taylor scheme* for pure jump SDEs by

$$\begin{aligned} \mathbf{Y}_{n+1}^\Delta &= \mathbf{Y}_n^\Delta + \sum_{k=0}^{2\gamma-1} \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \dots \int_{t_n}^{s_1} \int_{\mathcal{E}} (L^{-1})^k \mathbf{c}(t_n, \mathbf{Y}_n^\Delta, v^0) \\ &\quad \times p_\varphi(dv^0, ds^0) \dots p_\varphi(dv^k, ds^k), \end{aligned} \quad (8.9.22)$$

for $n \in \{0, 1, \dots, n_T - 1\}$ and $\gamma \in \{0.5, 1, 1.5, \dots\}$. Here we have used the notation

$$(L^{-1})^k \mathbf{c}(t_n, \mathbf{Y}_n^\Delta, v^0) = \begin{cases} \mathbf{c}(t_n, \mathbf{Y}_n^\Delta, v^0) & \text{when } k = 0 \\ L_{v^1}^{-1} \mathbf{c}(t_n, \mathbf{Y}_n^\Delta, v^0) & \text{when } k = 1 \\ L_{v^k}^{-1} \left(\dots (L_{v^1}^{-1} \mathbf{c}(t_n, \mathbf{Y}_n^\Delta, v^0)) \right) & \text{when } k \in \{1, \dots\}, \end{cases} \quad (8.9.23)$$

where the increment operator L^{-1} is defined in (4.3.6).

The following three results show that for SDEs driven by pure jump processes, Lipschitz and growth conditions are sufficient for Theorem 6.4.1 to guarantee the corresponding order of strong convergence.

Lemma 8.9.1 *Assume that the jump coefficient satisfies the Lipschitz condition*

$$|\mathbf{c}(t, \mathbf{x}, u) - \mathbf{c}(t, \mathbf{y}, u)| \leq K |\mathbf{x} - \mathbf{y}|, \quad (8.9.24)$$

for $t \in [0, T]$, $\mathbf{x}, \mathbf{y} \in \Re^d$ and $u \in \mathcal{E}$, with some constant $K \in (0, \infty)$. Then for any $\gamma \in \{0.5, 1, 1.5, \dots\}$ and $k \in \{0, 1, 2, \dots, 2\gamma - 1\}$ the k th coefficient $(L^{-1})^k \mathbf{c}(t, \mathbf{x}, u)$ of the strong order γ Taylor scheme, satisfies the Lipschitz condition

$$\left| (L^{-1})^k \mathbf{c}(t, \mathbf{x}, u) - (L^{-1})^k \mathbf{c}(t, \mathbf{y}, u) \right| \leq C_k |\mathbf{x} - \mathbf{y}|, \quad (8.9.25)$$

for $t \in [0, T]$, $\mathbf{x}, \mathbf{y} \in \Re^d$, $u \in \mathcal{E}^k$ and some constant $C_k \in (0, \infty)$ which only depends on k .

Proof: We prove the assertion (8.9.25) by induction with respect to k . For $k = 0$, by the Lipschitz condition (8.9.24) we obtain

$$\left| (L^{-1})^0 \mathbf{c}(t, \mathbf{x}, u) - (L^{-1})^0 \mathbf{c}(t, \mathbf{y}, u) \right| = \left| \mathbf{c}(t, \mathbf{x}, u) - \mathbf{c}(t, \mathbf{y}, u) \right| \leq K |\mathbf{x} - \mathbf{y}|.$$

For $k = l + 1$, by the induction hypothesis, Jensen's inequality, and the Lipschitz condition (8.9.24) we obtain

$$\begin{aligned} & \left| (L^{-1})^{l+1} \mathbf{c}(t, \mathbf{x}, u) - (L^{-1})^{l+1} \mathbf{c}(t, \mathbf{y}, u) \right| \\ &= \left| (L^{-1})^l \mathbf{c}(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x}, v), u) - (L^{-1})^l \mathbf{c}(t, \mathbf{x}, u) \right. \\ &\quad \left. - (L^{-1})^l \mathbf{c}(t, \mathbf{y} + \mathbf{c}(t, \mathbf{y}, v), u) + (L^{-1})^l \mathbf{c}(t, \mathbf{y}, u) \right| \\ &\leq C_l \left| \mathbf{x} - \mathbf{y} + (\mathbf{c}(t, \mathbf{x}, v) - \mathbf{c}(t, \mathbf{y}, v)) \right| + C_l |\mathbf{x} - \mathbf{y}| \\ &\leq 2C_l |\mathbf{x} - \mathbf{y}| + C_l K |\mathbf{x} - \mathbf{y}| \\ &\leq C_{l+1} |\mathbf{x} - \mathbf{y}|, \end{aligned}$$

which completes the proof of Lemma 8.9.1. \square

Lemma 8.9.2 *Assume that the jump coefficient satisfies the growth condition*

$$|\mathbf{c}(t, \mathbf{x}, u)|^2 \leq \tilde{K} (1 + |\mathbf{x}|^2) \quad (8.9.26)$$

for $t \in [0, T]$ and $\mathbf{x} \in \Re^d$ and $u \in \mathcal{E}$, with some constant $\tilde{K} \in (0, \infty)$. Then for any $\gamma \in \{0.5, 1, 1.5, \dots\}$ and $k \in \{0, 1, 2, \dots, 2\gamma - 1\}$ the k th coefficient

$(L^{-1})^k \mathbf{c}(t, \mathbf{x}, \mathbf{u})$ of the strong order γ Taylor scheme, satisfies the growth condition

$$\left| (L^{-1})^k \mathbf{c}(t, \mathbf{x}, \mathbf{u}) \right|^2 \leq \tilde{C}_k (1 + |\mathbf{x}|^2) \quad (8.9.27)$$

for $t \in [0, T]$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $\mathbf{u} \in \mathcal{E}^k$ and some constant $\tilde{C}_k \in (0, \infty)$ which only depends on k .

Proof: We prove the assertion of Lemma 8.9.2 by induction with respect to k . For $k = 0$, by applying the growth condition (8.9.26) we obtain

$$\left| (L^{-1})^0 \mathbf{c}(t, \mathbf{x}, \mathbf{u}) \right|^2 = \left| \mathbf{c}(t, \mathbf{x}, \mathbf{u}) \right|^2 \leq \tilde{K} (1 + |\mathbf{x}|^2).$$

For $k = l+1$, by the induction hypotheses, Jensen's inequality and the growth condition (8.9.26) we obtain

$$\begin{aligned} \left| (L^{-1})^{l+1} \mathbf{c}(t, \mathbf{x}, \mathbf{u}) \right|^2 &= \left| (L^{-1})^l \mathbf{c}(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x}, v), \mathbf{u}) - (L^{-1})^l \mathbf{c}(t, \mathbf{x}, \mathbf{u}) \right|^2 \\ &\leq 2 \left(\tilde{C}_l (1 + |\mathbf{x} + \mathbf{c}(t, \mathbf{x}, v)|^2) + \tilde{C}_l (1 + |\mathbf{x}|^2) \right) \\ &\leq 2 \left(\tilde{C}_l (1 + 2(|\mathbf{x}|^2 + |\mathbf{c}(t, \mathbf{x}, v)|^2)) + \tilde{C}_l (1 + |\mathbf{x}|^2) \right) \\ &\leq \tilde{C}_{l+1} (1 + |\mathbf{x}|^2), \end{aligned}$$

which completes the proof of Lemma 8.9.2. \square

Lemma 8.9.3 Let us assume that

$$E(|\mathbf{X}_0|^2) < \infty \quad (8.9.28)$$

and the jump coefficient satisfies the Lipschitz condition

$$|\mathbf{c}(t, \mathbf{x}, u) - \mathbf{c}(t, \mathbf{y}, u)| \leq K_1 |\mathbf{x} - \mathbf{y}| \quad (8.9.29)$$

and the growth condition

$$|\mathbf{c}(t, \mathbf{x}, u)|^2 \leq K_2 (1 + |\mathbf{x}|^2) \quad (8.9.30)$$

for $t \in [0, T]$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, and $u \in \mathcal{E}$, with constants $K_1, K_2 \in (0, \infty)$. Then for any $\gamma \in \{0.5, 1, 1.5, \dots\}$ and $k \in \{0, 1, 2, \dots, 2\gamma - 1\}$ the k th coefficient $(L^{-1})^k \mathbf{c}(t, \mathbf{x}, u)$ of the strong order γ Taylor scheme satisfies the integrability condition

$$(L^{-1})^k \mathbf{c}(\cdot, \mathbf{x}, \cdot) \in \mathcal{H}_k,$$

for $\mathbf{x} \in \mathbb{R}^d$, where \mathcal{H}_k is the set of adapted stochastic process $\mathbf{g} = \{\mathbf{g}(t), t \in [0, T]\}$ such that

$$E \left(\int_0^T \int_{\mathcal{E}} \int_0^{s_k} \int_{\mathcal{E}} \cdots \int_0^{s_2} |g(s, v^1, \dots, v^k, \omega)|^2 \varphi(dv^1) ds_1 \dots \varphi(dv^k) ds_k \right) < \infty.$$

Proof: By Lemma 8.9.2 for any $\gamma \in \{0.5, 1, 1.5, \dots\}$ and $k \in \{0, 1, 2, \dots, 2\gamma - 1\}$ the k th coefficient $(L^{-1})^k \mathbf{c}(t, \mathbf{x}, \mathbf{u})$ of the strong order γ Taylor scheme satisfies the growth condition

$$\left| (L^{-1})^k \mathbf{c}(t, \mathbf{x}, \mathbf{u}) \right|^2 \leq \tilde{C}_k (1 + |\mathbf{x}|^2) \quad (8.9.31)$$

for $t \in [0, T]$, $\mathbf{x}, \mathbf{y} \in \Re^d$ and $\mathbf{u} \in \mathcal{E}^k$, with the constant $\tilde{C}_k \in (0, \infty)$. Therefore, for any $\gamma \in \{0.5, 1, 1.5, \dots\}$ and $k \in \{0, 1, 2, \dots, 2\gamma - 1\}$, by condition (8.9.31) and Fubini's theorem we obtain

$$\begin{aligned} & E \left(\int_0^T \int_{\mathcal{E}} \int_0^{s_k} \int_{\mathcal{E}} \cdots \int_0^{s_1} \int_{\mathcal{E}} \left| (L^{-1})^k \mathbf{c}(t, \mathbf{X}_{s_0}, u^0) \right|^2 \varphi(du^0) ds_0 \cdots \varphi(du^k) ds_k \right) \\ & \leq E \left(\int_0^T \int_{\mathcal{E}} \int_0^{s_k} \int_{\mathcal{E}} \cdots \int_0^{s_1} \int_{\mathcal{E}} \tilde{C}_k (1 + |\mathbf{X}_{s_0}|^2) \varphi(du^0) ds_0 \cdots \varphi(du^k) ds_k \right) \\ & \leq \tilde{C}_k \frac{(T\lambda)^k}{k!} + \tilde{C}_k \int_0^T \int_0^{s_k} \cdots \int_0^{s_1} E \left(\sup_{0 \leq z \leq T} |\mathbf{X}_z|^2 \right) ds_0 \cdots ds_k < \infty. \end{aligned}$$

The last passage holds, since conditions (8.9.28), (8.9.29) and (8.9.30) ensure that

$$E \left(\sup_{z \in [0, T]} |\mathbf{X}_z|^2 \right) < \infty,$$

see Theorem 1.9.3. This completes the proof of Lemma 8.9.3. \square

We emphasize that in the case of pure jump SDEs, unlike the more general case of jump diffusions, no extra differentiability of the jump coefficient c is required when deriving higher order approximations.

Theorem 8.9.4. *For given $\gamma \in \{0.5, 1, 1.5, 2, \dots\}$, let $\mathbf{Y}^\Delta = \{\mathbf{Y}_t^\Delta, t \in [0, T]\}$ be the strong order γ Taylor scheme (8.9.22) for the SDE (8.9.21) corresponding to a regular time discretization $(t)_\Delta$ with maximum step size $\Delta \in (0, 1)$. We assume the jump coefficient $\mathbf{c}(t, \mathbf{x}, v)$ satisfies both the Lipschitz condition (8.9.24) and the growth condition (8.9.26). Moreover, suppose that*

$$E(|\mathbf{X}_0|^2) < \infty \quad \text{and} \quad \sqrt{E(|\mathbf{X}_0 - \mathbf{Y}_0^\Delta|^2)} \leq K_1 \Delta^\gamma.$$

Then the estimate

$$\sqrt{E(\max_{0 \leq n \leq n_T} |\mathbf{X}_n - \mathbf{Y}_n^\Delta|^2)} \leq K \Delta^\gamma$$

holds, where the constant K does not depend on Δ .

Proof: The proof of Theorem 8.9.4 is a direct consequence of the convergence Theorem 6.4.1 for jump diffusions presented in Chap. 6. This must be the case, as by Lemma 8.9.1, 8.9.2 and 8.9.3, the coefficients of the strong order γ Taylor scheme (8.9.22) satisfy the conditions required by convergence Theorem 6.4.1. Note that the differentiability condition in Theorem 6.4.1, being $f_{-\alpha} \in \mathcal{C}^{1,2}$ for all $\alpha \in \mathcal{A}_\gamma \cup B(\mathcal{A}_\gamma)$, is not required for pure jump SDEs. This condition is used in the proof of the general convergence Theorem 6.4.1, only for the derivation of the Wagner-Platen expansion. In the pure jump case, as shown above, one needs only measurability of the jump coefficient c to derive the corresponding Wagner-Platen expansion. \square

Theorem 8.9.4 states that the strong order γ Taylor scheme for pure jump SDEs achieves a strong order of convergence equal to γ . In fact Theorem 8.9.4 states a strong convergence of order γ not only at the endpoint T , but uniformly over all time discretization points. Thus, by including enough terms from the Wagner-Platen expansion (8.9.16) we are able to construct schemes of any given strong order of convergence $\gamma \in \{0.5, 1, 1.5, \dots\}$. Note that Theorem 8.9.4 applies to solutions of multi-dimensional pure jump SDEs, which is an interesting fact.

For the mark-independent pure jump SDE (8.9.3) driven by a single Poisson process, the strong order γ Taylor scheme (8.9.22) reduces to

$$\mathbf{Y}_{n+1}^\Delta = \mathbf{Y}_n^\Delta + \sum_{k=1}^{2\gamma} (\tilde{\Delta})^k \mathbf{f}(\mathbf{Y}_n^\Delta) \binom{\Delta N_n}{k} \quad (8.9.32)$$

for $n \in \{0, 1, \dots, n_T - 1\}$, with $\mathbf{f}(\mathbf{x}) = \mathbf{x}$, where the operator $\tilde{\Delta}$ is defined in (8.9.15). In this case the generation of the multiple stochastic integrals involved is straightforward since only one Poisson distributed random variable at each time step is required, as we have seen in (4.1.23). This allows the above schemes to be easily implemented. Such an implementation is less complex than in the case of genuine jump diffusion SDEs. As such, there exist significant advantages that one can exploit when deriving higher order strong Taylor schemes for pure jump SDEs.

8.10 Exercises

8.1. Consider the pure jump SDE

$$dX_t = (a + b X_{t-}) dN_t,$$

where $N = \{N_t, t \geq 0\}$ is a Poisson process with intensity $\lambda < \infty$. Write down the corresponding strong order 1.0 Taylor scheme.

8.2. For the SDE in the previous exercise derive the strong order 2.0 Taylor scheme.

Estimating Discretely Observed Diffusions

In many areas of application the estimation of parameters in SDEs is an important practical task. This estimation is almost like inverting the problem of scenario simulation and can benefit from the application of Wagner-Platen expansions. We have already mentioned that it is important to apply scenario simulation when checking empirically the usefulness of a proposed estimation methods. This chapter introduces estimation techniques for discretely observed diffusion processes. Transform functions are applied to the data in order to obtain estimators of both the drift and diffusion coefficients. Consistency and asymptotic normality of the resulting estimators is investigated. Power transforms are used to estimate the parameters of affine diffusions, for which explicit estimators are obtained.

9.1 Maximum Likelihood Estimation

There exists a rich literature on parameter estimation for diffusion processes. An account of modern methods in optimal estimation is given, for instance, in [Heyde \(1997\)](#). Estimation of continuously observable diffusions is studied, for instance, in [Kutoyants \(1984\)](#), [Prakasa Rao \(1999\)](#) or [Liptser & Shiryaev \(2001\)](#). When a diffusion process is continuously observed, *maximum likelihood estimation* yields estimators that are consistent, asymptotically normal and efficient. For this estimation method the parameters are typically only allowed to enter the drift coefficient of the diffusion. In financial applications it is mainly the diffusion coefficient that is of particular interest. Furthermore, in practice it is usually only possible to observe the diffusion process at discrete time points. Inference for discretely observed diffusions can be based, for instance, on an approximation of the likelihood function. If the transition densities of the diffusion are known explicitly, see Chap. 2, then the likelihood function can be used to estimate the parameters. For the resulting maximum likelihood estimators the properties of consistency and asymptotic normal-

ity have been studied, for instance, in [Dacunha-Castelle & Florens-Zmirou \(1986\)](#). When the transition densities are unknown, a possible approach is to approximate the log-likelihood function based on continuous observations. This technique has the problem that the estimators that result are inconsistent if the time between observations is fixed, see [Florens-Zmirou \(1989\)](#). However, this problem can be solved by suitable modifications, see [Bibby & Sørensen \(1995\)](#) and [Sørensen \(2001\)](#). If the time between observations is sufficiently small, then this method works for the estimation of drift parameters, see [Kloeden, Platen, Schurz & Sørensen \(1996\)](#), as we will describe below.

There has been substantial research into alternative methods for the estimation of discretely observed diffusions, many of which are reviewed in [Prakasa Rao \(1999\)](#). One main direction has been the development of various approaches to the approximation of the transition density and hence to the likelihood function, see, for instance, [Pedersen \(1995\)](#), [Poulsen \(1999\)](#), [Brandt & Santa-Clara \(2002\)](#), [Elerain, Chib & Shephard \(2001\)](#), [Eraker \(2001\)](#) and [Ait-Sahalia \(2002\)](#). Interesting work relevant to the estimation for parameters in SDEs with jumps has been presented in [Ait-Sahalia & Jacod \(2009\)](#) and [Bladt & Sørensen \(2009\)](#).

Parameter Estimation of the Drift

Often a financial modeler can use a phenomenological or theoretical argument to justify the use of an SDE with a particular structure but needs to estimate the appropriate parameter values from observations of the actual quantity being modeled. Such parameter estimates are, strictly speaking, random variables. They depend strongly on the observed trajectories. Something needs to be known or assumed about the probability distribution of their deviations from the true value of the parameter. For example, if one has an SDE with a solution that has a stationary probability density, then by ergodicity one might expect that the parameter values obtained from the observation of a single trajectory over a finite sufficiently long time interval would converge to the true parameter values as the length of the observation interval increases. One could then use appropriate limit theorems to obtain an indication of the reliability of such estimates.

To illustrate the basic ideas of parameter estimation for solutions of SDEs we shall consider a scalar SDE with additive noise

$$dX_t = \theta a(X_t) dt + dW_t. \quad (9.1.1)$$

Here θ is the parameter to be estimated and the function $a = a(x)$ is possibly nonlinear. The maximum likelihood estimate $\hat{\theta}_T$ determined from observations of a trajectory of a solution process (9.1.1) over the time interval $t \in [0, T]$ is the value $\tilde{\theta}$, which maximizes the *likelihood ratio*

$$L_T(\tilde{\theta}) = \exp \left\{ \frac{1}{2} \tilde{\theta}^2 \int_0^T a(X_t)^2 dt - \tilde{\theta} \int_0^T a(X_t) dX_t \right\} \quad (9.1.2)$$

of the process $X = \{X_t, t \in [0, T]\}$ with respect to the Wiener process $W = \{W_t, t \in [0, T]\}$. This result can be derived using the *Girsanov Theorem*, mentioned in Sect. 2.7. It says that the right hand side of (9.1.2) equals the Radon-Nikodym derivative $\frac{dP_X}{dP_W}$ of the probability measures P_X and P_W , corresponding to the processes X and W , under which these are integrated as Wiener processes, respectively. It can be justified heuristically by approximating (9.1.1) with the Euler scheme

$$Y_{n+1} = Y_n + \theta a(Y_n) \Delta + \Delta W_n \quad (9.1.3)$$

for $n \in \{0, 1, \dots, N-1\}$, where $\Delta = \frac{T}{N}$. The increments $\Delta W_0, \Delta W_1, \dots, \Delta W_{N-1}$ of the Wiener process are independent $N(0, 1)$ distributed random variables. Their joint probability density is given by

$$\begin{aligned} p_W^{(N)} &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\Delta}} \exp\left\{-\frac{1}{2\Delta} (\Delta W_n)^2\right\} \\ &= \frac{1}{(2\pi\Delta)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2\Delta} \sum_{n=0}^{N-1} (\Delta W_n)^2\right\}. \end{aligned} \quad (9.1.4)$$

Writing $\Delta Y_n = Y_{n+1} - Y_n$ for $n \in \{0, 1, \dots, N-1\}$ we can determine the joint probability density $p_Y^{(N)}$ for $\Delta Y_0, \Delta Y_1, \dots, \Delta Y_{N-1}$ independent $N(0, \Delta)$ distributed as

$$p_Y^{(N)} = \frac{1}{(2\pi\Delta)^{\frac{N}{2}}} \exp\left\{-\frac{1}{2\Delta} \sum_{n=0}^{N-1} (\Delta Y_n)^2\right\}. \quad (9.1.5)$$

The Radon-Nikodym derivative of the discrete process $Y = \{Y_n, n \in \{0, 1, \dots, N-1\}\}$ with respect to the Wiener process W is simply the ratio

$$\begin{aligned} \frac{p_Y^{(N)}}{p_W^{(N)}} &= \exp\left\{-\frac{1}{2\Delta} \sum_{n=0}^{N-1} ((\Delta Y_n)^2 - (\Delta W_n)^2)\right\} \\ &= \exp\left\{-\frac{1}{2} \theta^2 \sum_{n=0}^{N-1} a(Y_n)^2 \Delta - \theta \sum_{n=0}^{N-1} a(Y_n) \Delta W_n\right\} \\ &= \exp\left\{\frac{1}{2} \theta^2 \sum_{n=0}^{N-1} a(Y_n)^2 \Delta - \theta \sum_{n=0}^{N-1} a(Y_n) \Delta Y_n\right\}. \end{aligned} \quad (9.1.6)$$

Taking limit as $N \rightarrow \infty$, we see that the term in the exponential converges to the difference of the integrals in the formula (9.1.2).

Maximum Likelihood Estimator

We differentiate the likelihood ratio (9.1.2) with respect to θ and solve the equation

$$\frac{\partial L_T(\theta)}{\partial \theta} = 0 \quad (9.1.7)$$

to obtain the *maximum likelihood estimator*

$$\hat{\theta}_T = \frac{\int_0^T a(X_t) dX_t}{\int_0^T a(X_t)^2 dt}. \quad (9.1.8)$$

This is a random variable depending on the particular sample path of the process X that is observed over the time interval $[0, T]$. From (9.1.1) and (9.1.8) we have

$$\hat{\theta}_T - \theta = \frac{\int_0^T a(X_t) dW_t}{\int_0^T a(X_t)^2 dt}, \quad (9.1.9)$$

where θ is the true value of the parameter. According to Sect. 2.7, if

$$E \left(\int_0^T a(X_t)^2 dt \right) < \infty \quad (9.1.10)$$

for all T and if the SDE (9.1.1) has a stationary solution which is ergodic with density \bar{p} , found by solving the stationary Kolmogorov forward equation, then it follows from the ergodicity of X that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T a(X_t) dW_t \xrightarrow{\text{a.s.}} 0 \quad (9.1.11)$$

and

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T a(X_t)^2 dt \xrightarrow{\text{a.s.}} \int_{\mathbb{R}} a(x)^2 \bar{p}(x) dx. \quad (9.1.12)$$

We can thus conclude from (9.1.9) that

$$\lim_{T \rightarrow \infty} \hat{\theta}_T \xrightarrow{\text{a.s.}} \theta. \quad (9.1.13)$$

Moreover, a version of the Central Limit Theorem tells us that $T^{\frac{1}{2}}(\hat{\theta}_T - \theta)$ converges in distribution as $T \rightarrow \infty$ to an $N(0, \sigma^2)$ -distributed random variable with variance

$$\sigma^2 = \left(\int_{\mathbb{R}} a(x)^2 \bar{p}(x) dx \right)^{-1}. \quad (9.1.14)$$

We can use this information to determine confidence intervals for θ and to estimate an approximate value of T for a desired confidence level. The above maximum likelihood estimation provides us with a basic estimation method, which requires the knowledge of the transition density of the stochastic process.

9.2 Discretization of Estimators

In this section estimators of drift parameters for diffusion processes are obtained by using numerical methods. Here estimators are constructed by making discrete-time approximations to stochastic integrals in the maximum likelihood estimators for continuously observed diffusions. We refer here to results given in Kloeden et al. (1996).

Maximum Likelihood Estimators

Let us study parametric statistical models for d -dimensional diffusion processes defined by a class of SDEs of the form

$$d\mathbf{X}_t = (\mathbf{A}_t(\mathbf{X}_t) + \mathbf{B}_t(\mathbf{X}_t)\boldsymbol{\theta})dt + \mathbf{D}_t(\mathbf{X}_t)d\mathbf{W}_t. \quad (9.2.1)$$

Note that the parameter $\boldsymbol{\theta}$ appears linearly in the drift of the above SDE.

We assume that this Itô SDE has a unique strong solution for every vector parameter value $\boldsymbol{\theta}$ from some open set $\Theta \subseteq \mathbb{R}^k$, $k \in \{0, 1, \dots\}$. In (9.2.1) \mathbf{A}_t , \mathbf{B}_t and \mathbf{D}_t are vector or matrix functions depending for each t on \mathbf{X}_t . For simplicity, we will assume that they are continuous. The statistical parameter $\boldsymbol{\theta}$ is k -dimensional, and \mathbf{W} is a m -dimensional Wiener process, so \mathbf{B}_t is a $d \times k$ -matrix and \mathbf{D}_t is a $d \times m$ -matrix. The vector \mathbf{A}_t is d -dimensional. \mathbf{A}_t , \mathbf{B}_t and \mathbf{D}_t are assumed to be known and given by the problem under study. The drift parameter vector $\boldsymbol{\theta}$ is to be estimated from an observed sample path $\mathbf{X} = \{\mathbf{X}_t, t \in [0, T]\}$.

We assume that $\boldsymbol{\theta} \in \Theta$. Without loss of generality we can assume that $\mathbf{0} \in \Theta$. Since the functionals \mathbf{A}_t , \mathbf{B}_t and \mathbf{D}_t depend on \mathbf{X} through \mathbf{X}_t only, the solution of the SDE (9.2.1) is Markovian.

Suppose we have observed the process \mathbf{X} in the time interval $[0, T]$. Let P_θ^T denote the probability measure on the set of continuous functions from $[0, T]$ into \mathbb{R}^d corresponding to the solution of (9.2.1) for the parameter $\boldsymbol{\theta}$. The likelihood function for our observation $\mathbf{X} = \{\mathbf{X}_t, t \in [0, T]\}$ is the Radon-Nikodym derivative

$$L_T(\boldsymbol{\theta}) = \frac{dP_\theta^T}{dP_0^T} \quad (9.2.2)$$

provided P_θ^T is dominated by P_0^T for all $\boldsymbol{\theta} \in \Theta$, see Sect. 2.7. This is the case if the $d \times d$ -squared diffusion coefficient matrix

$$\mathbf{C}_t(\mathbf{X}) = \mathbf{D}_t(\mathbf{X})\mathbf{D}_t(\mathbf{X})^\top$$

is nonsingular for almost all $t \in [0, T]$, and $P_\theta^T(S_T^{i,\ell} < \infty) = 1$, $i, \ell \in \{1, 2, \dots, k\}$, for all $\boldsymbol{\theta} \in \Theta$, where \mathbf{S}_T is the $k \times k$ -information matrix

$$\mathbf{S}_T = \int_0^T \mathbf{B}_t(\mathbf{X}_t)^\top \mathbf{C}_t(\mathbf{X}_t)^{-1} \mathbf{B}_t(\mathbf{X}_t) dt. \quad (9.2.3)$$

Define the k -dimensional random *score vector*

$$\mathbf{H}_T = \int_0^T \mathbf{B}_t(\mathbf{X}_t)^\top \mathbf{C}_t(\mathbf{X}_t)^{-1} d\tilde{\mathbf{X}}_t, \quad (9.2.4)$$

where

$$\tilde{\mathbf{X}}_t = \mathbf{X}_t - \int_0^t \mathbf{A}_s(\mathbf{X}_s) ds.$$

Then the likelihood function (9.2.2) is given by

$$L_T(\boldsymbol{\theta}) = \exp \left\{ \boldsymbol{\theta}^\top \mathbf{H}_T - \frac{1}{2} \boldsymbol{\theta}^\top \mathbf{S}_T \boldsymbol{\theta} \right\}.$$

The *maximum likelihood estimator* obtained by maximizing $L_T(\boldsymbol{\theta})$ is

$$\hat{\boldsymbol{\theta}}_T = \mathbf{S}_T^{-1} \mathbf{H}_T, \quad (9.2.5)$$

which exists because \mathbf{S}_T is nonsingular under the conditions imposed.

Under natural regularity conditions one can apply a Central Limit Theorem for martingales to the score martingale $\mathbf{H}_T - \mathbf{S}_T \boldsymbol{\theta}$ to prove that the maximum likelihood estimator is consistent and asymptotically normally distributed, see Sørensen (1991).

Drift Estimation for a Linear SDE

Consider the one-dimensional SDE

$$dX_t = (\theta_1 X_t + \theta_2) dt + c dW_t \quad (9.2.6)$$

for $t \in [0, T]$, where c is known. Its solution is an Ornstein-Uhlenbeck process, see Sect. 1.7. For the model (9.2.6) we find

$$A_t(X_t) = 0, \quad \mathbf{B}_t(X_t) = (X_t, 1), \quad D_t(X_t) = c \quad \text{and} \quad C_t(X_t) = c^2.$$

This yields

$$\begin{aligned} S_T^{1,1} &= \frac{1}{c^2} \int_0^T X_t^2 dt, \quad S_T^{2,1} = S_T^{1,2} = \frac{1}{c^2} \int_0^T X_t dt, \quad S_T^{2,2} = \frac{T}{c^2}, \\ H_T^1 &= \frac{1}{c^2} \int_0^T X_t dt \quad \text{and} \quad H_T^2 = \frac{1}{c^2} (X_T - X_0). \end{aligned}$$

By (9.2.5) the maximum likelihood estimators follow in the form

$$\begin{aligned} \hat{\theta}_{1,T} &= \frac{1}{N_T} \left(T \int_0^T X_t dX_t - (X_T - X_0) \int_0^T X_t dt \right) \\ &= \frac{1}{N_T} \left(\frac{1}{2} T (X_T^2 - X_0^2 - c^2 T) - (X_T - X_0) \int_0^T X_t dt \right) \quad (9.2.7) \end{aligned}$$

and

$$\begin{aligned}\hat{\theta}_{2,T} &= \frac{1}{N_T} \left((X_T - X_0) \int_0^T X_t^2 dt - \int_0^T X_t dX_t \int_0^T X_t dt \right) \\ &= \frac{1}{N_T} \left((X_T - X_0) \int_0^T X_t^2 dt - \frac{1}{2} (X_T^2 - X_0^2 - c^2 T) \int_0^T X_t dt \right) \quad (9.2.8)\end{aligned}$$

with

$$N_T = T \int_0^T X_t^2 dt - \left(\int_0^T X_t dt \right)^2, \quad (9.2.9)$$

where we have used Itô's formula.

Discrete-Time Observation

In practice, trajectories are not observed continuously. Rather the process is observed at a finite number of times $t_0 < t_1 < \dots < t_n$. The exact likelihood function corresponding to such data is the product of the corresponding transition densities, which can only rarely be found explicitly. A simple estimation procedure for such data is to use the continuous time maximum likelihood estimator (9.2.5) with $t_0 = 0$ and $t_n = T$ and with suitable approximations to the integrals in \mathbf{H}_T and \mathbf{S}_T given by (9.2.3) and (9.2.4). If the spacing between consecutive observation times are small, some of the properties of the continuous time maximum likelihood estimator are usually preserved, although its discrete-time version will be biased to some extent, as shown in Florens-Zmirou (1989) and Le Breton (1976).

It is well-known how to approximate numerically the Riemann integrals in \mathbf{S}_T and \mathbf{H}_T by Riemann sums or quadrature formulae like the trapezoidal rule. Also, the stochastic integral in \mathbf{H}_T can be approximated by a finite sum. However, it is preferable to replace the stochastic integral by Riemann and Stieltjes integrals when possible, and then to approximate these. This can in most cases be achieved for a one-dimensional diffusion by a direct use of Itô's formula, as shown in the example above. However, when \mathbf{X} is multi-dimensional it can only be performed in special cases.

In the following we will generate an approximate solution for the SDE (9.2.6) using strong schemes as described in Chap. 5. From these simulated sample paths we will estimate the parameters by approximate maximum likelihood estimators. Here we approximate time integrals by the trapezoidal rule. Other quadrature formulae could be used as well but usually make no major difference.

For the SDE (9.2.6) of an Ornstein-Uhlenbeck process we simulate exactly according to Chap. 2 a path over the interval $[0, T]$ with $T = 100$ with parameters $\theta_1 = -1.0$, $\theta_2 = 1.0$ and $c = 1.0$ starting at $X_0 = 0$. A simulated trajectory is shown in Fig. 9.2.1.

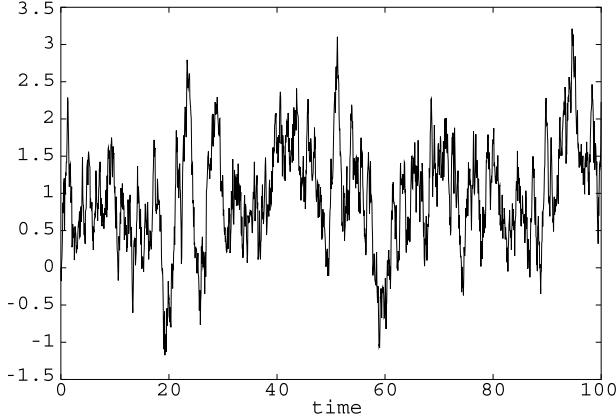


Fig. 9.2.1. A trajectory for $\theta_1 = -1.0$, $\theta_2 = 1.0$, $c = 1.0$ and $X_0 = 0$

Approximate Estimators

To estimate the parameters θ_1 and θ_2 we apply the estimators given by the last expressions in (9.2.7) and (9.2.8) and obtain, after approximation of the relevant time integrals by the trapezoidal formula, the *approximate estimators*

$$\hat{\theta}_{1,T}^{\Delta} = \frac{1}{2 N_T^{\Delta}} \left(T (X_T^2 - X_0^2 - c^2 T) - (X_T - X_0) \sum_{n=0}^{n_T-1} (X_{t_{n+1}} + X_{t_n}) \Delta \right) \quad (9.2.10)$$

and

$$\begin{aligned} \hat{\theta}_{2,T}^{\Delta} = & \frac{1}{2 N_T^{\Delta}} \left((X_T - X_0) \sum_{n=0}^{n_T-1} (X_{t_{n+1}}^2 + X_{t_n}^2) \Delta \right. \\ & \left. - \frac{1}{2} (X_T^2 - X_0^2 - c^2 T) \sum_{n=0}^{n_T-1} (X_{t_{n+1}} + X_{t_n}) \Delta \right) \quad (9.2.11) \end{aligned}$$

with

$$N_T^{\Delta} = \frac{T}{2} \sum_{n=0}^{n_T-1} (X_{t_{n+1}}^2 + X_{t_n}^2) \Delta - \left(\frac{1}{2} \sum_{n=0}^{n_T-1} (X_{t_{n+1}} + X_{t_n}) \Delta \right)^2. \quad (9.2.12)$$

Figs. 9.2.2 and 9.2.3 show that the approximate estimators $\hat{\theta}_{1,T}^{\Delta}$ and $\hat{\theta}_{2,T}^{\Delta}$ converge after initial oscillations to the corresponding values of θ_1 and θ_2 , respectively. It is important to note that the maximum likelihood method as applied above, typically only allows us to estimate parameters in the drift coefficient of the SDE when based on a measure transformation as previously indicated. However, parameters also in the diffusion coefficient can be estimated if the transition density is known and permits a suitable estimator.

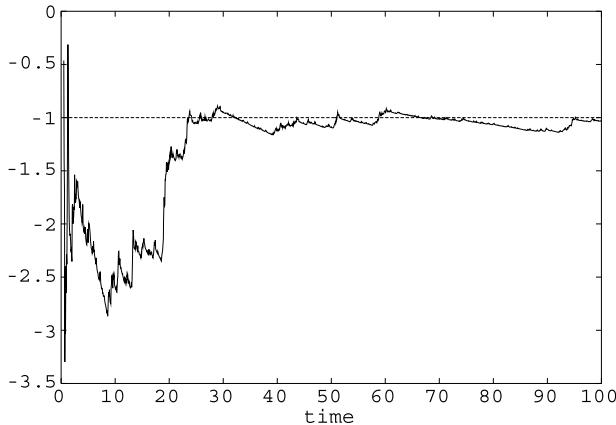


Fig. 9.2.2. A trajectory for $\hat{\theta}_{1,T}^A$, where $\theta_1 = -1.0$

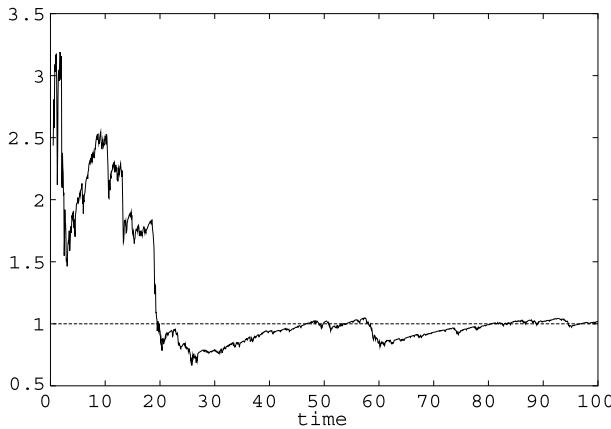


Fig. 9.2.3. A trajectory for $\hat{\theta}_{2,T}^A$, where $\theta_2 = 1.0$

9.3 Transform Functions for Diffusions

In this and the following two sections we will describe results obtained in Kelly, Platen & Sørensen (2004). The method allows also for the estimation of parameters located in the diffusion coefficient.

Estimating Functions

An alternative approach that avoids the use of transition densities is the use of *estimating functions*. These are functions of both the parameter and the observed data, and can be constructed to provide reliable estimators. Surveys of recent results on estimating functions are given in Heyde (1997), Sørensen (1997) and Bibby, Jacobsen & Sørensen (2003). The advantages associated with using estimating functions are highlighted in Heyde (1997). Es-

timating functions are invariant under one-to-one transforms of the data and these functions can be combined more simply than the estimators themselves. Bibby & Sørensen (1995) studied martingale estimating functions obtained from the derivative of the continuous time log-likelihood function by correcting for the discretization bias when subtracting its compensator. The resulting estimating function, known as a linear martingale estimating function, depends on the conditional moments of the diffusion process. Quadratic martingale estimating functions also involving second order conditional moments were obtained by Bibby & Sørensen (1996) from a Gaussian approximation to the likelihood function. Kessler & Sørensen (1999) considered estimating functions based on eigenfunctions for which the conditional moments are explicit. Sørensen (2000) considered more general estimating functions, known as prediction-based estimating functions. Here conditional moments are approximated by expressions involving only unconditional moments. This type of estimating functions, is a useful alternative to martingale estimating functions, for instance, in stochastic volatility models and certain interest rate term structure models. Christensen, Poulsen & Sørensen (2001) compared *optimal martingale estimating functions* and the approximate maximum likelihood method, mentioned earlier, for the estimation of the parameters in a model of the short rate to techniques such as the *generalized method of moments*, see Hansen (1982), and *indirect inference*, see Gouriéroux, Monfort & Renault (1993) and Gallant & Tauchen (1996). It was found that optimal martingale estimating functions and the approximate maximum likelihood method reduce bias, true standard errors and bias in estimated standard errors when compared to the aforementioned techniques. In Bibby (1994) martingale estimating functions are combined to estimate parameters in both the drift and the diffusion coefficient using techniques discussed in Heyde (1997).

The approach that will be presented below is based on the estimating function approach. The objective is to obtain a simple yet general estimation method, which provides flexibility in the estimation of discretely observed diffusion processes via the use of *transform functions*. Unlike some of the aforementioned techniques, particular information about the conditional and unconditional moments of the diffusion process are not needed. Furthermore, parameters in the diffusion coefficients can be estimated and stationarity of the observed process is not required. The Wagner-Platen expansion will be used to expand the increments of functions of the observed diffusion process.

Diffusion Model

Consider a class of one-dimensional diffusion processes defined by the following SDE

$$dX_t = b(t, X_t; \boldsymbol{\theta}) dt + \sigma(t, X_t; \boldsymbol{\theta}) dW_t \quad (9.3.1)$$

for $t \in [0, T]$. The initial value $X_0 = x_0$ is assumed to be \mathcal{A}_0 -measurable. Here $W = \{W_t, t \in [0, T]\}$ denotes a standard Wiener process given on the

filtered probability space $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P_\theta)$, where the filtration $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0, T]}$ satisfies the usual conditions. We assume that the SDE (9.3.1) has a unique strong solution for all parameter values $\boldsymbol{\theta}$ in a given open subset $\Theta \subseteq \mathbb{R}^p$, $p \in \{1, 2, \dots\}$. The drift and diffusion coefficient functions $b(\cdot, \cdot; \boldsymbol{\theta}) : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ and $\sigma(\cdot, \cdot; \boldsymbol{\theta}) : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$, respectively, are assumed to be known with the exception of the parameter vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top \in \Theta$.

It is our aim to estimate the unknown parameter vector $\boldsymbol{\theta}$ from observations of the diffusion process $X = \{X_t, t \in [0, T]\}$. For simplicity, an equidistant time discretization with observation times τ_n , where $0 = \tau_0 < \tau_1 < \dots < \tau_n < \tau_{n+1} < \dots$, is assumed to be such that the time step size $\Delta = \tau_n - \tau_{n-1} \in (0, 1)$. For $t \in [0, T]$ we introduce the integer n_t as the largest integer n for which τ_n does not exceed t , that is

$$n_t = \max\{n \in \{0, 1, \dots\} : \tau_n \leq t\} = \left[\frac{t}{\Delta} \right], \quad (9.3.2)$$

where $[x]$ denotes the integer part of the real number x .

Transform Function

To provide sufficient flexibility for the estimation we consider, at the observation times $\tau_0, \tau_1, \tau_2, \dots$, the original data

$$X_{\tau_0}, X_{\tau_1}, X_{\tau_2}, \dots \quad (9.3.3)$$

and the transformed data

$$U(\tau_0, X_{\tau_0}; \lambda_i), U(\tau_1, X_{\tau_1}; \lambda_i), U(\tau_2, X_{\tau_2}; \lambda_i), \dots \quad (9.3.4)$$

for $i \in \{1, 2, \dots, p\}$. Here $U(\cdot, \cdot; \cdot) : [0, T] \times \mathbb{R} \times \Lambda \rightarrow \mathbb{R}$ is a smooth real valued function with respect to $t \in [0, T]$ and $x \in \mathbb{R}$, where $\Lambda \subseteq \mathbb{R}$. The idea here is that by transforming the originally observed data via a known smooth function, one can more easily and reliably estimate certain parameters.

The function $U(\cdot, \cdot; \lambda_i)$ for $i \in \{1, 2, \dots, p\}$, is called the *i*th transform function, and is used to transform the data in a manner that allows us to obtain useful estimates of the unknown parameters. In principle, for each $i \in \{1, 2, \dots, p\}$, a different function could be used to estimate the parameters. For fixed $\lambda_i \in \Lambda$ we obtain, by the Itô formula, the following SDE for the transformed data

$$dU(t, X_t; \lambda_i) = L_\theta^0 U(t, X_t; \lambda_i) dt + L_\theta^1 U(t, X_t; \lambda_i) dW_t \quad (9.3.5)$$

for $t \in [0, T]$. Here we have used the operators

$$L_\theta^0 u(t, x) = \left(\frac{\partial}{\partial t} u(t, x) + b(t, x; \theta) \frac{\partial}{\partial x} u(t, x) + \frac{1}{2} \sigma^2(t, x; \theta) \frac{\partial^2}{\partial x^2} u(t, x) \right) \quad (9.3.6)$$

and

$$L_\theta^1 u(t, x) = \sigma(t, x; \theta) \frac{\partial}{\partial x} u(t, x). \quad (9.3.7)$$

For $n \in \{1, 2, \dots\}$, $i \in \{1, 2, \dots, p\}$ and $\lambda_i \in \Lambda$ we introduce the *normalized difference*

$$D_{\lambda_i, n, \Delta} = \frac{1}{\tau_n - \tau_{n-1}} (U(\tau_n, X_{\tau_n}; \lambda_i) - U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)) \quad (9.3.8)$$

and the *normalized squared increment*

$$Q_{\lambda_i, n, \Delta} = \frac{1}{\tau_n - \tau_{n-1}} (U(\tau_n, X_{\tau_n}; \lambda_i) - U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))^2. \quad (9.3.9)$$

The time between observations is assumed to be small. Therefore, it makes sense to use some approximation for the increments of U . By a truncated Wagner-Platen expansion, see (4.4.7), the increment of U in (9.3.8) and (9.3.9) can be expressed in terms of multiple stochastic integrals. Thus, we obtain

$$\begin{aligned} & U(\tau_n, X_{\tau_n}; \lambda_i) - U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) \\ &= L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) (W_{\tau_n} - W_{\tau_{n-1}}) + L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) (\tau_n - \tau_{n-1}) \\ &+ L_\theta^1 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) \frac{1}{2} ((W_{\tau_n} - W_{\tau_{n-1}})^2 - (\tau_n - \tau_{n-1})) \\ &+ L_\theta^0 L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) \frac{(\tau_n - \tau_{n-1})^2}{2} \\ &+ L_\theta^1 L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) \int_{\tau_{n-1}}^{\tau_n} \int_{\tau_{n-1}}^s dz dW_s \\ &+ L_\theta^0 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) \int_{\tau_{n-1}}^{\tau_n} \int_{\tau_{n-1}}^s dW_z ds \\ &+ R_{\lambda_i, n, \theta}(\tau_n, \tau_{n-1}, X_{\tau_n}), \end{aligned} \quad (9.3.10)$$

where $R_{\lambda_i, n, \theta}(\tau_n, \tau_{n-1}, X_{\tau_n})$ is the corresponding remainder term. By Lemma 4.5.1 the first term in (9.3.10) has mean zero and is the leading term of the expansion. Note that the third, fifth and sixth terms also have mean zero but are of a higher order than the first term. In (9.3.10) the order of the first term is $\sqrt{\Delta}$, that of the second and third term is Δ and that of the fourth term is Δ^2 . The order of the fifth and sixth term is $\Delta^{\frac{3}{2}}$. The remainder term has mean and variance of order Δ^3 .

A Class of Estimating Functions

Using (9.3.8) and (9.3.9) we can construct estimating functions by exploiting the structure of the first and second term of the above increment (9.3.10). To do this we define

$$\mathbf{F}_n(\boldsymbol{\theta}) = (\mathbf{F}_n^{(1)}(\boldsymbol{\theta})^\top, \mathbf{F}_n^{(2)}(\boldsymbol{\theta})^\top)^\top$$

with

$$\mathbf{F}_n^{(j)}(\boldsymbol{\theta})^\top = (F_{1,n}^{(j)}(\boldsymbol{\theta}), \dots, F_{q,n}^{(j)}(\boldsymbol{\theta}))$$

for $j \in \{1, 2\}$, where

$$F_{i,n}^{(1)}(\boldsymbol{\theta}) = D_{\lambda_i, n, \Delta} - L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i)$$

and

$$F_{i,n}^{(2)}(\boldsymbol{\theta}) = Q_{\lambda_i, n, \Delta} - (L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))^2,$$

for $\boldsymbol{\theta} \in \Theta$ and suitably chosen values of $\lambda_i \in \Lambda$, $i \in \{1, 2, \dots, q\}$. It is not necessary that the number of λ_i 's is the same for $F_n^{(1)}(\boldsymbol{\theta})$ and $F_n^{(2)}(\boldsymbol{\theta})$ and they need not have the same value for the two functions. This assumption only simplifies the exposition.

A class of *estimating functions* is then given by the sum

$$\mathbf{K}(\boldsymbol{\theta}, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} \mathbf{M}(\boldsymbol{\theta}) \mathbf{F}_n(\boldsymbol{\theta}), \quad (9.3.11)$$

where the $p \times 2q$ matrix valued *weighting function* $\mathbf{M}(\boldsymbol{\theta}) = \mathbf{M}(\boldsymbol{\theta}, \tau_{n-1}, X_{\tau_{n-1}}, \Delta)$ is free to be chosen appropriately. Throughout the section the dependence of a weighting matrix and its elements on τ_{n-1} , $X_{\tau_{n-1}}$ and Δ will be suppressed in our notation. The estimating function $\mathbf{K}(\tilde{\boldsymbol{\theta}}, t, \Delta)$, where $\tilde{\boldsymbol{\theta}}$ is taken to be the true parameter value, has expectation of order Δ . Thus, when the observation interval Δ is sufficiently small, the expectation of $\mathbf{K}(\tilde{\boldsymbol{\theta}}, t, \Delta)$ is approximately zero. Essentially, the approach adopted here is to approximate the conditional moments of the transformed diffusion process using the expansion in (9.3.10). By this argument we obtain the p -dimensional *estimating equation*

$$\mathbf{K}(\boldsymbol{\theta}, t, \Delta) = \mathbf{0} \quad (9.3.12)$$

for $t > \tau_1$. If the corresponding system of p equations has a unique solution, then we obtain an estimator $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_p)^\top$ for $\boldsymbol{\theta}$.

Optimal Weighting

The estimating function (9.3.11) is slightly biased since we neglect remainder terms. To determine the *optimal weighting matrix* $\mathbf{M}^*(\boldsymbol{\theta})$, one can consider the unbiased estimating function

$\mathbf{K}^*(\boldsymbol{\theta}, t, \Delta)$ obtained by compensating $\mathbf{K}(\boldsymbol{\theta}, t, \Delta)$ such that

$$\mathbf{K}^*(\boldsymbol{\theta}, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} \mathbf{M}(\boldsymbol{\theta}) (\mathbf{F}_n(\boldsymbol{\theta}) - \bar{\mathbf{F}}_n(\boldsymbol{\theta})). \quad (9.3.13)$$

If one denotes by E_θ expectation with respect to the underlying probability measure P_θ , then

$$\bar{\mathbf{F}}_n(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} (\mathbf{F}_n(\boldsymbol{\theta}) \mid X_{\tau_{n-1}})$$

is the compensator for $\mathbf{F}_n(\boldsymbol{\theta})$ and is of order Δ . The optimal choice for the weighting matrix $\mathbf{M}(\boldsymbol{\theta})$ in the unbiased estimating equation (9.3.13), in the sense of Godambe & Heyde (1987), can be derived using the method outlined in Heyde (1997). For details regarding the case of diffusion processes we refer to Sørensen (1997), where the optimal weighting matrix is given by

$$\mathbf{M}^*(\boldsymbol{\theta}) = \mathbf{B}^*(\boldsymbol{\theta}) \mathbf{V}^*(\boldsymbol{\theta})^{-1}. \quad (9.3.14)$$

Here $\mathbf{V}^*(\boldsymbol{\theta})$ is the $2q \times 2q$ conditional covariance matrix

$$\begin{aligned} \mathbf{V}^*(\boldsymbol{\theta}) &= \mathbf{V}^*(\boldsymbol{\theta}, \tau_{n-1}, X_{\tau_{n-1}}) \\ &= E_{\boldsymbol{\theta}} ((\mathbf{F}_n(\boldsymbol{\theta}) - \bar{\mathbf{F}}_n(\boldsymbol{\theta}))(\mathbf{F}_n(\boldsymbol{\theta}) - \bar{\mathbf{F}}_n(\boldsymbol{\theta}))^\top \mid X_{\tau_{n-1}}) \end{aligned}$$

and $\mathbf{B}^*(\boldsymbol{\theta}) = (\mathbf{B}^{*(1)}(\boldsymbol{\theta}), \mathbf{B}^{*(2)}(\boldsymbol{\theta}))$ where $\mathbf{B}^{*(k)}(\boldsymbol{\theta})$, $k \in \{1, 2\}$, denotes the $p \times q$ matrix, where the (i, j) th entry is

$$B_{i,j}^{*(k)}(\boldsymbol{\theta}) = B^{*(k)}(\boldsymbol{\theta}, \tau_{n-1}, X_{\tau_{n-1}})_{i,j} = E_{\boldsymbol{\theta}} \left(\frac{\partial}{\partial \theta_i} \left[F_{j,n}^{(k)}(\boldsymbol{\theta}) - \bar{F}_{j,n}^{(k)}(\boldsymbol{\theta}) \right] \mid X_{\tau_{n-1}} \right).$$

The values of the λ_i 's should be chosen in such a way that the conditional covariance matrix $\mathbf{V}^*(\boldsymbol{\theta})$ is invertible.

Keeping only the leading terms, we obtain

$$\tilde{\mathbf{K}}(\boldsymbol{\theta}, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} \tilde{\mathbf{M}}(\boldsymbol{\theta}) \mathbf{F}_n(\boldsymbol{\theta}), \quad (9.3.15)$$

where $\tilde{\mathbf{M}}(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta}) \mathbf{V}(\boldsymbol{\theta})^{-1}$. Here $\mathbf{B}(\boldsymbol{\theta}) = (\mathbf{B}^{(1)}(\boldsymbol{\theta}), \mathbf{B}^{(2)}(\boldsymbol{\theta}))$ and the (i, j) th entry of the $p \times q$ matrices $\mathbf{B}^{(1)}(\boldsymbol{\theta})$ and $\mathbf{B}^{(2)}(\boldsymbol{\theta})$ are

$$B^{(1)}(\boldsymbol{\theta})_{i,j} = \frac{\partial}{\partial \theta_i} L_{\theta}^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) \quad (9.3.16)$$

and

$$B^{(2)}(\boldsymbol{\theta})_{i,j} = \frac{\partial}{\partial \theta_i} (L_{\theta}^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j))^2, \quad (9.3.17)$$

respectively. Moreover,

$$\mathbf{V}(\boldsymbol{\theta}) = \left\{ \begin{array}{l} \mathbf{V}^{11}(\boldsymbol{\theta}) \mathbf{V}^{12}(\boldsymbol{\theta}) \\ \mathbf{V}^{21}(\boldsymbol{\theta}) \mathbf{V}^{22}(\boldsymbol{\theta}) \end{array} \right\},$$

where the (i, j) th entry of the $q \times q$ matrices $\mathbf{V}^{11}(\boldsymbol{\theta})$, $\mathbf{V}^{22}(\boldsymbol{\theta})$ and $\mathbf{V}^{12}(\boldsymbol{\theta})$ are

$$V^{11}(\boldsymbol{\theta})_{i,j} = \frac{1}{\Delta} L_{\theta}^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) L_{\theta}^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j), \quad (9.3.18)$$

$$V^{22}(\boldsymbol{\theta})_{i,j} = 2 \left[L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) \right]^2, \quad (9.3.19)$$

and

$$\begin{aligned} V^{12}(\boldsymbol{\theta})_{i,j} &= 2 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) \\ &\quad + 2 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) L_\theta^1 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j) \\ &\quad + L_\theta^1 L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) [L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j)]^2, \end{aligned} \quad (9.3.20)$$

while $\mathbf{V}^{21}(\boldsymbol{\theta}) = (\mathbf{V}^{12}(\boldsymbol{\theta}))^\top$.

The weighting matrix $\tilde{\mathbf{M}}(\boldsymbol{\theta})$ in (9.3.15) is optimal in what is referred to as the fixed sample sense, see Godambe & Heyde (1987) and Heyde (1997). This means, the weighting matrix results in an estimating function (9.3.15), that is, to the order of approximation used, and closest within the class of estimating functions of the form (9.3.11) to the corresponding, usually unknown, score function. In this sense, this gives under appropriate assumptions the most efficient estimator for a fixed number of observations within the class of estimators considered. For example, if the diffusion process is ergodic, it can be shown that a fixed sample optimal martingale estimating function is also asymptotically optimal, also known as *Heyde-optimal*, see Heyde (1997). Heyde-optimality results in an estimator that has the smallest asymptotic confidence intervals within the class of estimators considered. The estimating functions proposed in this section are approximations of martingale estimating functions to the order Δ .

Optimal Estimating Equations

For the given transform functions, with parameters $\lambda_i \in \Lambda$, $i \in \{1, 2, \dots, q\}$, we have now obtained a p -dimensional *optimal estimating equation*

$$\tilde{\mathbf{K}}(\boldsymbol{\theta}, t, \Delta) = \mathbf{0}$$

for $t \geq \tau_1$, see (9.3.15). Assuming that the resulting system of p equations has a unique solution, we obtain for the particular SDE (9.3.1) an estimator $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_p)^\top$ for the parameter vector $\boldsymbol{\theta}$. Note that the vector of estimators $\hat{\boldsymbol{\theta}}$ depends on t , Δ , $\lambda_1, \dots, \lambda_p$ and the observed data. Appropriate values of λ_i for $i \in \{1, 2, \dots, p\}$ can be found by exploiting asymptotic properties of the estimating functions as will be described later. The choice of the λ_i for $i \in \{1, 2, \dots, p\}$ determines the resulting system of equations.

A simpler, although less efficient, estimation procedure can be used when the parameter $\boldsymbol{\theta}$ can be written as $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})^\top$. Here it is assumed that the p_1 -dimensional parameter $\boldsymbol{\alpha}$ appears only in the drift coefficient, while the diffusion coefficient depends only on the p_2 -dimensional parameter $\boldsymbol{\beta}$. In this case we first estimate $\boldsymbol{\beta}$ by solving

$$\tilde{\mathbf{H}}(\hat{\boldsymbol{\beta}}, t, \Delta) = \mathbf{0},$$

where

$$\tilde{\mathbf{H}}(\boldsymbol{\beta}, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} \mathbf{B}^{(2)}(\boldsymbol{\beta}) \mathbf{V}^{22}(\boldsymbol{\beta})^{-1} \mathbf{F}_n^{(2)}(\boldsymbol{\beta}) \quad (9.3.21)$$

with the $p_2 \times q$ matrix $\mathbf{B}^{(2)}(\boldsymbol{\beta})$ given by

$$\mathbf{B}^{(2)}(\boldsymbol{\beta})_{i,j} = \frac{\partial}{\partial \beta^i} (L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j))^2, \quad (9.3.22)$$

and $\mathbf{V}^{22}(\boldsymbol{\beta}) = \mathbf{V}^{22}(\boldsymbol{\theta})$ given by (9.3.19). Note that $\tilde{\mathbf{H}}(\boldsymbol{\beta}, t, \Delta)$ does not depend on $\boldsymbol{\alpha}$. Next estimate $\boldsymbol{\alpha}$ by solving $\tilde{\mathbf{G}}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, t, \Delta) = \mathbf{0}$, where $\hat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$ previously obtained, and

$$\tilde{\mathbf{G}}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} \mathbf{B}^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \mathbf{V}^{11}(\boldsymbol{\alpha}, \boldsymbol{\beta})^{-1} \mathbf{F}_n^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \quad (9.3.23)$$

with the $p_1 \times q$ matrix $\mathbf{B}^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ given by

$$\mathbf{B}^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta})_{i,j} = \frac{\partial}{\partial \alpha^i} L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_j). \quad (9.3.24)$$

and $\mathbf{V}^{11}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbf{V}^{11}(\boldsymbol{\theta})$ given by (9.3.18). The estimating functions $\tilde{\mathbf{G}}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t, \Delta)$ and $\tilde{\mathbf{H}}(\boldsymbol{\beta}, t, \Delta)$ are, to the order of approximation used, optimal within the classes of estimating functions

$$\mathbf{G}(\boldsymbol{\alpha}, \boldsymbol{\beta}, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} \mathbf{M}^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \mathbf{F}_n^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \quad (9.3.25)$$

and

$$\mathbf{H}(\boldsymbol{\beta}, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} \mathbf{M}^{(2)}(\boldsymbol{\beta}) \mathbf{F}_n^{(2)}(\boldsymbol{\beta}) \quad (9.3.26)$$

for estimating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. The optimal martingale estimating function for the form (9.3.13) is the optimal combination of the optimal martingale estimating functions to which (9.3.21) and (9.3.23) are approximations, see Heyde (1997) and Bibby (1994).

9.4 Estimation of Affine Diffusions

Affine Diffusions

We now recall a specific class of *affine diffusions* that will aid in highlighting the features of the methodology described above. So called affine diffusions

are popular for modeling in finance, see [Duffie & Kan \(1994\)](#), [Filipović \(2001\)](#) and Sect. 2.3. Consider the affine SDE

$$dX_t = (\theta_1 + \theta_2 X_t) dt + \sqrt{\theta_3 + \theta_4 X_t} dW_t \quad (9.4.1)$$

for $t \geq 0$, where the drift function

$$b(t, x; \theta) = \theta_1 + \theta_2 x$$

is affine, as is the squared diffusion coefficient function

$$\sigma^2(t, x; \theta) = \theta_3 + \theta_4 x.$$

In the following, the parameter vector $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^\top \in \Re^4$ with $p = 4$ shall be chosen such that the process $X = \{X_t, t \in [0, T]\}$ is *ergodic*. This happens when either

$$\theta_4 = 0, \quad \theta_2 < 0 \quad \text{and} \quad \theta_3 > 0 \quad (9.4.2)$$

or

$$\theta_4 > 0, \quad \theta_2 < 0 \quad \text{and} \quad \frac{2}{\theta_4} \left(\theta_1 - \frac{\theta_2 \theta_3}{\theta_4} \right) \geq 1, \quad (9.4.3)$$

see [Platen & Heath \(2006\)](#). In the first case, the Ornstein-Uhlenbeck process emerges, see (1.7.5). The process X lives then on the entire real line and the stationary distribution of the process is Gaussian with mean $-\frac{\theta_1}{\theta_2}$ and variance $-\frac{\theta_3}{(2\theta_2)}$. In the latter case the process X lives on the interval (y_0, ∞) with $y_0 = -\frac{\theta_3}{\theta_4}$. The stationary density for such an ergodic affine diffusion is of the form

$$\bar{p}(x) = \frac{\left(\frac{-2\theta_2}{\theta_4}\right)^{\frac{2}{\theta_4} \left(\theta_1 - \frac{\theta_2 \theta_3}{\theta_4}\right)} \left(x + \frac{\theta_3}{\theta_4}\right)^{\frac{2}{\theta_4} \left(\theta_1 - \frac{\theta_2 \theta_3}{\theta_4}\right) - 1} \exp\left\{\frac{2\theta_2}{\theta_4} \left(x + \frac{\theta_3}{\theta_4}\right)\right\}}{\Gamma\left(\frac{2}{\theta_4} \left(\theta_1 - \frac{\theta_2 \theta_3}{\theta_4}\right)\right)} \quad (9.4.4)$$

for $x \in (y_0, \infty)$, where $\Gamma(\cdot)$ denotes the well-known Gamma function. Note that the stationary density in (9.4.4) is a shifted Gamma distribution. In this case the stationary mean is

$$\int_{y_0}^{\infty} x \bar{p}(x) dx = -\frac{\theta_1}{\theta_2} \quad (9.4.5)$$

and the stationary second moment has the form

$$\int_{y_0}^{\infty} x^2 \bar{p}(x) dx = -\frac{(2\theta_1 + \theta_4) \theta_1 - \theta_3 \theta_2}{2(\theta_2)^2}. \quad (9.4.6)$$

Power Transform Functions

To illustrate the transform function method for affine diffusions we need to specify a class of transform functions. Let us consider the *power transform function*, which is one of the most tractable transforms. We set

$$U(t, x; \lambda) = x^\lambda \quad (9.4.7)$$

for $t \in [0, T]$, $x \in (y_0, \infty)$ with $\lambda > 0$. Setting $\boldsymbol{\alpha} = (\theta_1, \theta_2)$ and $\boldsymbol{\beta} = (\theta_3, \theta_4)$ and letting $\mathbf{M}^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\mathbf{M}^{(2)}(\boldsymbol{\beta})$ be the identity matrix we obtain from (9.3.25) and (9.3.26), the estimating functions

$$H_i(\boldsymbol{\beta}, t, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} F_{i,n}^{(2)}(\boldsymbol{\beta}) \quad (9.4.8)$$

with

$$F_{i,n}^{(2)}(\boldsymbol{\beta}) = (Q_{\lambda_i, n, \Delta} - (L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))^2)$$

for $i \in \{3, 4\}$ and

$$G_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \Delta) = \frac{1}{n_t} \sum_{n=1}^{n_t} F_{i,n}^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \quad (9.4.9)$$

where

$$F_{i,n}^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (D_{\lambda_i, n, \Delta} - L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))$$

for $i \in \{1, 2\}$. For the affine diffusions we have by (9.3.6)

$$\begin{aligned} L_\theta^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) &= (\theta_1 + \theta_2 X_{\tau_{n-1}}) \lambda_i X_{\tau_{n-1}}^{\lambda_i - 1} \\ &\quad + \frac{1}{2} (\theta_3 + \theta_4 X_{\tau_{n-1}}) \lambda_i (\lambda_i - 1) X_{\tau_{n-1}}^{\lambda_i - 2} \end{aligned} \quad (9.4.10)$$

and by (9.3.7)

$$L_\theta^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) = \lambda_i X_{\tau_{n-1}}^{\lambda_i - 1} \sqrt{\theta_3 + \theta_4 X_{\tau_{n-1}}}. \quad (9.4.11)$$

Estimating Functions

We obtain from (9.3.9), (9.3.26), (9.4.8) and (9.4.11) the estimating function

$$H_i(\boldsymbol{\beta}, t, \Delta) = A_\Delta^{0,1,0}(\lambda_i) - \theta_3 (\lambda_i)^2 A_\Delta^{2(\lambda_i - 1)} - \theta_4 (\lambda_i)^2 A_\Delta^{2\lambda_i - 1} \quad (9.4.12)$$

for $i \in \{3, 4\}$ for two different values $\lambda_3, \lambda_4 > 0$. Here we have used the notation

$$A_\Delta^{r,k,j}(\lambda_i) = \frac{1}{n_t} \sum_{n=1}^{n_t} (X_{\tau_{n-1}})^r (Q_{\lambda_i, n, \Delta})^k (D_{\lambda_i, n, \Delta})^j \quad (9.4.13)$$

and $A_{\Delta}^r = A_{\Delta}^{r,0,0}(\lambda_i)$, which refers to an equidistant time discretization of step size Δ .

Similarly, we obtain from (9.3.8), (9.3.25), (9.4.9) and (9.4.10) the estimating function

$$G_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, t, \Delta) = C_{\Delta}(\lambda_i, \theta_3, \theta_4) - \theta_1 \lambda_i A_{\Delta}^{\lambda_i-1} - \theta_2 \lambda_i A_{\Delta}^{\lambda_i} \quad (9.4.14)$$

with

$$C_{\Delta}(\lambda_i, \theta_3, \theta_4) = A_{\Delta}^{0,0,1}(\lambda_i) - \frac{\lambda_i(\lambda_i - 1)}{2} (\theta_3 A_{\Delta}^{\lambda_i-2} + \theta_4 A_{\Delta}^{\lambda_i-1}) \quad (9.4.15)$$

for $i \in \{1, 2\}$ for two different values $\lambda_1, \lambda_2 > 0$.

It follows from (9.3.10), properties of multiple stochastic integrals, and the existence of all moments, of positive order, of X such that

$$\begin{aligned} \lim_{n_t \rightarrow \infty} E(H_i(\boldsymbol{\beta}, t, \Delta)) &= \lim_{n_t \rightarrow \infty} E_{\theta} \left(\frac{1}{n_t} \sum_{n=1}^{n_t} \frac{1}{\Delta} \left[(L_{\theta}^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))^2 \Delta^2 \right. \right. \\ &\quad + (L_{\theta}^1 L_{\theta}^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i))^2 \frac{\Delta^2}{2} \\ &\quad + L_{\theta}^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) L_{\theta}^1 L_{\theta}^0 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) \frac{\Delta^2}{2} \\ &\quad \left. \left. + L_{\theta}^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) L_{\theta}^0 L_{\theta}^1 U(\tau_{n-1}, X_{\tau_{n-1}}; \lambda_i) \frac{\Delta^2}{2} \right] \right) \\ &\quad + \Delta^2 E_{\theta}(R_1(\boldsymbol{\beta}, t, \Delta, \lambda_i)) \\ &= \Delta \int_{y_0}^{\infty} \left((L_{\theta}^0 U(1, y; \lambda_i))^2 + \frac{1}{2} (L_{\theta}^1 L_{\theta}^1 U(1, y; \lambda_i))^2 \right. \\ &\quad + \frac{1}{2} L_{\theta}^1 U(1, y; \lambda_i) L_{\theta}^1 L_{\theta}^0 U(1, y; \lambda_i) \\ &\quad \left. + \frac{1}{2} L_{\theta}^1 U(1, y; \lambda_i) L_{\theta}^0 L_{\theta}^1 U(1, y; \lambda_i) \right) \bar{p}(y) dy \\ &\quad + \Delta^2 E_{\theta}(R_2(\boldsymbol{\beta}, 1, \Delta, \lambda_i)). \end{aligned} \quad (9.4.16)$$

Here $\bar{p}(y)$ is given by (9.4.4), and $E_{\theta}(R_j(\boldsymbol{\beta}, t, \Delta, \lambda_i))$, for $j \in \{1, 2\}$ and $i \in \{3, 4\}$ are finite. Similarly, we obtain

$$\begin{aligned} \lim_{n_t \rightarrow \infty} E_{\theta}(G_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, t, \Delta)) \\ = \Delta \int_{y_0}^{\infty} \frac{1}{2} (L_{\theta}^0 L_{\theta}^0 U(1, y; \lambda_i)) \bar{p}(y) dy + \Delta^{\frac{3}{2}} E_{\theta}(R_4(\boldsymbol{\alpha}, \boldsymbol{\beta}, 1, \Delta, \lambda_i)), \end{aligned} \quad (9.4.17)$$

where $E_{\theta}(R_4(\boldsymbol{\alpha}, \boldsymbol{\beta}, t, \Delta, \lambda_i))$, for $i \in \{1, 2\}$ are finite.

Estimating Equations

Note that if the time between observations Δ tends to zero, then the expectation of the functions H_i and G_i will approach zero. By setting the estimating functions to zero we obtain the linear system of four estimating equations

$$0 = A_{\Delta}^{0,1,0}(\lambda_i) - \hat{\theta}_3(\lambda_i)^2 A_{\Delta}^{2(\lambda_i-1)} - \hat{\theta}_4(\lambda_i)^2 A_{\Delta}^{2\lambda_i-1} \quad (9.4.18)$$

for $i \in \{3, 4\}$ and

$$0 = C_{\Delta}(\lambda_i, \hat{\theta}_3, \hat{\theta}_4) - \hat{\theta}_1 \lambda_i A_{\Delta}^{\lambda_i-1} - \hat{\theta}_2 \lambda_i A_{\Delta}^{\lambda_i} \quad (9.4.19)$$

for $i \in \{1, 2\}$.

As discussed previously, the values of λ_i , $i \in \{1, 2, 3, 4\}$ should ideally be chosen such that the estimator bias is minimized. Here we will choose them based on a more practical consideration. Intuitively, the values chosen for the λ_i , $i \in \{1, 2, 3, 4\}$, should remain small, since large values of λ_i would result in transform functions that produce unstable estimating functions with terms that may increase rapidly over time. Furthermore, simple explicit solutions of the system of equations (9.4.18) and (9.4.19) can be obtained by choosing small integer values. A convenient choice is $\lambda_1 = \lambda_3 = 1$, $\lambda_2 = \lambda_4 = 2$. Using these values for λ_i for $i \in \{1, 2, 3, 4\}$ we obtain the following four equations for the estimators,

$$\begin{aligned} \hat{\theta}_1 + A_{\Delta}^1 \hat{\theta}_2 &= C_{\Delta}(1, \hat{\theta}_3, \hat{\theta}_4) \\ A_{\Delta}^1 \hat{\theta}_1 + A_{\Delta}^2 \hat{\theta}_2 &= \frac{1}{2} C_{\Delta}(2, \hat{\theta}_3, \hat{\theta}_4) \\ \hat{\theta}_3 + A_{\Delta}^1 \hat{\theta}_4 &= A_{\Delta}^{0,1,0}(1) \\ A_{\Delta}^2 \hat{\theta}_3 + A_{\Delta}^3 \hat{\theta}_4 &= \frac{1}{4} A_{\Delta}^{0,1,0}(2). \end{aligned} \quad (9.4.20)$$

This system has the explicit solution

$$\begin{aligned} \hat{\theta}_1 &= C_{\Delta}(1, \hat{\theta}_3, \hat{\theta}_4) - A_{\Delta}^1 \hat{\theta}_2 \\ \hat{\theta}_2 &= \frac{\frac{1}{2} C_{\Delta}(2, \hat{\theta}_3, \hat{\theta}_4) - A_{\Delta}^1 C_{\Delta}(1, \hat{\theta}_3, \hat{\theta}_4)}{A_{\Delta}^2 - (A_{\Delta}^1)^2} \\ \hat{\theta}_3 &= A_{\Delta}^{0,1,0}(1) - A_{\Delta}^1 \hat{\theta}_4 \\ \hat{\theta}_4 &= \frac{\frac{1}{4} A_{\Delta}^{0,1,0}(2) - A_{\Delta}^2 A_{\Delta}^{0,1,0}(1)}{A_{\Delta}^3 - A_{\Delta}^1 A_{\Delta}^2}. \end{aligned} \quad (9.4.21)$$

This means, we have obtained explicit expressions for the estimators of the given class of affine diffusions using power transform functions. The illustrated transform function method can be extended to other classes of diffusions including nonergodic and multi-dimensional diffusions and the case with jumps. If no explicit solution of the system of estimating equations is available, then numerical solution techniques can be applied to identify their solution.

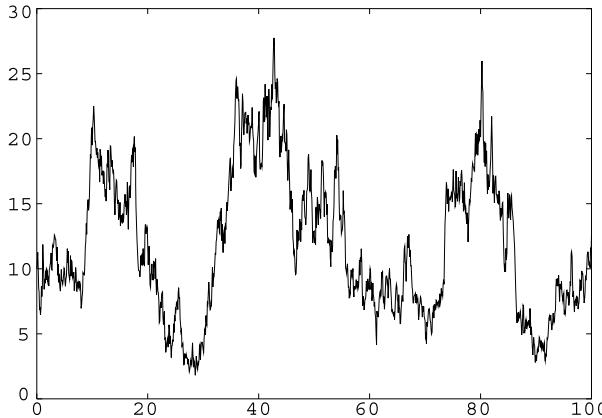


Fig. 9.4.1. Sample path of a square root process with $\theta_1 = \theta_4 = 1$, $\theta_2 = -0.1$ and $\theta_3 = 0$

Estimation of a Square Root Process

To illustrate the practical applicability of the proposed transform function method, we consider an example for the above affine diffusion given in (9.4.1) for the case $\theta_1 = \theta_4$ and $\theta_3 = 0$, which results in a square root process of dimension four. Sample paths of this process can be simulated using the exact method described in Chap. 2.

The affine diffusion was simulated with 1,200 steps over the time period $[0, T]$, with the parameters set to $\theta_1 = 1$, $\theta_2 = -0.1$, $\theta_3 = 0$, $\theta_4 = 1$, $T = 100$ and $X_0 = 10$, see Fig. 9.4.1. The parameters were then estimated from the almost exact simulated path shown in Fig. 9.4.1 by application of the estimators given in (9.4.21), using the small time step size $\Delta = 0.0833$. Consequently, we expect the estimator to be consistent.

The evolution of the estimators through time is shown in Figs. 9.4.2 and 9.4.3. The estimator of the diffusion parameter $\theta_4 = \theta_1$ is relatively stable as can be seen from Fig. 9.4.2. There is still substantial variability over time in the drift parameter estimator $\hat{\theta}_2$ as it evolves. It needs a rather long period of time to obtain a reasonably accurate estimate for θ_2 . This underlines the general fact that it is extremely difficult to estimate drift parameters in diffusions. Very long periods of observation are needed to identify with reasonable accuracy the value of a drift parameter.

9.5 Asymptotics of Estimating Functions

Stationary Diffusions

The above transform function method is designed to encompass both stationary and nonstationary diffusion processes. It is advantageous to analyze the

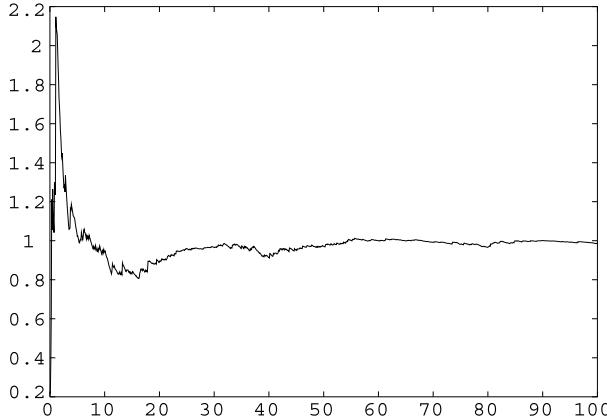


Fig. 9.4.2. Estimate of the diffusion parameter $\hat{\theta}_4 = \hat{\theta}_1$

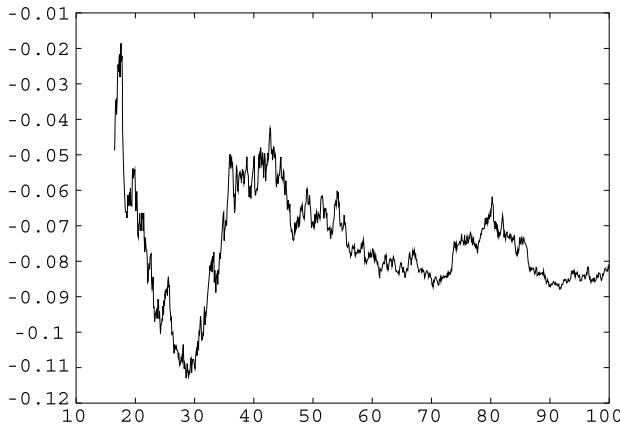


Fig. 9.4.3. Estimate for the drift parameter $\hat{\theta}_2$

asymptotic behavior and bias of the parameter estimators for some given class of stationary diffusion processes. We assume in this section that X is ergodic, see Sect. 2.7, described by the SDE (9.3.1) with time homogeneous coefficient functions

$$b(t, x; \boldsymbol{\theta}) = b(x, \boldsymbol{\theta}) \quad (9.5.1)$$

and

$$\sigma(t, x; \boldsymbol{\theta}) = \sigma(x; \boldsymbol{\theta}) \quad (9.5.2)$$

for $t \in [0, T]$, $x \in \mathfrak{R}$ and $\boldsymbol{\theta} \in \Theta$. We use as state space the interval (ℓ, r) where $-\infty \leq \ell < r \leq \infty$. For given parameter vector $\boldsymbol{\theta} \in \Theta$, the density of the scale measure $s : (\ell, r) \rightarrow [0, T]$ is given by the expression

$$s(x; \boldsymbol{\theta}) = \exp \left\{ -2 \int_{y_0}^x \frac{b(y; \boldsymbol{\theta})}{\sigma^2(y; \boldsymbol{\theta})} dy \right\} \quad (9.5.3)$$

for $x \in (\ell, r)$ with some reference value $y_0 \in (\ell, r)$, see Sect. 2.7. If the following two conditions

$$\int_{y_0}^r s(x; \boldsymbol{\theta}) dx = \int_{\ell}^{y_0} s(x; \boldsymbol{\theta}) dx = \infty \quad (9.5.4)$$

and

$$\int_{\ell}^r \frac{1}{s(x; \boldsymbol{\theta}) \sigma^2(x; \boldsymbol{\theta})} dx < \infty \quad (9.5.5)$$

are satisfied, then X is ergodic with stationary density

$$\bar{p}(x; \boldsymbol{\theta}) = \frac{C(\boldsymbol{\theta})}{\sigma^2(x; \boldsymbol{\theta})} \exp \left\{ 2 \int_{y_0}^x \frac{b(u; \boldsymbol{\theta})}{\sigma^2(u; \boldsymbol{\theta})} du \right\} \quad (9.5.6)$$

for $x \in (\ell, r)$ and $\boldsymbol{\theta} \in \Theta$, see (2.7.34). The constant $C(\boldsymbol{\theta})$ results from the normalization condition

$$\int_{\ell}^r \bar{p}(x; \boldsymbol{\theta}) dx = 1. \quad (9.5.7)$$

Conditions on the Diffusion

To prove the existence, consistency and asymptotic normality of the estimators we introduce the following conditions and notation where, essentially, we follow Sørensen (1999). We denote by $\tilde{\boldsymbol{\theta}}$ the true parameter value, where $\tilde{\boldsymbol{\theta}}$ is an interior point of Θ . The true probability measure is denoted by $P_{\tilde{\boldsymbol{\theta}}}$. Furthermore, let $p(\Delta, x, y; \tilde{\boldsymbol{\theta}})$ be the true transition density of the observed diffusion process X for a transition from x to y over a time period of length $\Delta > 0$. Throughout the remainder of this section we take Δ to be fixed. We consider estimating functions of the form

$$\mathbf{G}_t(\boldsymbol{\theta}) = \frac{1}{n_t} \sum_{n=1}^{n_t} \mathbf{g}(\Delta, X_{\tau_{n-1}}, X_{\tau_n}; \boldsymbol{\theta}), \quad (9.5.8)$$

for $t \in [0, T]$ and where \mathbf{G}_t and $\mathbf{g} = (g_1, g_2, \dots, g_p)^\top$ are p -dimensional. Furthermore, we assume that X is stationary and impose the condition that X is *geometrically α -mixing*. For a definition of this concept, see for instance Doukhan (1994). For a given ergodic diffusion process X , there exist a number of relatively simple criteria ensuring α -mixing with exponentially decreasing mixing coefficients. We cite the following rather weak set of conditions used in Genon-Catalot, Jeantheau & Laredo (2000) on the coefficients b and σ that are sufficient to ensure *geometric α -mixing* of X .

Condition 9.5.1

- (i) The function b is continuously differentiable and σ is twice continuously differentiable with respect to $x \in (\ell, r)$, $\sigma(x; \tilde{\theta}) > 0$ for all $x \in (\ell, r)$, and there exists a constant $K > 0$ such that $|b(x; \tilde{\theta})| \leq K(1 + |x|)$ and $\sigma^2(x; \tilde{\theta}) \leq K(1 + |x|^2)$ for all $x \in (\ell, r)$.
- (ii) $\sigma(x; \tilde{\theta})\bar{p}(x; \tilde{\theta}) \rightarrow 0$ as $x \downarrow \ell$ and $x \uparrow r$.
- (iii) $1/\gamma(x; \tilde{\theta})$ has a finite limit as $x \downarrow \ell$ and $x \uparrow r$, where

$$\gamma(x; \tilde{\theta}) = \frac{\partial \sigma(x; \tilde{\theta})}{\partial x} - \frac{2b(x; \tilde{\theta})}{\sigma(x; \tilde{\theta})}.$$

Each pair of neighboring observations $(X_{\tau_{n-1}}, X_{\tau_n})$ has the joint probability density

$$q_{\tilde{\theta}}^{\Delta}(x, y) = \bar{p}(x; \tilde{\theta}) p(\Delta, x, y; \tilde{\theta})$$

on $(\ell, r)^2$. For a function $\mathbf{f} : (\ell, r)^2 \rightarrow \mathbb{R}^k$, $k \in \{1, 2, \dots\}$, where we assume that the following integral exists, we introduce the vector valued functional

$$q_{\tilde{\theta}}^{\Delta}(\mathbf{f}) = \int_{\ell}^r \int_{\ell}^r \mathbf{f}(x, y) p(\Delta, x, y; \tilde{\theta}) \bar{p}(x; \tilde{\theta}) dy dx.$$

For our purposes we cannot assume that the estimating function in (9.5.8) is unbiased. Instead we make the following assumption.

Condition 9.5.2 *There exists a unique parameter value $\bar{\theta}$ that is an interior point of Θ such that*

$$q_{\tilde{\theta}}^{\Delta}(\mathbf{g}(\Delta, \bar{\theta})) = \mathbf{0}.$$

Note that the equation in Condition 9.5.2 is a vector equation with zero-vector on the right hand side. We can now impose conditions on the estimating function (9.5.8) similar to those by Barndorff-Nielsen & Sørensen (1994) and Sørensen (1999).

Condition 9.5.3

- (i) The function $g_i(\Delta, x, y; \cdot) : \Theta \rightarrow \mathbb{R}$ is twice continuously differentiable with respect to $\theta \in \Theta$ for all $x, y \in (\ell, r)$, and $i \in \{1, 2, \dots, p\}$.
- (ii) The function $g_i(\Delta, \cdot, \cdot; \theta) : (\ell, r) \times (\ell, r) \rightarrow \mathbb{R}$ is such that there exists a $\delta > 0$ with $q_{\tilde{\theta}}^{\Delta}(g_i(\Delta, \theta)^{2+\delta}) < \infty$ for all $\theta \in \Theta$ and $i \in \{1, 2, \dots, p\}$.
- (iii) For the partial derivatives $\frac{\partial}{\partial \theta_j} g_i(\Delta, x, y; \theta)$ and $\frac{\partial^2}{\partial \theta_j \partial \theta_k} g_i(\Delta, x, y; \theta)$, $i, j, k \in \{1, 2, \dots, p\}$, there exists for every $\theta^* \in \Theta$ a neighborhood $t_N(\theta^*) \subset \Theta$ of θ^* and a non-negative random variable $L(\theta^*)$ with $E_{\tilde{\theta}}(L(\theta^*)) < \infty$ such that $|\frac{\partial}{\partial \theta_j} g_i(\Delta, x, y; \theta)| \leq L(\theta^*)$ and $|\frac{\partial^2}{\partial \theta_j \partial \theta_k} g_i(\Delta, x, y; \theta)| \leq L(\theta^*)$ for all $\theta \in t_N(\theta^*)$, $(x, y) \in (\ell, r)^2$, and $i, j, k \in \{1, 2, \dots, p\}$.
- (iv) The $p \times p$ matrix

$$\mathbf{A}(\tilde{\theta}, \bar{\theta}) = \left\{ q_{\tilde{\theta}}^{\Delta} \left(\frac{\partial}{\partial \theta_j} g_i(\Delta, \cdot, \cdot; \bar{\theta}) \right) \right\}_{i,j=1}^p$$

is invertible.

Properties of Estimators

The following theorem can be found in [Kelly et al. \(2004\)](#).

Theorem 9.5.4. (Kelly-Platen-Sørensen) *Suppose Conditions 9.5.2 and 9.5.3 are satisfied. Then for $T > \Delta$, there exists an estimator $\hat{\theta}_T$ that solves the system of estimating equations*

$$\mathbf{G}_T(\hat{\theta}_T) = \mathbf{0}$$

with a probability tending to one as $n_T \rightarrow \infty$. Moreover, we have the limit in $P_{\tilde{\theta}}$ -probability

$$\lim_{n_T \rightarrow \infty} \hat{\theta}_T \xrightarrow{P_{\tilde{\theta}}} \bar{\theta}$$

and under $P_{\tilde{\theta}}$ the limit in distribution

$$\lim_{n_T \rightarrow \infty} \sqrt{n_T}(\hat{\theta}_T - \bar{\theta}) \xrightarrow{d} \mathbf{R},$$

where

$$\mathbf{R} \sim N\left(\mathbf{0}, \mathbf{A}(\tilde{\theta}, \bar{\theta})^{-1} \mathbf{v}(\tilde{\theta}, \bar{\theta})(\mathbf{A}(\tilde{\theta}, \bar{\theta})^{-1})^\top\right)$$

is a p -dimensional, zero mean Gaussian distributed random variable with covariance matrix $\mathbf{A}(\tilde{\theta}, \bar{\theta})^{-1} \mathbf{v}(\tilde{\theta}, \bar{\theta})(\mathbf{A}(\tilde{\theta}, \bar{\theta})^{-1})^\top$, where

$$\begin{aligned} \mathbf{v}(\tilde{\theta}, \bar{\theta}) &= \mathbf{q}_{\tilde{\theta}}^\Delta (\mathbf{g}(\Delta, \bar{\theta}) \mathbf{g}(\Delta, \bar{\theta})^\top) \\ &+ \sum_{k=1}^{\infty} \left\{ E_{\tilde{\theta}} (\mathbf{g}(\Delta, X_{\tau_0}, X_{\tau_1}; \bar{\theta}) \mathbf{g}(\Delta, X_{\tau_k}, X_{\tau_{k+1}}; \bar{\theta})^\top) \right. \\ &\quad \left. + E_{\tilde{\theta}} (\mathbf{g}(\Delta, X_{\tau_k}, X_{\tau_{k+1}}; \bar{\theta}) \mathbf{g}(\Delta, X_{\tau_0}, X_{\tau_1}; \bar{\theta})^\top) \right\}. \end{aligned} \quad (9.5.9)$$

Under the conditions imposed, the covariances in the infinite sum in (9.5.9) tend to zero exponentially fast as $k \rightarrow \infty$. So the sum converges rapidly and can usually be well approximated by a finite sum with relatively few terms.

The theorem can be derived in complete analogy with the proof of a similar theorem in [Sørensen \(1999\)](#).

9.6 Estimating Jump Diffusions

Parametric Jump Diffusion Model

The fact that jumps play an important role in the modeling of finance and insurance has led to an increasing array of literature which approaches also the question of parameter estimation for jump diffusions. For instance, [Sørensen \(1991\)](#) and [Ait-Sahalia \(2004\)](#) estimate via the maximum likelihood method

the presence and impact of jumps. Eraker, Johannes & Polson (2003) apply a Bayesian approach by using Markov chain Monte Carlo methods to infer jump parameters from index data. Chernov, Gallant, Ghysels & Tauchen (2003) apply the efficient method of moments. Barndorff-Nielsen & Shephard (2006) and Wörner (2006) study the behavior of interesting statistics, including power variations.

In the following we provide a first impression on the estimation of parameters for jump diffusions following Ait-Sahalia (2004). For simplicity, let us consider the Merton model, see Sect. 1.7 and Merton (1976). The logarithm of an asset price is given by the SDE

$$dX_t = \mu dt + \sigma dW_t + dY_t \quad (9.6.1)$$

with the compound Poisson process

$$Y_t = \sum_{i=1}^{N_t} J_i \quad (9.6.2)$$

modeling its jumps for $t \geq 0$ with $X_0 \in \mathfrak{R}$. Here $W = \{W_t, t \geq 0\}$ denotes a standard Wiener process and $N = \{N_t, t \geq 0\}$ an independent Poisson process with intensity $\lambda > 0$. Let the jump size J_i in the log-asset price of the i th jump at time τ_i , $i \in \{1, 2, \dots\}$, be independent Gaussian distributed with mean β and variance η . By application of the Itô formula the corresponding SDE for the asset price

$$S_t = S_0 \exp\{X_t\} \quad (9.6.3)$$

is of the form

$$dS_t = S_{t-} \left(\left(\mu + \frac{\sigma^2}{2} \right) dt + \sigma dW_t + (\exp\{\Delta Y_t\} - 1) dN_t \right) \quad (9.6.4)$$

for $t \geq 0$, where $\Delta Y_t = Y_t - Y_{t-}$ and, thus, $\Delta Y_{\tau_i} = J_i$.

The parameter vector in our problem is $\boldsymbol{\theta} = (\mu, \sigma^2, \lambda, \beta, \eta)^\top$. Assume that $\boldsymbol{\theta}$ is a vector in a bounded set $\Theta \subset \mathfrak{R}^5$.

Transition Density and Moments

As was pointed out previously, the transition density for the Merton model can be explicitly described. In particular, the log-returns of this simple model are independent and identically distributed. Recall that the distribution of the Poisson process is discrete and it holds

$$P(N_t = k | \boldsymbol{\theta}) = \exp\{-\lambda t\} \frac{(\lambda t)^k}{k!} \quad (9.6.5)$$

for $t \geq 0$ and $k \in \{1, 2, \dots\}$. When conditioning on the number of jumps between 0 and t and applying the Bayes formula, we obtain

$$P(X_t \leq x \mid X_0 = x_0, \boldsymbol{\theta}) = \sum_{k=0}^{\infty} P(X_t \leq x \mid X_0 = x_0, N_t = k, \boldsymbol{\theta}) P(N_t = k \mid \boldsymbol{\theta}) \quad (9.6.6)$$

for $x \in \Re$. In the event $N_t = k$ there have been k jumps in $[0, t]$ at the times $0 < \tau_1 < \tau_2 < \dots < \tau_k \leq t$. Since $J_i = \Delta Y_{\tau_i} \sim N(\beta, \eta)$ is Gaussian, the transition density of X_t given $X_0 = x_0$ is of the form

$$\begin{aligned} p(x \mid t, x_0, \boldsymbol{\theta}) &= \sum_{k=0}^{\infty} p(t, x \mid X_0 = x_0, N_t = k, \boldsymbol{\theta}) P(N_t = k \mid \boldsymbol{\theta}) \\ &= \sum_{k=0}^{\infty} \frac{\exp\{-\lambda t\} (\lambda t)^k}{\sqrt{2\pi(k\eta + t\sigma^2)} k!} \exp\left\{-\frac{(x - x_0 - \mu t - k\beta)^2}{2(k\eta + t\sigma^2)}\right\}, \end{aligned} \quad (9.6.7)$$

see Press (1967). When setting the parameters, care must be taken to ensure that the maximum likelihood for the above mixture of normals remains bounded, see Honoré (1998).

Press (1967) calculated the first four moments from the above transition density in the form:

$$\begin{aligned} m_t^1 &= E(X_t \mid \boldsymbol{\theta}) = x_0 + t(\mu + \beta\lambda) \\ M(t, \theta, 2) &= E((X_t - m_t^1)^2 \mid \boldsymbol{\theta}) = t(\sigma^2 + (\beta^2 + \eta)\lambda) \\ M(t, \theta, 3) &= E((X_t - m_t^1)^3 \mid \boldsymbol{\theta}) = t\lambda\beta(\beta^2 + 3\eta) \\ M(t, \theta, 4) &= E((X_t - m_t^1)^4 \mid \boldsymbol{\theta}) = t(\beta^4\lambda + 6\beta^2\eta\lambda + 3\eta^2\lambda) \\ &\quad + 3t^2(\sigma^2 + (\beta^2 + \eta)\lambda)^2. \end{aligned} \quad (9.6.8)$$

These moments are useful when calculating estimators.

General Method of Moments

We are now presenting a general method of moments (GMM) framework that will also accommodate the maximum likelihood method. We use an equidistant time discretization $(t)_\Delta$ with $t_n = n\Delta$ and time step size $\Delta > 0$, where $n_t = \max\{n \in \{0, 1, \dots\} : t_n \leq t\}$. We denote by

$$Z_{t_n} = X_{t_n} - X_{t_{n-1}} \quad (9.6.9)$$

the log-return of the asset price over the period from t_{n-1} until t_n , $n \in \{1, 2, \dots\}$. Recall that the log-returns are independent under the above Merton model.

For the estimation of the d -dimensional parameter vector $\boldsymbol{\theta}$ we introduce a vector of m moment conditions $\mathbf{h}(z, \Delta, \boldsymbol{\theta})$, $m \geq d$, where the vector function

\mathbf{h} is continuously differentiable in $\boldsymbol{\theta}$ but otherwise remains rather flexible. We then form a vector of sample averages

$$\mathbf{m}_T(\boldsymbol{\theta}) = \frac{1}{n_T} \sum_{n=1}^{n_T} \mathbf{h}(Z_{t_n}, \Delta, \boldsymbol{\theta}). \quad (9.6.10)$$

This allows us to consider the quadratic form

$$Q_T(\boldsymbol{\theta}) = \mathbf{m}_T(\boldsymbol{\theta})^\top \mathbf{G}_T \mathbf{m}_T(\boldsymbol{\theta}), \quad (9.6.11)$$

where \mathbf{G}_T is an $m \times m$ positive definite weight matrix converging in probability to a positive definite limit \mathbf{G} . In the case when the system is exactly identified, that is $m = d$, then the choice of G_T does not matter because minimizing $Q_T(\boldsymbol{\theta})$ amounts to setting $\mathbf{m}_T(\boldsymbol{\theta})$ to zero.

The estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ is obtained by minimizing $Q_T(\boldsymbol{\theta})$ in (9.6.11). To ensure consistency of $\hat{\boldsymbol{\theta}}$, the vector \mathbf{h} is assumed to satisfy the condition

$$E(\mathbf{h}(Z_{t_n}, \Delta, \boldsymbol{\theta}_0)) = \mathbf{0} \quad (9.6.12)$$

for $n \in \{1, 2, \dots\}$, where $\boldsymbol{\theta}_0 \in \Theta$ is the true parameter vector of the observed dynamics. We will denote by $\dot{\mathbf{h}}$ the gradient of \mathbf{h} with respect to $\boldsymbol{\theta}$. By application of standard statistical arguments, see for instance Hansen (1982), and subject to some regularity conditions it follows that $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to an $N(\mathbf{0}, \boldsymbol{\Omega})$ distributed random vector with

$$\boldsymbol{\Omega}^{-1} = \frac{1}{\Delta} \mathbf{D}^\top \mathbf{G} \mathbf{D} (\mathbf{D}^\top \mathbf{G} \mathbf{S} \mathbf{G} \mathbf{D})^{-1} \mathbf{D}^\top \mathbf{G} \mathbf{D}, \quad (9.6.13)$$

where

$$\mathbf{D} = E(\dot{\mathbf{h}}(Z_{t_1}, \Delta, \boldsymbol{\theta}_0)) \quad (9.6.14)$$

is an $m \times d$ matrix and

$$\mathbf{S} = E(\mathbf{h}(Z_{t_1}, \Delta, \boldsymbol{\theta}_0) \mathbf{h}(Z_{t_1}, \Delta, \boldsymbol{\theta}_0)^\top) \quad (9.6.15)$$

is an $m \times m$ matrix.

Note that the weight \mathbf{G}_T can be designed optimally to minimize the variance Ω . If G_T is chosen optimally with $m = d$, then one has $\mathbf{G} = \mathbf{S}^{-1}$ and (9.6.13) reduces to

$$\boldsymbol{\Omega}^{-1} = \frac{1}{\Delta} \mathbf{D}' \mathbf{S}^{-1} \mathbf{D}. \quad (9.6.16)$$

Furthermore, the moments (9.6.8) allow for the case when \mathbf{h} is a vector of polynomials the calculation of \mathbf{D} and \mathbf{S} . In this manner one can apply the GMM to construct estimators for any of the parameters. Of course, the form of the estimators depends on which parameters are considered to be unknown and which particular choices for \mathbf{h} are made, see Ait-Sahalia (2004).

Maximum Likelihood Estimation

We remark that the maximum likelihood method can be interpreted as a special case of the GMM: Let us denote by

$$\ell(z, \Delta, \boldsymbol{\theta}) = \ln(p(x_0 + z \mid \Delta, x_0, \boldsymbol{\theta})) \quad (9.6.17)$$

the log-likelihood function. As pointed out in [Ait-Sahalia \(2004\)](#), this corresponds to the choice of the score vector

$$h(z, \Delta, \boldsymbol{\theta}) = -\dot{\ell}(z, \Delta, \boldsymbol{\theta}) \quad (9.6.18)$$

and we obtain \mathbf{S} and \mathbf{D} from (9.6.14) and (9.6.15), respectively.

The asymptotic variance of the resulting maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{MLE}$ takes the form

$$\text{AVAR}_{MLE}(\boldsymbol{\theta}) = \Delta(\mathbf{D} \mathbf{S}^{-1} \mathbf{D})^{-1} = \Delta \mathbf{D}^{-1}. \quad (9.6.19)$$

Furthermore, it has been shown in [Ait-Sahalia \(2004\)](#) that when the asset price has jumps, the maximum likelihood estimator of σ^2 has for $\Delta \rightarrow 0$ the same asymptotic distribution as if no jumps were present. This means for instance, when using high-frequency data the presence of jumps will not hinder the estimation of the volatility.

Maximum likelihood estimation is usually the theoretically preferred methodology, however, already for the above simple model one cannot explicitly compute the maximum likelihood estimators. Numerically it can be obtained which can become rather time consuming. Therefore, alternative estimators, which are directly based on certain moments and easily computable, are of particular interest. For details on such GMM estimators we refer to [Ait-Sahalia \(2004\)](#). Further interesting work related to the estimation of jump diffusions can be found in [Ait-Sahalia, Fan & Peng \(2009\)](#), [Ait-Sahalia & Jacod \(2009\)](#) and [Bladt & Sørensen \(2005\)](#).

9.7 Exercises

9.1. Write down for the Vasicek interest rate model with SDE

$$dr_t = \gamma(\bar{r} - r_t) dt + \beta dW_t$$

a maximum likelihood estimator for the speed of adjustment parameter γ , when the interest rate reference level \bar{r} and the diffusion coefficient β are known constants.

9.2. Describe a simple way of estimating the volatility parameter σ for the Black-Scholes model for an asset price process $X = \{X(t), t \in [0, T]\}$ with SDE

$$dX_t = X_t(a dt + \sigma dW_t)$$

for $t \in [0, T]$, $X_0 > 0$.

Filtering

A very powerful approach that allows us to extract, in an adaptive manner, information from observed date is that of filtering. The aim of this chapter is to introduce *filtering* of information about hidden variables that evolve over time. These variables may follow continuous time hidden Markov chains or may satisfy certain hidden SDEs. Their observation is considered to be perturbed by the noise of Wiener or other processes. Approximate discrete-time filters driven by observation processes will be constructed for different purposes.

10.1 Kalman-Bucy Filter

In financial modeling or other applications one envisages in some situations hidden quantities that can only be observed indirectly through noisy observations or that only become fully revealed after some time. The procedure of determining an estimate of the hidden state of a system from noise contaminated observations is known as filtering. Filtering is becoming a crucial step in advanced estimation and calibration. It has been employed for solving optimal stochastic control problems in portfolio optimization, for risk measurement and for hedging, and will become a valuable tool in future generations of financial and other technologies. It can be used in many areas of application, such as the tracking of the growth rate or the inflation rate of an economy. To be of practical use, a filter needs to be robust and implementable on-line. This leads, in general, to challenging computational problems and requires a deep understanding of the underlying mathematical theory. There are special situations where adequate filters can be rather directly obtained and applied. The Kalman-Bucy filter, see [Kalman & Bucy \(1961\)](#), for linear Gaussian systems represents one of the cases where one can determine a filter explicitly without having to solve an SDE.

Linear Filter Problem

For simplicity, we shall give first a description of the *Kalman-Bucy filter* for state and observation equations with constant coefficients. In a filtered probability space $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ suppose that the *hidden state* X_t at time t forms a d -dimensional Gaussian process satisfying the linear state SDE

$$d\mathbf{X}_t = \mathbf{A} \mathbf{X}_t dt + \mathbf{B} d\mathbf{W}_t \quad (10.1.1)$$

for $t \in [0, T]$ with $\mathbf{X}_0 \in \mathbb{R}^d$, where \mathbf{A} is a $d \times d$ matrix, \mathbf{B} a $d \times m$ matrix and $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \in [0, T]\}$ an m -dimensional standard Wiener process. To ensure that the solution \mathbf{X}_t of (10.1.1) is Gaussian, the initial value \mathbf{X}_0 has also to be a Gaussian random variable. Then the mean vector and covariance matrix of \mathbf{X}_t satisfy some ordinary differential equations, see (1.7.23) and (1.7.24).

Suppose that the *observed process* $\mathbf{Y} = \{\mathbf{Y}_t = (Y_t^1, \dots, Y_t^r)^\top, t \in [0, T]\}$ is an r -dimensional process, where $1 \leq r \leq d$, and is related to the hidden state process $\mathbf{X} = \{\mathbf{X}_t = (X_t^1, \dots, X_t^d)^\top, t \in [0, T]\}$ by the observation SDE

$$d\mathbf{Y}_t = \mathbf{H} \mathbf{X}_t dt + \boldsymbol{\Gamma} d\mathbf{W}_t^* \quad (10.1.2)$$

for $t \in [0, T]$ with $\mathbf{Y}_0 = 0$, where \mathbf{H} is a $d \times e$ matrix, $\boldsymbol{\Gamma}$ an $e \times n$ matrix and $\mathbf{W}^* = \{\mathbf{W}_t^* = (W_t^{*,1}, \dots, W_t^{*,n})^\top, t \in [0, T]\}$ an n -dimensional standard Wiener process, which is independent of the Wiener process \mathbf{W} .

For each $t \in [0, T]$, let $\mathcal{A}_t = \sigma\{\mathbf{X}_0, \mathbf{Y}_s, \mathbf{W}_s : s \in [0, t]\}$ be the sigma-algebra generated by \mathbf{X}_0 , \mathbf{Y}_s and \mathbf{W}_s for $s \in [0, t]$, which expresses the *total information* up to time t . Furthermore, $\mathcal{Y}_t = \sigma\{\mathbf{Y}_s, s \in [0, t]\}$ is the sigma-algebra generated by the observations \mathbf{Y}_s for $s \in [0, t]$ and provides the *observable information* up to time t . Thus we have $\mathcal{Y}_t \subset \mathcal{A}_t$ for $t \in [0, T]$.

Kalman-Bucy Filter

The *Kalman-Bucy filter* is defined as the conditional expectation

$$\hat{\mathbf{X}}_t = E(\mathbf{X}_t | \mathcal{Y}_t), \quad (10.1.3)$$

which is the least-squares estimate of \mathbf{X}_t , given the observed information with

$$E\left(\left|\mathbf{X}_t - \hat{\mathbf{X}}_t\right|^2\right) \leq E\left(\left|\mathbf{X}_t - \mathbf{Z}_t\right|^2\right) \quad (10.1.4)$$

for all e -dimensional \mathcal{Y}_t -measurable random variables \mathbf{Z}_t . The Kalman-Bucy filter $\hat{\mathbf{X}}_t$ is \mathcal{Y}_t -measurable and hence \mathcal{A}_t -measurable. It is a Gaussian random variable with *error covariance matrix*

$$\mathbf{C}_t = E\left(\left(\mathbf{X}_t - \hat{\mathbf{X}}_t\right)\left(\mathbf{X}_t - \hat{\mathbf{X}}_t\right)^\top\right) \quad (10.1.5)$$

satisfies the *matrix Riccati equation*

$$\frac{d\mathbf{C}_t}{dt} = \mathbf{A}\mathbf{C}_t + \mathbf{C}_t\mathbf{A}^\top + \mathbf{B}\mathbf{B}^\top - \mathbf{C}_t\mathbf{H}^\top(\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top)^{-1}\mathbf{H}\mathbf{C}_t \quad (10.1.6)$$

for $t \in [0, T]$ with initial value $\mathbf{C}_0 = E(\mathbf{X}_0\mathbf{X}_0^\top)$. This equation differs from the linear equation (1.7.12) for the second moment $\mathbf{P}(t) = E(\mathbf{X}_t\mathbf{X}_t^\top)$ by its last term, which is a quadratic term. The Kalman-Bucy filter $\hat{\mathbf{X}}_t$ can be shown to satisfy the *filter SDE*

$$d\hat{\mathbf{X}}_t = \left(\mathbf{A} - \mathbf{C}_t\mathbf{H}^\top(\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top)^{-1}\mathbf{H} \right) \hat{\mathbf{X}}_t dt + \mathbf{C}_t\mathbf{H}^\top(\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top)^{-1} d\mathbf{Y}_t \quad (10.1.7)$$

for $t \in [0, T]$, see [Kallianpur \(1980\)](#) or [Liptser & Shiryaev \(1977\)](#). Here the observation process \mathbf{Y} appears as the driving process. Note that the stochastic differential in (10.1.7) with respect to \mathbf{Y} is defined as an Itô differential, see Sect. 1.4.

Filter as Least Square Estimate

The derivation of the equations (10.1.6) and (10.1.7) is based on the fact that the least-squares estimate of an element of a Hilbert space with respect to a subspace is its orthogonal projection onto the subspace. In the given case the appropriate Hilbert space consists of the square integrable processes $\mathbf{f} = \{\mathbf{f}_s, s \in [0, t]\}$ adapted to the *total filtration* $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0, T]}$ and the subspace is formed by those processes that are adapted to the *observation filtration* $\underline{\mathcal{Y}} = (\mathcal{Y}_t)_{t \in [0, T]}$. The *inner product* here is

$$(\mathbf{f}, \mathbf{g}) = E \left(\int_0^t \mathbf{f}_s^\top \mathbf{g}_s ds \right), \quad (10.1.8)$$

with \mathbf{f} and \mathbf{g} *orthogonal* if $(\mathbf{f}, \mathbf{g}) = 0$. The filter $\hat{\mathbf{X}}_t$ and its *filter error*

$$\varepsilon_t = \mathbf{X}_t - \hat{\mathbf{X}}_t \quad (10.1.9)$$

are orthogonal in this sense, and also in the stronger pointwise sense, such that

$$E \left(\left(\mathbf{X}_t - \hat{\mathbf{X}}_t \right)^\top \hat{\mathbf{X}}_t \right) = 0$$

for each $t \in [0, T]$, see [Kallianpur \(1980\)](#). Note that the Riccati equation (10.1.6) must, in general, be solved numerically, which can be performed off-line. It only involves the known coefficients of the state and observation SDEs (10.1.1) and (10.1.2). The coefficients of the filter equation (10.1.7) are, therefore, known. This SDE can be solved on-line as the observations become available. Generally, solving a filter SDE must be performed numerically using a strong discrete-time approximation. However, in a linear Gaussian situation

this can be avoided since (10.1.3) satisfies an explicit formula, as illustrated below.

It has been shown that the Kalman-Bucy filter is robust in the sense that the resulting filter value changes only slightly for small changes in the parameters. This is highly important in practice since the actual noise parameters are often not exactly known.

Constant Hidden Level

Let us consider the simple scalar case $d = e = n = 1$, $m = 0$ with Gaussian random but constant hidden level $X_t = X_0$, for $t \in [0, T]$, where $E(X_0) = 0$ and

$$E(X_0^2) = \sigma^2. \quad (10.1.10)$$

Then the state SDE is trivial with

$$dX_t = 0 \quad (10.1.11)$$

for $t \in [0, T]$ and we have $A = B = 0$. Assume that the observation SDE (10.1.2) is of the form

$$dY_t = X_t dt + dW_t^* \quad (10.1.12)$$

with $Y_0 = 0$ and $H = \Gamma = 1$. Then we obtain the Riccati equation

$$dC_t = -(C_t)^2 dt \quad (10.1.13)$$

for $t \in [0, T]$ with initial value $C_0 = \sigma^2$. Fortunately, this Riccati equation has an explicit solution

$$C_t = \frac{\sigma^2}{1 + \sigma^2 t}. \quad (10.1.14)$$

In the given example one can show that the filter has the explicit form

$$\hat{X}_t = C_t Y_t \quad (10.1.15)$$

for $t \in [0, T]$ since this is the explicit solution of the corresponding filter SDE (10.1.7).

To see this, let us introduce the quantity

$$J_t = \exp \left\{ \int_0^t C_s ds \right\}. \quad (10.1.16)$$

Then one obtains from (10.1.7) and (10.1.16) by the Itô formula the SDE

$$d(\hat{X}_t J_t) = C_t J_t dY_t \quad (10.1.17)$$

for $t \in [0, T]$. Note that by (10.1.13) and (10.1.16) one has

$$\begin{aligned} \frac{d(C_t J_t)}{dt} &= J_t \frac{dC_t}{dt} + C_t \frac{dJ_t}{dt} \\ &= J_t (C_t)^2 - J_t (C_t)^2 = 0 \end{aligned} \quad (10.1.18)$$

for all $t \in [0, T]$. On the other hand, one has by the Itô formula

$$(C_t J_t) Y_t = (C_0 J_0) Y_0 + \int_0^t (C_s J_s) dY_s + \int_0^t Y_s d(C_s J_s) \quad (10.1.19)$$

for $t \in [0, T]$. Using the fact that $\hat{X}_0 = Y_0 = 0$ we obtain by using (10.1.19), (10.1.18) and (10.1.17) the equation

$$\begin{aligned} C_t J_t Y_t &= \int_0^t C_s J_s dY_s \\ &= \int_0^t d(\hat{X}_s J_s) \\ &= \hat{X}_t J_t \end{aligned} \quad (10.1.20)$$

for $t \in [0, T]$. Consequently, it follows from (10.1.20) that the Kalman-Bucy filter has in the given case with constant hidden Gaussian level by (10.1.14) the form

$$\hat{X}_t = C_t Y_t = \frac{\sigma^2}{1 + \sigma^2 t} Y_t \quad (10.1.21)$$

for $t \in [0, T]$, as was stated in (10.1.15). Note that for short observation time $t \ll 1$ we obtain approximately the filter

$$\hat{X}_t \approx \sigma^2 Y_t.$$

In this case the least-squares estimate of X_t at time t equals approximately $\sigma^2 Y_t$ and is thus almost proportional to the variance of the hidden level. This means, for higher variance of the hidden constant level the level of the filter is more extreme.

In using the Kalman-Bucy filter one essentially estimates adaptively the conditional probability distribution of the hidden state $X_t = X_0$ via the estimate \hat{X}_t .

Simulation of a Kalman-Bucy Filter

Let us illustrate the application of a Kalman-Bucy filter for the above example with constant hidden level. For visualization purposes we fix the hidden levels in the following eleven cases, where X_0 is set to $-2, -1.5, -1, -0.5, -0.25, 0, 0.25, 0.5, 1, 1.5, 2$. We interpret these levels as outcomes of a Gaussian distribution with mean zero and variance $\sigma^2 = 1.2$. Then we simulate a Wiener path W^* and compute for each of the levels the corresponding observation Y_t for

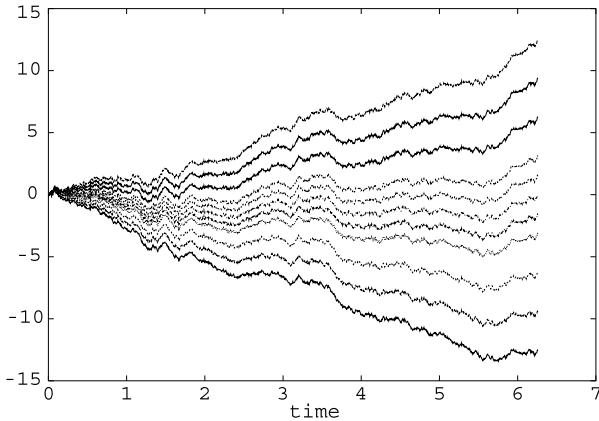


Fig. 10.1.1. Trajectories for observations for different hidden levels

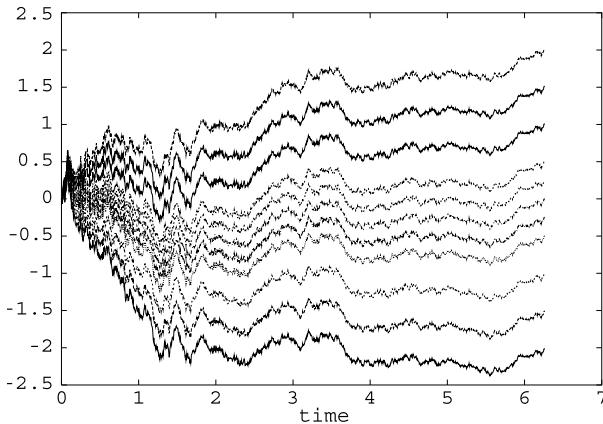


Fig. 10.1.2. Trajectories of the Kalman-Bucy filter for different hidden levels

$t \in [0, 6.25]$. These observed paths for the 11 cases are plotted in Fig. 10.1.1. They show the trends caused by the corresponding different hidden levels. It then remains to calculate, according to (10.1.21), the paths for the corresponding Kalman-Bucy filter. These are displayed in Fig. 10.1.2. They demonstrate clearly the functioning of the filter. After some initial time the Kalman-Bucy filter \hat{X}_t has approximately detected in each of the cases the hidden level.

10.2 Hidden Markov Chain Filters

The above Kalman-Bucy filter is exploiting the rather simple linear Gaussian dynamics. Sometimes, this is not a realistic assumption for a given situation. The aim of this section is to construct approximate discrete-time filters for

hidden, continuous time finite-state Markov chains with observations that are perturbed by the noise of a Wiener process.

The Wonham Filter Problem

The systematic construction and investigation of filters for hidden Markov chains goes back to [Wonham \(1965\)](#), [Zakai \(1969\)](#) and Fujisaki, Kallianpur & Kunita ([1972](#)). Later, the question of finding discrete-time approximations for the optimal filter was considered by [Clark & Cameron \(1980\)](#) and [Newton \(1986a, 1991\)](#).

Now, we introduce filters for hidden, continuous time, finite state Markov chains. Let $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ be the underlying filtered probability space and suppose that the *hidden state process* $\xi = \{\xi_t, t \in [0, T]\}$ is a continuous time, homogeneous Markov chain on the finite state space $\mathcal{X} = \{a_1, a_2, \dots, a_d\}$, $d \in \{0, 1, \dots\}$. This could model, for instance, the hidden trend of the logarithms of a group of commodity prices. Its d -dimensional probability vector $\mathbf{p}(t) = (p_1(t), \dots, p_d(t))^\top$ at time t , with components

$$p_i(t) = P(\xi_t = a_i) \quad (10.2.1)$$

for each $a_i \in \mathcal{X}$, satisfies then the vector ODE

$$\frac{d\mathbf{p}(t)}{dt} = \mathbf{A}\mathbf{p}(t), \quad (10.2.2)$$

where \mathbf{A} is the intensity matrix. In addition, suppose that the m -dimensional *observation process* $\mathbf{W} = \{\mathbf{W}_t, t \in [0, T]\}$ is the solution of the SDE

$$d\mathbf{W}_t = \mathbf{h}(\xi_t) dt + d\mathbf{W}_t^* \quad (10.2.3)$$

for $t \in [0, T]$ with $\mathbf{W}_0 = \mathbf{W}_0^* \in \Re^m$. This type of disturbance of a signal by a Wiener process is called a *Wonham filter problem*, see [Wonham \(1965\)](#).

The observation process could, for instance, represent the short rate that a central bank chooses. In the SDE (10.2.3) the noise process $\mathbf{W}^* = \{\mathbf{W}_t^*, t \in [0, T]\}$ with $\mathbf{W}_0^* = 0$ is an m -dimensional standard Wiener process with respect to the real world probability measure P . The Wiener process \mathbf{W}^* is assumed to be independent of the hidden state process ξ . Finally, let

$$\mathcal{Y}_t = \sigma \{\mathbf{W}(s), s \in [0, t]\}$$

denote the observation sigma-algebra generated by the observations \mathbf{W}_s for $s \in [0, t]$. This means, $\underline{\mathcal{Y}} = (\mathcal{Y}_t)_{t \in [0, T]}$ is the filtration that represents the release of observed information, whereas $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0, T]}$ with $\mathcal{A}_t = \sigma\{\xi_s, \mathbf{W}_s : s \in [0, t]\}$ expresses the evolution of the total information generated in the market.

Our task is to filter as much information about the hidden state process ξ as we can from the observation process \mathbf{W} . With this aim we shall evaluate

for a given function $g : \mathcal{X} \rightarrow \mathbb{R}$ the *Wonham filter*, which is the conditional expectation

$$\hat{g}(\xi_T) = E(g(\xi_T) | \mathcal{Y}_T)$$

with respect to the real world probability P . The function $g(\cdot)$ could be, for instance, an indicator function $\mathbf{1}_{\{\xi_T=a_i\}}$, which yields $\hat{g}(\xi_T)$ as a probability. It could also be a power function $(\xi_T)^q$, which leads to $\hat{g}(\xi_T)$ describing the q th moment of the hidden Markov chain.

Zakai Equation

By application of the Girsanov transformation, see Sect. 2.7, one obtains a probability measure \dot{P} , where

$$d\dot{P} = L_T^{-1} dP \quad (10.2.4)$$

with

$$L_T = \exp \left\{ -\frac{1}{2} \int_0^T |\mathbf{h}(\xi_s)|^2 ds + \int_0^T \mathbf{h}(\xi_s)^\top d\mathbf{W}_s \right\} \quad (10.2.5)$$

such that W is a Wiener process with respect to \dot{P} . Here $L_T = \frac{dP}{d\dot{P}}$ is the corresponding Radon-Nikodym derivative. Note that we express in this situation the real world probability measure P in terms of the new probability measure \dot{P} for which \mathbf{W} is a standard vector Wiener process by (10.2.3).

Let us introduce the *unnormalized conditional probability* $\sigma(\xi_t)^i$ for the state $a_i \in \mathcal{X}$ at time t by the conditional expectation

$$\sigma(\xi_t)^i = \dot{E}(\mathbf{1}_{\{\xi_t=a_i\}} L_t | \mathcal{Y}_t) \quad (10.2.6)$$

with respect to the new probability measure \dot{P} for $i \in \{1, 2, \dots, d\}$ and $t \in [0, T]$. It follows from Bayes' rule, see Sect. 2.7, which is in filtering also known as *Kallianpur-Striebel formula*, see Fujisaki, Kallianpur & Kunita (1972), that the conditional probabilities of ξ_t given in \mathcal{Y}_t are

$$P(\xi_t = a_i | \mathcal{Y}_t) = E(\mathbf{1}_{\{\xi_t=a_i\}} | \mathcal{Y}_t) = \frac{\sigma(\xi_t)^i}{\sum_{k=1}^d \sigma(\xi_t)^k} \quad (10.2.7)$$

for $a_i \in \mathcal{X}$ and $t \in [0, T]$. Here the d -dimensional process $\sigma(\xi) = \{\sigma(\xi_t) = (\sigma(\xi_t)^1, \dots, \sigma(\xi_t)^d)^\top, t \in [0, T]\}$ of the unnormalized conditional probabilities satisfies the *Zakai equation*

$$\sigma(\xi_t) = \mathbf{p}(0) + \int_0^t \mathbf{A} \sigma(\xi_s) ds + \sum_{j=1}^m \int_0^t \mathbf{H}_j \sigma(\xi_s) d\mathbf{W}_s^j \quad (10.2.8)$$

for $t \in [0, T]$. This is a homogeneous linear Itô SDE. In (10.2.8) \mathbf{H}_j is the $d \times d$ diagonal matrix with (i, i) th component $h_j(a_i)$ for $i \in \{1, 2, \dots, d\}$ and $j \in \{1, 2, \dots, m\}$.

The least-squares estimate at time t for $g(\xi_t)$ with respect to the given observations at time t , that is with respect to the sigma-algebra \mathcal{Y}_t , is then the *Wonham filter* and given by the conditional expectation

$$\hat{g}(\xi_t) = E(g(\xi_t) \mid \mathcal{Y}_t) = \frac{\sum_{k=1}^d g(a_k) \sigma(\xi_t)^k}{\sum_{k=1}^d \sigma(\xi_t)^k} \quad (10.2.9)$$

for $t \in [0, T]$.

Quasi-exact Filters

Let us consider the following d-dimensional multiplicative noise SDE

$$d\mathbf{X}_t = \mathbf{A}\mathbf{X}_t dt + \sum_{k=1}^m \mathbf{H}_k \mathbf{X}_t dW_t^k, \quad (10.2.10)$$

with a solution that is representing a vector geometric Brownian motion, where $\mathbf{X} = \{\mathbf{X}_t = (X_t^1, X_t^2, \dots, X_t^d)^\top, t \in [0, \infty)\}$, $\mathbf{A} = [a^{i,j}]_{i,j=1}^d$ and $\mathbf{H}_k = [H_k^{i,j}]_{i,j=1}^d$, $k \in \{1, 2, \dots, m\}$. Here, W^k , $k \in \{1, 2, \dots, m\}$, are the elements of the vector SDE (10.2.3), that describes the observation process.

It turns out that if the matrices $\mathbf{A}, \mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_m$ are constant and commute, that is, if

$$\mathbf{A}\mathbf{H}_k = \mathbf{H}_k\mathbf{A} \quad \text{and} \quad \mathbf{H}_k\mathbf{H}_n = \mathbf{H}_n\mathbf{H}_k \quad (10.2.11)$$

for all $k, n \in \{1, 2, \dots, m\}$, then an explicit solution of the SDE (10.2.10) can be expressed by

$$\mathbf{X}_t = \Psi_t \mathbf{X}_0, \quad (10.2.12)$$

for $t \in [0, \infty)$, see (2.4.9). Here, Ψ_t is the matrix exponential

$$\Psi_t = \exp \left\{ \mathbf{A}t - \frac{1}{2} \sum_{\ell=1}^m (\mathbf{H}_\ell)^2 t + \sum_{r=1}^m \mathbf{H}_r W_t^r \right\}, \quad (10.2.13)$$

for $t \in [0, \infty)$.

The above derivation shows that an SDE of the type (10.2.8) has an explicit solution if the matrices $\mathbf{A}, \mathbf{H}_1, \dots, \mathbf{H}_m$ commute. Note that $\mathbf{H}_1, \dots, \mathbf{H}_m$ in (10.2.8) are diagonal matrices, and thus commute with each other. However, the matrix \mathbf{A} is not commuting with the other matrices. Therefore, we do not have an exact explicit solution of the Zakai equation (10.2.8). Nevertheless, as illustrated in Platen & Rendek (2009b), if we formally take the matrix exponential (10.2.13) in the product (10.2.12), then one obtains a proxy of the solution of the corresponding Zakai equation. It turns out that this quasi-exact solution provides in many cases an excellent approximation of the exact solution. This is a practically very valuable observation. The solution can be exploited to solve approximately and efficiently the Wonham filter

problem. What, of course, needs to be done is to verify for given matrices $\mathbf{A}, \mathbf{H}_1, \dots, \mathbf{H}_m$ and initial vector \mathbf{X}_0 that the quasi-exact solution is indeed close to the exact solution. This can be checked via simulation by comparing the proposed approximation with a very accurately obtained numerical approximation using an extremely small time step size, as will be described below.

By analogy we then assume that a similar quality of approximation is obtained also when using real data.

Approximate Filter

To compute the Wonham filter (10.2.9) one has to solve the Itô SDE (10.2.8), which does not have an explicit solution, see Chap. 2.

In practice, it is impossible to detect \mathbf{W} continuously on $[0, T]$. Using, for instance, intraday data in finance one could approximate increments of observations of \mathbf{W} in integral form of the type

$$\int_{t_0}^{\tau_1} dW_s^j, \dots, \int_{t_n}^{\tau_{n+1}} dW_s^j, \dots, \int_{t_0}^{\tau_1} \int_{t_0}^{s_2} dW_{s_1}^j dW_{s_2}^k, \dots$$

for each $j \in \{1, 2, \dots, m\}$, $\tau_n = n\Delta$ and $n \in \{0, 1, \dots\}$. We shall see later on that with such integral observations it is possible to construct strong discrete-time approximations \mathbf{Y}^Δ with time step Δ of the solution $\boldsymbol{\sigma}(\xi)$ of the Zakai equation (10.2.8). This allows then for the given function g to form the *approximate Wonham filter*

$$\hat{g}^\Delta(\xi_t) = \frac{\sum_{k=1}^d g(a_k) Y_t^{\Delta,k}}{\sum_{k=1}^d Y_t^{\Delta,k}} \quad (10.2.14)$$

for $t \in [0, T]$.

As in Chap. 5 we shall say that a discrete-time approximation \mathbf{Y}^Δ with time step size Δ converges on the time interval $[0, T]$ with strong order $\gamma > 0$ to the solution \mathbf{X} of the corresponding SDE if there exists a finite constant K , not depending on Δ , and a $\delta_0 \in (0, 1)$ such that

$$\dot{E} \left(\left| \boldsymbol{\sigma}(\xi_{\tau_n}) - \mathbf{Y}_{\tau_n}^\Delta \right| \right) \leq K \Delta^\gamma \quad (10.2.15)$$

for all $\Delta \in (0, \delta_0)$ and $\tau_n \in [0, T]$. Note that the expectation in (10.2.15) is taken with respect to the probability measure \dot{P} under which the observation process \mathbf{W} is a Wiener process.

Analogously, we say that an approximate Markov chain filter $\hat{g}^\Delta(\xi_{\tau_{n_t}})$ with time step size Δ converges on the time interval $[0, T]$ with strong order $\gamma > 0$ to the optimal filter $\hat{g}(\xi_{\tau_{n_t}})$ for a given test function g if there exists a finite constant K , not depending on Δ , and a $\delta_0 \in (0, 1)$ such that

$$E \left(|\hat{g}(\xi_{\tau_{n_t}}) - \hat{g}^\Delta(\xi_{\tau_{n_t}})| \right) \leq K \Delta^\gamma \quad (10.2.16)$$

for all $\Delta \in (0, \delta_0)$ and $t \in [0, T]$. In contrast with (10.2.15) the expectation in (10.2.16) is taken with respect to the original probability measure P . In Kloeden, Platen & Schurz (1993) the following convergence result was derived.

Theorem 10.2.1. (Kloeden-Platen-Schurz) *An approximate Markov chain filter $\hat{g}^\Delta(\xi_{\tau_n})$ with time step size Δ converges for $t \in [0, T]$ with strong order $\gamma > 0$ to the optimal filter $\hat{g}(\xi_{\tau_n})$ for a given bounded function g if the discrete-time approximation \mathbf{Y}^Δ used converges on $[0, T]$ to the solution $\sigma(\xi)$ of the Zakai equation (10.2.8) with the same strong order γ .*

Proof: In view of the criterion (10.2.16) and the Bayes rule, see Sect. 2.7, we need to estimate the strong approximation error

$$\begin{aligned} F_{\tau_n}^\Delta(g) &= E(|\hat{g}(\xi_{\tau_n}) - \hat{g}^\Delta(\xi_{\tau_n})|) \\ &= \dot{E}(L_{\tau_n} |\hat{g}(\xi_{\tau_n}) - \hat{g}^\Delta(\xi_{\tau_n})|) \end{aligned} \quad (10.2.17)$$

for all $\tau_n \in [0, T]$. We shall write

$$G_{\tau_n}(f) = \sum_{k=1}^d f(a_k) X_{\tau_n}^k \quad (10.2.18)$$

and

$$G_{\tau_n}^\Delta(f) = \sum_{k=1}^d f(a_k) Y_{\tau_n}^{\Delta,k} \quad (10.2.19)$$

for any bounded function $f : \mathcal{X} \rightarrow \mathbb{R}$, $\Delta \in (0, \delta_0)$ and $\tau_n \in [0, T]$. Then, similarly to Picard (1984), we can use (10.2.6), (10.2.9) and (10.2.14) to rewrite the error (10.2.17) in the form

$$\begin{aligned} F_{\tau_n}^\Delta(g) &= \dot{E}(G_{\tau_n}(1) |\hat{g}(\xi_{\tau_n}) - \hat{g}^\Delta(\xi_{\tau_n})|) \\ &= \dot{E}\left(G_{\tau_n}(1) \left| \frac{1}{G_{\tau_n}(1)} (G_{\tau_n}(g) - G_{\tau_n}^\Delta(g) + \hat{g}^\Delta(\xi_{\tau_n}) (G_{\tau_n}^\Delta(1) - G_{\tau_n}(1))) \right| \right) \\ &\leq \dot{E}(|G_{\tau_n}(g) - G_{\tau_n}^\Delta(g)|) + \dot{E}(|\hat{g}^\Delta(\xi_{\tau_n})| |G_{\tau_n}^\Delta(1) - G_{\tau_n}(1)|) \\ &\leq K_1 \sum_{k=1}^d \dot{E}(|Y_{\tau_n}^{\Delta,k} - \sigma(\xi_{\tau_n})^k|). \end{aligned} \quad (10.2.20)$$

Finally, using (10.2.15) in (10.2.20) gives the estimate

$$F_{\tau_n}^\Delta(g) \leq K_2 \Delta^\gamma$$

and hence the desired strong order of convergence γ . \square

Explicit Filters

Now, we derive discrete-time strong approximations \mathbf{Y}^Δ that converge with a given strong order $\gamma > 0$ to the solution $\sigma(\xi)$ of the Zakai equation (10.2.8), which can be used to build a corresponding approximate filter. A systematic presentation of such discrete-time strong approximations is provided in Chap. 5.

Given an equidistant time discretization of the interval $[0, T]$ with step size $\Delta = \frac{T}{N}$ for some $N \in \{1, 2, \dots\}$, we define the partition sigma-algebra

$$\mathcal{P}_N^1 = \sigma\{\Delta W_{i-1}^j : i \in \{1, 2, \dots, N\}, j \in \{1, 2, \dots, m\}\} \quad (10.2.21)$$

as the sigma-algebra generated by the increments

$$\Delta W_0^j = \int_0^\Delta dW_s^j, \dots, \Delta W_{N-1}^j = \int_{(N-1)\Delta}^{N\Delta} dW_s^j \quad (10.2.22)$$

for all $j \in \{1, 2, \dots, m\}$. Thus, \mathcal{P}_N^1 contains the information about the increments of \mathbf{W} for the given time discretization.

The simplest discrete-time approximation is obtained from the Euler scheme, see (5.3.13). It has for the Zakai equation (10.2.8) the form

$$\mathbf{Y}_{\tau_{n+1}}^\Delta = [\mathbf{I} + \mathbf{A} \Delta + \mathbf{G}_n] \mathbf{Y}_{\tau_n}^\Delta \quad (10.2.23)$$

with

$$\mathbf{G}_n = \sum_{j=1}^m \mathbf{H}_j \Delta W_n^j \quad (10.2.24)$$

and initial value $\mathbf{Y}_0 = \sigma(\xi_0)$, where \mathbf{I} is the $d \times d$ unit matrix. The scheme (10.2.23) converges with strong order $\gamma = 0.5$ under the given assumptions. For a general SDE this is the maximum order of strong convergence that can be achieved under the partition sigma-algebra \mathcal{P}_N^1 , as was shown in Clark & Cameron (1980). However, the diffusion commutativity property, see (5.3.27), of the Zakai equation (10.2.8) follows from the diagonal structure of its volatility matrices. This allows the strong order $\gamma = 1.0$ to be attained with the information given by \mathcal{P}_N^1 .

The Milstein scheme, which is of strong order $\gamma = 1.0$, see (5.3.18), has for the Zakai equation (10.2.8) the form

$$\mathbf{Y}_{\tau_{n+1}}^\Delta = \left(\mathbf{I} + \underline{\mathbf{A}} \Delta + \mathbf{G}_n \left(\mathbf{I} + \frac{1}{2} \mathbf{H}_n \right) \right) \mathbf{Y}_{\tau_n}^\Delta, \quad (10.2.25)$$

where

$$\underline{\mathbf{A}} = \mathbf{A} - \frac{1}{2} \sum_{j=1}^m \mathbf{H}_j^2. \quad (10.2.26)$$

Newton (1986a) searched for a scheme which is asymptotically the “best” in the class of order $\gamma = 1.0$ schemes in the sense that it has the smallest leading

error coefficient in an error estimate similar to (10.2.15). He obtained the scheme

$$\begin{aligned} \mathbf{Y}_{\tau_{n+1}}^{\Delta} &= \left(\mathbf{I} + \underline{\mathbf{A}} \Delta + \mathbf{G}_n + \frac{\Delta^2}{2} \mathbf{A}^2 + \frac{\Delta}{2} \mathbf{A} \mathbf{G}_n \right. \\ &\quad \left. - \frac{\Delta}{2} \mathbf{G}_n \mathbf{A} + \mathbf{G}_n \underline{\mathbf{A}} \Delta + \frac{1}{2} \mathbf{G}_n^2 + \frac{1}{6} \mathbf{G}_n^3 \right) \mathbf{Y}_{\tau_n}^{\Delta}, \end{aligned} \quad (10.2.27)$$

which is called asymptotically efficient under \mathcal{P}_N^1 . A scheme with similar properties was proposed in [Burrage \(1998\)](#).

One can obtain higher order strong convergence for $\gamma > 1$ by exploiting additional information about the observation process \mathbf{W} such as that contained in the integral observations

$$\Delta Z_0^j = \int_0^{\Delta} \int_0^s dW_r^j ds, \quad \dots, \quad \Delta Z_{N-1}^j = \int_{(N-1)\Delta}^{N\Delta} \int_{(N-1)\Delta}^s dW_r^j ds \quad (10.2.28)$$

for all $j \in \{1, 2, \dots, m\}$. We shall define as partition sigma-algebra

$$\mathcal{P}_N^{1.5} = \sigma \left\{ \Delta Z_{i-1}^j, \Delta W_{i-1}^j : i \in \{1, 2, \dots, N\}, j \in \{1, 2, \dots, m\} \right\} \quad (10.2.29)$$

the sigma-algebra generated by \mathcal{P}_N^1 together with the multiple stochastic integrals $\Delta Z_0^j, \dots, \Delta Z_{N-1}^j$ for all $j \in \{1, 2, \dots, m\}$.

The order 1.5 strong Taylor scheme, described in (5.3.35) uses, due to the diffusion commutativity property of the Zakai equation (10.2.8), only the information contained in $\mathcal{P}_N^{1.5}$. It takes the form

$$\begin{aligned} \mathbf{Y}_{\tau_{n+1}}^{\Delta} &= \left(\mathbf{I} + \underline{\mathbf{A}} \Delta + \mathbf{G}_n + \frac{\Delta^2}{2} \mathbf{A}^2 + \mathbf{A} \mathbf{M}_n \right. \\ &\quad \left. - \mathbf{M}_n \mathbf{A} + \mathbf{G}_n \underline{\mathbf{A}} \Delta + \frac{1}{2} \mathbf{G}_n^2 + \frac{1}{6} \mathbf{G}_n^3 \right) \mathbf{Y}_{\tau_n}^{\Delta}, \end{aligned} \quad (10.2.30)$$

where

$$\mathbf{M}_n = \sum_{j=1}^m \mathbf{H}_j \Delta Z_n^j. \quad (10.2.31)$$

Note that one obtains the strong order 1.0 scheme (10.2.27) from (10.2.30) if one replaces the ΔZ_n^j by their conditional expectations under \mathcal{P}_N^1 with respect to the probability measure \dot{P} , that is if one substitutes $\frac{1}{2} \mathbf{G}_n \Delta$ for \mathbf{M}_n in (10.2.30).

In order to form a scheme of strong order $\gamma = 2.0$ one needs the information from the observation process expressed in the partition sigma-algebra

$$\begin{aligned} \mathcal{P}_N^2 &= \sigma \left\{ \Delta W_{i-1}^{j_1}, \Delta Z_{i-1}^{j_1}, J_{(j_1, j_2, 0), i-1}, J_{(j_1, 0, j_2), i-1} : i \in \{1, 2, \dots, N\}, \right. \\ &\quad \left. j \in \{1, 2, \dots, m\} \right\} \end{aligned} \quad (10.2.32)$$

which is generated by $\mathcal{P}_N^{1.5}$ together with the multiple Stratonovich integrals

$$\begin{aligned} J_{(j_1, j_2, 0), n} &= \int_{\tau_n}^{\tau_{n+1}} \int_{\tau_n}^{s_3} \int_{\tau_n}^{s_2} \circ dW_{s_1}^{j_1} \circ dW_{s_2}^{j_2} ds_3, \\ J_{(j_1, 0, j_2), n} &= \int_{\tau_n}^{\tau_{n+1}} \int_{\tau_n}^{s_3} \int_{\tau_n}^{s_2} \circ dW_{s_1}^{j_1} ds_2 \circ dW_{s_3}^{j_2} \end{aligned} \quad (10.2.33)$$

for all $n \in \{0, 1, \dots, N-1\}$ and $j_1, j_2 \in \{1, 2, \dots, m\}$. One may be able to extract these Stratonovich integrals from the observation of intraday financial data in practical filtering situations. Using this information one can then apply the order 2.0 strong Taylor scheme (5.3.41) for the Zakai equation (10.2.8) to obtain the order $\gamma = 2.0$ strong approximation

$$\begin{aligned} \mathbf{Y}_{\tau_{n+1}}^\Delta &= \left[\mathbf{I} + \underline{\mathbf{A}} \Delta \left(\mathbf{I} + \frac{1}{2} \underline{\mathbf{A}} \Delta \right) - \mathbf{M}_n \underline{\mathbf{A}} + \underline{\mathbf{A}} \mathbf{M}_n \right. \\ &\quad + \mathbf{G}_n \left(\mathbf{I} + \underline{\mathbf{A}} \Delta + \frac{1}{2} \mathbf{G}_n \left(\mathbf{I} + \frac{1}{3} \mathbf{G}_n \left(\mathbf{I} + \frac{1}{4} \mathbf{G}_n \right) \right) \right) \\ &\quad + \sum_{j_1, j_2=1}^m \left(\underline{\mathbf{A}} \mathbf{H}_{j_2} \mathbf{H}_{j_1} J_{(j_1, j_2, 0), n} + \mathbf{H}_{j_2} \underline{\mathbf{A}} \mathbf{H}_{j_1} J_{(j_1, 0, j_2), n} \right. \\ &\quad \left. \left. + \mathbf{H}_{j_2} \mathbf{H}_{j_1} \underline{\mathbf{A}} (\Delta J_{(j_1, j_2)} - J_{(j_1, j_2, 0)} - J_{(j_1, 0, j_2)}) \right) \right] \mathbf{Y}_{\tau_n}^\Delta. \end{aligned} \quad (10.2.34)$$

We remark that the corresponding orders of strong convergence of the explicit schemes described above follow from Chap. 5.

Implicit Filters

Explicit discrete-time strong approximations can sometimes behave numerically unstable, as was discussed in Sect. 7.3 and will be the topic of Chap. 14. In particular, this applies to cases when one faces multiplicative noise as in the Zakai equation (10.2.8). In such a situation control can be lost over the propagation of errors when using explicit schemes, and the approximation is rendered useless. One can then instead try to use an implicit discrete-time strong scheme to obtain a numerically stable approximation. Here it is useful to apply some of the implicit strong schemes from Chap. 7 to integrate the Zakai equation (10.2.8). Fortunately, since the Zakai equation is linear they can be rearranged algebraically to express the next iterate just in terms of its predecessor. Still, a matrix has to be inverted, which can be performed upfront.

After rearranging we get from the family of drift-implicit Euler schemes the algorithm

$$\mathbf{Y}_{\tau_{n+1}}^\Delta = (\mathbf{I} - \theta \underline{\mathbf{A}} \Delta)^{-1} [\mathbf{I} + (1 - \theta) \underline{\mathbf{A}} \Delta + \mathbf{G}_n] \mathbf{Y}_{\tau_n}^\Delta, \quad (10.2.35)$$

where $\theta \in [0, 1]$ denotes the degree of implicitness, see Sect. 7.2. As usual, the Euler scheme (10.2.35) converges with strong order $\gamma = 0.5$.

The family of drift-implicit Milstein schemes with diffusion commutativity condition, all of which converge with strong order $\gamma = 1.0$, gives us the algorithm

$$\mathbf{Y}_{\tau_{n+1}}^\Delta = (\mathbf{I} - \theta \underline{\mathbf{A}} \Delta)^{-1} \left[\mathbf{I} + (1 - \theta) \underline{\mathbf{A}} \Delta + \mathbf{G}_n \left(\mathbf{I} + \frac{1}{2} \mathbf{G}_n \right) \right] \mathbf{Y}_{\tau_n}^\Delta, \quad (10.2.36)$$

see Sect. 7.2. We can continue to construct similar higher order schemes for the Zakai equation. In principle, to each explicit scheme there corresponds a family of drift-implicit schemes obtained from making implicit the terms involving the nonrandom multiple stochastic integrals such as $I_{(0)} = \Delta$ and $I_{(0,0)} = \frac{1}{2} \Delta^2$.

As an example we mention the *strong order 1.5 drift-implicit Taylor scheme*

$$\begin{aligned} \mathbf{Y}_{\tau_{n+1}}^\Delta &= \left(\mathbf{I} - \frac{1}{2} \underline{\mathbf{A}} \Delta \right)^{-1} \left[\mathbf{I} + \frac{1}{2} \underline{\mathbf{A}} \Delta + \mathbf{G}_n \underline{\mathbf{A}} \Delta - \mathbf{M}_n \underline{\mathbf{A}} + \underline{\mathbf{A}} \mathbf{M}_n \right. \\ &\quad \left. + \mathbf{G}_n \left(\mathbf{I} + \frac{1}{2} \mathbf{G}_n \left(\mathbf{I} + \frac{1}{3} \mathbf{G}_n \right) \right) \right] \mathbf{Y}_{\tau_n}^\Delta, \end{aligned} \quad (10.2.37)$$

which makes the drift implicit in the strong order 1.5 Taylor scheme (5.3.40).

10.3 Filtering a Mean Reverting Process

In this section a hidden Markov chain model with mean reverting observations will be considered as a model for some financial time series, for instance, that of a commodity growth rate. We present results from Elliott, Fischer & Platen (1999b), where the Wonham filter for the state of the hidden Markov chain is constructed. Additionally a number of auxiliary filters are used that allow the parameters of the model to be estimated using the *Expectation Maximization (EM) algorithm*. A simulation study demonstrates the feasibility of this approach.

Mean Reverting Model

With the development of finer structured, multi-factor models for asset prices, interest rates and other financial quantities, the on-line calibration of these models is becoming an important challenge.

We consider a model for the observations which is linear mean reverting to a stochastic level determined by a hidden, continuous time Markov chain. Mean reverting diffusion processes are commonly used to model various types of phenomena in finance. We consider a two factor, mean reverting model,

where the mean reversion level changes according to a continuous time finite state Markov chain $\xi = \{\xi_t, t \in [0, T]\}$. This could be a model for the logarithm of a commodity price, see [Miltersen & Schwartz \(1998\)](#), where an unobservable reference growth rate ξ_t changes from time to time. If investors change their behavior suddenly, for instance, when a bull market for commodities has come to an end, then the reference growth rate of the log-price also changes. Such an effect may be modeled well by a hidden Markov chain.

Let us take this as a motivation and concentrate on the underlying mathematics. We derive a finite-dimensional filter for the unobservable state of the Markov chain $\xi = \{\xi_t, t \in [0, T]\}$ based on observations of a mean reverting diffusion process $r = \{r_t, t \in [0, T]\}$, which we interpret here as a log-price process. Additionally, various auxiliary filters will be presented that allow us to estimate the parameters of the hidden Markov chain.

Let the *log price* $r = \{r_t, t \in [0, T]\}$ be described by the SDE

$$dr_t = \gamma(\xi_t - r_t) dt + \varsigma dW_t^* \quad (10.3.1)$$

for $t \in [0, T]$ with $r_0 \in \Re$. Here $W^* = \{W_t^*, t \in [0, T]\}$ is a standard Wiener process on a filtered probability space $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$, where P is the real world probability measure. Suppose that the *unobserved reference level*, that is, the level $\xi = \{\xi_t, t \in [0, T]\}$ follows a continuous time, finite state Markov chain, where $T > 0$ is a finite time horizon. W^* is assumed to be independent of ξ . The parameters γ and ς are positive constants. The model in (10.3.1) is described under the real world probability measure P . To estimate the hidden reference growth rate ξ_t in an adaptive manner we employ a filter.

We consider the situation where the mean reverting process $r = \{r_t, t \in [0, T]\}$ is observed and inferences are to be made about the hidden process ξ and certain parameters. We follow the techniques used in [Elliott et al. \(1999b\)](#) to derive the filter for ξ and to construct additional filters by introducing a change of measure.

Unobserved Reference Level

In order to model the hidden reference level ξ_t it is convenient to consider a d -state continuous time Markov chain $\mathbf{X} = \{\mathbf{X}_t, t \in [0, T]\}$ that is identical to ξ after a transformation of the state space. We choose the state space for X as the set $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ of unit vectors in \Re^d , where we have $\mathbf{e}_1 = (1, 0, 0, \dots, 0)^\top$, $\mathbf{e}_2 = (0, 1, 0, \dots, 0)^\top$ and so on, see [Elliott, Aggoun & Moore \(1995\)](#). Then we can write

$$\xi_t = \mathbf{X}_t^\top \mathbf{a}, \quad (10.3.2)$$

using the vector $\mathbf{a} = (a_1, \dots, a_d)^\top \in \Re^d$ with $a_i \in \mathcal{X} = \{a_1, a_2, \dots, a_d\}$, $i \in \{1, 2, \dots, d\}$, of possible states. Here $\mathbf{X}_t^\top \mathbf{a}$ is the scalar product of the vectors \mathbf{X}_t and \mathbf{a} . We also introduce the vector of probabilities

$$\mathbf{p}(t) = E(\mathbf{X}_t). \quad (10.3.3)$$

Let $\mathbf{A} = \{\mathbf{A}_t = [a_t^{i,j}]_{i,j=1}^d, t \in [0, T]\}$ be the family of transition intensity matrices associated with the continuous time Markov chain \mathbf{X} , so that $\mathbf{p}(t)$ satisfies the Kolmogorov equation

$$\frac{d\mathbf{p}(t)}{dt} = \mathbf{A}_t \mathbf{p}(t), \quad (10.3.4)$$

with given initial probability vector $p(0)$.

In our two factor model the transition intensity matrix \mathbf{A}_t governs the dynamics of the reference level ξ_t , see equation (10.3.2). This level feeds into the equation (10.3.1) as a mean reversion level. Thus, one notes that the observed quantity r_t is conditionally Gaussian, when conditioned on the independent path of the Markov chain that describes the reference level. This is an important observation that will be exploited later on.

We now introduce some filtrations. Let \mathcal{Y}_t be the observation sigma-algebra generated by r_s for $s \in [0, t]$, that is $\mathcal{Y}_t = \sigma\{r_s, s \in [0, t]\}$. Set $\mathcal{F}_t = \sigma\{\mathbf{X}_s, s \in [0, t]\}$, $\mathcal{A}_t = \sigma\{X_s, r_s : s \in [0, t]\}$ and write $\underline{\mathcal{Y}} = (\mathcal{Y}_t)_{t \in [0, T]}$, $\underline{\mathcal{F}} = (\mathcal{F}_t)_{t \in [0, T]}$ and $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0, T]}$ for the corresponding right-continuous, augmented filtrations.

Define the vector valued process $\mathbf{M} = \{\mathbf{M}_t, t \in [0, T]\}$ by setting

$$\mathbf{M}_t = \mathbf{X}_t - \mathbf{X}_0 - \int_0^t \mathbf{A}_s \mathbf{X}_s ds. \quad (10.3.5)$$

It is not hard to show that \mathbf{M} is an $\underline{\mathcal{F}}$ -martingale under P , see Elliott et al. (1995). This gives a semimartingale representation for the continuous time Markov chain \mathbf{X} in the form

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t \mathbf{A}_s \mathbf{X}_s ds + \mathbf{M}_t \quad (10.3.6)$$

for $t \in [0, T]$.

Change of Measure

To obtain the filters and estimators required for estimating the parameters of the model we now introduce a change of measure, see Sect. 2.7.

Assume that there is a probability measure \dot{P} on $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}})$ under which ξ is a finite state Markov chain with the transition intensity matrix family \mathbf{A} as described before. The process $W = \{W_t = \frac{r_t}{\zeta}, t \in [0, T]\}$ is a Wiener process under \dot{P} , independent of ξ . In the following, the expected value of a random variable Z under the measure \dot{P} shall be denoted by $\dot{E}(Z)$. Let $L = \{L_t, t \in [0, T]\}$ be the process defined by

$$L_t = \exp \left\{ \int_0^t \nu (\mathbf{X}_s^\top \mathbf{a} - r_s) dW_s - \frac{1}{2} \int_0^t \nu^2 (\mathbf{X}_s^\top \mathbf{a} - r_s)^2 ds \right\}, \quad (10.3.7)$$

where $\nu = \frac{\gamma}{\xi}$. Define a measure P on $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}})$ by setting its Radon-Nikodym derivative with respect to \dot{P} as

$$\frac{dP}{d\dot{P}} \Big|_{\mathcal{A}_T} = L_T. \quad (10.3.8)$$

By (10.3.1) we have the relation

$$dW_t^* = \varsigma^{-1}(dr_t - \gamma(\mathbf{X}_t^\top \mathbf{a} - r_t) dt), \quad (10.3.9)$$

where $W_0^* = 0$. Then Girsanov's theorem, see Sect. 2.7, implies that the process W^* is an $(\underline{\mathcal{A}}, P)$ -Wiener process, independent of ξ . That is, under P , the processes \mathbf{X} and r have the dynamics given by (10.3.6) and (10.3.1), respectively.

The reason for this change of measure is that under the measure \dot{P} , calculations are easier because the observable process $W = \{W_t = \frac{r_t}{\xi}, t \in [0, T]\}$ is interpretable simply as a Wiener process. The fact that the measure P , which reflects real world probabilities, is a different measure does not matter because we can here use a version of the Bayes rule, see Sect. 2.7, to convert calculations made under one probability measure into a corresponding quantity calculated under another equivalent probability measure.

General Filter

Let $\phi = \{\phi_t, t \in [0, T]\}$ be an \mathcal{A} -adapted process. We shall write $\hat{\phi} = \{\hat{\phi}_t, t \in [0, T]\}$ for the *\mathcal{Y} -optional projection* of the process ϕ under the measure P , that is

$$\hat{\phi}_t = E(\phi_t | \mathcal{Y}_t) \quad (10.3.10)$$

P -a.s. We call $\hat{\phi}$ the *filter* for ϕ .

The \mathcal{Y} -optional projection of the process ϕL under the measure \dot{P} will be denoted by $\sigma(\phi) = \{\sigma(\phi_t), t \in [0, T]\}$. Now, following Elliott et al. (1995) by applying a form of the Bayes rule implies that

$$\hat{\phi}_t = \frac{\sigma(\phi_t)}{\sigma(1)} \quad (10.3.11)$$

P -a.s. for $t \in [0, T]$. To formulate a general filter consider a process $H = \{H_t, t \in [0, T]\}$ of the form

$$H_t = H_0 + \int_0^t \alpha_s ds + \int_0^t \beta_s^\top d\mathbf{M}_s + \int_0^t \delta_s dW_s^*. \quad (10.3.12)$$

Here α and δ are \mathcal{A} -predictable, square integrable processes and β is an \mathcal{A} -predictable, square integrable, d -dimensional vector process. Different specifications of α , β and δ will allow us to obtain various filters. Using equation (10.3.6) and the Itô formula for semimartingales, see (1.5.15), we obtain

$$\begin{aligned}
H_t \mathbf{X}_t &= H_0 \mathbf{X}_0 + \int_0^t \alpha_s \mathbf{X}_{s-} ds + \int_0^t \mathbf{X}_{s-} (\boldsymbol{\beta}_s^\top d\mathbf{M}_s) + \int_0^t \delta_s \mathbf{X}_{s-} dW_s^* \\
&\quad + \int_0^t H_{s-} \mathbf{A}_s \mathbf{X}_s ds + \int_0^t H_{s-} d\mathbf{M}_s + \sum_{0 < s \leq t} (\boldsymbol{\beta}_s^\top \Delta \mathbf{X}_s) \Delta \mathbf{X}_s. \quad (10.3.13)
\end{aligned}$$

Our aim will be to find the filter for $H_t \mathbf{X}_t$. From equation (10.3.11) it is sufficient to determine the evolution of $\sigma(H_t \mathbf{X}_t)$.

Theorem 10.3.1. *The recursive equation for the process $\sigma(H\mathbf{X})$ is given by*

$$\begin{aligned}
\sigma(H_t \mathbf{X}_t) &= \sigma(H_0 \mathbf{X}_0) + \int_0^t \sigma(\alpha_s \mathbf{X}_{s-}) ds + \int_0^t \mathbf{A}_s \sigma(H_{s-} \mathbf{X}_s) ds \\
&\quad + \sum_{i,j=1}^d \int_0^t (\sigma(\boldsymbol{\beta}_s^j \mathbf{X}_{s-} - \boldsymbol{\beta}_s^i \mathbf{X}_{s-})^\top \mathbf{e}_i) a_s^{j,i} ds (\mathbf{e}_j - \mathbf{e}_i) \\
&\quad + \int_0^t \varsigma^{-1} (\nu \mathbf{B}_s \sigma(H_{s-} \mathbf{X}_s) + \sigma(\delta_s \mathbf{X}_{s-})) dr_s, \quad (10.3.14)
\end{aligned}$$

for $t \in [0, T]$, where $\mathbf{B}_s = [B_s^{i,j}]_{i,j=1}^d$ is the $d \times d$ diagonal matrix with

$$B_s^{i,i} = a^i - r_s. \quad (10.3.15)$$

The proof is similar to that of Theorem 8.3.2 in Elliott et al. (1995).

As an example, if we take $H_t = H_0 = 1$ in Theorem 10.3.1, that is if we choose α , $\boldsymbol{\beta}$, and δ to be zero, then we find that

$$\sigma(\mathbf{X}_t) = E(\mathbf{X}_0) + \int_0^t \mathbf{A}_s \sigma(\mathbf{X}_s) ds + \int_0^t \varsigma^{-1} \nu \mathbf{B}_s \sigma(\mathbf{X}_s) dr_s. \quad (10.3.16)$$

Suppose $\mathbf{1} = (1, \dots, 1)^\top$ denotes the d -vector of ones, then $\mathbf{X}_t^\top \mathbf{1} = 1$. Using (10.3.11) the recursive equation for the filter $\hat{\mathbf{X}}_t$ for \mathbf{X}_t has then the form

$$\hat{\mathbf{X}}_t = E(\mathbf{X}_t | \mathcal{Y}_t) = \frac{\sigma(\mathbf{X}_t)}{\sigma(\mathbf{X}_t)^\top \mathbf{1}}, \quad (10.3.17)$$

for $t \in [0, T]$. More generally, by choosing α , $\boldsymbol{\beta}$ and δ appropriately one can compute a wide range of filters. At the end of this section we derive the equations for a number of filters.

Parameter Estimation via the EM Algorithm

Recall that $a_t^{j,i}$ is the j, i th element of the intensity matrix \mathbf{A}_t , $t \in [0, T]$. Consider now the case when the transition intensity matrix $\mathbf{A}_t = \mathbf{A}$ is constant

and unknown, and the vector \mathbf{a} of possible values for the reference level is also unknown. In this situation we can use the, so called, *Expectation Maximization (EM) algorithm*, see [Dembo & Zeitouni \(1986\)](#), to estimate these unknown parameters after observing the process r for a given time period, say up to time t . One of the basic properties of the transition intensity matrix \mathbf{A} is that for each $i \in \{1, 2, \dots, d\}$ one has $\sum_{j=1}^d a^{i,j} = 0$, so that there is no need to estimate the elements $a^{i,i}$ for $i \in \{1, 2, \dots, d\}$ if the others are already estimated.

We shall use the notation θ for the set of parameters that we want to estimate. These are, for instance, the transition intensities $a^{i,j}$ and reference levels a_i , which we are estimating, so $\theta = \{a^{i,j}, a_i : i, j \in \{1, 2, \dots, d\}\}$. The first step in the EM algorithm is to choose an initial guess θ_0 for the given parameter set. The EM algorithm is then applied to produce an estimate θ_1 for the parameters. This is repeated iteratively to produce a sequence of estimates $(\theta_n)_{n \in \mathcal{N}}$. Each iteration consists of two steps:

The first is the *expectation step* or E-step. In this step we start with θ_n , the n th iteration of estimated parameters. If $\tilde{\theta}$ is some set of possible parameter values we shall denote by $P_{\tilde{\theta}}$ the probability measure induced by the values $\tilde{\theta}$ on $(\Omega, \mathcal{A}_T, \mathcal{A})$, and E_n shall denote expectation under the measure P_{θ_n} . We then compute the conditional expectation of the log-likelihood function

$$Q(\tilde{\theta}, \theta_n) = E_n \left(\ln \left(\frac{dP_{\tilde{\theta}}}{dP_{\theta_n}} \right) \middle| \mathcal{Y}_t \right).$$

The second step is the *maximization step* or M-step. It consists of maximizing $Q(\tilde{\theta}, \theta_n)$ with respect to $\tilde{\theta}$ to obtain the new estimate θ_{n+1} . In this way the EM algorithm maximizes iteratively the likelihood that the estimated parameters are the true underlying parameters. For a detailed discussion of the EM algorithm we refer to [Dembo & Zeitouni \(1986\)](#).

To apply the EM algorithm in our case we begin by finding an expression for $\frac{dP_{\tilde{\theta}}}{dP_{\theta_n}}|_{\mathcal{A}_t}$, where $\tilde{\theta} = \{\tilde{a}^{i,j}, \tilde{a}_i : i, j \in \{1, 2, \dots, d\}\}$ and $\theta_n = \{a_n^{i,j}, a_{n,i} : i, j \in \{1, 2, \dots, d\}\}$ are two possible parameter sets.

Let \mathcal{J}_t^{ij} denote the number of jumps that the process X makes from state e_i to e_j in the interval $[0, t]$. Using Theorem T3 in Chapter VI of [Brémaud \(1981\)](#) in conjunction with the Girsanov Theorem, see Sect. 2.7, one can calculate that

$$\begin{aligned} \frac{dP_{\tilde{\theta}}}{dP_{\theta_n}} \Big|_{\mathcal{A}_t} &= \exp \left\{ \int_0^t \nu \zeta^{-1} (\mathbf{X}_s^\top \tilde{\mathbf{a}} - \mathbf{a}_n) dr_s \right. \\ &\quad \left. - \frac{1}{2} \int_0^t \nu^2 \left((\mathbf{X}_s^\top \tilde{\mathbf{a}} - r_s)^2 - (\mathbf{X}_s^\top \mathbf{a}_n - r_s)^2 \right) ds \right\} \\ &\times \prod_{i,j=1, i \neq j}^d \exp \left\{ \int_0^t \ln \left(\frac{\tilde{a}^{j,i}}{a_n^{j,i}} \right) d\mathcal{J}_s^{ij} - \int_0^t (\tilde{a}^{j,i} - a_n^{j,i}) (\mathbf{X}_s^\top e_i) ds \right\}, \end{aligned} \quad (10.3.18)$$

where $\tilde{\mathbf{a}} = (\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_d)^\top$ and $\mathbf{a}_n = (a_{n,1}, a_{n,2}, \dots, a_{n,d})^\top$.

Using the fact that $\mathbf{X}_s^\top \tilde{\mathbf{a}} = \sum_{i=1}^d \tilde{a}_i (\mathbf{X}_s^\top \mathbf{e}_i)$ we find

$$\begin{aligned} Q(\tilde{\theta}, \theta_n) &= E_n \left(\ln \left(\frac{dP_{\tilde{\theta}}}{dP_{\theta_n}} \Big| \mathcal{A}_t \right) \Big| \mathcal{Y}_t \right) \\ &= \sum_{i=1}^d \nu \varsigma^{-1} \tilde{a}_i E_n \left(\int_0^t (\mathbf{X}_s^\top \mathbf{e}_i) dr_s \Big| \mathcal{Y}_t \right) \\ &\quad - \frac{1}{2} \nu^2 E_n \left(\int_0^t ((\mathbf{X}_s^\top \tilde{\mathbf{a}})^2 - 2r_s (\mathbf{X}_s^\top \tilde{\mathbf{a}})) ds \Big| \mathcal{Y}_t \right) \\ &\quad + \sum_{i,j=1, i \neq j}^d \left(\ln(\tilde{a}^{j,i}) E_n(\mathcal{J}_t^{ij} \Big| \mathcal{Y}_t) \right. \\ &\quad \left. - \tilde{a}^{j,i} E_n \left(\int_0^t (\mathbf{X}_s^\top \mathbf{e}_i) ds \Big| \mathcal{Y}_t \right) \right) + R(\theta_n), \end{aligned} \quad (10.3.19)$$

where $R(\theta_n)$ is independent of $\tilde{\theta}$. Equation (10.3.19) can be further simplified using the fact that

$$(\mathbf{X}_s^\top \tilde{\mathbf{a}})^2 = \sum_{i=1}^d (\tilde{a}_i)^2 (\mathbf{X}_s^\top \mathbf{e}_i),$$

so

$$\begin{aligned} Q(\tilde{\theta}, \theta_n) &= \sum_{i=1}^d \nu \varsigma^{-1} \tilde{a}_i E_n \left(\int_0^t (\mathbf{X}_s^\top \mathbf{e}_i) dr_s \Big| \mathcal{Y}_t \right) \\ &\quad - \sum_{i=1}^d \frac{1}{2} \nu^2 (\tilde{a}_i)^2 E_n \left(\int_0^t (\mathbf{X}_s^\top \mathbf{e}_i) ds \Big| \mathcal{Y}_t \right) \\ &\quad + \sum_{i=1}^d \nu^2 \tilde{a}_i E_n \left(\int_0^t r_s (\mathbf{X}_s^\top \mathbf{e}_i) ds \Big| \mathcal{Y}_t \right) \\ &\quad + \sum_{i,j=1, i \neq j}^d \left(\ln(\tilde{a}^{j,i}) E_n(\mathcal{J}_t^{ij} \Big| \mathcal{Y}_t) \right. \\ &\quad \left. - \tilde{a}^{j,i} E_n \left(\int_0^t (\mathbf{X}_s^\top \mathbf{e}_i) ds \Big| \mathcal{Y}_t \right) \right) + R(\theta_n). \end{aligned} \quad (10.3.20)$$

Given the concavity of $Q(\tilde{\theta}, \theta_n)$ with respect to $\tilde{\theta}$ we maximize it with respect to $\tilde{\theta}$ by equating the partial derivatives of equation (10.3.20) in $\tilde{a}^{j,i}$ and \tilde{a}_i to zero. This implies that the next parameter set $\theta_{n+1} = \{a_{n+1}^{i,j}, a_{n+1}^i : i, j \in \{1, 2, \dots, d\}\}$ generated by the EM algorithm is given by

$$a_{n+1}^{j,i} = \left(E_n(\mathcal{J}_t^{ij} | \mathcal{Y}_t) \right) \left(E_n \left(\int_0^t (\mathbf{X}_s^\top \mathbf{e}_i) ds | \mathcal{Y}_t \right) \right)^{-1} \quad (10.3.21)$$

and

$$a_{n+1}^i = \frac{\gamma^{-1} E_n \left(\int_0^t (\mathbf{X}_s^\top \mathbf{e}_i) dr_s | \mathcal{Y}_t \right) + E_n \left(\int_0^t r_s (\mathbf{X}_s^\top \mathbf{e}_i) ds | \mathcal{Y}_t \right)}{E_n \left(\int_0^t (\mathbf{X}_s^\top \mathbf{e}_i) ds | \mathcal{Y}_t \right)}. \quad (10.3.22)$$

We evaluate the expected values in equations (10.3.21) and (10.3.22) by applying special cases of the general filter derived previously. For the remainder of this section it is convenient to use the less cumbersome notation, where we write

$$\mathcal{O}_t^i = \int_0^t (\mathbf{X}_s^\top \mathbf{e}_i) ds, \quad (10.3.23)$$

$$\mathcal{K}_t^i = \int_0^t (\mathbf{X}_s^\top \mathbf{e}_i) dr_s, \quad (10.3.24)$$

and

$$\mathcal{I}_t^i = \int_0^t r_s (\mathbf{X}_s^\top \mathbf{e}_i) ds. \quad (10.3.25)$$

Previously, we used Theorem 10.3.1 to derive a recursive equation for $\sigma(\mathcal{O}_t^i \mathbf{X}_t)$, which gives $\sigma(\mathcal{O}_t^i)$ since $\sigma(\mathcal{O}_t^i) = (\sigma(\mathcal{O}_t^i \mathbf{X}_t), \mathbf{1})$. The particular equations are given at the end of this section. The same procedure can be used for the processes $\sigma(\mathcal{J}_t^{ij})$, $\sigma(\mathcal{K}_t^i)$ and $\sigma(\mathcal{I}_t^i)$. From equation (10.3.11) we see, using equations (10.3.21) and (10.3.22), that the next set of parameter values is given by

$$a_{n+1}^{j,i} = \frac{\sigma(\mathcal{J}_t^{ij})}{\sigma(\mathcal{O}_t^i)} \quad (10.3.26)$$

and

$$a_{n+1}^i = \frac{\gamma^{-1} \sigma(\mathcal{K}_t^i) + \sigma(\mathcal{I}_t^i)}{\sigma(\mathcal{O}_t^i)}. \quad (10.3.27)$$

Extended Parameter Estimation

The parameter estimation can be extended to also estimate the speed of adjustment parameter γ . However, the order of the hidden Markov chain, that is, the number of points in its state space, is a delicate estimation problem, which is discussed, for example, in Baras & Fineso (1991) and Setiawaty (1999). The filtering methodology that we describe is based on a change of measure technique, which by its nature does not provide direct estimates of diffusion coefficients. The transform method presented in Sect. 9.3 provides a possibility to estimate these parameters. Alternatively, in particular cases one can estimate these, for instance, from the square root of the slope of the empirical quadratic variation of the observed process. The procedure used to estimate

the parameters $a^{i,j}$ for $i, j \in \{1, 2, \dots, d\}$ remains then unchanged. Including γ as a parameter that we wish to estimate implies that equation (10.3.18) takes a different form. The Radon-Nikodym derivative of the probability measure induced by the updated parameter values $\tilde{\theta} = \{\tilde{\gamma}, \tilde{a}^{i,j}, \tilde{a}_i : i, j \in \{1, 2, \dots, d\}\}$ with respect to the probability measure induced by the old parameter values $\theta_n = \{\gamma_n, a_n^{i,j}, a_{n,i} : i, j \in \{1, 2, \dots, d\}\}$ is then

$$\frac{dP_{\tilde{\theta}}}{dP_{\theta_n}} \Big|_{\mathcal{A}_t} = \exp \left\{ \int_0^t (\tilde{\gamma} \varsigma^{-2} (\mathbf{X}_s^\top \tilde{\mathbf{a}} - r_s) - \gamma_n \varsigma^{-2} (\mathbf{X}_s^\top \mathbf{a}_n - r_s)) dr_s - \frac{1}{2} \int_0^t (\tilde{\gamma}^2 \varsigma^{-2} ((\mathbf{X}_s^\top \tilde{\mathbf{a}}) - r_s)^2 - \gamma_n^2 \varsigma^{-2} (\mathbf{X}_s^\top \mathbf{a}_n - r_s)^2) ds \right\} R_1. \quad (10.3.28)$$

Here $\tilde{\mathbf{a}} = (\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_d)^\top$, $\mathbf{a}_n = (a_{n,1}, a_{n,2}, \dots, a_{n,d})^\top$ with R_1 denoting a term that is independent of $\tilde{\gamma}$ and $\tilde{\mathbf{a}}$. Now

$$\begin{aligned} Q(\tilde{\theta}, \theta_n) \varsigma^2 &= \tilde{\gamma} \left(E_n \left(\int_0^t \mathbf{X}_s^\top \tilde{\mathbf{a}} dr_s \mid \mathcal{Y}_t \right) - E_n \left(\int_0^t r_s dr_s \mid \mathcal{Y}_t \right) \right) \\ &\quad - \frac{1}{2} \tilde{\gamma}^2 \left(E_n \left(\int_0^t \mathbf{X}_s^\top \tilde{\mathbf{a}}^2 ds \mid \mathcal{Y}_t \right) - 2 E_n \left(\int_0^t r_s \mathbf{X}_s^\top \tilde{\mathbf{a}} ds \mid \mathcal{Y}_t \right) \right. \\ &\quad \left. + E_n \left(\int_0^t r_s^2 ds \mid \mathcal{Y}_t \right) \right) + R_2, \end{aligned} \quad (10.3.29)$$

recalling that E_n means expectation with respect to P_{θ_n} . Here R_2 is independent of both $\tilde{\gamma}$ and $\tilde{\mathbf{a}}$. Employing $\hat{\phi}$ as shorthand notation for the \mathcal{Y} -optional projection of the process ϕ under the measure P_{θ_n} and using the processes \mathcal{O}^i , \mathcal{K}^i and \mathcal{I}^i , see equations (10.3.23) to (10.3.25), we proceed as previously and write equation (10.3.29) as

$$\begin{aligned} Q(\tilde{\theta}, \theta_n) \varsigma^2 &= \tilde{\gamma} \left(\sum_{i=1}^d \tilde{a}_i \hat{\mathcal{K}}_t^i - \int_0^t r_s dr_s \right) \\ &\quad - \frac{1}{2} \tilde{\gamma}^2 \left(\sum_{i=1}^d (\tilde{a}_i)^2 \hat{\mathcal{O}}_t^i - 2 \sum_{i=1}^d \tilde{a}_i \hat{\mathcal{I}}_t^i + \int_0^t r_s^2 ds \right). \end{aligned} \quad (10.3.30)$$

We maximize $Q(\tilde{\theta}, \theta_n)$ by equating to zero the partial derivatives with respect to \tilde{a}_i and $\tilde{\gamma}$ of equation (10.3.30). After some simplification one finds that the new estimate for γ is given by

$$\gamma_{n+1} = \frac{\tilde{C}}{\tilde{D}}, \quad (10.3.31)$$

where

$$\tilde{C} = \int_0^t r_s dr_s - \sum_{i=1}^d \frac{\hat{\mathcal{I}}_t^i \hat{\mathcal{K}}_t^i}{\hat{\mathcal{O}}_t^i} \quad (10.3.32)$$

and

$$\tilde{D} = \sum_{i=1}^d \frac{(\hat{\mathcal{I}}_t^i)^2}{\hat{\mathcal{O}}_t^i} - \int_0^t r_s^2 ds. \quad (10.3.33)$$

The new estimate for a_i is now

$$a_{n+1,i} = \frac{\gamma_{n+1}^{-1} \hat{\mathcal{K}}_t^i + \hat{\mathcal{I}}_t^i}{\hat{\mathcal{O}}_t^i}. \quad (10.3.34)$$

Simulation Example

For illustration, the filter methods derived above are now applied in a simulation study. The results computed use discrete-time approximate filters as proposed in the previous section. The approximate filters were calculated using the Milstein scheme, see (5.3.19) with a sufficiently small time step size to ensure high accuracy and numerical stability. Alternative methods, which provide potentially more numerical stability can also be applied, as will be discussed in Sect. 10.4.

The mean reverting process r was simulated using a total of 800,000 discretization points. In the simulation the hidden process ξ is a three state continuous time Markov chain, that is, we have $d = 3$. The intensity matrix for the process ξ is assumed to be

$$\mathbf{A} = \begin{bmatrix} -0.8 & 0.8 & 0 \\ 1.5 & -3.0 & 1.5 \\ 0 & 0.4 & -0.4 \end{bmatrix} \quad (10.3.35)$$

The other parameter values were taken to be $\varsigma = 1.2$, $\gamma = 8.0$, $a_1 = -0.3$, $a_2 = 1.9$, $a_3 = 2.4$. The initial values for the EM algorithm to start with, are set to

$$\mathbf{A} = \begin{bmatrix} -1.0 & 1.0 & 0 \\ 0.5 & -1.0 & 0.5 \\ 0 & 1.0 & -1.0 \end{bmatrix} \quad (10.3.36)$$

with $\gamma = 5.0$, $a_1 = 0.0$, $a_2 = 1.0$ and $a_3 = 2.0$.

The EM algorithm was run over 50 iterations. Figs. 10.3.1 to 10.3.4 show the evolution of the estimates of the parameters in dependence on the number of iterations and include the true values for reference.

The simulation was conducted over the extremely long time interval $[0, 20000]$ to ensure that each state in the Markov chain was visited sufficiently often. The accurate estimation of the elements of the intensity matrix \mathbf{A} depends on the detection of numerous jumps in the Markov chain. With shorter time series one already may be close to the true value, though no guarantee can be given. In the first part of the simulation with about 2,000 discretization points over the rather short time interval $[0, 10]$ the EM algorithm visually appeared to converge rather fast, however, the estimates could still vary substantially.

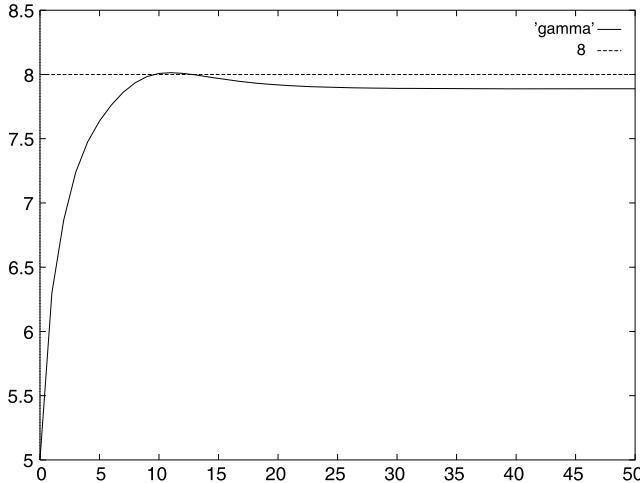


Fig. 10.3.1. The evolution of the estimates of $\gamma = 8.0$

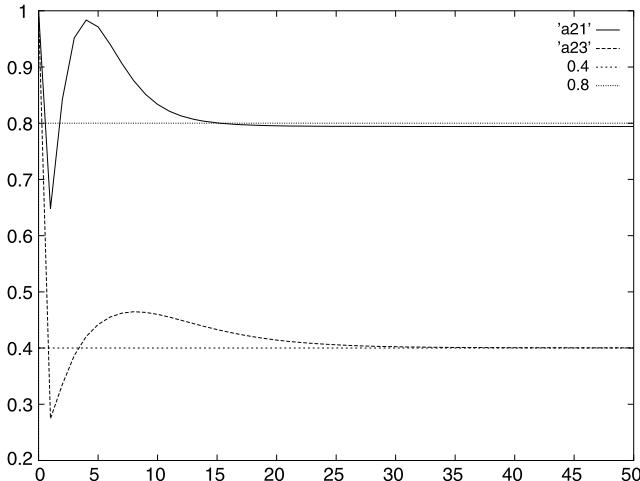


Fig. 10.3.2. The evolution of the estimates of $a^{2,1} = 0.8$ and $a^{3,2} = 0.4$

The Filter for \mathcal{O}^i

Let us provide further results on particular filters. When applying Theorem 10.3.1 with

$$H_t = \mathcal{O}_t^i, \quad H_0 = 0, \quad \alpha_s = (\mathbf{X}_s^\top \mathbf{e}_i), \quad \beta_s = \mathbf{0} \quad \text{and} \quad \delta_s = 0,$$

and using the fact that

$$(\mathbf{X}_s^\top \mathbf{e}_i) \mathbf{X}_s = (\mathbf{X}_s^\top \mathbf{e}_i) \mathbf{e}_i \tag{10.3.37}$$

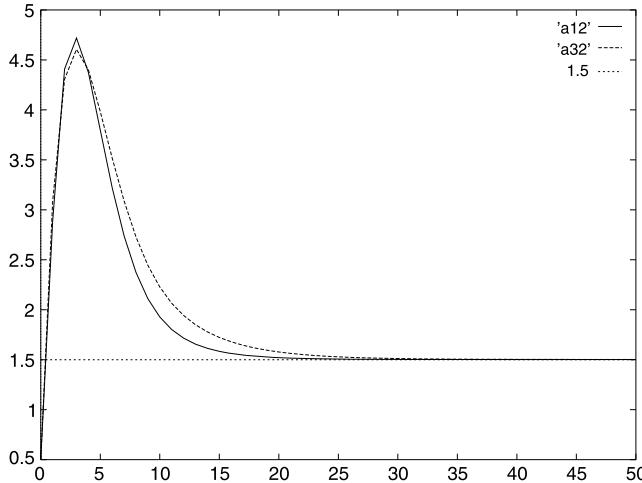


Fig. 10.3.3. The evolution of the estimates of $a^{2,1} = a^{2,3} = 1.5$

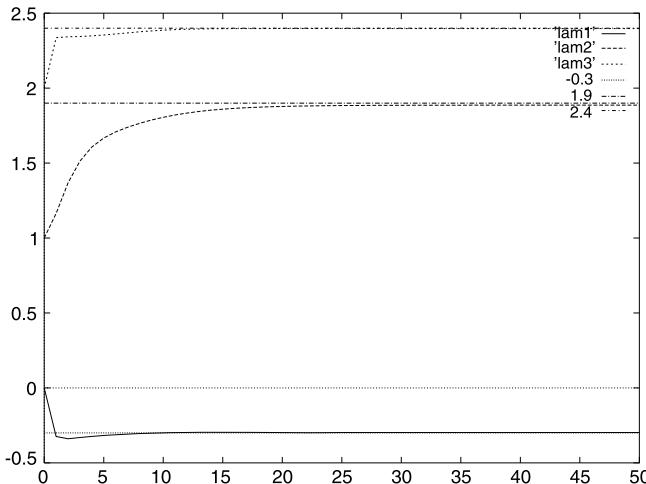


Fig. 10.3.4. The evolution of the estimates of a_i , $i \in \{1, 2, 3\}$

we find that

$$\begin{aligned} \sigma(\mathcal{O}_t^i \mathbf{X}_t) &= \int_0^t (\sigma(\mathbf{X}_s)^\top \mathbf{e}_i) \mathbf{e}_i ds + \int_0^t \mathbf{A} \sigma(\mathcal{O}_s^i \mathbf{X}_s) ds \\ &\quad + \int_0^t \zeta^{-1} \nu \mathbf{B}_s \sigma(\mathcal{O}_s^i \mathbf{X}_s) dr_s. \end{aligned} \tag{10.3.38}$$

We can evaluate equation (10.3.38) because equation (10.3.16) gives us $\sigma(\mathbf{X}_s)$. We also have the unnormalized filter for \mathcal{O}^i

$$\sigma(\mathcal{O}_t^i) = \sigma(\mathcal{O}_t^i \mathbf{X}_t)^\top \mathbf{1} \quad (10.3.39)$$

and the filter for \mathcal{O}^i

$$E(\mathcal{O}_t^i | \mathcal{Y}_t) = \frac{\sigma(\mathcal{O}_t^i)}{\sigma(\mathbf{X}_t)^\top \mathbf{1}}. \quad (10.3.40)$$

The Filter for \mathcal{J}^{ij}

Using Elliott et al. (1995) we can write

$$\mathcal{J}_t^{ij} = \int_0^t (\mathbf{X}_{s-}^\top \mathbf{e}_i) a^{j,i} ds + \int_0^t (\mathbf{X}_{s-}^\top \mathbf{e}_i) \mathbf{e}_j^\top d\mathbf{M}_s. \quad (10.3.41)$$

Then we can apply Theorem 10.3.1 by taking

$$H_t = \mathcal{J}_t^{ij}, \quad H_0 = 0, \quad \alpha_s = (\mathbf{X}_{s-}^\top \mathbf{e}_i) a^{j,i}, \quad \delta_s = 0 \quad \text{and} \quad \beta_s = (\mathbf{X}_{s-}^\top \mathbf{e}_i) \mathbf{e}_j.$$

Since $(\mathbf{X}_{s-}^\top \mathbf{e}_i) \mathbf{X}_{s-} = (\mathbf{X}_{s-}^\top \mathbf{e}_i) \mathbf{e}_i$ we find

$$\begin{aligned} \sum_{k,\ell=1}^d ((\beta_s^\ell \mathbf{X}_{s-} - \beta_s^k \mathbf{X}_{s-})^\top \mathbf{e}_k) a^{\ell,k} (\mathbf{e}_\ell - \mathbf{e}_k) &= \left((\mathbf{X}_{s-}^\top \mathbf{e}_i) \mathbf{e}_i^\top \mathbf{e}_i \right) a^{j,i} (\mathbf{e}_j - \mathbf{e}_i) \\ &= (\mathbf{X}_{s-}^\top \mathbf{e}_i) a^{j,i} (\mathbf{e}_j - \mathbf{e}_i). \end{aligned} \quad (10.3.42)$$

Applying Theorem 10.3.1 and using equations (10.3.42) and (10.3.37) we find the recursive equation for $\sigma(\mathcal{J}_t^{ij} \mathbf{X}_t)$ in the form

$$\begin{aligned} \sigma(\mathcal{J}_t^{ij} \mathbf{X}_t) &= \int_0^t a^{j,i} (\sigma(\mathbf{X}_s)^\top \mathbf{e}_i) \mathbf{e}_j ds + \int_0^t \mathbf{A} \sigma(\mathcal{J}_s^{ij} \mathbf{X}_s) ds \\ &\quad + \int_0^t \varsigma^{-1} \nu \mathbf{B}_s \sigma(\mathcal{J}_s^{ij} \mathbf{X}_s) dr_s. \end{aligned} \quad (10.3.43)$$

The unnormalized filter for \mathcal{J}^{ij} is given by

$$\sigma(\mathcal{J}_t^{ij}) = \sigma(\mathcal{J}_t^{ij} \mathbf{X}_t)^\top \mathbf{1}, \quad (10.3.44)$$

and the normalized filter for \mathcal{J}^{ij} follows as

$$E(\mathcal{J}_t^{ij} | \mathcal{Y}_t) = \frac{\sigma(\mathcal{J}_t^{ij})}{\sigma(\mathbf{X}_t)^\top \mathbf{1}}. \quad (10.3.45)$$

The Filter for \mathcal{K}^i

Using equations (10.3.1) and (10.3.37) we see that

$$\mathcal{K}_t^i = \int_0^t \gamma (a_i - r_s) (\mathbf{X}_s^\top \mathbf{e}_i) ds + \int_0^t \varsigma (\mathbf{X}_s^\top \mathbf{e}_i) dW_s^*. \quad (10.3.46)$$

If we apply Theorem 10.3.1 with

$$H_t = \mathcal{K}_t^i, \quad H_0 = 0, \quad \alpha_s = \gamma (a_i - r_s) (\mathbf{X}_s^\top \mathbf{e}_i), \quad \delta_s = \varsigma (\mathbf{X}_s^\top \mathbf{e}_i) \quad \text{and} \quad \beta_s = \mathbf{0}$$

and use equation (10.3.37), then we find

$$\begin{aligned} \sigma(\mathcal{K}_t^i \mathbf{X}_t) &= \int_0^t \gamma (a_i - r_s) (\sigma(\mathbf{X}_s)^\top \mathbf{e}_i) \mathbf{e}_i ds + \int_0^t \mathbf{A} \sigma(\mathcal{K}_s^i \mathbf{X}_s) ds \\ &\quad + \int_0^t (\varsigma^{-1} \nu \mathbf{B}_s \sigma(\mathcal{K}_s^i \mathbf{X}_s) + (\sigma(\mathbf{X}_s)^\top \mathbf{e}_i) \mathbf{e}_i) dr_s. \end{aligned} \quad (10.3.47)$$

Again, the unnormalized filter for \mathcal{K}^i is given by

$$\sigma(\mathcal{K}_t^i) = \sigma(\mathcal{K}_t^i \mathbf{X}_t)^\top \mathbf{1}, \quad (10.3.48)$$

and the normalized filter for \mathcal{K}^i is obtained as

$$E(\mathcal{K}_t^i | \mathcal{Y}_t) = \frac{\sigma(\mathcal{K}_t^i)}{\sigma(\mathbf{X}_t)^\top \mathbf{1}}. \quad (10.3.49)$$

The Filter for \mathcal{I}^i

By applying Theorem 10.3.1 with

$$H_t = \mathcal{I}_t^i, \quad H_0 = 0, \quad \alpha_s = r_s (\mathbf{X}_s^\top \mathbf{e}_i), \quad \beta_s = \mathbf{0} \quad \text{and} \quad \delta_s = 0$$

and using equation (10.3.37) we see that

$$\begin{aligned} \sigma(\mathcal{I}_t^i \mathbf{X}_t) &= \int_0^t r_s (\sigma(\mathbf{X}_s)^\top \mathbf{e}_i) \mathbf{e}_i ds + \int_0^t \mathbf{A} \sigma(\mathcal{I}_s^i \mathbf{X}_s) ds \\ &\quad + \int_0^t \varsigma^{-1} \nu \mathbf{B}_s \sigma(\mathcal{I}_s^i \mathbf{X}_s) dr_s. \end{aligned} \quad (10.3.50)$$

Once again we have the unnormalized filter for \mathcal{I}^i given by

$$\sigma(\mathcal{I}_t^i) = \sigma(\mathcal{I}_t^i \mathbf{X}_t)^\top \mathbf{1}, \quad (10.3.51)$$

and the normalized filter for \mathcal{I}^i is then of the form

$$E(\mathcal{I}_t^i | \mathcal{Y}_t) = \frac{\sigma(\mathcal{I}_t^i)}{\sigma(\mathbf{X}_t)^\top \mathbf{1}}. \quad (10.3.52)$$

10.4 Balanced Implicit Method in Filtering

In practice, the numerical stability of discrete-time approximations for filters is highly important. In this section we study the application of the balanced implicit method in hidden Markov chain filtering when the time between observations is rather large. This potentially resolves some problems that often arise from the strong numerical solution of the Zakai equation. Comparisons are conducted between the balanced implicit method and other strong approximation methods.

Wonham Filter

As we have already seen in previous sections, strong discrete-time methods for the approximate solution of SDEs provide essential tools in hidden Markov chain filtering. Numerical instabilities of explicit or drift-implicit strong approximations may prevent their use in important practical situations. There are various drift-implicit schemes that perform well on some SDEs for certain parameter ranges and sufficiently small step sizes, as discussed in Chap. 5 and Sect. 10.2 and as will be analyzed in Chap. 14. However, numerical instabilities caused by the presence of multiplicative noise in the Zakai SDE are, in general, not overcome by most of these methods. For larger time step sizes numerical instabilities potentially arise.

The following study is aimed at the numerical solution of the Zakai equation (10.2.8) for hidden Markov chain filtering. This critical task has to be effectively and reliably performed to make a hidden Markov chain filter successful. Observations are, in practice, only available at discrete-time instants. Therefore, a discrete-time approximate solution of the Zakai equation has to be obtained. The results of this study were obtained in Fischer & Platen (1999). They demonstrate that the balanced implicit method, see Sect. 7.3, is a valuable tool in filtering.

We will introduce two problems with some practical background involving the discrete-time numerical solution of hidden Markov chain filters. In each of the problems we study a different type of dynamics for the observation process. In the first one, the signal is corrupted by a Wiener process, as in Sect. 10.2. For the second one a mean reverting process perturbs the signal, similar to Sect. 10.3.

To avoid confusion we first recall our notation. The unobservable signal process $\xi = \{\xi_t, t \in [0, T]\}$ is a time homogeneous, continuous time, real valued Markov chain with the set $\mathcal{X} = \{a_1, a_2, \dots, a_d\}$ of states. The observation process $W = \{W_t, t \in [0, T]\}$ is given by the relation

$$W_t = \int_0^t \xi_s ds + W_t^*, \quad (10.4.1)$$

for $t \in [0, T]$. One can say that W represents the time integral over the signal process that is corrupted by a Wiener process W^* . We estimate the hidden state of the Markov chain ξ by observing only the values of the process W .

As previously, it is convenient to consider a d state continuous time Markov chain $\mathbf{X} = \{\mathbf{X}_t, t \in [0, T]\}$ that is identical to ξ under a transformation of the state space. We choose as the state space for \mathbf{X} the set $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ of unit vectors in \Re^d , with $\mathbf{e}_1 = (1, 0, 0, \dots, 0)^\top$, $\mathbf{e}_2 = (0, 1, 0, \dots, 0)^\top$ and so on. Then we write,

$$\xi_t = \mathbf{X}_t^\top \mathbf{a} \quad (10.4.2)$$

with $\mathbf{a} = (a_1, a_2, \dots, a_d)^\top \in \Re^d$. Let $\mathbf{A} = [a^{i,j}]_{i,j=1}^d$ be the constant intensity matrix associated with the homogeneous, continuous time Markov chain \mathbf{X} , so that $\mathbf{p}(t) = E(\mathbf{X}_t)$ satisfies the vector ordinary differential equation

$$\frac{d\mathbf{p}(t)}{dt} = \mathbf{A} \mathbf{p}(t), \quad (10.4.3)$$

for $t \in [0, T]$, with given initial probability vector $\mathbf{p}(0)$. \mathbf{H} is the diagonal $d \times d$ matrix that has the elements of the vector \mathbf{a} as diagonal elements and is zero elsewhere.

Denote by \mathcal{Y}_t the observation sigma-algebra generated by W up to time t . The Wonham filter for \mathbf{X} at time t is then given as

$$\hat{\mathbf{X}}_t = E(\mathbf{X}_t \mid \mathcal{Y}_t).$$

As we have seen in Sect. 10.2, the theoretical solution to the problem of calculating $\hat{\mathbf{X}}_t$ involves the unnormalized filter for the conditional distribution of \mathbf{X} , which is denoted by $\sigma(\mathbf{X}) = \{\sigma(\mathbf{X}_t), t \in [0, T]\}$ and satisfies the Zakai equation

$$d\sigma(\mathbf{X}_t) = \mathbf{A} \sigma(\mathbf{X}_t) dt + \mathbf{D} \sigma(\mathbf{X}_t) dW_t \quad (10.4.4)$$

for $t \in [0, T]$. The Wonham filter for \mathbf{X} is then computed as

$$\hat{\mathbf{X}}_t = E(\mathbf{X}_t \mid \mathcal{Y}_t) = \frac{\sigma(\mathbf{X}_t)}{\sigma(\mathbf{X}_t)^\top \mathbf{1}} \quad (10.4.5)$$

for $t \in [0, T]$, see (10.2.9).

Approximate Wonham Filter

The numerical problem consists of solving the SDE (10.4.4) for a given path of the observation process W . This means, we have to seek a strong discrete-time approximation for $\sigma(\mathbf{X}_t)$ if we are supplied with discrete observations of W .

We assume that the observation process is only given at the discretization times $0 = \tau_0 < \tau_1 < \dots < \tau_N = T$. Then one can apply strong discrete-time schemes to approximately solve the SDE (10.4.4). Candidates for these methods are the Euler scheme, the Milstein scheme and other explicit and implicit strong schemes.

To demonstrate how the balanced implicit method deals with potential numerical instabilities caused by multiplicative noise, several simulation studies will be carried out. At first, for the filter equation (10.4.4) the balanced implicit method is compared with three possible alternative schemes. The alternative schemes considered are the Euler scheme, the Milstein scheme and the drift-implicit Milstein scheme, which are all described in Chap. 7.

For the Wonham filter, that is the solution of the SDE (10.4.4), the Euler scheme yields the stochastic difference equation

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \mathbf{A} \mathbf{Y}_n \Delta + \mathbf{H} \mathbf{Y}_n \Delta W_n, \quad (10.4.6)$$

where $\Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}$, $\Delta = \tau_{n+1} - \tau_n$ and $n \in \{0, 1, \dots, N-1\}$. The increments of the observation process play the role of the driving noise process.

The drift-implicit Milstein scheme, when applied to the filter equation (10.4.4), allows for different degrees of implicitness $\theta \in [0, 1]$ and is given by the algorithm

$$\begin{aligned} \mathbf{Y}_{n+1} = & (\mathbf{I} - \theta \mathbf{A} \Delta)^{-1} (\mathbf{Y}_n + (1 - \theta) \mathbf{A} \mathbf{Y}_n \Delta \\ & + \mathbf{H} \mathbf{Y}_n \Delta W_n + \frac{1}{2} \mathbf{H}^2 \mathbf{Y}_n ((\Delta W_n)^2 - \Delta)). \end{aligned} \quad (10.4.7)$$

The Milstein scheme itself is then given by equation (10.4.7) with the choice $\theta = 0$.

To construct a balanced implicit method (7.3.9)–(7.3.10) we must specify the functions c^0 and c^1 in (7.3.10). When implementing numerical schemes to approximately solve an SDE, problems arise when the coefficients of $\sigma(\mathbf{X}_t)$ in the Itô differentials become large in magnitude. With this in mind we choose $c^0 = 0$ and $c^1 = |\mathbf{H}|$ in the balanced implicit method which yields

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n \left(\frac{\mathbf{I} + \mathbf{A} \Delta + \mathbf{H} \Delta W_n}{1 + |\mathbf{H}| |\Delta W_n|} \right). \quad (10.4.8)$$

The Case of No Mean Reversion

For our numerical simulations the parameters of the Markov chain X , or equivalently ξ , are chosen to give a realistic multiplicative noise term in the Zakai equation (10.4.4). The simulated hidden Markov chain X is chosen to have three states, with the vector \mathbf{a} taken to be $\mathbf{a} = (5, 0, -5)^\top$. Experiments have shown that using more states does not change the nature of the results that we obtain. The intensity matrix \mathbf{A} of the hidden Markov chain is chosen to be of the simple form

$$\mathbf{A} = \begin{bmatrix} -1.0 & 1.0 & 0 \\ 0.5 & -1.0 & 0.5 \\ 0 & 1.0 & -1.0 \end{bmatrix}. \quad (10.4.9)$$

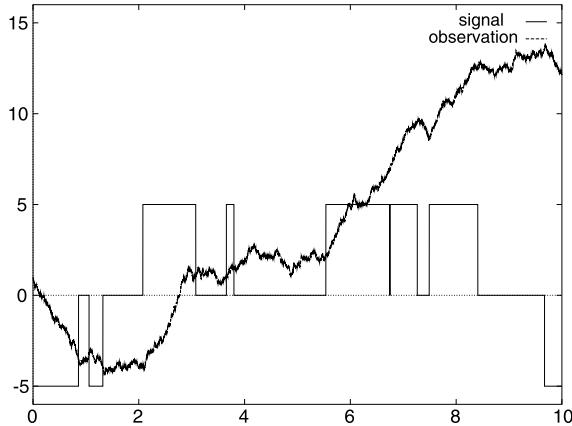


Fig. 10.4.1. Simulation of the signal and observation processes

This describes how the Markov chain jumps with prescribed intensities to neighboring states. For instance, the intensity to jump from level 0 to level 5 is 0.5 per unit of time.

Now, let us investigate the approximate calculation of the Wonham filter. We simulate the signal and observation processes over the time interval $[0, T]$ with $T = 10$. The simulated output can be seen in Fig. 10.4.1.

Consider the probability

$$q_t^i = \hat{\mathbf{X}}_t^\top \mathbf{e}_i = E \left(\mathbf{X}_t^\top \mathbf{e}_i \mid \mathcal{Y}_t \right) \quad (10.4.10)$$

which, for $i \in \{1, 2, 3\}$ is the filtered probability that the hidden Markov chain \mathbf{X} is in the i th state. For illustration let us consider q_t^1 , which corresponds to the level $\xi = 5$, that is $i = 1$.

To obtain the quantity q_t^1 we have to solve the SDE (10.4.4). The solution has to be calculated numerically using any of the above mentioned strong schemes on the basis of the discrete-time observations of W . When the step size Δ was chosen extremely small with about $\frac{1}{1000}$, then all four numerical schemes produced virtually identical results. Fig. 10.4.2 displays a plot of q_t^1 as calculated by the semi-drift-implicit Milstein scheme with $\Delta = \frac{1}{1000}$ and $\theta = 0.5$, showing that the high levels of q_t^1 track well the periods when the hidden state indeed equals its highest level, $\xi = 5$. However, such a fine time discretization for the observation process is often not available in real world financial filtering.

Differences between the schemes become apparent when the time step size Δ is chosen to be larger. Using the same realizations of the observation and signal processes that were given in Fig. 10.4.1, the Figs. 10.4.3 to 10.4.6 display the plots of the filtered probability q_t^1 , see (10.4.10), calculated by the four different schemes when using the larger step size $\Delta = \frac{1}{20}$. For this step size we can see that the only acceptable scheme appears to be the balanced implicit

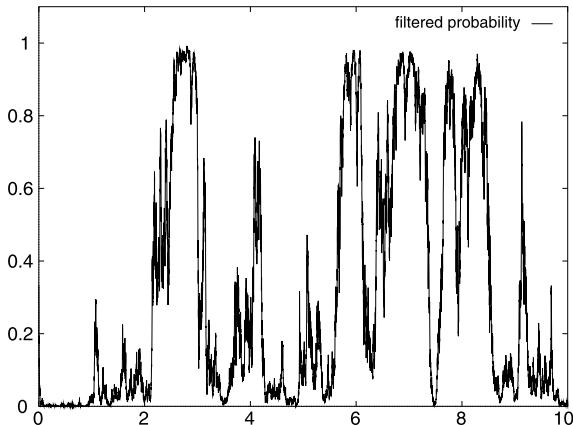


Fig. 10.4.2. q_t^1 - obtained by the drift-implicit Milstein scheme with $\theta = 0.5$ and $\Delta = \frac{1}{1000}$

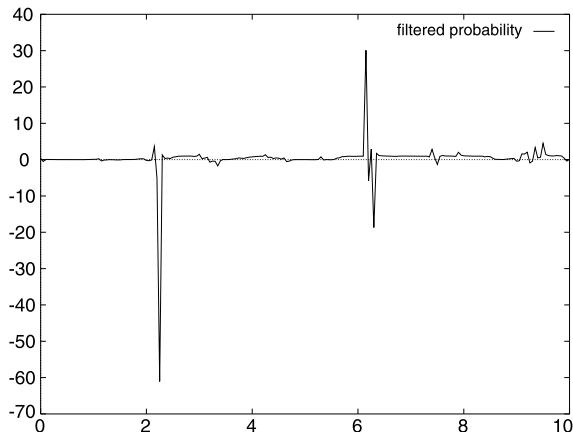


Fig. 10.4.3. q_t^1 - obtained by the Euler method with $\Delta = \frac{1}{20}$

method. In the other cases we even obtain negative “probabilities” and other unrealistic estimates as filter values. The balanced implicit method makes the filter work even for the given larger time step size, which is a remarkable fact.

Mean Reverting Observation Process

The second problem that we consider refers to a linearly mean reverting observation process \bar{W} . Using mostly the same notation as in the previous example, the observation process \bar{W} is here assumed to satisfy the linear mean reverting SDE

$$d\bar{W}_t = \gamma (\xi_t - \bar{W}_t) dt + dW_t^* \quad (10.4.11)$$

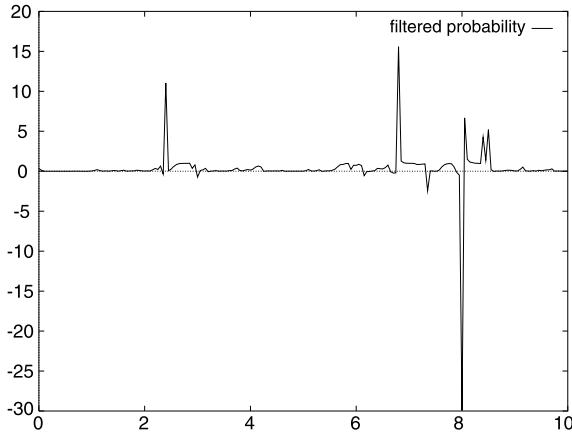


Fig. 10.4.4. q_t^1 - obtained by the Milstein scheme with $\Delta = \frac{1}{20}$

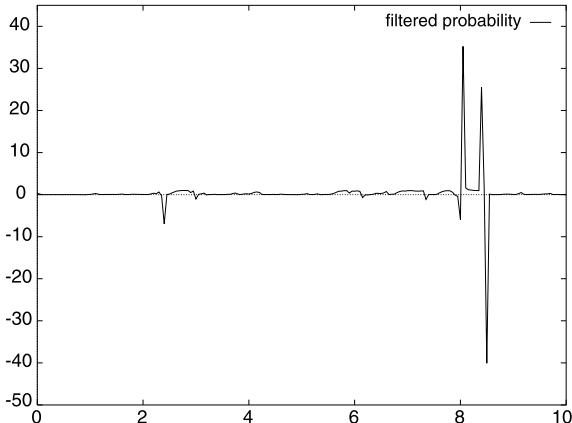


Fig. 10.4.5. q_t^1 - obtained by the semi-drift-implicit Milstein scheme with $\Delta = \frac{1}{20}$

with speed of adjustment parameter $\gamma > 0$. Using a similar approach to that described in Sect. 10.3, the unnormalized filter, here denoted by $\bar{\sigma}(\mathbf{X})$, for the hidden Markov chain \mathbf{X} can be obtained, which satisfies the Zakai equation

$$d\bar{\sigma}(\mathbf{X}_t) = A \bar{\sigma}(\mathbf{X}_t) dt + \gamma (\mathbf{H} - \mathbf{I} \bar{W}_t) \bar{\sigma}(\mathbf{X}_t) d\bar{W}_t \quad (10.4.12)$$

for $t \in [0, T]$. As before, the filter $\hat{\mathbf{X}}_t$ for \mathbf{X}_t is recovered by using equation (10.4.5) replacing $\bar{\sigma}$ by σ , that is

$$\hat{\mathbf{X}}_t = E(\mathbf{X}_t | \mathcal{Y}_t) = \frac{\bar{\sigma}(\mathbf{X}_t)}{\bar{\sigma}(\mathbf{X}_t) \mathbf{1}}$$

for $t \in [0, T]$.

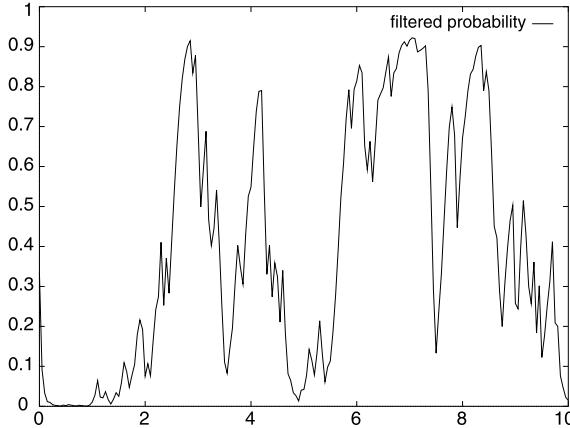


Fig. 10.4.6. q_t^1 - obtained by the balanced implicit method with $\Delta = \frac{1}{20}$

The Euler method, when approximating the solution of the Zakai equation (10.4.12), has the form

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \mathbf{A} \mathbf{Y}_n \Delta + \gamma (\mathbf{H} - \mathbf{I} \bar{W}_n) \mathbf{Y}_n \Delta \bar{W}_n, \quad (10.4.13)$$

where $\Delta \bar{W}_n = \bar{W}_{\tau_{n+1}} - \bar{W}_{\tau_n}$, $\bar{W}_n = \bar{W}_{\tau_n}$, $\Delta = \tau_{n+1} - \tau_n$ and $n \in \{0, 1, \dots, N-1\}$. The drift-implicit Milstein scheme that approximates the solution of (10.4.12) yields

$$\begin{aligned} \mathbf{Y}_{n+1} &= (\mathbf{I} - \theta \mathbf{A} \Delta)^{-1} (\mathbf{Y}_n + (1 - \theta) \mathbf{A} \mathbf{Y}_n \Delta + \gamma (\mathbf{H} - \mathbf{I} \bar{W}_n) \mathbf{Y}_n \Delta \bar{W}_n \\ &\quad + \frac{1}{2} \gamma^2 (\mathbf{H} - \mathbf{I} \bar{W}_n)^2 \mathbf{Y}_n ((\Delta \bar{W}_n)^2 - \Delta)), \end{aligned} \quad (10.4.14)$$

where $\theta \in [0, 1]$. With the choice $\theta = 0$ the algorithm (10.4.14) yields the Milstein scheme when applied to (10.4.12).

Using the same reasoning as in the previous example we implement the balanced implicit method (7.3.9)–(7.3.10) by choosing $c^0 = 0$ and $c_n^1 = |\gamma(\mathbf{H} - \mathbf{I} \bar{W}_n)|$ such that

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n \left(\frac{\mathbf{I} + \mathbf{A} \Delta + \gamma (\mathbf{H} - \mathbf{I} \bar{W}_n) \Delta \bar{W}_n}{1 + |\gamma(\mathbf{H} - \mathbf{I} \bar{W}_n)| |\Delta \bar{W}_n|} \right). \quad (10.4.15)$$

In this example we choose ξ to have the following set of ten states $\{0, 1, \dots, 9\}$, with transition intensity matrix \mathbf{A} such that ξ leaves each state with intensity 1, that is $a^{i,i} = -1$. It jumps only to the neighboring states with equal probability, that is with intensity 0.5 up or down if there are two neighboring states and otherwise with intensity 1 to the neighboring state. We choose the speed of adjustment parameter $\gamma = 8$ in the mean reversion term of the SDE (10.4.11).

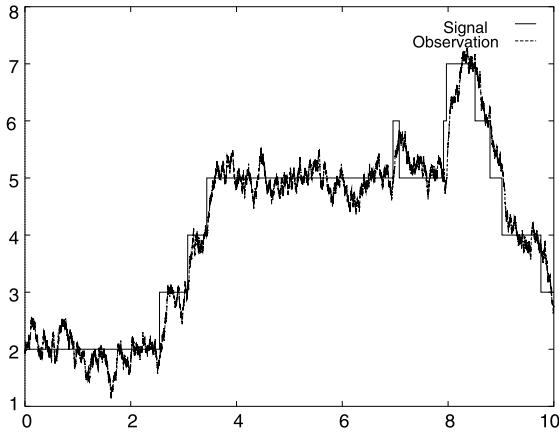


Fig. 10.4.7. Simulation of the signal and observation processes

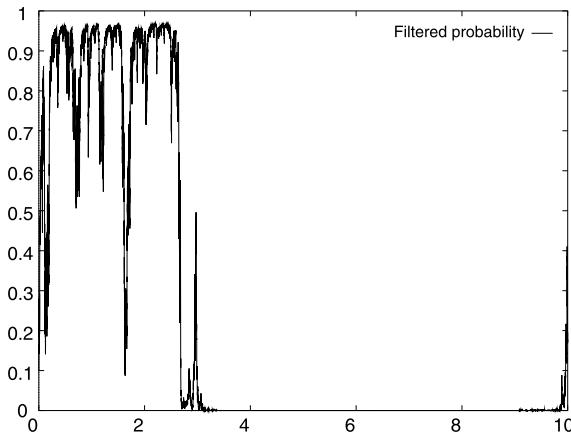


Fig. 10.4.8. q_t^3 - obtained by the drift-implicit Milstein scheme with $\theta = 0.5$ and $\Delta = \frac{1}{2000}$

The numerical experiments that were carried out in the case without mean reversion are repeated in the mean reversion case, with the minor change that all of the time step sizes are now halved. In this second study we again first inspect the trajectories generated by a single realization of the signal and observation processes. For the time step size $\Delta = \frac{1}{2000}$, Fig. 10.4.7 shows the simulated paths of the signal process ξ and the observation process \bar{W} . The generated observation process is a model for some quantity that is fluctuating around the hidden reference level, which is here allowed to take ten different values.

To compare the output from the various numerical schemes we consider $q_t^3 = E(\mathbf{X}_t^\top \mathbf{e}_3 | \mathcal{Y}_t)$, which is the filtered probability that X is in the third

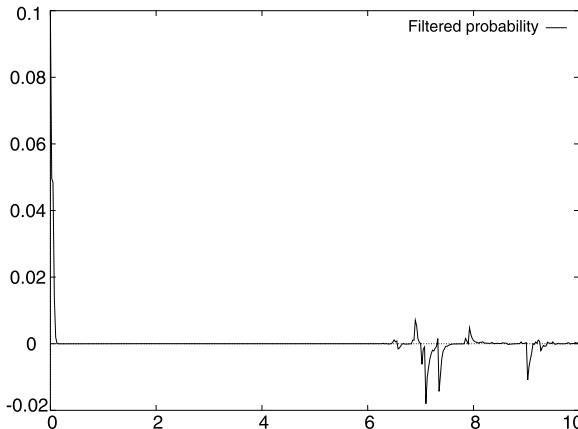


Fig. 10.4.9. q_t^3 - obtained by the Euler method with $\Delta = \frac{1}{40}$

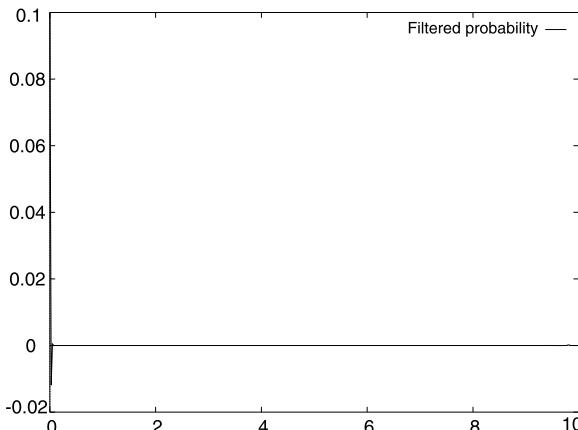


Fig. 10.4.10. q_t^3 - obtained by the Milstein scheme with $\Delta = \frac{1}{40}$

state, which has the value $\xi = 2$. Fig. 10.4.8 shows the plot of the approximate value of q_t^3 computed by the drift-implicit scheme with $\theta = 0.5$ with an extremely fine time step size $\Delta = \frac{1}{2000}$. We interpret this trajectory as being close to the exact trajectory of q_t^3 . Note in fact the correct detection of the periods when the hidden state was indeed equal to 2.

Figs. 10.4.9 to 10.4.12 display the filter q_t^3 calculated by using the four alternative schemes when the observation time step size is much larger, namely $\Delta = \frac{1}{40}$. Here only the balanced implicit method produces a useful filter. For this rough step size the trajectory of this filter still approximates surprisingly well the trajectory that was computed with the much finer step size in Fig. 10.4.8. The other numerical schemes clearly fail for the larger step size due to their insufficient numerical stability in filtering.

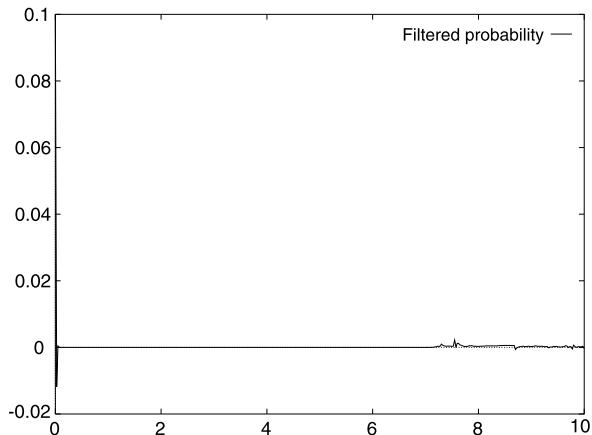


Fig. 10.4.11. q_t^3 - obtained by the semi-drift-implicit Milstein scheme with $\Delta = \frac{1}{40}$

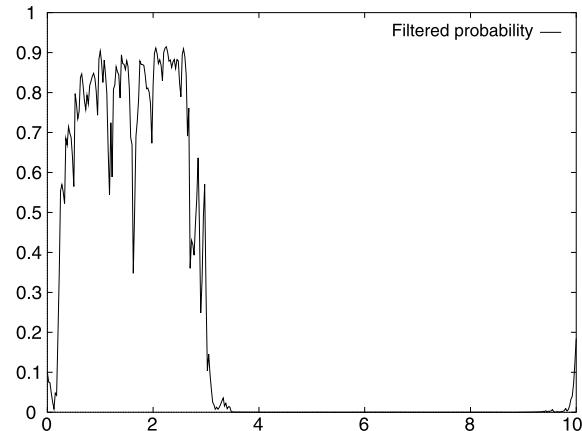


Fig. 10.4.12. q_t^3 - obtained by the balanced implicit method with $\Delta = \frac{1}{40}$

The balanced implicit method can potentially provide the necessary reliability that needs to be guaranteed. It extends considerably the class of problems that are amenable to the application of stochastic numerical techniques.

10.5 A Benchmark Approach to Filtering in Finance

This section discusses the use of the benchmark approach in finance for filtering and pricing in incomplete markets, when there are unobserved factors to be estimated. The proposed filtering framework is applicable also in cases when there does not exist an equivalent risk neutral probability measure. The results presented were obtained in [Platen & Rungaldier \(2005\)](#).

Financial Modeling and Filtering

In financial modeling it is sometimes the case that not all quantities, which determine the dynamics of security prices, can be fully observed. For instance, there is an increasing literature on models that describe regime switching. Some of the factors that impact the market dynamics may be hidden, say the kind of the current regime. However, these unobserved factors may be essential to satisfactorily build a market model for the type of dynamics that one empirically observes. This then leads to a filtering problem. Filtering methods, as those described previously, determine the so called *filter distribution* of the unobserved factors, given the available information. This distribution then allows us to compute the expectation of quantities, for instance, derivative prices that are dependent on unobserved factors.

There is a growing literature in the area of filtering in finance. To mention a few publications one can refer to [Elliott & van der Hoek \(1997\)](#), [Fischer, Platen & Runggaldier \(1999\)](#), [Elliott et al. \(1999b\)](#), [Fischer & Platen \(1999\)](#), [Landen \(2000\)](#), [Gombani & Runggaldier \(2001\)](#), [Frey & Runggaldier \(2001\)](#), [Elliott & Platen \(2001\)](#), [Bhar, Chiarella & Runggaldier \(2002, 2004\)](#), [Chiarella, Pasquali & Runggaldier \(2001\)](#) and [Crisan & Rozovski \(2010\)](#). The areas of application include optimal asset allocation, interest rate and credit risk term structure modeling, estimation of risk premia, volatility estimation and hedging under partial observation.

A key problem that arises in the literature for most filtering applications in finance is the determination of a suitable risk neutral equivalent probability measure for the pricing of derivatives under various information structures. The resulting derivative prices usually depend significantly on the chosen measure, as described in [Runggaldier \(2004\)](#). Given the observed quantities and their historical evolution, filtering can be based on the real world probability measure. Since risk neutral pricing may be too restrictive it is important to explore alternative filtering approaches in finance that are based on the real world probability measure and allow consistent derivative pricing under partial information, as well as other filtering tasks to be performed.

This section applies the benchmark approach to filtering, along the lines of [Platen & Runggaldier \(2005, 2007\)](#), where the benchmark is chosen as the growth optimal portfolio (GOP), see Chap. 3 and [Kelly \(1956\)](#), [Long \(1990\)](#) and [Platen \(2002\)](#). Given an information structure, one naturally obtains in this approach a *fair* price system, where benchmarked prices are martingales and, thus, their benchmarked present value equals their expected future benchmarked value. This avoids the delicate involvement of putative equivalent risk neutral probability measures under incomplete information in an incomplete market. We have seen in Chap. 3, under the benchmark approach all nonnegative portfolios, when expressed in units of the GOP, turn out to be supermartingales with respect to the given real world probability measure. Consequently, complex mathematical issues, which result from measure transformations under various information structures, can be avoided. Moreover, in

cases when no equivalent risk neutral probability measure exists, the benchmark approach allows us still to price contingent claims consistently. The previous section highlights how filters can be numerically calculated. Now, we employ filters in a financial setting.

Factor Model

To build a financial market model with a sufficiently rich structure and computational tractability we introduce a multi-factor model with $n \geq 2$ factors Z^1, Z^2, \dots, Z^n , forming the vector process

$$\mathbf{Z} = \left\{ \mathbf{Z}_t = (Z_t^1, \dots, Z_t^k, Z_t^{k+1}, \dots, Z_t^n)^\top, t \in [0, T] \right\}. \quad (10.5.1)$$

We shall assume that not all of the factors are observable. More precisely, only the first k factors are directly observed, while the remaining $n - k$ are not. Here k is an integer with $1 \leq k < n$ that we shall suppose to be fixed during most of this section. However, later on we shall discuss the implications of a varying k . For fixed k we shall consider the following subvectors of \mathbf{Z}_t

$$\mathbf{Y}_t = (Y_t^1, \dots, Y_t^k)^\top = (Z_t^1, \dots, Z_t^k)^\top$$

and

$$\mathbf{X}_t = (X_t^1, \dots, X_t^{n-k})^\top = (Z_t^{k+1}, \dots, Z_t^n)^\top \quad (10.5.2)$$

with \mathbf{Y}_t representing the *observed* and \mathbf{X}_t the hidden or *unobserved factors*. To be specific, we assume that \mathbf{Y}_t includes as components the observed security prices. These are given by $d + 1$ *primary security account* processes S^0, S^1, \dots, S^d , $d \in \{1, 2, \dots, k - 1\}$, as was the case in Chap. 3. We assume that a primary security account holds only units of one security and the income or loss accrued from holding the units of this security is always reinvested. In the case of shares this models the usual ownership of productive units. Here S^0 is the *savings account* process $S^0 = \{S_t^0, t \in [0, T]\}$, where T is a fixed time horizon. Note that in our framework the market can be *incomplete*. We shall identify S_t^j with Y_t^j for $j \in \{1, 2, \dots, d\}$ and the short rate r_t with Y_t^{d+1} for $t \in [0, T]$. This means, we consider the short rate to be observable.

Let there be given a filtered probability space $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$, where $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0, T]}$ is a given filtration to which all the processes will be adapted. We assume that the hidden factors satisfy the SDE

$$d\mathbf{X}_t = \mathbf{a}_t(\mathbf{Z}_t) dt + \mathbf{b}_t(\mathbf{Z}_t) d\mathbf{W}_t + \mathbf{g}_{t-}(\mathbf{Z}_{t-}) d\mathbf{M}_t \quad (10.5.3)$$

and the observed factors the SDE

$$d\mathbf{Y}_t = \mathbf{A}_t(\mathbf{Z}_t) dt + \mathbf{B}_t(\mathbf{Y}_t) d\mathbf{V}_t + \mathbf{G}_{t-}(\mathbf{Y}_{t-}) d\mathbf{N}_t \quad (10.5.4)$$

for $t \in [0, T]$ with given vector $\mathbf{Z}_0 = (Y_0^1, \dots, Y_0^k, X_0^1, \dots, X_0^{n-k})^\top$ of initial values. Here

$$\mathbf{W} = \left\{ \mathbf{W}_t = (W_t^1, \dots, W_t^k, W_t^{k+1}, \dots, W_t^n)^\top, t \in [0, T] \right\} \quad (10.5.5)$$

is an n -dimensional $(\underline{\mathcal{A}}, P)$ -Wiener process and

$$\mathbf{V}_t = (W_t^1, \dots, W_t^k)^\top \quad (10.5.6)$$

is the subvector of its first k components. Notice, to avoid the degeneracy of the filter, we let $\mathbf{B}_t(\mathbf{Y}_t)$ and $\mathbf{G}_{t-}(\mathbf{Y}_{t-})$ depend only on \mathbf{Y}_t and \mathbf{Y}_{t-} , respectively. The process $\mathbf{M} = \{\mathbf{M}_t = (M_t^1, \dots, M_t^k, M_t^{k+1}, \dots, M_t^n)^\top, t \in [0, T]\}$ is an n -dimensional $(\underline{\mathcal{A}}, P)$ -jump martingale defined as follows: Consider n counting processes N^1, \dots, N^n having no common jumps. These are at time $t \in [0, T]$ characterized by the corresponding vector of intensities $\boldsymbol{\lambda}_t(\mathbf{Z}_t) = (\lambda_t^1(\mathbf{Z}_t), \dots, \lambda_t^n(\mathbf{Z}_t))^\top$, where

$$\lambda_t^i(\mathbf{Z}_t) = \tilde{\lambda}_t^i(\mathbf{Y}_t) \quad (10.5.7)$$

for $t \in [0, T]$ and $i \in \{1, 2, \dots, k\}$. This means that we assume, without loss of generality, that the jump intensities of the first k counting processes are observed. The i th $(\underline{\mathcal{A}}, P)$ -jump martingale is then defined by the stochastic differential

$$dM_t^i = dN_t^i - \lambda_t^i(\mathbf{Z}_{t-}) dt \quad (10.5.8)$$

for $t \in [0, T]$ and $i \in \{1, 2, \dots, n\}$. In the SDE (10.5.3)

$$\mathbf{N}_t = (N_t^1, \dots, N_t^k)^\top \quad (10.5.9)$$

denotes the vector of the first k counting processes at time $t \in [0, T]$. Concerning the coefficients in the SDE (10.5.3)–(10.5.4), we assume that the vectors $\mathbf{a}_t(\mathbf{Z}_t)$, $\mathbf{A}_t(\mathbf{Z}_t)$, $\boldsymbol{\lambda}_t(\mathbf{Z}_t)$ and the matrices $\mathbf{b}_t(\mathbf{Z}_t)$, $\mathbf{B}_t(\mathbf{Y}_t)$, $\mathbf{g}_t(\mathbf{Z}_t)$ and $\mathbf{G}_t(\mathbf{Y}_t)$ are such that a unique strong solution of (10.5.3) exists that does not explode until time T , see Chap. 1. We shall also assume that the $k \times k$ -matrix $\mathbf{B}_t(\mathbf{Y}_t)$ is invertible for all $t \in [0, T]$. Finally, $\mathbf{g}_t(\mathbf{Z}_t)$ may be any bounded function and the $k \times k$ -matrix $\mathbf{G}_t(\mathbf{Y}_t)$ is assumed to be a given function of \mathbf{Y}_t that is invertible for each $t \in [0, T]$. This latter assumption implies that, since there are no common jumps among the components of \mathbf{N}_t , by observing a jump of \mathbf{Y}_t we can establish which of the processes N^i , $i \in \{1, 2, \dots, k\}$, has jumped.

Finite Dimensional Filter

In addition to the filtration $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0, T]}$, which represents the *total information*, we shall also consider the subfiltration

$$\underline{\mathcal{Y}}^k = (\mathcal{Y}_t^k)_{t \in [0, T]} \subseteq \underline{\mathcal{A}}, \quad (10.5.10)$$

where $\mathcal{Y}_t^k = \sigma\{\mathbf{Y}_s = (Z_s^1, \dots, Z_s^k)^\top, s \in [0, t]\}$ represents the *observed information* at time $t \in [0, T]$. Thus $\underline{\mathcal{Y}}^k$ provides the structure of the evolution

of the actually available information in the market, which depends on the specification of the degree k of observed information.

We shall be interested in the conditional distribution of \mathbf{X}_t , given \mathcal{Y}_t^k , that we call, according to standard terminology, the *filter distribution* at time $t \in [0, T]$. There exist general filter equations for the dynamics described by the SDEs given in (10.5.3)–(10.5.4), see Liptser & Shiryaev (1977). It turns out that these are SDEs for the conditional expectations of integrable functions of the unobserved factors \mathbf{X}_t , given \mathcal{Y}_t^k . Notice that, in particular, $\exp\{\nu \cdot \mathbf{X}_t\}$ is, for given $\nu \in \mathbb{R}^k$ and with i denoting the imaginary unit, a bounded and, thus, integrable function of \mathbf{X}_t . Its conditional expectation leads to the conditional characteristic function of the distribution of \mathbf{X}_t , given \mathcal{Y}_t^k . The latter characterizes completely the entire filter distribution. Considering conditional expectations of integrable functions of \mathbf{X}_t is, thus, not a restriction for the identification of filter equations.

To keep the filter reasonably tractable we assume that the SDEs given in (10.5.3)–(10.5.4) are such that the corresponding filter distributions admit a representation of the form

$$P(Z_t^{k+1} \leq Z^{k+1}, \dots, Z_t^n \leq Z^n \mid \mathcal{Y}_t^k) = F_{Z_t^{k+1}, \dots, Z_t^n}(Z^{k+1}, \dots, Z^n \mid \zeta_t^1, \dots, \zeta_t^q) \quad (10.5.11)$$

for all $t \in [0, T]$. This means, that we have a *finite-dimensional filter*, characterized by the filter state process

$$\zeta = \left\{ \zeta_t = (\zeta_t^1, \dots, \zeta_t^q)^\top, t \in [0, T] \right\}, \quad (10.5.12)$$

which is an \mathcal{Y}_t^k -adapted process with finite dimension $q \geq 1$. We shall denote by $\tilde{\mathbf{Z}}_t^k$ the resulting $(k + q)$ -vector of *observables*

$$\tilde{\mathbf{Z}}_t^k = (Y_t^1, \dots, Y_t^k, \zeta_t^1, \dots, \zeta_t^q)^\top, \quad (10.5.13)$$

which consists of the k observed factors and the q components of the filter state process. Furthermore, the filter state ζ_t satisfies an SDE of the form

$$d\zeta_t = \mathbf{J}_t(\tilde{\mathbf{Z}}_t^k) dt + \mathbf{D}_{t-}(\tilde{\mathbf{Z}}_{t-}^k) d\mathbf{Y}_t \quad (10.5.14)$$

with $\mathbf{J}_t(\cdot)$ denoting a q -vector valued function and $\mathbf{D}_t(\cdot)$ a $q \times k$ -matrix valued function, $t \in [0, T]$.

There are various models of the type (10.5.3)–(10.5.4) that admit a finite-dimensional filter with ζ_t satisfying an equation of the form (10.5.14). The following two subsections will recall, in a generalized form, the two classical models we have discussed previously. These are the *conditionally Gaussian model*, which leads to a generalized Kalman-Bucy filter, see Sect. 10.1; and the *finite-state jump model*, which is related to hidden Markov chain filters that we studied in previous sections of this chapter. Various combinations of these models have finite-dimensional filters and can be applied in the present setting.

Conditionally Gaussian Filter Model

Assume that in the system of SDEs (10.5.3)–(10.5.4) the functions $\mathbf{a}_t(\cdot)$ and $\mathbf{A}_t(\cdot)$ are linear and that $\mathbf{b}_t(\mathbf{Z}_t) = \mathbf{b}_t$ is a deterministic function, while $\mathbf{g}_t(\mathbf{Z}_t) = \mathbf{G}_t(\mathbf{Y}_t) = \mathbf{0}$ for $t \in [0, T]$. This means that the model (10.5.3)–(10.5.4) takes the form

$$\begin{aligned} d\mathbf{X}_t &= [\mathbf{a}_t^0 + \mathbf{a}_t^1 \mathbf{X}_t + \mathbf{a}_t^2 \mathbf{Y}_t] dt + \mathbf{b}_t d\mathbf{W}_t \\ d\mathbf{Y}_t &= [\mathbf{A}_t^0 + \mathbf{A}_t^1 \mathbf{X}_t + \mathbf{A}_t^2 \mathbf{Y}_t] dt + \mathbf{B}_t(\mathbf{Y}_t) d\mathbf{V}_t, \end{aligned} \quad (10.5.15)$$

for $t \in [0, T]$ with given Gaussian initial values \mathbf{X}_0 and \mathbf{Y}_0 . Here \mathbf{a}_t^0 and \mathbf{A}_t^0 are column vectors of dimensions $(n - k)$ and k respectively, and $\mathbf{a}_t^1, \mathbf{a}_t^2, \mathbf{b}_t, \mathbf{A}_t^1, \mathbf{A}_t^2, \mathbf{B}_t(\mathbf{Y}_t)$ are matrices of appropriate dimensions. Recall that \mathbf{W} is an n -dimensional $(\underline{\mathcal{A}}, P)$ -Wiener process and \mathbf{V} the vector of its first k components.

In this case the filter distribution is a Gaussian distribution with vector mean $\hat{\mathbf{X}}_t = (\hat{X}_t^1, \dots, \hat{X}_t^{n-k})^\top$, where

$$\hat{X}_t^i = E(X_t^i | \mathcal{Y}_t^k) \quad (10.5.16)$$

and covariance matrix $\mathbf{C}_t = [C_t^{\ell,i}]_{\ell,i=1}^{n-k}$, where

$$C_t^{\ell,i} = E((X_t^\ell - \hat{X}_t^\ell)(X_t^i - \hat{X}_t^i)^\top | \mathcal{Y}_t^k) \quad (10.5.17)$$

for $t \in [0, T]$ and $\ell, i \in \{1, 2, \dots, n - k\}$. The dependence of $\hat{\mathbf{X}}_t$ and \mathbf{C}_t on k is for simplicity suppressed in our notation. The above filter can be obtained from a generalization of the well-known Kalman-Bucy filter, see Chapter 10 in Liptser & Shiryaev (1977) and Sect. 10.1, namely

$$\begin{aligned} d\hat{\mathbf{X}}_t &= [\mathbf{a}_t^0 + \mathbf{a}_t^1 \hat{\mathbf{X}}_t + \mathbf{a}_t^2 \mathbf{Y}_t] dt + [\bar{\mathbf{b}}_t \mathbf{B}_t(\mathbf{Y}_t)^\top + \mathbf{C}_t(\mathbf{A}_t^1)^\top] \\ &\quad \times (\mathbf{B}_t(\mathbf{Y}_t) \mathbf{B}_t(\mathbf{Y}_t)^\top)^{-1} [d\mathbf{Y}_t - (\mathbf{A}_t^0 + \mathbf{A}_t^1 \hat{\mathbf{X}}_t + \mathbf{A}_t^2 \mathbf{Y}_t) dt] \\ d\mathbf{C}_t &= \left\{ \mathbf{a}_t^1 \mathbf{C}_t + \mathbf{C}_t(\mathbf{a}_t^1)^\top + (\mathbf{b}_t \mathbf{b}_t^\top) - [\bar{\mathbf{b}}_t \mathbf{B}_t(\mathbf{Y}_t)^\top + \mathbf{C}_t(\mathbf{A}_t^1)^\top] \right. \\ &\quad \left. \times (\mathbf{B}_t(\mathbf{Y}_t) \mathbf{B}_t(\mathbf{Y}_t)^\top)^{-1} [\bar{\mathbf{b}}_t \mathbf{B}_t(\mathbf{Y}_t)^\top + \mathbf{C}_t(\mathbf{A}_t^1)^\top]^\top \right\} dt, \end{aligned} \quad (10.5.18)$$

where $\bar{\mathbf{b}}_t$ is the k -dimensional vector formed by the first k components of \mathbf{b}_t , $t \in [0, T]$. We recall that $\mathbf{B}_t(\mathbf{Y}_t)$ is assumed to be invertible. Obviously, this generalizes the results presented in Sect. 10.1

Although for $t \in [0, T]$, \mathbf{C}_t is defined as a conditional expectation, it follows from (10.5.18) that if $\mathbf{B}_t(\mathbf{Y}_t)$ does not depend on the observable factors \mathbf{Y}_t , then \mathbf{C}_t can be precomputed off-line. Notice that the computation of \mathbf{C}_t is contingent upon the knowledge of the coefficients in the second equation

of (10.5.18). These coefficients are given deterministic functions of time, except for $\mathbf{B}_t(\mathbf{Y}_t)$ which depends also on observed factors. The value of $\mathbf{B}_t(\mathbf{Y}_t)$ becomes known only at time t . However, this is sufficient to determine the solution of (10.5.18) at time t . The model (10.5.15) is in fact a *conditionally Gaussian filter model*, where the filter process ζ is given by the vector process $\hat{\mathbf{X}} = \{\hat{\mathbf{X}}_t, t \in [0, T]\}$ and the upper triangular array of the elements of the matrix process $\mathbf{C} = \{\mathbf{C}_t, t \in [0, T]\}$ with $q = (n - k) \frac{[3 + (n - k)]}{2}$. Note by (10.5.17) that the matrix \mathbf{C}_t is symmetric. Obviously, in the case when $\mathbf{B}_t(\mathbf{Y}_t)$ does not depend on \mathbf{Y}_t for all $t \in [0, T]$, then we have a *Gaussian filter model*.

Finite-State Jump Filter Model

Here we assume that the unobserved factors form a continuous time, $(n - k)$ -dimensional jump process $\mathbf{X} = \{\mathbf{X}_t = (X_t^1, \dots, X_t^{n-k})^\top, t \in [0, T]\}$, which can take a finite number M of values. More precisely, given a \mathbf{Z}_t -dependent matrix $\mathbf{g}_t(\mathbf{Z}_t)$ and an intensity vector $\boldsymbol{\lambda}_t(\mathbf{Z}_t) = (\lambda_t^1(\mathbf{Z}_t), \dots, \lambda_t^n(\mathbf{Z}_t))^\top$ at time $t \in [0, T]$ for the vector counting process $\bar{\mathbf{N}} = \{\bar{\mathbf{N}}_t = (\bar{N}_t^1, \dots, \bar{N}_t^n)^\top, t \in [0, T]\}$, we consider the particular case of the model equations (10.5.3)–(10.5.4), where in the \mathbf{X} -dynamics we have $\mathbf{a}_t(\mathbf{Z}_t) = \mathbf{g}_t(\mathbf{Z}_t)\boldsymbol{\lambda}_t(\mathbf{Z}_t)$ and $\mathbf{b}_t(\mathbf{Z}_t) = \mathbf{0}$. Thus, by (10.5.3)–(10.5.4) and (10.5.8) we have

$$d\mathbf{X}_t = \mathbf{g}_{t-}(\mathbf{Z}_{t-}) d\bar{\mathbf{N}}_t \quad (10.5.19)$$

for $t \in [0, T]$. Notice that the process \mathbf{X} of unobserved factors is here a pure jump process and is therefore piecewise constant. On the other hand, for the vector \mathbf{Y}_t of observed factors at time t we assume that it satisfies the equation (10.5.3) with $\mathbf{G}_t(\mathbf{Y}_t) = \mathbf{0}$. This means that the process of observed factors \mathbf{Y} is only perturbed by continuous noise and does not jump.

In this example, the filter distribution is completely characterized by the vector of conditional probabilities $\mathbf{p}(t) = (p^1(t), \dots, p^M(t))^\top$, where M is the number of possible states $\boldsymbol{\eta}^1, \dots, \boldsymbol{\eta}^M$ of the vector \mathbf{X}_t and

$$p(t)^j = P(\mathbf{X}_t = \boldsymbol{\eta}^j \mid \mathcal{Y}_t^k), \quad (10.5.20)$$

for $t \in [0, T]$ and $j \in \{1, 2, \dots, M\}$. Let $\tilde{a}_t^{i,j}(\mathbf{y}, \boldsymbol{\eta}^h)$ denote the transition kernel for \mathbf{X} at time t to jump from state i into state j given $\mathbf{Y}_t = \mathbf{y}$ and $\mathbf{X}_t = \boldsymbol{\eta}^h$, see Liptser & Shiryaev (1977). The components of the vector $\mathbf{p}(t)$ satisfy the following dynamics

$$\begin{aligned} dp(t)^j &= (\tilde{\mathbf{a}}_t(\mathbf{Y}_t, \mathbf{p}(t))^\top \mathbf{p}(t))^j dt + p(t)^j [\mathbf{A}_t(\mathbf{Y}_t, \boldsymbol{\eta}^j) - \tilde{\mathbf{A}}_t(\mathbf{Y}_t, \mathbf{p}(t))] \\ &\quad \times (\mathbf{B}_t(\mathbf{Y}_t) \mathbf{B}_t(\mathbf{Y}_t)^\top)^{-1} [d\mathbf{Y}_t - \tilde{\mathbf{A}}_t(\mathbf{Y}_t, \mathbf{p}(t)) dt], \end{aligned} \quad (10.5.21)$$

where

$$\begin{aligned}
(\tilde{\mathbf{a}}_t(\mathbf{Y}_t, \mathbf{p}(t))^\top \mathbf{p}(t))^j &= \sum_{i=1}^M \left(\sum_{h=1}^M \bar{a}_t^{i,j}(\mathbf{Y}_t, \boldsymbol{\eta}^h) p(t)^h \right) p(t)^i \\
\mathbf{A}_t(\mathbf{Y}_t, \boldsymbol{\eta}^j) &= \mathbf{A}_t(\mathbf{Y}_t, \mathbf{X}_t) \Big|_{X_t=\eta^j} \\
\tilde{\mathbf{A}}_t(\mathbf{Y}_t, \mathbf{p}(t)) &= \sum_{j=1}^M \mathbf{A}_t(\mathbf{Y}_t, \boldsymbol{\eta}^j) p(t)^j
\end{aligned} \tag{10.5.22}$$

for $t \in [0, T]$, $j \in \{1, 2, \dots, M\}$. The filter state process $\zeta = \{\zeta_t = (\zeta_t^1, \dots, \zeta_t^q)^\top, t \in [0, T]\}$ for the finite state jump model is thus given by setting

$$\zeta_t^j = p^j(t) \tag{10.5.23}$$

for $j \in \{1, 2, \dots, q\}$ with $q = M - 1$. Since the probabilities add to one, we need only $M - 1$ probabilities to characterize the filter.

Markovian Representation

As in the two previous examples we have, in general, to deal with the quantity $E(\mathbf{A}_t(\mathbf{Z}_t) \mid \mathcal{Y}_t^k)$, assuming that it exists. This is the conditional expectation of the coefficient $\mathbf{A}_t(\mathbf{Z}_t) = \mathbf{A}_t(Y_t^1, \dots, Y_t^k, X_t^1, \dots, X_t^{n-k})$ with respect to the filter distribution at time t for the unobserved factors \mathbf{X}_t , see (10.5.3)–(10.5.4). Since the filter is characterized by the filter state process ζ , we obtain for this conditional expectation the representation

$$\tilde{\mathbf{A}}_t(\tilde{\mathbf{Z}}_t^k) = E(\mathbf{A}_t(\mathbf{Z}_t) \mid \mathcal{Y}_t^k), \tag{10.5.24}$$

where the vector $\tilde{\mathbf{Z}}_t^k$ is as defined in (10.5.13).

Notice that, in the case of the conditionally Gaussian model, the expression $\tilde{\mathbf{A}}_t(\tilde{\mathbf{Z}}_t^k)$ takes the particular form

$$\tilde{\mathbf{A}}_t(\tilde{\mathbf{Z}}_t^k) = \mathbf{A}_t^0 + \mathbf{A}_t^1 \hat{\mathbf{X}}_t + \mathbf{A}_t^2 \mathbf{Y}_t. \tag{10.5.25}$$

Furthermore, for the finite-state jump model, $\tilde{\mathbf{A}}_t(\tilde{\mathbf{Z}}_t^k)$ can be represented as

$$\tilde{\mathbf{A}}_t(\tilde{\mathbf{Z}}_t^k) = \tilde{\mathbf{A}}_t(\mathbf{Y}_t, \mathbf{p}(t)) = \sum_{j=1}^M \mathbf{A}_t(\mathbf{Y}_t, \boldsymbol{\eta}^j) p(t)^j \tag{10.5.26}$$

for $t \in [0, T]$, see (10.5.22).

We can now formulate the following generalization of Theorem 7.12 in Liptser & Shiryaev (1977), which provides a representation of the SDE for the observed factors.

Proposition 10.5.1 (Platen-Runggaldier) Let $\mathbf{A}_t(\mathbf{Z}_t)$ and the invertible matrix $\mathbf{B}_t(\mathbf{Y}_t)$ in (10.5.3)–(10.5.4) be such that

$$\int_0^T E(|\mathbf{A}_t(\mathbf{Z}_t)|) dt < \infty \quad \text{and} \quad \int_0^T \mathbf{B}_t(\mathbf{Y}_t) \mathbf{B}_t(\mathbf{Y}_t)^\top dt < \infty \quad (10.5.27)$$

P -a.s. Then there exists a k -dimensional $\underline{\mathcal{Y}}^k$ -adapted Wiener process $\tilde{\mathbf{V}} = \{\tilde{\mathbf{V}}_t, t \in [0, T]\}$ such that the process $\mathbf{y} = \{\mathbf{Y}_t, t \in [0, T]\}$ of observed factors in (10.5.3)–(10.5.4) satisfies the SDE

$$d\mathbf{Y}_t = \tilde{\mathbf{A}}_t(\tilde{\mathbf{Z}}_t^k) dt + \mathbf{B}_t(\mathbf{Y}_t) d\tilde{\mathbf{V}}_t + \mathbf{G}_{t-}(\mathbf{Y}_{t-}) d\mathbf{N}_t \quad (10.5.28)$$

with $\tilde{\mathbf{A}}_t(\tilde{\mathbf{Z}}_t^k)$ as in (10.5.24).

Proof: Denote by \mathbf{Y}^c the continuous part of the observation process \mathbf{Y} , that is

$$\mathbf{Y}_t^c = \mathbf{Y}_t - \sum_{\tau_j \leq t} \mathbf{G}_{\tau_j-}(\mathbf{Y}_{\tau_j-}) \Delta \mathbf{N}_{\tau_j}, \quad (10.5.29)$$

where the τ_j denote the jump times of $\mathbf{N} = \{\mathbf{N}_t, t \in [0, T]\}$, and $\Delta \mathbf{N}_{\tau_j} = \mathbf{N}_{\tau_j} - \mathbf{N}_{\tau_j-}$ is the vector $(\Delta N_{\tau_j-}^1, \dots, \Delta N_{\tau_j-}^k)^\top$. Let us now define the k -dimensional $\underline{\mathcal{Y}}^k$ -adapted process $\tilde{\mathbf{V}} = \{\tilde{\mathbf{V}}_t, t \in [0, T]\}$ by

$$\mathbf{B}_t(\mathbf{Y}_t) d\tilde{\mathbf{V}}_t = d\mathbf{Y}_t^c - \tilde{\mathbf{A}}_t(\tilde{\mathbf{Z}}_t^k) dt. \quad (10.5.30)$$

From (10.5.3)–(10.5.4), (10.5.29) and (10.5.30) it follows that

$$d\tilde{\mathbf{V}}_t = d\mathbf{V}_t + \mathbf{B}_t(\mathbf{Y}_t)^{-1} [\mathbf{A}_t(\mathbf{Z}_t) - \tilde{\mathbf{A}}_t(\tilde{\mathbf{Z}}_t^k)] dt. \quad (10.5.31)$$

From this we find, by the Itô formula with $\boldsymbol{\nu} \in \Re^k$ a row vector and i the imaginary unit, that

$$\begin{aligned} \exp \left\{ i\boldsymbol{\nu} (\tilde{\mathbf{V}}_t - \tilde{\mathbf{V}}_s) \right\} &= 1 + i\boldsymbol{\nu} \int_s^t \exp \left\{ i\boldsymbol{\nu} (\tilde{\mathbf{V}}_u - \tilde{\mathbf{V}}_s) \right\} d\mathbf{V}_u \\ &\quad + i\boldsymbol{\nu} \int_s^t \exp \left\{ i\boldsymbol{\nu} (\tilde{\mathbf{V}}_u - \tilde{\mathbf{V}}_s) \right\} \mathbf{B}_u^{-1}(\mathbf{Y}_u) (\mathbf{A}_u(\mathbf{Z}_u) - \tilde{\mathbf{A}}_u(\tilde{\mathbf{Z}}_u^k)) du \\ &\quad - \frac{\boldsymbol{\nu} \boldsymbol{\nu}^\top}{2} \int_s^t \exp \left\{ i\boldsymbol{\nu} (\tilde{\mathbf{V}}_u - \tilde{\mathbf{V}}_s) \right\} du. \end{aligned} \quad (10.5.32)$$

By recalling that \mathbf{V} is an $\underline{\mathcal{Y}}^k$ -measurable Wiener process, one notices that

$$E \left(\int_s^t \exp \left\{ i\boldsymbol{\nu} (\tilde{\mathbf{V}}_u - \tilde{\mathbf{V}}_s) \right\} d\mathbf{V}_u \mid \mathcal{Y}_s^k \right) = \mathbf{0} \quad (10.5.33)$$

and that, by our assumptions and by the boundedness of $\exp \{i\boldsymbol{\nu} (\tilde{\mathbf{V}}_u - \tilde{\mathbf{V}}_s)\}$,

$$\begin{aligned}
E \left(\int_s^t \exp \left\{ \boldsymbol{\nu} (\tilde{\mathbf{V}}_u - \tilde{\mathbf{V}}_s) \right\} \mathbf{B}_u^{-1}(\mathbf{Y}_u) \left(\mathbf{A}_u(\mathbf{Z}_u) - \tilde{\mathbf{A}}_u(\tilde{\mathbf{Z}}_u^k) \right) du \mid \mathcal{Y}_s^k \right) = \\
E \left(\int_s^t \exp \left\{ \boldsymbol{\nu} (\tilde{\mathbf{V}}_u - \tilde{\mathbf{V}}_s) \right\} \mathbf{B}_u^{-1}(\mathbf{Y}_u) E \left(\left(\mathbf{A}_u(\mathbf{Z}_u) - \tilde{\mathbf{A}}_u(\tilde{\mathbf{Z}}_u^k) \right) \mid \mathcal{Y}_u^k \right) du \mid \mathcal{Y}_s^k \right) \\
= \mathbf{0}. \quad (10.5.34)
\end{aligned}$$

Taking conditional expectations on the left and the right hand sides of (10.5.32) we end up with the equation

$$E \left(\exp \left\{ \boldsymbol{\nu} [(\tilde{\mathbf{V}}_t - \tilde{\mathbf{V}}_s)] \right\} \mid \mathcal{Y}_s^k \right) = 1 - \frac{\boldsymbol{\nu} \boldsymbol{\nu}^\top}{2} \int_s^t E \left(\exp \left\{ \boldsymbol{\nu} (\tilde{\mathbf{V}}_u - \tilde{\mathbf{V}}_s) \right\} \mid \mathcal{Y}_s^k \right) du, \quad (10.5.35)$$

which has the solution

$$E \left(\exp \left\{ \boldsymbol{\nu} (\tilde{\mathbf{V}}_t - \tilde{\mathbf{V}}_s) \right\} \mid \mathcal{Y}_s^k \right) = \exp \left\{ - \frac{\boldsymbol{\nu} \boldsymbol{\nu}^\top}{2} (t-s) \right\} \quad (10.5.36)$$

for $0 \leq s \leq t \leq T$. We can conclude that $(\tilde{\mathbf{V}}_t - \tilde{\mathbf{V}}_s)$ is a k -dimensional vector of independent \mathcal{Y}_t^k -measurable Gaussian random variables, each with variance $(t-s)$ and independent of \mathcal{Y}_s^k . By Lévy's theorem, see Sect. 1.3, $\tilde{\mathbf{V}}$ is thus a k -dimensional $\underline{\mathcal{Y}}^k$ -adapted standard Wiener process. \square

Markovian Model for Observables

Instead of the original factors $\mathbf{Z}_t = (Y_t^1, \dots, Y_t^k, X_t^1, \dots, X_t^{n-k})^\top = (Z_t^1, \dots, Z_t^n)^\top$, where $\mathbf{X}_t = (X_t^1, \dots, X_t^{n-k})^\top$ is unobserved, we may now base our analysis on the components of the vector $\tilde{\mathbf{Z}}_t^k = (Y_t^1, \dots, Y_t^k, \zeta_t^1, \dots, \zeta_t^q)^\top$, see (10.5.13), that are all observed. Just as was the case with $\mathbf{Z} = \{\mathbf{Z}_t, t \in [0, T]\}$, also the vector process $\tilde{\mathbf{Z}}^k = \{\tilde{\mathbf{Z}}_t^k, t \in [0, T]\}$ has a Markovian dynamics. In fact, replacing $d\mathbf{Y}_t$ in (10.5.14) by its expression resulting from (10.5.28), we obtain

$$\begin{aligned}
d\zeta_t &= \left[\mathbf{J}_t(\tilde{\mathbf{Z}}_t^k) + \mathbf{D}_t(\tilde{\mathbf{Z}}_t^k) \tilde{\mathbf{A}}_t(\tilde{\mathbf{Z}}_t^k) \right] dt + \mathbf{D}_t(\tilde{\mathbf{Z}}_t^k) \mathbf{B}_t(\mathbf{Y}_t) d\tilde{\mathbf{V}}_t \\
&\quad + \mathbf{D}_{t-}(\tilde{\mathbf{Z}}_{t-}^k) \mathbf{G}_{t-}(\mathbf{Y}_{t-}) dN_t \\
&= \tilde{\mathbf{J}}_t(\tilde{\mathbf{Z}}_t^k) dt + \tilde{\mathbf{D}}_t(\tilde{\mathbf{Z}}_t^k) d\tilde{\mathbf{V}}_t + \tilde{\mathbf{G}}_{t-}(\tilde{\mathbf{Z}}_{t-}^k) dN_t, \quad (10.5.37)
\end{aligned}$$

whereby we implicitly define the vector $\tilde{\mathbf{J}}_t(\tilde{\mathbf{Z}}_t^k)$ and the matrices $\tilde{\mathbf{D}}_t(\tilde{\mathbf{Z}}_t^k)$ and $\tilde{\mathbf{G}}_t(\tilde{\mathbf{Z}}_t^k)$ for compact notation.

From equations (10.5.28) and (10.5.37) we immediately obtain the following result.

Corollary 10.5.2 *The dynamics of the vector $\tilde{\mathbf{Z}}_t^k = (\mathbf{Y}_t, \zeta_t)$ can be expressed by the system of SDEs*

$$\begin{aligned} d\mathbf{Y}_t &= \tilde{\mathbf{A}}_t(\tilde{\mathbf{Z}}_t^k) dt + \mathbf{B}_t(\mathbf{Y}_t) d\tilde{\mathbf{V}}_t + \mathbf{G}_{t-}(\mathbf{Y}_{t-}) d\mathbf{N}_t \\ d\zeta_t &= \tilde{\mathbf{J}}_t(\tilde{\mathbf{Z}}_t^k) dt + \tilde{\mathbf{D}}_t(\tilde{\mathbf{Z}}_t^k) d\tilde{\mathbf{V}}_t + \tilde{\mathbf{G}}_{t-}(\tilde{\mathbf{Z}}_{t-}^k) d\mathbf{N}_t. \end{aligned} \quad (10.5.38)$$

From Corollary 10.5.2 it follows that the process $\tilde{\mathbf{Z}}^k = \{\tilde{\mathbf{Z}}_t^k, t \in [0, T]\}$ is Markovian.

Due to the existence of Markovian filter dynamics we have our original Markovian factor model, see (10.5.3)–(10.5.4), projected into a Markovian model for the observed quantities. Here the driving observable noise $\tilde{\mathbf{V}}$ is an (\mathcal{Y}^k, P) -Wiener process and the observable counting process \mathbf{N} is generated by the first k components N^1, N^2, \dots, N^k of the given n counting processes.

Given k , for efficient notation, we write for the vector of observables $\tilde{\mathbf{Z}}_t^k = \bar{\mathbf{Z}}_t = (\bar{Z}_t^1, \bar{Z}_t^2, \dots, \bar{Z}_t^{k+q})^\top$ the corresponding system of SDEs in the form

$$\begin{aligned} d\bar{Z}_t^\ell &= \theta^\ell(t, \bar{Z}_t^1, \bar{Z}_t^2, \dots, \bar{Z}_t^{k+q}) dt + \sum_{r=1}^k \beta^{\ell,r}(t, \bar{Z}_t^1, \bar{Z}_t^2, \dots, \bar{Z}_t^{k+q}) d\tilde{V}_t^r \\ &\quad + \sum_{r=1}^k \gamma^{\ell,r}(t, \bar{Z}_{t-}^1, \bar{Z}_{t-}^2, \dots, \bar{Z}_{t-}^{k+q}) dN_t^r \end{aligned} \quad (10.5.39)$$

for $t \in [0, T]$ and $\ell \in \{1, 2, \dots, k+q\}$. The functions, θ^ℓ , $\beta^{\ell,r}$ and $\gamma^{\ell,r}$ follow directly from $\tilde{\mathbf{A}}$, \mathbf{B} , \mathbf{G} , $\tilde{\mathbf{C}}$, $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{G}}$ appearing in (10.5.38).

We also have as an immediate consequence of the Markovianity of $\tilde{\mathbf{Z}}^k = \bar{\mathbf{Z}}$, as well as property (10.5.11), the following result.

Corollary 10.5.3 *Given $k \in \{1, 2, \dots, n-1\}$, any expectation of the form*

$$E(u(t, \mathbf{Z}_t) \mid \mathcal{Y}_t^k) < \infty$$

for a given function $u : [0, T] \times \Re^n \rightarrow \Re$ can be expressed as

$$E(u(t, \mathbf{Z}_t) \mid \mathcal{Y}_t^k) = \tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k) = \tilde{u}^k(t, \bar{\mathbf{Z}}_t) \quad (10.5.40)$$

with a suitable function $\tilde{u}^k : [0, T] \times \Re^{k+q} \rightarrow \Re$.

Relation (10.5.40) in Corollary 10.5.3 is of importance, for instance, for contingent claim pricing, as we shall see later on.

Primary Security Accounts and Portfolios

Recall that we have $d+1$ primary security account processes S^0, \dots, S^d with $d < k$, all of which are observable. This means, the vector process $\mathbf{S} = \{\mathbf{S}_t =$

$(S_t^0, \dots, S_t^d)^\top, t \in [0, T]\}$ is $\underline{\mathcal{Y}}^k$ -adapted. As previously assumed, the primary security accounts form part of the observable factors. In particular, we have set

$$Y_t^j = \bar{Z}_t^j = S_t^j$$

for $j \in \{1, 2, \dots, d\}$ and

$$Y_t^{d+1} = \bar{Z}_t^{d+1} = r_t$$

for $t \in [0, T]$.

Since the $d + 1$ primary security account processes coincide with the observable factors Y^1, \dots, Y^{d+1} , we may write their dynamics in a form corresponding to (10.5.39). To this effect let for $i \in \{1, 2, \dots, k\}$, by analogy to (10.5.8),

$$d\tilde{M}_t^i = dN_t^i - \tilde{\lambda}_t^i(\bar{Z}_t) dt \quad (10.5.41)$$

be the compensated i th $(\underline{\mathcal{Y}}^k, P)$ -jump martingale relative to the filtration $\underline{\mathcal{Y}}^k$. Here, with some abuse of notation, we have denoted by $\tilde{\lambda}_t^i(\bar{Z}_t)$ the compensating jump intensity for N^i with respect to $\underline{\mathcal{Y}}^k$. For simple notation, in what follows we shall often use \bar{Z}_t for \tilde{Z}_t^k , see (10.5.39). Let us now rewrite (10.5.39) more concisely in vector form as

$$d\bar{Z}_t = \bar{\alpha}(t, \bar{Z}_t) dt + \beta(t, \bar{Z}_t) d\tilde{V}_t + \gamma(t-, \bar{Z}_{t-}) d\tilde{M}_t \quad (10.5.42)$$

with

$$\bar{\alpha}(t, \bar{Z}_t) = \alpha(t, \bar{Z}_t) + \gamma(t-, \bar{Z}_{t-}) \tilde{\lambda}_{t-}(\bar{Z}_{t-}), \quad (10.5.43)$$

where \tilde{M} and $\tilde{\lambda}$ are the k -vectors with components \tilde{M}^i and $\tilde{\lambda}^i$, respectively. Here $\alpha(t, \bar{Z}_t)$ is a $(k + q)$ -column vector and $\beta(t, \bar{Z}_t)$ as well as $\gamma(t, \bar{Z}_t)$ are $((k + q) \times k)$ -matrices.

Since we have assumed $d < k$, the primary security accounts do not necessarily span the entire observable uncertainty of the market. It is therefore reasonable to assume that among the driving random processes \tilde{V}^i for $i \in \{1, 2, \dots, k\}$ and \tilde{M}^ℓ for $\ell \in \{1, 2, \dots, k\}$, those that directly drive the evolution of the asset prices S_t^j , $j \in \{1, 2, \dots, d\}$, are exactly d in number. Think for instance, of asset price models with stochastic volatility, where the volatilities are driven by stochastic processes that are independent from those that directly drive the evolution of the asset prices. We shall, thus, assume that for $j \in \{1, 2, \dots, d\}$ the dynamics of the j th primary security account is given by the SDE

$$dS_t^j = \bar{\alpha}^j(t, \bar{Z}_t) dt + \sum_{i=1}^{h_1} \beta^{j,i}(t, \bar{Z}_t) d\tilde{V}_t^i + \sum_{\ell=1}^{h_2} \gamma^{j,\ell}(t-, \bar{Z}_{t-}) d\tilde{M}_t^\ell \quad (10.5.44)$$

for $t \in [0, T]$, where $h_1 + h_2 = d$.

For efficient notation we now rewrite the SDE (10.5.44) in the form

$$dS_t^j = S_{t-}^j \left\{ r_t dt + \sum_{i=1}^{h_1} b_t^{j,i} \left(d\tilde{V}_t^i + \theta_t^i dt \right) + \sum_{\ell=h_1+1}^d b_{t-}^{j,\ell} \left(d\tilde{M}_t^{\ell-h_1} + \theta_{t-}^\ell dt \right) \right\} \quad (10.5.45)$$

for $t \in [0, T]$ with $S_t^j > 0$, $j \in \{0, 1, \dots, d\}$. Here we set $S_0^0 = 1$ and $b_t^{0,i} = 0$ for $t \in [0, T]$ and $i \in \{0, 1, \dots, d\}$, where r_t is the short rate. Above we have in (10.5.45) for $i \in \{1, 2, \dots, h_1\}$ the *volatility*

$$b_t^{j,i} = \frac{\beta^{j,i}(t, \bar{\mathbf{Z}}_t)}{S_t^j} \quad (10.5.46)$$

and for $i \in \{h_1 + 1, \dots, d\}$ the *jump coefficient*

$$b_{t-}^{j,i} = \frac{\gamma^{j,i-h_1}(t-, \bar{\mathbf{Z}}_{t-})}{S_{t-}^j} \quad (10.5.47)$$

for $t \in [0, T]$ and $j \in \{1, 2, \dots, d\}$. We assume that the matrix $\mathbf{b}_t = [b_t^{j,i}]_{j,i=1}^d$ is *invertible* for all $t \in [0, T]$. This allows us to write the *market price of risk* vector $\boldsymbol{\theta}_t = (\theta_t^1, \dots, \theta_t^d)^\top$ in the form

$$\boldsymbol{\theta}_t = \mathbf{b}_t^{-1} [\mathbf{a}_t - r_t \mathbf{1}] \quad (10.5.48)$$

for $t \in [0, T]$. Note that we suppress here the dependence on k . Here $\mathbf{1} = (1, \dots, 1)^\top$ is the unit vector and $\mathbf{a}_t = (a_t^1, \dots, a_t^d)^\top$ is the *appreciation rate* vector with

$$a_t^j = \frac{\bar{\alpha}^j(t, \bar{\mathbf{Z}}_t)}{S_t^j} \quad (10.5.49)$$

for $t \in [0, T]$ and $j \in \{1, 2, \dots, d\}$.

We say that an $\underline{\mathcal{Y}}^k$ -predictable stochastic process $\boldsymbol{\delta} = \{\boldsymbol{\delta}_t = (\delta_t^0, \dots, \delta_t^d)^\top, t \in [0, T]\}$ is a *self-financing strategy*, if $\boldsymbol{\delta}$ is \mathbf{S} -integrable. The corresponding *portfolio* S_t^δ has at time t the value

$$S_t^\delta = \sum_{j=0}^d \delta_t^j S_t^j \quad (10.5.50)$$

and has the dynamics

$$dS_t^\delta = \sum_{j=0}^d \delta_{t-}^j dS_t^j \quad (10.5.51)$$

for all $t \in [0, T]$.

Growth Optimal Portfolio

We denote by $S_t^{\delta_*}$ the value of the GOP for the above market, where the available information is described by the filtration $\underline{\mathcal{Y}}^k$. The corresponding SDE for the discounted GOP $\bar{S}_t^{\delta_*} = \frac{S_t^{\delta_*}}{S_0^0}$ was given in Chap. 3 in the form

$$\begin{aligned} d\bar{S}_t^{\delta_*} &= \bar{S}_{t-}^{\delta_*} \left\{ \sum_{i=1}^{h_1} \theta_t^i (\theta_t^i dt + d\tilde{V}_t^i) \right. \\ &\quad \left. + \sum_{i=h_1+1}^d \frac{\theta_{t-}^i}{\tilde{\lambda}_{t-}^{i-h_1}(\bar{Z}_{t-}) - \theta_{t-}^i} \left(\theta_{t-}^i dt + d\tilde{M}_t^{i-h_1} \right) \right\} \end{aligned} \quad (10.5.52)$$

for $t \in [0, T]$ with $S_0^{\delta_*} = 1$. Here \tilde{M}_t^i denotes the i th component at time t of the jump martingale \tilde{M} defined in (10.5.41).

For $j \in \{0, 1, \dots, d\}$, the j th benchmarked primary security account $\hat{S}^j = \{\hat{S}_t^j, t \in [0, T]\}$ has at time t the value

$$\hat{S}_t^j = \frac{S_t^j}{S_t^{\delta_*}}, \quad (10.5.53)$$

and satisfies by (10.5.45), (10.5.52) and application of the Itô formula the SDE

$$\begin{aligned} d\hat{S}_t^j &= \hat{S}_{t-}^j \left\{ \sum_{i=1}^{h_1} (b_t^{j,i} - \theta_t^i) d\tilde{V}_t^i \right. \\ &\quad \left. + \sum_{i=h_1+1}^d \left(b_{t-}^{j,i} \left(1 - \frac{\theta_{t-}^i}{\tilde{\lambda}_{t-}^{i-h_1}(\bar{Z}_{t-})} \right) - \frac{\theta_{t-}^i}{\tilde{\lambda}_{t-}^{i-h_1}(\bar{Z}_{t-})} \right) d\tilde{M}_t^{i-h_1} \right\} \end{aligned} \quad (10.5.54)$$

for $t \in [0, T]$ and $j \in \{0, 1, \dots, d\}$, see Chap. 3.

Furthermore, by application of the Itô formula it can be shown that the benchmarked portfolio $\hat{S}^\delta = \{\hat{S}_t^\delta, t \in [0, T]\}$ with

$$\hat{S}_t^\delta = \frac{S_t^\delta}{S_t^{\delta_*}} \quad (10.5.55)$$

satisfies the SDE

$$\begin{aligned} d\hat{S}_t^\delta &= \hat{S}_{t-}^\delta \left[\sum_{i=1}^{h_1} \left(\sum_{j=1}^d \frac{\delta_t^j \hat{S}_t^j}{\hat{S}_t^\delta} b_t^{j,i} - \theta_t^i \right) d\tilde{V}_t^i + \sum_{i=h_1+1}^d \left\{ \left(\sum_{j=1}^d \frac{\delta_{t-}^j \hat{S}_{t-}^j}{\hat{S}_{t-}^\delta} b_{t-}^{j,i} + 1 \right) \right. \right. \\ &\quad \times \left. \left. \left(1 - \frac{\theta_{t-}^i}{\tilde{\lambda}_{t-}^{i-h_1}(\bar{Z}_{t-})} \right) - 1 \right\} d\tilde{M}_t^{i-h_1} \right] \end{aligned} \quad (10.5.56)$$

for $t \in [0, T]$.

Note again that the j th benchmarked primary security account \hat{S}^j and all benchmarked portfolios are driftless and, thus, (\mathcal{Y}^k, P) -local martingales. Due to a result in Ansel & Stricker (1994) any nonnegative benchmarked portfolio

process can be shown to be an $(\underline{\mathcal{Y}}^k, P)$ -supermartingale, see [Platen \(2002\)](#). Hence, it is impossible for a nonnegative portfolio to generate with strictly positive probability any strictly positive wealth from zero initial capital. This means that there is no strong arbitrage in the sense of Definition [3.3.2](#).

We emphasize that the benchmarked security prices under the benchmark approach are, in general, not $(\underline{\mathcal{Y}}^k, P)$ -martingales. Note also that we may complete the market with nonredundant securities, thereby introducing, for example, European call and put options on the underlying securities, see, for instance, [Mancini \(2002\)](#). We shall, however, assume here that the market is not necessarily complete and there may not exist an equivalent risk neutral probability measure. For examples of reasonable and realistic models, for which there does not exist an equivalent risk neutral probability measure, we refer to Chap. 3 or [Heath & Platen \(2002a, 2002b, 2002c\)](#). We stress the fact that under real world pricing we operate with the GOP as numéraire portfolio and the real world probability measure as pricing measure, as explained in Chap. 3 and will be emphasized in the following section.

Derivative Pricing

As previously in Chap. 3 we say that a price process $V = \{V_t, t \in [0, T]\}$ is *fair* for the information represented by $\underline{\mathcal{Y}}^k$ if its benchmarked value $\hat{V}_t = \frac{V_t}{S_t^{\delta^*}}$ forms an $(\underline{\mathcal{Y}}^k, P)$ -martingale. We recall that $\underline{\mathcal{Y}}_t^k$ describes the information, which is available at time t , whereas \mathcal{A}_t is the complete information at time t that determines the original model dynamics including also the unobserved factors.

Notice that, for any time instant τ , the value $S_\tau^{\delta^*}$ of the GOP is a tradable portfolio, see (10.5.50), with $\underline{\mathcal{Y}}^k$ -predictable strategy $\underline{\delta}$, of primary security accounts that are themselves observable. It follows, as already mentioned earlier, that the GOP $S_\tau^{\delta^*}$ is $\underline{\mathcal{Y}}_t^k$ -measurable.

For a given maturity date τ , which is assumed to be an $\underline{\mathcal{Y}}^k$ -stopping time, we consider a *contingent claim* $U(\tau, \mathbf{Y}_\tau)$ as a function of τ and the corresponding values of observed factors \mathbf{Y}_τ , where we assume that

$$E \left(\left| \frac{U(\tau, \mathbf{Y}_\tau)}{S_\tau^{\delta^*}} \right| \middle| \underline{\mathcal{Y}}_t^k \right) < \infty \quad (10.5.57)$$

a.s. for all $t \in [0, \tau]$. Note that there is no reason to let the payoff function depend on any other factors than the observed factors. Otherwise, the payoff would not be verifiable at time τ .

Since, as mentioned before, $S_\tau^{\delta^*}$ is $\underline{\mathcal{Y}}_\tau^k$ -measurable, it can be considered as a function of \mathbf{Z}_s for $s \leq \tau$. Furthermore, since \mathbf{Y}_τ is a subvector of \mathbf{Z}_τ and $\mathbf{Z} = \{Z_t, t \in [0, T]\}$ is a Markov process, we can define the process $u = \{u(t, \mathbf{Z}_t), t \in [0, T]\}$ as

$$u(t, \mathbf{Z}_t) = E \left(\frac{S_t^{\delta^*}}{S_\tau^{\delta^*}} U(\tau, \mathbf{Y}_\tau) \mid \mathcal{A}_t \right) \quad (10.5.58)$$

for $t \in [0, \tau]$, which at time t exploits the complete information characterized by the sigma-algebra \mathcal{A}_t . Next, we consider

$$\tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k) = E(u(t, \mathbf{Z}_t) \mid \mathcal{Y}_t^k) \quad (10.5.59)$$

for $t \in [0, \tau]$, which by Corollary 10.5.3 can be computed on the basis of the previous filtering results. Combining (10.5.58) with (10.5.59) and using the fact that $S_t^{\delta_*}$ is \mathcal{Y}_t^k -measurable, we obtain

$$\frac{\tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k)}{S_t^{\delta_*}} = E\left(\frac{U(\tau, \mathbf{Y}_\tau)}{S_\tau^{\delta_*}} \mid \mathcal{Y}_t^k\right) \quad (10.5.60)$$

for $t \in [0, \tau]$. This means, the benchmarked value $\frac{\tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k)}{S_t^{\delta_*}}$ forms for $t \in [0, \tau]$ a $(P, \underline{\mathcal{Y}}^k)$ -martingale. Obviously, it is the only $(P, \underline{\mathcal{Y}}^k)$ -martingale that coincides at time τ with $\frac{U(\tau, \mathbf{Y}_\tau)}{S_\tau^{\delta_*}}$. Thus $\tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k)$ is the *fair price* at time t of the claim $U(\tau, \mathbf{Y}_\tau)$ for the information represented by $\underline{\mathcal{Y}}^k$.

The above concept of real world pricing under partial information generalizes real world pricing, as introduced in Chap. 3 and, thus, the well-known concept of risk neutral pricing. It avoids not only the assumption on the existence of an equivalent risk neutral probability measure but also some delicate issues that arise from measure changes under different filtrations.

Note that the expression in (10.5.59) fits perfectly the one for the filtered factor model given in (10.5.40). The actual computation of the conditional expectation in (10.5.59) is, therefore, equivalent to the solution of the filtering problem for the unobserved factors.

The previous analysis demonstrates the usefulness of benchmarking with respect to the GOP when it comes to pricing under partial information. Moreover, the fair price $\tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k)$ can be computed according to (10.5.58) and (10.5.59), using the value process S^{δ_*} of the GOP.

Variance of Benchmarked Prices

As already mentioned earlier, the degree of available information is indexed by the parameter k . A larger value of k means that more factors are observed, thus, providing more information in $\underline{\mathcal{Y}}^k$.

Let us now investigate the impact of varying the degrees of information k concerning the factors $\mathbf{Z}_t = (Z_t^1, \dots, Z_t^n)^\top$ that underly our model dynamics, see (10.5.2)–(10.5.4). We use now the notation $\tilde{\mathbf{Z}}_t^k$ for the $k+q$ -vector of observables defined in (10.5.13), where we stress its dependence on k and recall that, by (10.5.38), the process $\tilde{\mathbf{Z}}^k$ is Markovian. Consider then a contingent claim

$$U(\tau, \mathbf{Y}_\tau) = U(\tau, Y_\tau^1, Y_\tau^2, \dots, Y_\tau^r) \quad (10.5.61)$$

for some fixed $r \in \{1, 2, \dots, k\}$, where we assume that the number of observed factors that influence the claim equals r . Let $\tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k)$ be the corresponding fair price at time t under the information \mathcal{Y}_t^k , as given by (10.5.59). Recall that, by (10.5.59), $\tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k)$ is the conditional expectation, under the real world probability measure, of $u(t, \mathbf{Z}_t)$ given \mathcal{Y}_t^k , which implies that the corresponding *conditional variance*

$$\text{Var}_t^k(u) = E \left(\left(u(t, \mathbf{Z}_t) - \tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k) \right)^2 \mid \mathcal{Y}_t^k \right) \quad (10.5.62)$$

at time $t \in [0, \tau]$ is the minimal value of the mean square error, conditional on \mathcal{Y}_t^k , corresponding to the deviation from $u(t, \mathbf{Z}_t)$ of any \mathcal{Y}_t^k -measurable random variable. This conditional variance is computed under the real world probability measure. It would not make sense if computed under any other probability measure since the market participants are affected by the real difference between $u(t, \mathbf{Z}_t)$ and $\tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k)$.

Note that for larger k we have more information available, which naturally should reduce the above conditional variance. In [Platen & Runggaldier \(2005\)](#) the following result is shown, which quantifies the reduction in conditional variance and can also be seen as a generalization of the celebrated Rao-Blackwell theorem towards filtering in incomplete markets.

Theorem 10.5.4. (Platen-Runggaldier) *For $m \in \{0, 1, \dots, n - k\}$ and $k \in \{1, 2, \dots, n - 1\}$ we have*

$$E \left(\text{Var}_t^{k+m}(u) \mid \mathcal{Y}_t^k \right) = \text{Var}_t^k(u) - R_t^{k+m}, \quad (10.5.63)$$

where

$$R_t^{k+m} = E \left(\left(\tilde{u}^{k+m}(t, \tilde{\mathbf{Z}}_t^{k+m}) - \tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k) \right)^2 \mid \mathcal{Y}_t^k \right) \quad (10.5.64)$$

for $t \in [0, \tau]$.

Proof: For $t \in [0, \tau]$ and $k \in \{1, 2, \dots, n - 1\}$ we have

$$\begin{aligned} & \left(u(t, \mathbf{Z}_t) - \tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k) \right)^2 \\ &= \left(u(t, \mathbf{Z}_t) - \tilde{u}^{k+m}(t, \tilde{\mathbf{Z}}_t^{k+m}) \right)^2 + \left(\tilde{u}^{k+m}(t, \tilde{\mathbf{Z}}_t^{k+m}) - \tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k) \right)^2 \\ &+ 2 \left(u(t, \mathbf{Z}_t) - \tilde{u}^{k+m}(t, \tilde{\mathbf{Z}}_t^{k+m}) \right) \left(\tilde{u}^{k+m}(t, \tilde{\mathbf{Z}}_t^{k+m}) - \tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k) \right). \end{aligned} \quad (10.5.65)$$

By taking conditional expectations with respect to \mathcal{Y}_t^k on both sides of the above equation it follows that

$$\begin{aligned} \text{Var}_t^k(u) &= E \left(\text{Var}_t^{k+m}(u) \mid \mathcal{Y}_t^k \right) + R_t^{k+m} + 2 E \left(\left(\tilde{u}^{k+m}(t, \tilde{\mathbf{Z}}_t^{k+m}) - \tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k) \right) \right. \\ &\quad \times E \left(\left(u(t, \mathbf{Z}_t) - \tilde{u}^{k+m}(t, \tilde{\mathbf{Z}}_t^{k+m}) \right) \mid \mathcal{Y}_t^{k+m} \right) \left. \mid \mathcal{Y}_t^k \right). \end{aligned} \quad (10.5.66)$$

Since the last term on the right hand side is equal to zero by definition, we obtain (10.5.63). \square

Hedging Strategy

To determine a hedging strategy in an incomplete market we have to use one of the hedging criteria for incomplete markets. It turns out that the real world pricing concept is related to *local risk minimization*, see Föllmer & Schweizer (1991) and DiMasi, Platen & Runggaldier (1995). To see this let us introduce the benchmarked pricing function

$$\hat{u}(t, \tilde{\mathbf{Z}}_t^k) = \frac{\tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k)}{S_t^{\delta^*}} \quad (10.5.67)$$

for $t \in [0, T]$. To express conveniently the evolution of $\hat{u}(t, \tilde{\mathbf{Z}}_t^k)$ we introduce for $i \in \{1, 2, \dots, k+q\}$ the operator

$$L^i \hat{u}(t, \tilde{\mathbf{Z}}_t^k) = \sum_{\ell=1}^{k+q} \beta^{\ell, i}(t, \bar{Z}_t^1, \dots, \bar{Z}_t^{k+q}) \frac{\partial \hat{u}(t, \tilde{\mathbf{Z}}_t^k)}{\partial \bar{Z}^\ell} \quad (10.5.68)$$

and the jump operator

$$\begin{aligned} \Delta_{\hat{u}}^i(t-, \tilde{\mathbf{Z}}_{t-}^k) &= \hat{u} \left(t, \bar{Z}_{t-}^1 + \gamma^{1,i} \left(t-, \bar{Z}_{t-}^1, \dots, \bar{Z}_{t-}^{k+q} \right), \dots, \right. \\ &\quad \left. \bar{Z}_{t-}^{k+q} + \gamma^{k+q,i} \left(t-, \bar{Z}_{t-}, \dots, \bar{Z}_{t-}^{k+q} \right) \right) \\ &\quad - \hat{u} \left(t-, \bar{Z}_{t-}^1, \dots, \bar{Z}_{t-}^{k+q} \right) \end{aligned} \quad (10.5.69)$$

for $t \in [0, T]$ with $\beta^{\ell, i}$ and $\gamma^{\ell, i}$ as in (10.5.39).

Assume that the above benchmarked pricing function $\hat{u}(\cdot, \cdot)$ in (10.5.67) is differentiable with respect to time and twice differentiable with respect to the observables. Then we obtain with (10.5.39), (10.5.68) and (10.5.69) by the Itô formula for the (\mathcal{Y}^k, P) -martingale $\hat{u} = \{\hat{u}(t, \tilde{\mathbf{Z}}_t^k), t \in [0, \tau]\}$ the *martingale representation*

$$\begin{aligned} \frac{U(\tau, \mathbf{Y}_\tau)}{S_\tau^{\delta^*}} &= \hat{u}(\tau, \tilde{\mathbf{Z}}_\tau^k) \\ &= \hat{u}(t, \tilde{\mathbf{Z}}_t^k) + \hat{I}_{t,\tau} + \hat{R}_{t,\tau}, \end{aligned} \quad (10.5.70)$$

with *hedgable part*, see (10.5.44),

$$\hat{I}_{t,\tau} = \sum_{\ell=1}^{h_1} \int_t^\tau L^\ell \hat{u}(s, \tilde{\mathbf{Z}}_s^k) d\tilde{V}_s^\ell + \sum_{\ell=1}^{h_2} \int_t^\tau \Delta_{\hat{u}}^\ell(s-, \tilde{\mathbf{Z}}_{s-}^k) d\tilde{M}_s^\ell \quad (10.5.71)$$

and *unhedgable part*

$$\hat{R}_{t,\tau} = \sum_{\ell=h_1+1}^k \int_t^\tau L^\ell \hat{u}(s, \tilde{\mathbf{Z}}_s^k) d\tilde{V}_s^\ell + \sum_{\ell=h_2+1}^k \int_t^\tau \Delta_{\hat{u}}^\ell(s-, \tilde{\mathbf{Z}}_{s-}^k) d\tilde{M}_s^\ell \quad (10.5.72)$$

for $t \in [0, \tau]$. Note that (10.5.70) is an $(\underline{\mathcal{Y}}^k, P)$ -martingale representation for the benchmarked contingent claim.

Let us search for a fair benchmarked portfolio process \hat{V}_{δ_U} , with self-financing *hedging strategy* δ_U that matches the hedgable part $\hat{I}_{t,\tau}$. This means, we compare the SDE (10.5.56) for $\hat{V}_{\delta_U}(t)$ with that of the hedgable part $\hat{I}_{t,\tau}$, see (10.5.71). Here we use for notational convenience the j th *fraction*

$$\pi_{\delta_U}^j(t) = \frac{\delta_U^j(t) \hat{S}_t^j}{\hat{V}_{\delta_U}(t)} \quad (10.5.73)$$

of the value of the corresponding hedging portfolio that has to be invested into the j th primary security account at time $t \in [0, \tau]$, $j \in \{0, 1, \dots, d\}$. By this comparison it follows that we need to satisfy for $i \in \{1, 2, \dots, h_1\}$ the equation

$$\sum_{j=1}^d \pi_{\delta_U}^j(t) b_t^{j,i} - \theta_t^i = \frac{L^i \hat{u}(t, \tilde{\mathbf{Z}}_t^k)}{\hat{V}_{\delta_U}(t)} \quad (10.5.74)$$

and for $i \in \{1, \dots, h_2\}$ the relation

$$\left(\sum_{j=1}^d \pi_{\delta_U}^j(t-) b_{t-}^{j,i} + 1 \right) \left(1 - \frac{\theta_{t-}^i}{\tilde{\lambda}_{t-}^{i-h_1}(\bar{\mathbf{Z}}_{t-})} \right) - 1 = \frac{\Delta_{\hat{u}}^i(t-, \tilde{\mathbf{Z}}_{t-}^k)}{\hat{V}_{\delta_U}(t-)} \quad (10.5.75)$$

for $t \in [0, \tau]$. The equations (10.5.74) and (10.5.75) lead with $\mathbf{e}_U(t) = (e_U^1(t), \dots, e_U^d(t))^\top$, where

$$e_U^\ell(t-) = \begin{cases} \frac{L^\ell \hat{u}(t-, \tilde{\mathbf{Z}}_{t-}^k)}{\hat{V}_{\delta_U}(t-)} + \theta_t^\ell & \text{for } \ell \in \{1, 2, \dots, h_1\} \\ \frac{\tilde{\lambda}_{t-}^{i-h_1}(\bar{\mathbf{Z}}_{t-}) \frac{\Delta_{\hat{u}}^i(t-, \tilde{\mathbf{Z}}_{t-}^k)}{\hat{V}_{\delta_U}(t-)} + \theta_{t-}^i}{\tilde{\lambda}_{t-}^{i-h_1}(\bar{\mathbf{Z}}_{t-}) - \theta_{t-}^i} & \text{for } \ell = h_1 + i \in \{h_1 + 1, \dots, d\}, \end{cases} \quad (10.5.76)$$

and $\boldsymbol{\pi}_{\delta_U}(t) = (\pi_{\delta_U}^1(t), \dots, \pi_{\delta_U}^d(t))^\top$ to the vector equation

$$\mathbf{e}_U(t-) = (\boldsymbol{\pi}_{\delta_U}^\top(t-) \mathbf{b}_{t-})^\top \quad (10.5.77)$$

for $t \in [0, \tau]$. Consequently, summarizing the above results we obtain the following statement.

Proposition 10.5.5 *The hedgable part of the contingent claim U can be replicated by the portfolio V_{δ_U} with fractions*

$$\boldsymbol{\pi}_{\delta_U}(t) = (\mathbf{e}_U(t)^\top \mathbf{b}_t^{-1})^\top \quad (10.5.78)$$

for $t \in [0, T]$.

The resulting hedging strategy can be interpreted as a local risk minimizing strategy in the sense of Föllmer & Schweizer (1991). The corresponding benchmarked version of the Föllmer-Schweizer decomposition has the form (10.5.70). Note that the driving martingales in the unhedgable part $\hat{R}_{t,\tau}$, see (10.5.72), are orthogonal to the martingales that drive the primary security accounts and, thus, the hedgable part $\hat{I}_{t,\tau}$, see (10.5.71). Obviously, to perform a local risk minimizing hedge the corresponding initial capital at time t equals the fair price $\tilde{u}^k(t, \tilde{\mathbf{Z}}_t^k)$, see (10.5.59), of the contingent claim.

10.6 Exercises

10.1. Name some type of finite-dimensional filters that can be applied in continuous time finance when observed and hidden quantities follow a continuous Gaussian process.

10.2. Name a class of finite-dimensional filters, where the observations are continuous and the hidden quantities jump.

10.3. Why, in practice, discrete-time strong numerical schemes are relevant in filtering.

10.4. Name a strong numerical scheme that is fully implicit and can potentially overcome numerical instabilities for a range of step sizes when approximating the Zakai equation for the Wonham filter.

10.5. For an independent constant hidden Gaussian growth rate η with mean $\mu > 0$ and variance $\sigma^2 > 0$ determine for an observed strictly, positive asset price process $S = \{S_t, t \in [0, T]\}$ with SDE

$$dS_t = S_t \left(\eta + \frac{1}{2} \right) dt + S_t dW_t^*$$

for $t \in [0, T]$ with $S_0 > 0$, the least-squares estimate

$$\hat{\eta}_t = E(\eta | \mathcal{Y}_t)$$

under the information given by the sigma-algebra $\mathcal{Y}_t = \sigma\{S_z : z \in [0, t]\}$, which is generated by the security S . Here $W^* = \{W_t^*, t \in [0, T]\}$ is an independent standard Wiener process.

10.6. What is the purpose of an Expectation Maximization algorithm.

Monte Carlo Simulation of SDEs

This chapter introduces what is commonly known as Monte Carlo simulation for stochastic differential equations. We explain that Monte Carlo simulation is a much simpler task than scenario simulation, discussed in the previous chapters. A weak convergence criterion will be introduced that allows us to classify various discrete-time approximations and numerical schemes for the purpose of Monte Carlo simulation. For simplicity, we focus on the case without jumps in this introductory chapter. The case with jumps is more complicated and will be described in Chaps. 12 and 13.

11.1 Introduction to Monte Carlo Simulation

There exists well developed literature on classical Monte Carlo methods, which, however, does not focus on stochastic differential equations. We mention, among others, [Hammersley & Handscomb \(1964\)](#) and [Fishman \(1996\)](#). Some of the general properties and principles of classical Monte Carlo methods can be exploited for the simulation of functionals of SDEs. However, by using the structure of the SDEs in weak discrete-time approximations, one can obtain a deeper insight into the simulation problem and, thus, more efficient methods than under the classical Monte Carlo approach. Monographs on Monte Carlo methods for SDEs include, for instance, [Kloeden & Platen \(1999\)](#), [Kloeden, Platen & Schurz \(2003\)](#), [Milstein \(1995a\)](#), [Jäckel \(2002\)](#) and [Glasserman \(2004\)](#). The specific structure of SDEs allows us to develop highly sophisticated simulation methods. In multi-factor models, when other methods fail or are difficult to implement, Monte Carlo methods can still provide a satisfactory result. In particular high-dimensional problems in finance are accessible via Monte Carlo simulation when finite difference and many other methods fail.

Weak Convergence Criterion

As for scenario simulation we introduce a criterion that allows us to classify different schemes. In Monte Carlo simulation it is the expectation of a certain payoff function that one wants to approximate. To introduce a suitable class of potential payoff functions let us denote by $\tilde{\mathcal{C}}_P(\mathbb{R}^d, \mathbb{R})$ the set of all polynomials $g : \mathbb{R}^d \rightarrow \mathbb{R}$. Consider the process $\mathbf{X} = \{\mathbf{X}_t, t \in [0, T]\}$, which is the exact solution of a given SDE, for instance,

$$d\mathbf{X}_t = \mathbf{a}(t, \mathbf{X}_t) dt + \sum_{j=1}^m \mathbf{b}^j(t, \mathbf{X}_t) dW_t^j \quad (11.1.1)$$

for $t \in [0, T]$ with $\mathbf{X}_0 \in \mathbb{R}^d$. We say that a discrete-time approximation \mathbf{Y}^Δ converges with weak order $\beta > 0$ to \mathbf{X} at time T as $\Delta \rightarrow 0$ if for each $g \in \tilde{\mathcal{C}}_P(\mathbb{R}^d, \mathbb{R})$ there exists a constant C_g , which does not depend on Δ and $\Delta_0 \in [0, 1]$ such that the following weak error $\mu(\Delta)$ satisfies the estimate

$$\mu(\Delta) = |E(g(\mathbf{X}_T)) - E(g(\mathbf{Y}_T^\Delta))| \leq C_g \Delta^\beta \quad (11.1.2)$$

for each $\Delta \in (0, \Delta_0)$. We call relation (11.1.2) the *weak convergence criterion*.

Systematic and Statistical Error

Under the weak convergence criterion (11.1.2) functionals of the form

$$u = E(g(\mathbf{X}_T))$$

are approximated via *weak approximations* \mathbf{Y}^Δ of the solution of the SDE (11.1.1). One can form a standard or *raw Monte Carlo estimate* using the sample average

$$u_{N,\Delta} = \frac{1}{N} \sum_{k=1}^N g(\mathbf{Y}_T^\Delta(\omega_k)), \quad (11.1.3)$$

with N independent simulated realizations $\mathbf{Y}_T^\Delta(\omega_1), \mathbf{Y}_T^\Delta(\omega_2), \dots, \mathbf{Y}_T^\Delta(\omega_N)$, where $\omega_k \in \Omega$ for $k \in \{1, 2, \dots, N\}$, of a discrete-time weak approximation \mathbf{Y}_T^Δ at time T . The *weak error* $\hat{\mu}_{N,\Delta}$ then has the form

$$\hat{\mu}_{N,\Delta} = u_{N,\Delta} - E(g(\mathbf{X}_T)), \quad (11.1.4)$$

which we decompose into a *systematic error* μ_{sys} and a *statistical error* μ_{stat} , such that

$$\hat{\mu}_{N,\Delta} = \mu_{\text{sys}} + \mu_{\text{stat}}. \quad (11.1.5)$$

Here we set

$$\begin{aligned}
\mu_{\text{sys}} &= E(\hat{\mu}_{N,\Delta}) \\
&= E\left(\frac{1}{N} \sum_{k=1}^N g(\mathbf{Y}_T^\Delta(\omega_k))\right) - E(g(\mathbf{X}_T)) \\
&= E(g(\mathbf{Y}_T^\Delta)) - E(g(\mathbf{X}_T)). \tag{11.1.6}
\end{aligned}$$

Thus, one has from (11.1.2) that

$$\mu(\Delta) = |\mu_{\text{sys}}|. \tag{11.1.7}$$

Obviously, the absolute systematic error $|\mu_{\text{sys}}|$ represents the weak error and is a critical variable under the weak convergence criterion (11.1.2).

For a large number N , of independent simulated realizations of \mathbf{Y}^Δ , we can conclude from the Central Limit Theorem that the statistical error μ_{stat} , becomes asymptotically Gaussian with mean zero and variance of the form

$$\text{Var}(\mu_{\text{stat}}) = \text{Var}(\hat{\mu}_{N,\Delta}) = \frac{1}{N} \text{Var}(g(\mathbf{Y}_T^\Delta)). \tag{11.1.8}$$

Note that we used in (11.1.8) the independence of the realizations for each ω_k . The expression (11.1.8) reveals a major disadvantage of the raw Monte Carlo method, being that the variance of the statistical error μ_{stat} decreases only with $\frac{1}{N}$. Consequently, the *deviation*

$$\text{Dev}(\mu_{\text{stat}}) = \sqrt{\text{Var}(\mu_{\text{stat}})} = \frac{1}{\sqrt{N}} \sqrt{\text{Var}(g(\mathbf{Y}_T^\Delta))} \tag{11.1.9}$$

of the statistical error decreases slowly at rate $\frac{1}{\sqrt{N}}$ as $N \rightarrow \infty$. This means that unless the random variable $g(\mathbf{Y}_T^\Delta)$ already has a small variance, one may need an extremely large number N , of sample paths to achieve a reasonably small confidence interval.

Confidence Intervals

The above deviation in (11.1.9) is proportional to the length of the interval with fixed confidence level $\alpha \in (0, 1)$. Thus, the length of a corresponding confidence interval for the systematic error, for instance, is only halved by a fourfold increase in the number N , of simulated realizations.

On the other hand, the Monte Carlo approach is very general and works under almost all circumstances. For certain high-dimensional functionals it is sometimes the only method for obtaining a reasonable numerical result. However, one pays a price for this generality since large sample sizes are, in general, required to achieve accurate estimates.

In Kloeden & Platen (1999) it was described how to form and observe confidence intervals during Monte Carlo simulation. Since we usually do not

know the variance of the raw Monte Carlo estimates, we can form batches of the independent outcomes of a simulation. Then the sample mean X_i , $i \in \{1, 2, \dots, n\}$, of each batch can be interpreted by the Central Limit Theorem as being approximately Gaussian with unknown variance. This then allows us to form Student t confidence intervals around the sample mean

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

involving the sample variance

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \hat{\mu}_n)^2.$$

The random variable $T_n = \frac{\hat{\mu}_n - \mu}{\sqrt{\frac{\hat{\sigma}_n^2}{n}}}$ is then, for $n > 3$, Student t distributed with $n-1$ degrees of freedom and $\mu = E(X_i)$. The $100(1-\alpha)\%$ confidence interval $(\hat{\mu}_n - a, \hat{\mu}_n + a)$ emerges from the relationship $P(|\hat{\mu}_n - \mu| < a) = P(|T_n|) < t_{1-\alpha, n-1} = 1 - \alpha$, where $a = t_{1-\alpha, n-1} \sqrt{\frac{\hat{\sigma}_n^2}{n}}$ and $t_{1-\alpha, n-1}$ is the $100(1-\alpha)\%$ quantile of the Student t distribution with $n-1$ degrees of freedom. Without going into further details, we will use this methodology for various simulation experiments to check for corresponding confidence intervals. However, in most experiments that we perform it is possible to reach, with the total number N , of simulated outcomes, a magnitude such that the length of the confidence intervals becomes practically negligible in the graphs we plot.

We note from (11.1.9) that the length of a confidence interval is not only proportional to $\frac{1}{\sqrt{N}}$ but is also proportional to the square root of the variance of the simulated functional. This indicates the possibility to overcome the typically large variance of raw Monte Carlo estimates. These are the estimates obtained without applying any change in the original simulation problem. One can often reformulate the random variable that one wants to construct into one that has the same mean but a much smaller variance. We describe later in Chap. 16 how to construct unbiased estimates $\tilde{u}_{N,\Delta}$ for the expectation $u = E(g(\mathbf{Y}_T^\Delta))$ with much smaller variances than the raw Monte Carlo estimate $u_{N,\Delta}$ given in (11.1.3). These methods are called *variance reduction techniques* and are described in detail later.

Tradeoff Between Step Size and Sample Size

When using exact simulations in a Monte Carlo method, as described in Chap. 2, then one has to be concerned only with the statistical error since there is no systematic error. In general, we have to deal with both the statistical and the systematic error. Duffie & Glynn (1995) show how to design a computationally efficient tradeoff between reducing the time step size Δ and

increasing the number N , of Monte Carlo simulations of the discrete-time approximate sample path \mathbf{Y}^Δ of \mathbf{X} .

Important here is that one takes the order β of weak convergence into account. We will see that an Euler scheme or a simplified Euler scheme usually has weak order $\beta = 1$. Higher order weak schemes are described later in this chapter. When approximating the solution of a d -dimensional SDE over the time period $[0, T]$ via Monte Carlo simulation, the number of calculations is roughly proportional to $d^2 TN/\Delta$. Typically, the dimension d and the time horizon T are given. Therefore, we are concerned with the relationship between the sample size N , and the time step size Δ . In [Duffie & Glynn \(1995\)](#) it was pointed out that the following rule makes sense under suitable conditions: With a scheme of weak order β it is sensible to increase N at the order $\Delta^{-2\beta}$. For instance, with the simplified Euler scheme ($\beta = 1$), the number of simulations N , should quadruple with each halving of the time step size. For a second weak order scheme ($\beta = 2$) the number of simulations N , should increase by a factor 16 for each halving of the time step size. This balances the impact of systematic and statistical error.

11.2 Weak Taylor Schemes

In this section we discuss discrete-time approximations of solutions of SDEs that are appropriate for the Monte Carlo simulation of functionals of diffusion processes. This means, we study primarily the weak order of convergence of numerical schemes. By truncating appropriately the Wagner-Platen expansion from Chap. 4 one obtains, so called, *weak Taylor schemes*. As with strong approximations, the desired order of weak convergence determines the kind of truncation that must be used. However, the truncations will be different from those required for the strong convergence of a comparable order. In general, weak Taylor schemes involve fewer terms than strong Taylor schemes, as we will see below. In the following sections we continue to use the abbreviations and notations introduced in the previous chapters for scenario simulation.

Euler and Simplified Weak Euler Scheme

We have already studied in Chap. 5 the Euler scheme in varying degrees of detail. Recall from (5.3.14) that for the general multi-dimensional case with $d, m \in \{1, 2, \dots\}$ the k th component of the Euler scheme for the SDE (11.1.1) has the form

$$Y_{n+1}^k = Y_n^k + a^k \Delta + \sum_{j=1}^m b^{k,j} \Delta W_n^j, \quad (11.2.1)$$

with $\Delta W_n^j = W_{t_{n+1}}^j - W_{t_n}^j$ and initial value $\mathbf{Y}_0 = \mathbf{X}_0$. Here we again suppress the dependence of the coefficients on \mathbf{Y}_n and τ_n .

The Euler scheme (11.2.1) corresponds to the truncated Wagner-Platen expansion (4.1.5), which contains only the ordinary time integral and the single Itô integrals with the Wiener processes as integrands. We shall see from a convergence result for weak Taylor schemes, as stated below, that the Euler scheme has order of weak convergence $\beta = 1.0$ if the drift and diffusion coefficient satisfy appropriate conditions. This means that the Euler scheme (11.2.1) is the weak order $\beta = 1.0$ Taylor scheme, which is consistent with the order of convergence that an Euler scheme achieves for a deterministic ODE.

For weak convergence we only need to approximate the probability measure induced by the process \mathbf{X} . Therefore, we can replace the Gaussian increments ΔW_n^j in (11.2.1) by other random variables $\Delta \hat{W}_n^j$ with similar moment properties to ΔW_n^j . Therefore, we obtain a simpler scheme by simply choosing more easily generated random variables. This leads to the *simplified Euler scheme*

$$Y_{n+1}^k = Y_n^k + a^k \Delta + \sum_{j=1}^m b^{k,j} \Delta \hat{W}_n^j, \quad (11.2.2)$$

where the $\Delta \hat{W}_n^j$ must be independent $\mathcal{A}_{\tau_{n+1}}$ -measurable random variables with moments satisfying the condition

$$\left| E\left(\Delta \hat{W}_n^j\right) \right| + \left| E\left(\left(\Delta \hat{W}_n^j\right)^3\right) \right| + \left| E\left((\Delta \hat{W}_n^j)^2\right) - \Delta \right| \leq K \Delta^2 \quad (11.2.3)$$

for some constant K and $j \in \{1, 2, \dots, m\}$. The simplest example of such a simplified random variable $\Delta \hat{W}_n^j$ to be used in (11.2.2) is a two-point distributed random variable with

$$P\left(\Delta \hat{W}_n^j = \pm \sqrt{\Delta}\right) = \frac{1}{2}. \quad (11.2.4)$$

Obviously, the above two-point distributed random variable satisfies the moment condition (11.2.3) required by the simplified Euler scheme (11.2.2). In particular, in the case when the drift and diffusion coefficients are constant, a random walk can be interpreted as being generated by a simplified Euler scheme. This applies also to tree methods, as will be shown in Chap. 17.

When the drift and diffusion coefficients are only Hölder continuous, then one can show, see [Mikulevicius & Platen \(1991\)](#), that the Euler scheme still converges weakly, but with some lower weak order $\beta < 1.0$.

Let us check numerically that the Euler and the simplified Euler schemes achieve an order of weak convergence $\beta = 1.0$. We suppose that $X_t \in \mathfrak{X}$ satisfies the following SDE

$$dX_t = a X_t dt + b X_t dW_t \quad (11.2.5)$$

for $t \in [0, T]$ with $X_0 > 0$. We use the default parameters: $X_0 = 1$, $a = 1.5$, $b = 0.01$ and $T = 1$ with test function $g(X) = x$ and $N = 16,000,000$ simulations. Furthermore, the time step size Δ ranges from 1 to $\frac{1}{25}$. The extremely large sample size is chosen to obtain confidence intervals of negligible length.

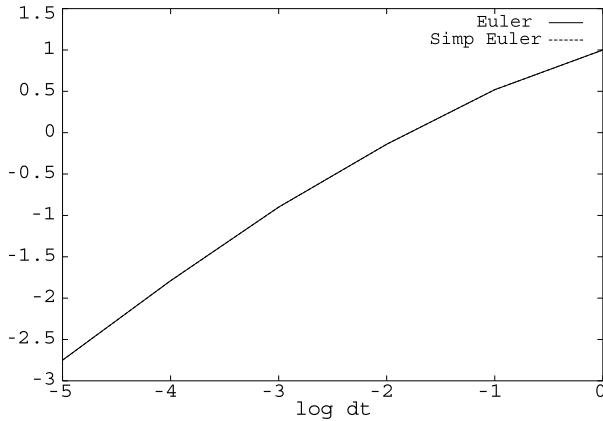


Fig. 11.2.1. Log-log plot of the weak error for an SDE with multiplicative noise for the Euler and simplified Euler schemes

In the log-log plot of Fig. 11.2.1 we show $\ln(\mu(\Delta))$ against $\ln(\Delta)$ for the Euler and simplified Euler scheme. The slope of this graph yields an estimate for the weak order of the scheme. We deserve that the weak errors are almost identical for both schemes, and how the error

$$\mu(\Delta) = |E(X_T) - E(Y_N)|$$

behaves if one decreases the step size Δ . We emphasize that there is no significant difference in the weak errors produced by these two methods.

Weak Order 2.0 Taylor Scheme

We can derive more accurate weak Taylor schemes by including further multiple stochastic integrals from the Wagner-Platen expansion (4.4.7). Under the weak convergence criterion the objective is now to include for a given weak order of convergence the terms with information about the probability measure of the underlying diffusion process into the expansion. As we will see, this is different to the choice of the terms that are necessary to approximate sample paths with some given strong order.

Let us consider the weak Taylor scheme of weak order $\beta = 2.0$. This scheme is obtained by adding all of the double stochastic integrals from the Wagner-Platen expansion (4.4.7) to the terms of the Euler scheme. This scheme was first proposed in Milstein (1978). In the autonomous, that is, the time independent, one-dimensional case $d = m = 1$ we have the *weak order 2.0 Taylor scheme*

$$\begin{aligned} Y_{n+1} &= Y_n + a \Delta + b \Delta W_n + \frac{1}{2} b b' \{ (\Delta W_n)^2 - \Delta \} + a' b \Delta Z_n \\ &\quad + \frac{1}{2} \left(a a' + \frac{1}{2} a'' b^2 \right) \Delta^2 + \left(a b' + \frac{1}{2} b'' b^2 \right) \{ \Delta W_n \Delta - \Delta Z_n \}. \end{aligned} \quad (11.2.6)$$

As for strong approximations, the random variable ΔZ_n represents the double Itô integral $I_{(1,0)}$, see Sect. 5.3. One can use relation (5.3.37) to generate the pair of correlated Gaussian random variables ΔW_n and ΔZ_n .

Under the weak convergence criterion one has considerably more freedom than under the strong convergence criterion for constructing random variables in a discrete-time weak approximation. For instance, it is possible to avoid the random variable ΔZ_n in the preceding scheme (11.2.6). One can show that it is sufficient to use a single random variable $\Delta \hat{W}_n$ with analogous moment properties to ΔW_n , and ΔZ_n replaced by $\frac{1}{2} \Delta \hat{W}_n \Delta$. To simplify our notation further, let us suppress from here on, the dependence of random variables on the number n of the time step.

In the above one-dimensional case we obtain then the *simplified weak order 2.0 Taylor scheme*

$$\begin{aligned} Y_{n+1} = Y_n + a \Delta + b \Delta \hat{W} + \frac{1}{2} b b' \left\{ (\Delta \hat{W})^2 - \Delta \right\} \\ + \frac{1}{2} \left(a' b + a b' + \frac{1}{2} b'' b^2 \right) \Delta \hat{W} \Delta + \frac{1}{2} \left(a a' + \frac{1}{2} a'' b^2 \right) \Delta^2, \end{aligned} \quad (11.2.7)$$

where $\Delta \hat{W}$ must satisfy the moment condition

$$\begin{aligned} \left| E(\Delta \hat{W}) \right| + \left| E((\Delta \hat{W})^3) \right| + \left| E((\Delta \hat{W})^5) \right| \\ + \left| E((\Delta \hat{W})^2) - \Delta \right| + \left| E((\Delta \hat{W})^4) - 3\Delta^2 \right| \leq K \Delta^3 \end{aligned} \quad (11.2.8)$$

for some constant K .

An $N(0, \Delta)$ Gaussian random variable satisfies the moment condition (11.2.8), as does a three-point distributed random variable $\Delta \hat{W}$ with

$$P(\Delta \hat{W} = \pm \sqrt{3\Delta}) = \frac{1}{6} \quad \text{and} \quad P(\Delta \hat{W} = 0) = \frac{2}{3}. \quad (11.2.9)$$

In the multi-dimensional case $d \in \{1, 2, \dots\}$ with $m = 1$, the k th component of the *weak order 2.0 Taylor scheme with scalar noise* is given by

$$\begin{aligned} Y_{n+1}^k = Y_n^k + a^k \Delta + b^k \Delta W + \frac{1}{2} L^1 b^k \{(\Delta W)^2 - \Delta\} \\ + \frac{1}{2} L^0 a^k \Delta^2 + L^0 b^k \{\Delta W \Delta - \Delta Z\} + L^1 a^k \Delta Z, \end{aligned} \quad (11.2.10)$$

where the operators L^0 and L^1 are defined in (5.3.1) and (5.3.3).

In the general multi-dimensional case $d, m \in \{1, 2, \dots\}$ the k th component of the *weak order 2.0 Taylor scheme* takes the form

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + a^k \Delta + \frac{1}{2} L^0 a^k \Delta^2 + \sum_{j=1}^m \left\{ b^{k,j} \Delta W^j + L^0 b^{k,j} I_{(0,j)} + L^j a^k I_{(j,0)} \right\} \\
&\quad + \sum_{j_1, j_2=1}^m L^{j_1} b^{k, j_2} I_{(j_1, j_2)}. \tag{11.2.11}
\end{aligned}$$

Under the weak convergence criterion one can substitute the multiple Itô integrals $I_{(0,j)}$ and $I_{(j,0)}$ with simpler random variables. This is a crucial advantage in Monte Carlo simulation compared to scenario simulation. In this way one obtains from (11.2.11) the following *simplified weak order 2.0 Taylor scheme* with k th component

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + a^k \Delta + \frac{1}{2} L^0 a^k \Delta^2 + \sum_{j=1}^m \left\{ b^{k,j} + \frac{1}{2} \Delta (L^0 b^{k,j} + L^j a^k) \right\} \Delta \hat{W}^j \\
&\quad + \frac{1}{2} \sum_{j_1, j_2=1}^m L^{j_1} b^{k, j_2} (\Delta \hat{W}^{j_1} \Delta \hat{W}^{j_2} + V_{j_1, j_2}). \tag{11.2.12}
\end{aligned}$$

Here the $\Delta \hat{W}^j$ are assumed to be independent random variables satisfying the moment condition (11.2.8) for $j \in \{1, 2, \dots, m\}$. The quantities V_{j_1, j_2} are independent two-point distributed random variables with

$$P(V_{j_1, j_2} = \pm \Delta) = \frac{1}{2} \tag{11.2.13}$$

for $j_2 \in \{1, \dots, j_1 - 1\}$,

$$V_{j_1, j_1} = -\Delta \tag{11.2.14}$$

and

$$V_{j_1, j_2} = -V_{j_2, j_1} \tag{11.2.15}$$

for $j_1 \in \{1, 2, \dots, m\}$ and $j_2 \in \{j_1 + 1, \dots, m\}$. Furthermore, one can choose the random variables $\Delta \hat{W}^j$ in (11.2.12) to be independent three point distributed as described in (11.2.9).

A convergence theorem at the end of this section formulates sufficient conditions under which the above schemes converge with weak order $\beta = 2.0$.

Weak Order 3.0 Taylor Scheme

Let us now consider weak Taylor schemes of weak order $\beta = 3.0$. As shown in Platen (1984) these need to include from the Wagner-Platen expansion (4.4.7) all of the multiple Itô integrals of multiplicity three, to ensure that the scheme converges with this weak order.

In the general multi-dimensional case $d, m \in \{1, 2, \dots\}$ the k th component of the *weak order 3.0 Taylor scheme*, proposed in Platen (1984), takes the form

$$\begin{aligned}
Y_{n+1}^k = & Y_n^k + a^k \Delta + \sum_{j=1}^m b^{k,j} \Delta W^j + \sum_{j=0}^m L^j a^k I_{(j,0)} + \sum_{j_1=0}^m \sum_{j_2=1}^m L^{j_1} b^{k,j_2} I_{(j_1,j_2)} \\
& + \sum_{j_1,j_2=0}^m L^{j_1} L^{j_2} a^k I_{(j_1,j_2,0)} + \sum_{j_1,j_2=0}^m \sum_{j_3=1}^m L^{j_1} L^{j_2} b^{k,j_3} I_{(j_1,j_2,j_3)}. \quad (11.2.16)
\end{aligned}$$

This scheme is of particular interest for various special cases of SDEs because it leads to the construction of efficient third order schemes.

In the scalar case $d = 1$ with scalar noise $m = 1$ the following *simplified weak order 3.0 Taylor scheme* can be obtained

$$\begin{aligned}
Y_{n+1} = & Y_n + a \Delta + b \Delta \tilde{W} + \frac{1}{2} L^1 b \left\{ (\Delta \tilde{W})^2 - \Delta \right\} \\
& + L^1 a \Delta \tilde{Z} + \frac{1}{2} L^0 a \Delta^2 + L^0 b \left\{ \Delta \tilde{W} \Delta - \Delta \tilde{Z} \right\} \\
& + \frac{1}{6} (L^0 L^0 b + L^0 L^1 a + L^1 L^0 a) \Delta \tilde{W} \Delta^2 \\
& + \frac{1}{6} (L^1 L^1 a + L^1 L^0 b + L^0 L^1 b) \left\{ (\Delta \tilde{W})^2 - \Delta \right\} \Delta \\
& + \frac{1}{6} L^0 L^0 a \Delta^3 + \frac{1}{6} L^1 L^1 b \left\{ (\Delta \tilde{W})^2 - 3\Delta \right\} \Delta \tilde{W}. \quad (11.2.17)
\end{aligned}$$

Here $\Delta \tilde{W}$ and $\Delta \tilde{Z}$ are correlated Gaussian random variables with

$$\Delta \tilde{W} \sim N(0, \Delta), \quad \Delta \tilde{Z} \sim N\left(0, \frac{1}{3} \Delta^3\right) \quad (11.2.18)$$

and covariance

$$E(\Delta \tilde{W} \Delta \tilde{Z}) = \frac{1}{2} \Delta^2. \quad (11.2.19)$$

For $d \in \{1, 2, \dots\}$ and $m = 1$ we can substitute the multiple stochastic integrals in the scheme (11.2.16) as follows:

$$\begin{aligned}
I_{(1)} &= \Delta \tilde{W}^1, \quad I_{(1,0)} \approx \Delta \tilde{Z}, \quad I_{(0,1)} \approx \Delta \Delta \tilde{W} - \Delta \tilde{Z} \\
I_{(1,1)} &\approx \frac{1}{2} \left((\Delta \tilde{W})^2 - \Delta \right), \quad I_{(0,0,1)} \approx I_{(0,1,0)} \approx I_{(1,0,0)} \approx \frac{1}{6} \Delta^2 \Delta \tilde{W}, \\
I_{(1,1,0)} &\approx I_{(1,0,1)} \approx I_{(0,1,1)} \approx \frac{1}{6} \Delta \left((\Delta \tilde{W})^2 - \Delta \right), \\
I_{(1,1,1)} &\approx \frac{1}{6} \Delta \tilde{W} \left((\Delta \tilde{W})^2 - 3\Delta \right),
\end{aligned}$$

where the $\Delta \tilde{W}$ and $\Delta \tilde{Z}$ are correlated Gaussian random variables as in (11.2.18) and (11.2.19). This then leads to a simplified weak order 3.0 scheme.

In Hofmann (1994) it has been shown for an SDE with additive noise and $m = \{1, 2, \dots\}$ that one can use approximate random variables of the form

$$I_{(0)} = \Delta, \quad I_{(j)} = \xi_j \sqrt{\Delta}, \quad I_{(0,0)} = \frac{\Delta^2}{2}, \quad I_{(0,0,0)} = \frac{\Delta^3}{6},$$

$$I_{(j,0)} \approx \frac{1}{2} \left(\xi_j + \varphi_j \frac{1}{\sqrt{3 \Delta}} \right) \Delta^{\frac{3}{2}}, \quad I_{(j,0,0)} \approx I_{(0,j,0)} \approx \frac{\Delta^{\frac{5}{2}}}{6} \xi_j$$

and

$$I_{(j_1, j_2, 0)} \approx \frac{\Delta^2}{6} \left(\xi_{j_1} \xi_{j_2} + \frac{V_{j_1, j_2}}{\Delta} \right). \quad (11.2.20)$$

In this case one can use independent four point distributed random variables ξ_1, \dots, ξ_m with

$$P \left(\xi_j = \pm \sqrt{3 + \sqrt{6}} \right) = \frac{1}{12 + 4\sqrt{6}}$$

and

$$P \left(\xi_j = \pm \sqrt{3 - \sqrt{6}} \right) = \frac{1}{12 - 4\sqrt{6}} \quad (11.2.21)$$

together with independent three-point distributed random variables $\varphi_1, \dots, \varphi_m$ as in (11.2.9) and independent two-point distributed random variables V_{j_1, j_2} as in (11.2.13)–(11.2.15). The resulting simplified scheme is then still of weak order $\beta = 3.0$.

For the multi-dimensional case $d, m \in \{1, 2, \dots\}$ with additive noise a *weak order 3.0 Taylor scheme* has for the k th component the form

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + a^k \Delta + \sum_{j=1}^m b^{k,j} \Delta \tilde{W}^j + \frac{1}{2} L^0 a^k \Delta^2 + \frac{1}{6} L^0 L^0 a^k \Delta^3 \\ &\quad + \sum_{j=1}^m \left[L^j a^k \Delta \tilde{Z}^j + L^0 b^{k,j} \left\{ \Delta \tilde{W}^j \Delta - \Delta \tilde{Z}^j \right\} \right. \\ &\quad \left. + \frac{1}{6} (L^0 L^0 b^{k,j} + L^0 L^j a^k + L^j L^0 a^k) \Delta \tilde{W}^j \Delta^2 \right] \\ &\quad + \frac{1}{6} \sum_{j_1, j_2=1}^m L^{j_1} L^{j_2} a^k \left\{ \Delta \tilde{W}^{j_1} \Delta \tilde{W}^{j_2} - I_{\{j_1=j_2\}} \Delta \right\} \Delta. \end{aligned} \quad (11.2.22)$$

Here the random variables $\Delta \tilde{W}^j$ and $\Delta \tilde{Z}^j$ for each $j \in \{1, 2, \dots, m\}$ are independent pairs of correlated Gaussian random variables as in the scheme (11.2.17). The third order of weak convergence for this scheme follows from the theorem that will be formulated at the end of this section.

Weak Order 4.0 Taylor Scheme

To construct the weak order 4.0 Taylor scheme we need to also include all of the fourth order multiple stochastic integrals from the Wagner-Platen expansion in Chap. 4. We shall write a^k as $b^{k,0}$ in order to simplify the resulting formulae. In the general multi-dimensional case $d, m \in \{1, 2, \dots\}$ we can write the k th component of the *weak order 4.0 Taylor scheme* in the form

$$Y_{n+1}^k = Y_n^k + \sum_{\ell=1}^4 \sum_{j_1, \dots, j_\ell=0}^m L^{j_1} \dots L^{j_{\ell-1}} b^{k,j_\ell} I_{(j_1, \dots, j_\ell)}. \quad (11.2.23)$$

This weak order $\beta = 4.0$ Taylor scheme appeared first in Platen (1984).

In the case of special SDEs, for instance, those with additive noise, one obtains highly accurate schemes when using (11.2.23). For the one-dimensional case $d = m = 1$ with additive noise the following *simplified weak order 4.0 Taylor scheme for additive noise* can be derived

$$\begin{aligned} Y_{n+1} = & Y_n + a \Delta + b \Delta \tilde{W} + \frac{1}{2} L^0 a \Delta^2 + L^1 a \Delta \tilde{Z} + L^0 b \left\{ \Delta \tilde{W} \Delta - \Delta \tilde{Z} \right\} \\ & + \frac{1}{3!} \left\{ L^0 L^0 b + L^0 L^1 a \right\} \Delta \tilde{W} \Delta^2 \\ & + L^1 L^1 a \left\{ 2 \Delta \tilde{W} \Delta \tilde{Z} - \frac{5}{6} (\Delta \tilde{W})^2 \Delta - \frac{1}{6} \Delta^2 \right\} \\ & + \frac{1}{3!} L^0 L^0 a \Delta^3 + \frac{1}{4!} L^0 L^0 L^0 a \Delta^4 \\ & + \frac{1}{4!} \left\{ L^1 L^0 L^0 a + L^0 L^1 L^0 a + L^0 L^0 L^1 a + L^0 L^0 L^0 b \right\} \Delta \tilde{W} \Delta^3 \\ & + \frac{1}{4!} \left\{ L^1 L^1 L^0 a + L^0 L^1 L^1 a + L^1 L^0 L^1 a \right\} \left\{ (\Delta \tilde{W})^2 - \Delta \right\} \Delta^2 \\ & + \frac{1}{4!} L^1 L^1 L^1 a \Delta \tilde{W} \left\{ (\Delta \tilde{W})^2 - 3\Delta \right\} \Delta. \end{aligned} \quad (11.2.24)$$

Here $\Delta \tilde{W}$ and $\Delta \tilde{Z}$ are correlated Gaussian random variables with $\Delta \tilde{W} \sim N(0, \Delta)$, $\Delta \tilde{Z} \sim N(0, \frac{1}{3} \Delta^3)$ and $E(\Delta \tilde{W} \Delta \tilde{Z}) = \frac{1}{2} \Delta^2$, see (11.2.18)–(11.2.19).

The convergence with weak order $\beta = 4.0$ of the above scheme will again follow from the theorem at the end of the section.

A Simulation Study

Let us analyze the above four weak Taylor schemes in a simulation study for an SDE with additive noise. In order to estimate the weak error (11.1.4)

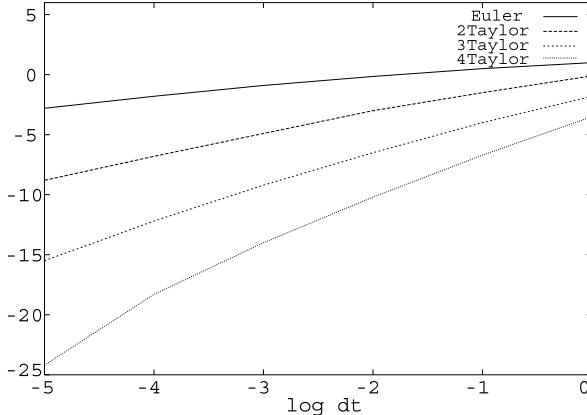


Fig. 11.2.2. Log-log plot of the weak errors for an SDE with additive noise

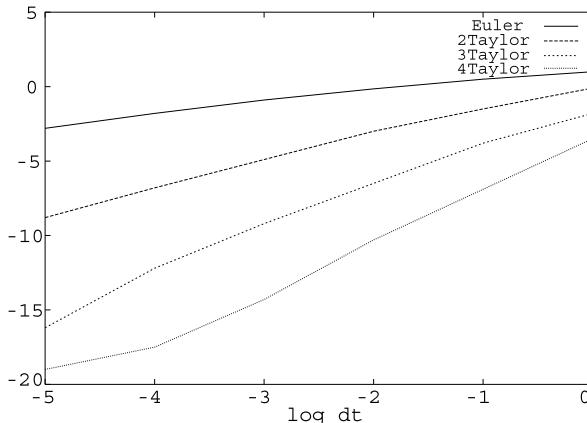


Fig. 11.2.3. Log-log plot of the weak error for an SDE with multiplicative noise

we run a sufficient number of simulations such that the systematic error μ_{sys} dominates the statistical error in (11.1.5) and the confidence intervals in the error plots become negligible.

For the SDE with additive noise

$$dX_t = a X_t dt + b dW_t \quad (11.2.25)$$

for $t \in [0, T]$ we use the default parameters $X_0 = 1$, $a = 1.5$, $b = 0.01$ and $T = 1$ and the test function $g(X) = x$. In Fig. 11.2.2 we show the log-log plot for the resulting weak errors. One notes that the slope of a curve in Fig. 11.2.2 corresponds to the weak order of the corresponding scheme, in particular, for smaller step sizes. At the end of this section we will formulate a convergence theorem which proves that these are the theoretically correct weak orders for these weak Taylor schemes.

We now test the schemes for another SDE. Consider the case with multiplicative noise

$$dX_t = a X_t dt + b X_t dW_t \quad (11.2.26)$$

for $t \in [0, T]$. Here we use the same parameter choices as above and $g(X) = x$. In Fig. 11.2.3 we show the log-log plot for the corresponding results. It can be seen that most schemes perform about as expected from theory.

Convergence Theorem

Let us now formulate a theorem that indicates the type of conditions that one needs to satisfy to obtain a given order of weak convergence.

In what follows we shall use the notations from Chap. 4 and Chap. 5, where we formulate the Wagner-Platen expansion and strong Taylor schemes. Recall that for a given function $f(t, \mathbf{x}) = \mathbf{x}$ we obtain the Itô coefficient functions

$$f_\alpha(t, \mathbf{x}) = L^{j_1} \cdots L^{j_{\ell-1}} b^{j_\ell}(\mathbf{x}) \quad (11.2.27)$$

for all $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d$ and all multi-indices $\alpha = (j_1, \dots, j_\ell) \in \mathcal{M}_m$, $m \in \{1, 2, \dots\}$. Here we set $b^0 = a$ and the differential operators L^j for $j \in \{0, 1, \dots, m\}$ are as given in Chap. 4. For a multi-index $\alpha = (j_1, \dots, j_\ell) \in \mathcal{M}_m$ we use the multiple Itô integral

$$I_{\alpha, t_n, t_{n+1}} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_\ell} \cdots \int_{t_n}^{s_2} dW_{s_1}^{j_1} \cdots dW_{s_{\ell-1}}^{j_{\ell-1}} dW_{s_\ell}^{j_\ell}, \quad (11.2.28)$$

where we set $dW_s^0 = ds$, see (4.2.15).

It has been shown in Platen (1984) that a weak Taylor scheme of order $\beta \in \{1.0, 2.0, 3.0, \dots\}$ is associated with the hierarchical set of multi-indices

$$\Gamma_\beta = \{\alpha \in \mathcal{M}_m : l(\alpha) \leq \beta\}, \quad (11.2.29)$$

Here $l(\alpha)$ denotes the number of components of the multi-index α . Obviously, the hierarchical set Γ_β contains all multi-indices α with a number of components $l(\alpha)$ not greater than β .

Let there be a given time discretization $0 = t_0 < t_1 < \dots < t_n < \dots$. In the general multi-dimensional case $d, m \in \{1, 2, \dots\}$ for $\beta \in \{1, 2, \dots\}$ we define the *weak Taylor scheme of order β* as the vector equation

$$\begin{aligned} \mathbf{Y}_{n+1} &= \mathbf{Y}_n + \sum_{\alpha \in \Gamma_\beta \setminus \{v\}} f_\alpha(t_n, \mathbf{Y}_n) I_{\alpha, t_n, t_{n+1}} \\ &= \sum_{\alpha \in \Gamma_\beta} f_\alpha(t_n, \mathbf{Y}_n) I_{\alpha, t_n, t_{n+1}}. \end{aligned} \quad (11.2.30)$$

Here the initial value $\mathbf{Y}_0 = \mathbf{X}_0$ is assumed to be a deterministic constant. For $\beta \in \{1.0, 2.0, 3.0, 4.0\}$ the scheme given by (11.2.30) coincides with the corresponding weak Taylor scheme presented earlier in the section.

Let $\mathcal{C}_P^\ell(\mathbb{R}^d, \mathbb{R})$ denote the space of ℓ times continuously differentiable functions $g : \mathbb{R}^d \rightarrow \mathbb{R}$ for which g and all of its partial derivatives of order up to and including ℓ have polynomial growth. This set is slightly more general than $C_P(\mathbb{R}^d, \mathbb{R})$ that we introduced in Sect. 11.1. We state the following theorem for an autonomous diffusion process \mathbf{X} with drift $\mathbf{a}(\mathbf{x})$ and diffusion coefficient $\mathbf{b}(\mathbf{x})$, which provides the order of weak convergence of the previously presented schemes.

Theorem 11.2.1. *For some $\beta \in \{1, 2, \dots\}$ and autonomous SDE for \mathbf{X} let \mathbf{Y}^Δ be a weak Taylor scheme of order β . Suppose that \mathbf{a} and \mathbf{b} are Lipschitz continuous with components $a^k, b^{k,j} \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ for all $k \in \{1, 2, \dots, d\}$ and $j \in \{0, 1, \dots, m\}$, and that the f_α satisfy a linear growth bound*

$$|f_\alpha(t, \mathbf{x})| \leq K (1 + |\mathbf{x}|), \quad (11.2.31)$$

for all $\alpha \in \Gamma_\beta$, $\mathbf{x} \in \mathbb{R}^d$ and $t \in [0, T]$, where $K < \infty$. Then for each $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ there exists a constant C_g , which does not depend on Δ , such that

$$\mu(\Delta) = \left| E(g(\mathbf{X}_T)) - E\left(g\left(\mathbf{Y}_T^\Delta\right)\right) \right| \leq C_g \Delta^\beta, \quad (11.2.32)$$

that is \mathbf{Y}^Δ converges with weak order β to \mathbf{X} at time T as $\Delta \rightarrow 0$.

The proof of this theorem can be found in Kloeden & Platen (1999). Note that Theorem 11.2.1 can also be applied to the nonautonomous case if one takes the time t as the first component of the process \mathbf{X} . We will later generalize this type of result to the case with jumps.

We emphasize the fact, which is expressed by the theorem, that the order of weak convergence does *not* depend on the specific choice of the test function g if the function is sufficiently smooth. The convergence of the first, second and higher order moments of the approximations \mathbf{Y}_T^Δ to those of \mathbf{X}_T is automatically covered by choosing appropriate polynomials as test function $g(\cdot)$.

11.3 Derivative-Free Weak Approximations

Higher order weak Taylor schemes require the determination and evaluation of derivatives of various orders of the drift and diffusion coefficients. As with strong schemes, we can construct derivative-free weak approximations, which avoid the use of such derivatives.

Explicit Weak Order 2.0 Scheme

For the autonomous case with scalar noise $m = 1$ and $d \in \{1, 2, \dots\}$ the following *explicit weak order 2.0 scheme* is suggested in Platen (1984)

$$\begin{aligned}\mathbf{Y}_{n+1} = \mathbf{Y}_n + \frac{1}{2} (\mathbf{a}(\bar{\mathbf{r}}) + \mathbf{a}) \Delta + \frac{1}{4} \left(\mathbf{b}(\bar{\mathbf{r}}^+) + \mathbf{b}(\bar{\mathbf{r}}^-) + 2\mathbf{b} \right) \Delta \hat{W} \\ + \frac{1}{4} \left(\mathbf{b}(\bar{\mathbf{r}}^+) - \mathbf{b}(\bar{\mathbf{r}}^-) \right) \left\{ (\Delta \hat{W})^2 - \Delta \right\} \Delta^{\frac{1}{2}}\end{aligned}\quad (11.3.1)$$

with supporting values

$$\bar{\mathbf{r}} = \mathbf{Y}_n + \mathbf{a} \Delta + \mathbf{b} \Delta \hat{W}$$

and

$$\bar{\mathbf{r}}^\pm = \mathbf{Y}_n + \mathbf{a} \Delta \pm \mathbf{b} \sqrt{\Delta}.$$

Here $\Delta \hat{W}$ is required to satisfy the moment condition (11.2.8). This means, $\Delta \hat{W}$, for instance, can be three-point distributed with

$$P(\Delta \hat{W} = \pm \sqrt{3\Delta}) = \frac{1}{6} \quad \text{and} \quad P(\Delta \hat{W} = 0) = \frac{2}{3}. \quad (11.3.2)$$

By comparing (11.3.1) with the corresponding simplified weak Taylor scheme (11.2.7), one notes that (11.3.1) avoids the derivatives that appear in (11.2.7) by using additional supporting values.

There is a generalization of the explicit weak order 2.0 scheme (11.3.1). For the autonomous case with $d, m \in \{1, 2, \dots\}$ the *multi-dimensional explicit weak order 2.0 scheme* has the vector form

$$\begin{aligned}\mathbf{Y}_{n+1} = \mathbf{Y}_n + \frac{1}{2} (\mathbf{a}(\bar{\mathbf{r}}) + \mathbf{a}) \Delta + \frac{1}{4} \sum_{j=1}^m \left[\left(\mathbf{b}^j(\bar{\mathbf{R}}_+^j) + \mathbf{b}^j(\bar{\mathbf{R}}_-^j) + 2\mathbf{b}^j \right) \Delta \hat{W}^j \right. \\ \left. + \sum_{\substack{r=1 \\ r \neq j}}^m \left(\mathbf{b}^j(\bar{U}_+^r) + \mathbf{b}^j(\bar{U}_-^r) - 2\mathbf{b}^j \right) \Delta \hat{W}^j \Delta^{-\frac{1}{2}} \right] \\ + \frac{1}{4} \sum_{j=1}^m \left[\left(\mathbf{b}^j(\bar{\mathbf{R}}_+^j) - \mathbf{b}^j(\bar{\mathbf{R}}_-^j) \right) \left\{ (\Delta \hat{W}^j)^2 - \Delta \right\} \right. \\ \left. + \sum_{\substack{r=1 \\ r \neq j}}^m \left(\mathbf{b}^j(\bar{U}_+^r) - \mathbf{b}^j(\bar{U}_-^r) \right) \left\{ \Delta \hat{W}^j \Delta \hat{W}^r + V_{r,j} \right\} \right] \Delta^{-\frac{1}{2}}\end{aligned}\quad (11.3.3)$$

with supporting values

$$\bar{\mathbf{r}} = \mathbf{Y}_n + \mathbf{a} \Delta + \sum_{j=1}^m \mathbf{b}^j \Delta \hat{W}^j, \quad \bar{\mathbf{R}}_\pm^j = \mathbf{Y}_n + \mathbf{a} \Delta \pm \mathbf{b}^j \sqrt{\Delta}$$

and

$$\bar{U}_\pm^j = \mathbf{Y}_n \pm \mathbf{b}^j \sqrt{\Delta}.$$

Here the random variables $\Delta \hat{W}^j$ and $V_{r,j}$ are defined as in (11.2.9) and (11.2.13)–(11.2.15).

For *additive noise* the second order weak scheme (11.3.3) reduces to the relatively simple algorithm

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \frac{1}{2} \left\{ \mathbf{a} \left(\mathbf{Y}_n + \mathbf{a} \Delta + \sum_{j=1}^m \mathbf{b}^j \Delta \hat{W}^j \right) + \mathbf{a} \right\} \Delta + \sum_{j=1}^m \mathbf{b}^j \Delta \hat{W}^j. \quad (11.3.4)$$

Explicit Weak Order 3.0 Schemes

In the autonomous case $d \in \{1, 2, \dots\}$ with $m = 1$ one obtains for additive noise the *explicit weak order 3.0 scheme* in vector form

$$\begin{aligned} \mathbf{Y}_{n+1} = & \mathbf{Y}_n + \mathbf{a} \Delta + \mathbf{b} \Delta \hat{W} + \frac{1}{2} \left(\mathbf{a}_\zeta^+ + \mathbf{a}_\zeta^- - \frac{3}{2} \mathbf{a} - \frac{1}{4} (\tilde{\mathbf{a}}_\zeta^+ + \tilde{\mathbf{a}}_\zeta^-) \right) \Delta \\ & + \sqrt{\frac{2}{\Delta}} \left(\frac{1}{\sqrt{2}} (\mathbf{a}_\zeta^+ - \mathbf{a}_\zeta^-) - \frac{1}{4} (\tilde{\mathbf{a}}_\zeta^+ - \tilde{\mathbf{a}}_\zeta^-) \right) \zeta \Delta \hat{Z} \\ & + \frac{1}{6} \left[\mathbf{a} \left(\mathbf{Y}_n + (\mathbf{a} + \mathbf{a}_\zeta^+) \Delta + (\zeta + \varrho) \mathbf{b} \sqrt{\Delta} \right) - \mathbf{a}_\zeta^+ - \mathbf{a}_\varrho^+ + \mathbf{a} \right] \\ & \times \left[(\zeta + \varrho) \Delta \hat{W} \sqrt{\Delta} + \Delta + \zeta \varrho \left\{ (\Delta \hat{W})^2 - \Delta \right\} \right] \quad (11.3.5) \end{aligned}$$

with

$$\mathbf{a}_\phi^\pm = \mathbf{a} \left(\mathbf{Y}_n + \mathbf{a} \Delta \pm \mathbf{b} \sqrt{\Delta} \phi \right)$$

and

$$\tilde{\mathbf{a}}_\phi^\pm = \mathbf{a} \left(\mathbf{Y}_n + 2 \mathbf{a} \Delta \pm \mathbf{b} \sqrt{2\Delta} \phi \right),$$

where ϕ is either ζ or ϱ . Here one can use two correlated Gaussian random variables $\Delta \hat{W} \sim N(0, \Delta)$ and $\Delta \hat{Z} \sim N(0, \frac{1}{3} \Delta^3)$ with $E(\Delta \hat{W} \Delta \hat{Z}) = \frac{1}{2} \Delta^2$, together with two independent two-point distributed random variables ζ and ϱ with

$$P(\zeta = \pm 1) = P(\varrho = \pm 1) = \frac{1}{2}.$$

In the autonomous case $d \in \{1, 2, \dots\}$ with general scalar noise $m = 1$ we have the following generalization of the scheme (11.3.5), which here is called the *explicit weak order 3.0 scheme for scalar noise*

$$\begin{aligned}
\mathbf{Y}_{n+1} = & \mathbf{Y}_n + \mathbf{a} \Delta + \mathbf{b} \Delta \hat{W} + \frac{1}{2} \mathbf{H}_a \Delta + \frac{1}{\Delta} \mathbf{H}_b \Delta \hat{Z} \\
& + \sqrt{\frac{2}{\Delta}} \mathbf{G}_a \zeta \Delta \hat{Z} + \frac{1}{\sqrt{2\Delta}} \mathbf{G}_b \zeta \left\{ (\Delta \hat{W})^2 - \Delta \right\} \\
& + \frac{1}{6} \mathbf{F}_a^{++} \left(\Delta + (\zeta + \varrho) \sqrt{\Delta} \Delta \hat{W} + \zeta \varrho \left\{ (\Delta \hat{W})^2 - \Delta \right\} \right) \\
& + \frac{1}{24} (\mathbf{F}_b^{++} + \mathbf{F}_b^{-+} + \mathbf{F}_b^{+-} + \mathbf{F}_b^{--}) \Delta \hat{W} \\
& + \frac{1}{24\sqrt{\Delta}} (\mathbf{F}_b^{++} - \mathbf{F}_b^{-+} + \mathbf{F}_b^{+-} - \mathbf{F}_b^{--}) \left\{ (\Delta \hat{W})^2 - \Delta \right\} \zeta \\
& + \frac{1}{24\Delta} (\mathbf{F}_b^{++} + \mathbf{F}_b^{--} - \mathbf{F}_b^{-+} - \mathbf{F}_b^{+-}) \left\{ (\Delta \hat{W})^2 - 3 \right\} \Delta \hat{W} \zeta \varrho \\
& + \frac{1}{24\sqrt{\Delta}} (\mathbf{F}_b^{++} + \mathbf{F}_b^{-+} - \mathbf{F}_b^{+-} - \mathbf{F}_b^{--}) \left\{ (\Delta \hat{W})^2 - \Delta \right\} \varrho \quad (11.3.6)
\end{aligned}$$

with

$$\mathbf{H}_g = \mathbf{g}^+ + \mathbf{g}^- - \frac{3}{2} \mathbf{g} - \frac{1}{4} (\tilde{\mathbf{g}}^+ + \tilde{\mathbf{g}}^-),$$

$$\mathbf{G}_g = \frac{1}{\sqrt{2}} (\mathbf{g}^+ - \mathbf{g}^-) - \frac{1}{4} (\tilde{\mathbf{g}}^+ - \tilde{\mathbf{g}}^-),$$

$$\begin{aligned}
\mathbf{F}_g^{+\pm} = & \mathbf{g} \left(\mathbf{Y}_n + (\mathbf{a} + \mathbf{a}^+) \Delta + \mathbf{b} \zeta \sqrt{\Delta} \pm \mathbf{b}^+ \varrho \sqrt{\Delta} \right) - \mathbf{g}^+ \\
& - \mathbf{g} \left(\mathbf{Y}_n + \mathbf{a} \Delta \pm \mathbf{b} \varrho \sqrt{\Delta} \right) + \mathbf{g},
\end{aligned}$$

$$\begin{aligned}
\mathbf{F}_g^{-\pm} = & \mathbf{g} \left(\mathbf{Y}_n + (\mathbf{a} + \mathbf{a}^-) \Delta - \mathbf{b} \zeta \sqrt{\Delta} \pm \mathbf{b}^- \varrho \sqrt{\Delta} \right) - \mathbf{g}^- \\
& - \mathbf{g} \left(\mathbf{Y}_n + \mathbf{a} \Delta \pm \mathbf{b} \varrho \sqrt{\Delta} \right) + \mathbf{g}
\end{aligned}$$

where

$$\mathbf{g}^\pm = \mathbf{g} \left(\mathbf{Y}_n + \mathbf{a} \Delta \pm \mathbf{b} \zeta \sqrt{\Delta} \right)$$

and

$$\tilde{\mathbf{g}}^\pm = \mathbf{g} \left(\mathbf{Y}_n + 2 \mathbf{a} \Delta \pm \sqrt{2} \mathbf{b} \zeta \sqrt{\Delta} \right),$$

with \mathbf{g} being equal to either \mathbf{a} or \mathbf{b} . The random variables $\Delta \hat{W}$, $\Delta \hat{Z}$, ζ and ϱ are as specified earlier for the scheme (11.3.5).

11.4 Extrapolation Methods

Extrapolation provides an efficient, yet simple way of obtaining a higher order weak approximation when using only lower order weak schemes.

Weak Order 2.0 Extrapolation

Only equidistant time discretizations of the time interval $[0, T]$ with $t_{n_T} = T$ will be used in what follows. As before, we denote the considered discrete-time approximation with time step size $\Delta > 0$ by Y^Δ , with value $Y_{t_n}^\Delta = Y_n^\Delta$ at the discretization times t_n , and the corresponding approximation with twice this step size by $Y^{2\Delta}$, and so on.

Suppose that we have simulated the functional

$$E(g(Y_T^\Delta))$$

for a weak order 1.0 approximation using, say, the Euler scheme (11.2.1) or the simplified Euler scheme (11.2.2) with step size Δ . Let us then repeat this raw Monte Carlo simulation with double the step size 2Δ , to obtain a raw Monte Carlo estimate of the functional

$$E(g(Y_T^{2\Delta})).$$

We can then combine these two functionals to obtain the *weak order 2.0 extrapolation*

$$V_{g,2}^\Delta(T) = 2E(g(Y_T^\Delta)) - E(g(Y_T^{2\Delta})), \quad (11.4.1)$$

proposed in [Talay & Tubaro \(1990\)](#). It is a stochastic generalization of the well-known *Richardson extrapolation*. This method works because the leading error terms of both parts of the difference on the right hand side of (11.4.1) offset each other. What remains is a higher order error term.

As an example, consider again the geometric Brownian motion process X satisfying the SDE

$$dX_t = a X_t dt + b X_t dW_t$$

with $X_0 = 0.1$, $a = 1.5$ and $b = 0.01$ on the time interval $[0, T]$ where $T = 1$. We use the Euler scheme (11.2.1) to simulate the Richardson extrapolation $V_{g,2}^\Delta(T)$ for $g(x) = x$ and $\Delta = 2^{-3}$. By using a sufficiently large number of trajectories we then determine the absolute systematic error

$$\mu(\Delta) = |V_{g,2}^\Delta(T) - E(g(X_T))|.$$

Then we repeat the calculations for step sizes $\Delta = 2^{-4}, 2^{-5}$ and 2^{-6} and plot the logarithm $\log_2(\mu(\Delta))$ of the error versus $\log_2(\Delta)$, shown in Fig. 11.4.1. This graph demonstrates that the Richardson extrapolation, is in the above example indeed a weak order $\beta = 2.0$ method.

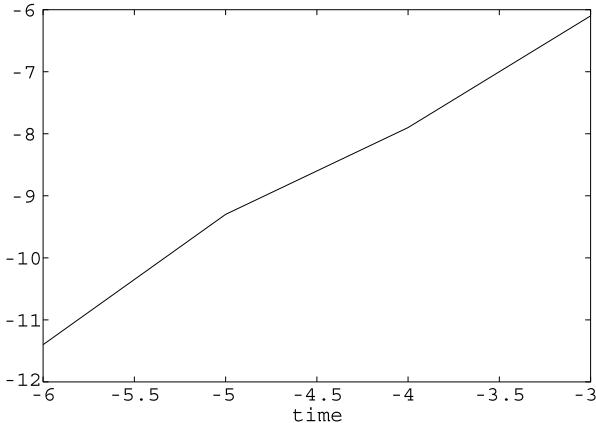


Fig. 11.4.1. Log-log plot for absolute systematic error of Richardson extrapolation

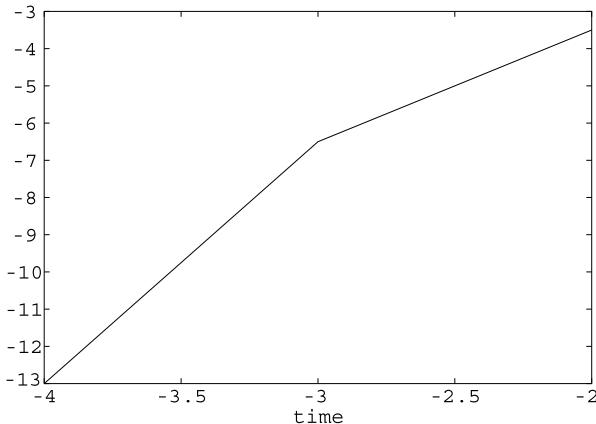


Fig. 11.4.2. Log-log plot for absolute systematic error of weak order 4.0 extrapolation

Higher Weak Order Extrapolation

If a weak method exhibits a certain leading error term representation, then a corresponding extrapolation method can be constructed, see [Kloeden & Platen \(1999\)](#). For instance, one can use a weak order $\beta = 2.0$ approximation Y^Δ and extrapolate it to obtain a fourth order weak approximation of the targeted functional. The *weak order 4.0 extrapolation* has the form

$$V_{g,4}^\Delta(T) = \frac{1}{21} \left[32 E(g(Y_T^\Delta)) - 12 E(g(Y_T^{2\Delta})) + E(g(Y_T^{4\Delta})) \right]. \quad (11.4.2)$$

Suitable weak order 2.0 approximations include the weak order 2.0 Taylor (11.2.6), simplified weak order 2.0 Taylor (11.2.7) or the explicit weak order 2.0 (11.3.1) schemes. In Fig. 11.4.2 we show the log-log plot for the absolute

weak error of the weak order 4.0 extrapolation applied to the above test example. Note the steep slope of the error.

Similarly, one can, in principle, extrapolate a weak order $\beta = 3.0$ approximation Y^Δ , such as the weak order 3.0 Taylor scheme (11.2.16), the simplified weak order 3.0 Taylor scheme (11.2.17) or the weak order 3.0 schemes (11.3.5) and (11.3.6). This then allows us to obtain a sixth weak order approximation. The *weak order 6.0 extrapolation* is defined by the method

$$\begin{aligned} V_{g,6}^\Delta(T) = \frac{1}{2905} & \left[4032 E(g(Y_T^\Delta)) - 1512 E(g(Y_T^{2\Delta})) \right. \\ & \left. + 448 E(g(Y_T^{3\Delta})) - 63 E(g(Y_T^{4\Delta})) \right]. \end{aligned} \quad (11.4.3)$$

The practical use of extrapolations of such a high order, depends strongly on the numerical stability of the underlying weak schemes. These weak methods need to yield numerically reliable results with the appropriate leading error coefficients for a wide range of step sizes. We emphasize that the idea of extrapolation is generally applicable for many weak approximation methods, for instance, for trees and also in the case of jumps.

11.5 Implicit and Predictor-Corrector Methods

This section will introduce weak implicit and predictor-corrector methods as rather efficient schemes for Monte Carlo simulation. For instance, predictor-corrector methods have proven to be useful in LIBOR market model simulation, see Joshi (2003). Also for Monte Carlo simulation the numerical stability of a scheme has a high priority, as will be demonstrated in Chap. 14. Higher weak order of convergence is a secondary objective.

Drift-Implicit Euler Scheme

The simplest implicit weak scheme is the *drift-implicit simplified Euler scheme*, which in the general multi-dimensional case for $d, m \in \{1, 2, \dots\}$ has the form

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \mathbf{a}(\tau_{n+1}, \mathbf{Y}_{n+1}) \Delta + \sum_{j=1}^m \mathbf{b}^j(\tau_n, \mathbf{Y}_n) \Delta \hat{W}^j, \quad (11.5.1)$$

where the random variables $\Delta \hat{W}^j$ for $j \in \{1, 2, \dots, m\}$ and $n \in \{1, 2, \dots\}$ are independent two-point distributed with

$$P(\Delta \hat{W}^j = \pm \sqrt{\Delta}) = \frac{1}{2}. \quad (11.5.2)$$

One can also form a *family of drift-implicit simplified Euler schemes*

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \{(1 - \theta) \mathbf{a}(\tau_n, \mathbf{Y}_n) + \theta \mathbf{a}(\tau_{n+1}, \mathbf{Y}_{n+1})\} \Delta + \sum_{j=1}^m \mathbf{b}^j(\tau_n, \mathbf{Y}_n) \Delta \hat{W}^j \quad (11.5.3)$$

with $\Delta \hat{W}^j$, $j \in \{1, 2, \dots, m\}$, as in (11.5.1). The parameter θ is the *degree of drift implicitness*. With $\theta = 0$ the scheme (11.5.3) reduces to the simplified Euler scheme (11.2.2), whereas with $\theta = 0.5$ it represents a stochastic generalization of the trapezoidal method. Under sufficient regularity one can show that the schemes (11.5.3) converge with weak order $\beta = 1.0$.

Fully Implicit Euler Scheme

For strong discrete-time approximations we noticed in Sect. 7.3 that making the diffusion terms implicit is not easily achieved because some Gaussian random variables usually appear in the denominator of the resulting scheme. However, the possible use of discrete random variables in weak simplified schemes allows us to construct fully implicit weak schemes, that is, algorithms where the approximation of the diffusion term also becomes implicit. This was not the case for strong schemes, where we finally suggested the use of a balanced implicit method (7.3.9)–(7.3.10) to achieve better numerical stability.

In general, implicit schemes require an algebraic equation to be solved at each time step. This imposes an additional computational burden. For certain SDEs the resulting implicit schemes can be directly solved, as is the case for the Black-Scholes dynamics.

In the one-dimensional autonomous case $d = m = 1$ the *fully implicit weak Euler scheme* has the form

$$Y_{n+1} = Y_n + \bar{a}(Y_{n+1}) \Delta + b(Y_{n+1}) \Delta \hat{W}, \quad (11.5.4)$$

where $\Delta \hat{W}$ is as in (11.5.2) and \bar{a} is some adjusted drift coefficient defined by

$$\bar{a} = a - b b'. \quad (11.5.5)$$

The drift adjustment is necessary, otherwise the approximation would not converge towards the solution of the given Itô SDE. In the general multi-dimensional case $d, m \in \{1, 2, \dots\}$ we have, in vector notation, a *family of implicit weak Euler schemes*

$$\begin{aligned} \mathbf{Y}_{n+1} = \mathbf{Y}_n &+ \{\theta \bar{\mathbf{a}}_\eta(\tau_{n+1}, \mathbf{Y}_{n+1}) + (1 - \theta) \bar{\mathbf{a}}_\eta(\tau_n, \mathbf{Y}_n)\} \Delta \\ &+ \sum_{j=1}^m \{\eta \mathbf{b}^j(\tau_{n+1}, \mathbf{Y}_{n+1}) + (1 - \eta) \mathbf{b}^j(\tau_n, \mathbf{Y}_n)\} \Delta \hat{W}^j, \end{aligned} \quad (11.5.6)$$

where the random variables $\Delta \hat{W}^j$ are as in (11.5.1) and the corrected drift coefficient $\bar{\mathbf{a}}_\eta$ is defined by

$$\bar{\mathbf{a}}_\eta = \mathbf{a} - \eta \sum_{j_1, j_2=1}^m \sum_{k=1}^d b^{k, j_1} \frac{\partial \mathbf{b}^{j_2}}{\partial x^k} \quad (11.5.7)$$

for $\theta, \eta \in [0, 1]$.

The choice $\theta = \eta = 1$ in (11.5.6) provides the fully implicit weak Euler scheme. For $\eta = 0.5$ the adjusted drift $\bar{a}_\eta = \underline{a}$ is the drift of the corresponding Stratonovich equation and for $\theta = 0.5$ the scheme (11.5.6) yields a stochastic generalization of the trapezoidal method. In [Kloeden & Platen \(1999\)](#) it has been shown that the above weak Euler schemes exhibit, under appropriate conditions, the weak order $\beta = 1.0$.

Implicit Weak Order 2.0 Taylor Scheme

In the autonomous one-dimensional case $d = m = 1$ the *implicit weak order 2.0 Taylor scheme* has the form

$$\begin{aligned} Y_{n+1} = Y_n + a(Y_{n+1}) \Delta + b \Delta \hat{W} \\ - \frac{1}{2} \left\{ a(Y_{n+1}) a'(Y_{n+1}) + \frac{1}{2} b^2(Y_{n+1}) a''(Y_{n+1}) \right\} \Delta^2 \\ + \frac{1}{2} b b' \left\{ (\Delta \hat{W})^2 - \Delta \right\} + \frac{1}{2} \left\{ -a' b + a b' + \frac{1}{2} b'' b^2 \right\} \Delta \hat{W} \Delta. \end{aligned} \quad (11.5.8)$$

Here $\Delta \hat{W}$ is $N(0, \Delta)$ -Gaussian or three-point distributed with

$$P(\Delta \hat{W} = \pm \sqrt{3\Delta}) = \frac{1}{6} \quad \text{and} \quad P(\Delta \hat{W} = 0) = \frac{2}{3}. \quad (11.5.9)$$

For the general multi-dimensional case with $d, m \in \{1, 2, \dots\}$, in [Milstein \(1995a\)](#) the following *family of implicit weak order 2.0 Taylor schemes* is proposed

$$\begin{aligned} \mathbf{Y}_{n+1} = \mathbf{Y}_n + \{\theta \mathbf{a}(\tau_{n+1}, \mathbf{Y}_{n+1}) + (1 - \theta) \mathbf{a}\} \Delta \\ + \frac{1}{2} \sum_{j_1, j_2=1}^m L^{j_1} \mathbf{b}^{j_2} \left(\Delta \hat{W}^{j_1} \Delta \hat{W}^{j_2} + V_{j_1, j_2} \right) \\ + \sum_{j=1}^m \left\{ \mathbf{b}^j + \frac{1}{2} (L^0 \mathbf{b}^j + (1 - 2\theta) L^j \mathbf{a}) \Delta \right\} \Delta \hat{W}^j \\ + \frac{1}{2} (1 - 2\theta) \{\beta L^0 \mathbf{a} + (1 - \beta) L^0 \mathbf{a}(\tau_{n+1}, \mathbf{Y}_{n+1})\} \Delta^2, \end{aligned} \quad (11.5.10)$$

where $\theta, \beta \in [0, 1]$ and the random variables $\Delta \hat{W}^j$ and V_{j_1, j_2} can be chosen as in (11.2.9) and (11.2.13)–(11.2.15).

For $\theta = 0.5$ the scheme (11.5.10) simplifies to

$$\begin{aligned} \mathbf{Y}_{n+1} = \mathbf{Y}_n + \frac{1}{2} \{ \mathbf{a}(\tau_{n+1}, \mathbf{Y}_{n+1}) + \mathbf{a} \} \Delta + \sum_{j=1}^m \mathbf{b}^j \Delta \hat{W}^j + \frac{1}{2} \sum_{j=1}^m L^0 \mathbf{b}^j \Delta \hat{W}^j \Delta \\ + \frac{1}{2} \sum_{j_1, j_2=1}^m L^{j_1} \mathbf{b}^{j_2} \left(\Delta \hat{W}^{j_1} \Delta \hat{W}^{j_2} + V_{j_1, j_2} \right). \end{aligned} \quad (11.5.11)$$

Note that the last two terms in (11.5.11) vanish for additive noise.

Implicit Weak Order 2.0 Scheme

One can avoid derivatives in the above implicit schemes. In the autonomous case $d \in \{1, 2, \dots\}$ with scalar noise, $m = 1$, in Platen (1995) the following *implicit weak order 2.0 scheme* was derived, where

$$\begin{aligned} \mathbf{Y}_{n+1} = \mathbf{Y}_n + \frac{1}{2} (\mathbf{a} + \mathbf{a}(\mathbf{Y}_{n+1})) \Delta + \frac{1}{4} \left(\mathbf{b}(\bar{\mathbf{r}}^+) + \mathbf{b}(\bar{\mathbf{r}}^-) + 2 \mathbf{b} \right) \Delta \hat{W} \\ + \frac{1}{4} \left(\mathbf{b}(\bar{\mathbf{r}}^+) - \mathbf{b}(\bar{\mathbf{r}}^-) \right) \left\{ \left(\Delta \hat{W} \right)^2 - \Delta \right\} \Delta^{-\frac{1}{2}} \end{aligned} \quad (11.5.12)$$

with supporting values

$$\bar{\mathbf{r}}^\pm = \mathbf{Y}_n + \mathbf{a} \Delta \pm \mathbf{b} \sqrt{\Delta}.$$

Here the random variable $\Delta \hat{W}$ can be chosen as in (11.5.8).

There is a multi-dimensional generalization of the second weak order scheme (11.5.12). For the autonomous case with $d, m \in \{1, 2, \dots\}$ the *implicit weak order 2.0 scheme* has the vector form

$$\begin{aligned} \mathbf{Y}_{n+1} = \mathbf{Y}_n + \frac{1}{2} (\mathbf{a} + \mathbf{a}(\mathbf{Y}_{n+1})) \Delta + \frac{1}{4} \sum_{j=1}^m \left[\mathbf{b}^j (\bar{\mathbf{R}}_+^j) + \mathbf{b}^j (\bar{\mathbf{R}}_-^j) + 2 \mathbf{b}^j \right. \\ \left. + \sum_{\substack{r=1 \\ r \neq j}}^m \left(\mathbf{b}^j (\bar{\mathbf{U}}_+^r) + \mathbf{b}^j (\bar{\mathbf{U}}_-^r) - 2 \mathbf{b}^j \right) \Delta^{-\frac{1}{2}} \right] \Delta \hat{W}^j \\ + \frac{1}{4} \sum_{j=1}^m \left[\left(\mathbf{b}^j (\bar{\mathbf{R}}_+^j) - \mathbf{b}^j (\bar{\mathbf{R}}_-^j) \right) \left\{ \left(\Delta \hat{W}^j \right)^2 - \Delta \right\} \right. \\ \left. + \sum_{\substack{r=1 \\ r \neq j}}^m \left(\mathbf{b}^j (\bar{\mathbf{U}}_+^r) - \mathbf{b}^j (\bar{\mathbf{U}}_-^r) \right) \left\{ \Delta \hat{W}^j \Delta \hat{W}^r + V_{r,j} \right\} \right] \Delta^{-\frac{1}{2}} \end{aligned} \quad (11.5.13)$$

with supporting values

$$\bar{\mathbf{R}}_\pm^j = \mathbf{Y}_n + \mathbf{a} \Delta \pm \mathbf{b}^j \sqrt{\Delta}$$

and

$$\bar{U}_\pm^j = \mathbf{Y}_n \pm \mathbf{b}^j \sqrt{\Delta}.$$

Here the random variables $\Delta \hat{W}^j$ and $V_{r,j}$ are as in (11.5.10). In Kloeden & Platen (1999) it is shown that the above second order weak schemes converge under appropriate conditions with weak order $\beta = 2.0$.

Weak Order 1.0 Predictor-Corrector Methods

Predictor-corrector methods are similar to implicit methods but do not require the solution of an algebraic equation at each step. They are used mainly because of their good numerical stability properties which they inherit from the implicit counterparts of their corrector. In addition, the difference that can be observed between the predicted and the corrected values at each time step provides an indicator for the local approximation error. Most of the following predictor-corrector methods were defined in Platen (1995).

In the autonomous one-dimensional case with scalar noise, $d = m = 1$, we have the *modified trapezoidal method of weak order* $\beta = 1.0$ with corrector

$$Y_{n+1} = Y_n + \frac{1}{2} \{a(\bar{Y}_{n+1}) + a\} \Delta + b \Delta \hat{W} \quad (11.5.14)$$

and predictor

$$\bar{Y}_{n+1} = Y_n + a \Delta + b \Delta \hat{W}. \quad (11.5.15)$$

Here the random variable $\Delta \hat{W}$ can be chosen to be Gaussian $N(0, \Delta)$ distributed or as two-point distributed random variable with

$$P(\Delta \hat{W} = \pm \sqrt{\Delta}) = \frac{1}{2}. \quad (11.5.16)$$

In the general multi-dimensional case for $d, m \in \{1, 2, \dots\}$ one can form the following *family of weak order 1.0 predictor-corrector methods* with corrector

$$\begin{aligned} \mathbf{Y}_{n+1} = \mathbf{Y}_n &+ \{\theta \bar{\mathbf{a}}_\eta(\tau_{n+1}, \bar{\mathbf{Y}}_{n+1}) + (1 - \theta) \bar{\mathbf{a}}_\eta(\tau_n, \mathbf{Y}_n)\} \Delta \\ &+ \sum_{j=1}^m \{\eta \mathbf{b}^j(\tau_{n+1}, \bar{\mathbf{Y}}_{n+1}) + (1 - \eta) \mathbf{b}^j(\tau_n, \mathbf{Y}_n)\} \Delta \hat{W}^j \end{aligned} \quad (11.5.17)$$

for $\theta, \eta \in [0, 1]$, where

$$\bar{\mathbf{a}}_\eta = \mathbf{a} - \eta \sum_{j_1, j_2=1}^m \sum_{k=1}^d b^{k, j_1} \frac{\partial \mathbf{b}^{j_2}}{\partial x^k}, \quad (11.5.18)$$

and with predictor

$$\bar{\mathbf{Y}}_{n+1} = \mathbf{Y}_n + \mathbf{a} \Delta + \sum_{j=1}^m \mathbf{b}^j \Delta \hat{W}^j, \quad (11.5.19)$$

where the $\Delta\hat{W}^j$ are as in (11.5.16). Note that the corrector (11.5.17) with $\eta > 0$ allows us to include some implicitness in the diffusion terms. This scheme provides an efficient and numerically reliable method. By performing Monte Carlo simulation with different parameter choices for θ and η , then comparing these, one obtains useful information about the numerical stability of the scheme for a given application.

Weak Order 2.0 Predictor-Corrector Methods

In the autonomous one-dimensional case with scalar noise, for $d = m = 1$, a *weak order 2.0 predictor-corrector method* is obtained by choosing as a corrector

$$Y_{n+1} = Y_n + \frac{1}{2} \{a(\bar{Y}_{n+1}) + a\} \Delta + \Psi_n \quad (11.5.20)$$

with

$$\Psi_n = b \Delta\hat{W} + \frac{1}{2} b b' \left\{ (\Delta\hat{W})^2 - \Delta \right\} + \frac{1}{2} \left(a b' + \frac{1}{2} b^2 b'' \right) \Delta\hat{W} \Delta$$

and as predictor

$$\bar{Y}_{n+1} = Y_n + a \Delta + \Psi_n + \frac{1}{2} a' b \Delta\hat{W} \Delta + \frac{1}{2} \left(a a' + \frac{1}{2} a'' b^2 \right) \Delta^2. \quad (11.5.21)$$

Here the random variable $\Delta\hat{W}$ can be $N(0, \Delta)$ Gaussian or three-point distributed with

$$P(\Delta\hat{W} = \pm\sqrt{3\Delta}) = \frac{1}{6} \quad \text{and} \quad P(\Delta\hat{W} = 0) = \frac{2}{3}. \quad (11.5.22)$$

For the general multi-dimensional case with $d, m \in \{1, 2, \dots\}$ this scheme can be generalized to a method with corrector

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \frac{1}{2} \{\mathbf{a}(\tau_{n+1}, \bar{\mathbf{Y}}_{n+1}) + \mathbf{a}\} \Delta + \boldsymbol{\Psi}_n, \quad (11.5.23)$$

where

$$\boldsymbol{\Psi}_n = \sum_{j=1}^m \left\{ \mathbf{b}^j + \frac{1}{2} L^0 \mathbf{b}^j \Delta \right\} \Delta\hat{W}^j + \frac{1}{2} \sum_{j_1, j_2=1}^m L^{j_1} \mathbf{b}^{j_2} \left(\Delta\hat{W}^{j_1} \Delta\hat{W}^{j_2} + V_{j_1, j_2} \right),$$

and predictor

$$\bar{\mathbf{Y}}_{n+1} = \mathbf{Y}_n + \mathbf{a} \Delta + \boldsymbol{\Psi}_n + \frac{1}{2} L^0 \mathbf{a} \Delta^2 + \frac{1}{2} \sum_{j=1}^m L^j \mathbf{a} \Delta\hat{W}^j \Delta. \quad (11.5.24)$$

Here the independent random variables $\Delta\hat{W}^j$ and V_{j_1, j_2} can be chosen as in (11.2.12).

In the autonomous case $d \in \{1, 2, \dots\}$ with scalar noise $m = 1$ a *derivative-free weak order 2.0 predictor-corrector method* has corrector

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \frac{1}{2} \{ \mathbf{a}(\bar{\mathbf{Y}}_{n+1}) + \mathbf{a} \} \Delta + \phi_n, \quad (11.5.25)$$

where

$$\phi_n = \frac{1}{4} \left(\mathbf{b}(\bar{\mathbf{Y}}^+) + \mathbf{b}(\bar{\mathbf{Y}}^-) + 2\mathbf{b} \right) \Delta \hat{W} + \frac{1}{4} \left(\mathbf{b}(\bar{\mathbf{Y}}^+) - \mathbf{b}(\bar{\mathbf{Y}}^-) \right) \left\{ (\Delta \hat{W})^2 - \Delta \right\} \Delta^{-\frac{1}{2}}$$

with supporting values

$$\bar{\mathbf{Y}}^\pm = \mathbf{Y}_n + \mathbf{a} \Delta \pm \mathbf{b} \sqrt{\Delta},$$

and predictor

$$\bar{\mathbf{Y}}_{n+1} = \mathbf{Y}_n + \frac{1}{2} \{ \mathbf{a}(\bar{\mathbf{Y}}) + \mathbf{a} \} \Delta + \phi_n \quad (11.5.26)$$

with the supporting value

$$\bar{\mathbf{Y}} = \mathbf{Y}_n + \mathbf{a} \Delta + \mathbf{b} \Delta \hat{W}.$$

Here the random variable $\Delta \hat{W}$ can be chosen as in (11.5.21).

There is also a generalization of the above scheme. For the autonomous case with $d, m \in \{1, 2, \dots\}$ the method has corrector

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \frac{1}{2} \{ \mathbf{a}(\bar{\mathbf{Y}}_{n+1}) + \mathbf{a} \} \Delta + \phi_n, \quad (11.5.27)$$

where

$$\begin{aligned} \phi_n = & \frac{1}{4} \sum_{j=1}^m \left[\mathbf{b}^j(\bar{\mathbf{R}}_+^j) + \mathbf{b}^j(\bar{\mathbf{R}}_-^j) + 2\mathbf{b}^j + \sum_{\substack{r=1 \\ r \neq j}}^m (\mathbf{b}^j(\bar{\mathbf{U}}_+^r) + \mathbf{b}^j(\bar{\mathbf{U}}_-^r) - 2\mathbf{b}^j) \Delta^{-\frac{1}{2}} \right] \Delta \hat{W}^j \\ & + \frac{1}{4} \sum_{j=1}^m \left[(\mathbf{b}^j(\bar{\mathbf{R}}_+^j) - \mathbf{b}^j(\bar{\mathbf{R}}_-^j)) \left\{ (\Delta \hat{W})^2 - \Delta \right\} \right. \\ & \left. + \sum_{\substack{r=1 \\ r \neq j}}^m (\mathbf{b}^j(\bar{\mathbf{U}}_+^r) - \mathbf{b}^j(\bar{\mathbf{U}}_-^r)) (\Delta \hat{W}^j \Delta \hat{W}^r + V_{r,j}) \right] \Delta^{-\frac{1}{2}}. \end{aligned}$$

Here one uses the supporting values

$$\bar{\mathbf{R}}_\pm^j = \mathbf{Y}_n + \mathbf{a} \Delta \pm \mathbf{b}^j \sqrt{\Delta} \quad \text{and} \quad \bar{\mathbf{U}}_\pm^j = \mathbf{Y}_n \pm \mathbf{b}^j \sqrt{\Delta},$$

and the predictor

$$\bar{\mathbf{Y}}_{n+1} = \mathbf{Y}_n + \frac{1}{2} \{ \mathbf{a}(\bar{\mathbf{Y}}) + \mathbf{a} \} \Delta + \phi_n \quad (11.5.28)$$

that has the supporting value

$$\tilde{\mathbf{Y}} = \mathbf{Y}_n + \mathbf{a} \Delta + \sum_{j=1}^m \mathbf{b}^j \Delta \hat{W}^j.$$

The independent random variables $\Delta \hat{W}^j$ and $V_{r,j}$ can be chosen as in (11.2.12).

In the above schemes one first computes at each step the predicted approximate value $\bar{\mathbf{Y}}_{n+1}$ and afterwards the corrected value \mathbf{Y}_{n+1} . The difference

$$\mathbf{Z}_{n+1} = \bar{\mathbf{Y}}_{n+1} - \mathbf{Y}_{n+1}$$

between these values provides us with information about the local approximation error at each time step. This information can be used on-line to introduce a time step size control during the simulation, if considered to be efficient. For instance, if the absolute mean of \mathbf{Z}_{n+1} is rather large, then one can change to a finer time step size.

For weak convergence results and further references on the above mentioned predictor-corrector methods we refer to [Kloeden & Platen \(1999\)](#) and [Platen \(1995\)](#).

11.6 Exercises

11.1. Verify that the two-point distributed random variable $\Delta \hat{W}$ with

$$P(\Delta \hat{W} = \pm \sqrt{\Delta}) = \frac{1}{2}$$

satisfies the moment conditions

$$\left| E(\Delta \hat{W}) \right| + \left| E((\Delta \hat{W})^3) \right| + \left| E((\Delta \hat{W})^2) - \Delta \right| \leq K \Delta^2.$$

11.2. Prove that the three-point distributed random variable $\Delta \tilde{W}$ with

$$P(\Delta \tilde{W} = \pm \sqrt{3\Delta}) = \frac{1}{6} \quad \text{and} \quad P(\Delta \tilde{W} = 0) = \frac{2}{3}$$

satisfies the moment conditions

$$\begin{aligned} \left| E(\Delta \tilde{W}) \right| + \left| E((\Delta \tilde{W})^3) \right| + \left| E((\Delta \tilde{W})^5) \right| + \\ + \left| E((\Delta \tilde{W})^2) - \Delta \right| + \left| E((\Delta \tilde{W})^4) - 3\Delta^2 \right| \leq K \Delta^3. \end{aligned}$$

11.3. For the Black-Scholes model with SDE

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

for $t \in [0, T]$ with $X_0 = 0$ and W a standard Wiener process, write down a simplified Euler scheme. Which weak order of convergence does this scheme achieve?

11.4. For the BS model in Exercise 11.3 describe the simplified weak order 2.0 Taylor method.

11.5. For the BS model in Exercise 11.3 construct a drift implicit Euler scheme.

11.6. For the BS model in Exercise 11.3 describe a fully implicit Euler scheme.

Regular Weak Taylor Approximations

In this chapter we present for the case with jumps regular weak approximations obtained directly from a truncated Wagner-Platen expansion. The desired weak order of convergence determines which terms of the stochastic expansion have to be included in the approximation. These weak Taylor schemes are different from the regular strong Taylor schemes presented in Chap. 6. We will see that the construction of weak schemes requires a separate analysis. A convergence theorem, which allows us to construct weak Taylor approximations of any given weak order, will be presented at the end of this chapter.

12.1 Weak Taylor Schemes

As in Chap. 6, we first consider the one-dimensional, $d = m = 1$, jump diffusion SDE

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t + \int_{\mathcal{E}} c(t, X_{t-}, v) p_\varphi(dv, dt), \quad (12.1.1)$$

for $t \in [0, T]$, with $X_0 \in \Re$, where $W = \{W_t, t \in [0, T]\}$ is a one-dimensional Wiener process and $p_\varphi(dv, dt)$ is a Poisson measure. Later, we also consider the autonomous d -dimensional jump diffusion SDE

$$d\mathbf{X}_t = \mathbf{a}(\mathbf{X}_t)dt + \mathbf{b}(\mathbf{X}_t)d\mathbf{W}_t + \int_{\mathcal{E}} \mathbf{c}(\mathbf{X}_{t-}, v) p_\varphi(dv, dt), \quad (12.1.2)$$

for $t \in [0, T]$, with $\mathbf{X}_0 \in \Re^d$. In the following, we will consider both the case of a scalar Wiener process, $m = 1$, and that of an m -dimensional Wiener process, $m \in \{1, 2, \dots\}$. Note that in the case of an autonomous multi-dimensional SDE, we can always cover the case of an SDE with time-dependent coefficients by considering the time t as first component of the vector process \mathbf{X} . Moreover, we remark that we will sometimes treat separately, the simpler case of a mark-independent jump coefficient $c(t, x, v) = c(t, x)$.

For the following discrete-time approximations, we consider a regular time discretization $(t)_\Delta$ with maximum time step size Δ , as defined in Chap. 6, which does not include the jump times of the Poisson measure. Recall that we use the step size $\Delta_n = t_{n+1} - t_n$ at the n th time step.

Euler Scheme

Due to the nature of the Wagner-Platen expansion the simplest, useful weak Taylor approximation is the Euler scheme (6.1.24), presented in Chap. 6 as a strong scheme. Nonetheless, we will prove at the end of this chapter that the Euler scheme attains the order of weak convergence $\beta = 1.0$, as opposed to the strong order $\gamma = 0.5$, derived in Chap. 6.

We recall that in the general multi-dimensional case the k th component of the Euler scheme is given by the formula

$$Y_{n+1}^k = Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c^k(\xi_i),$$

for $k \in \{1, 2, \dots, d\}$. Here we have used the abbreviations defined in (6.1.12)–(6.1.14), with ξ_i denoting an independent random variable that is distributed according to the measure $\frac{\varphi(dv)}{\varphi(\mathcal{E})}$. This scheme is identical to the Euler scheme presented in Chap. 6.

Order 2.0 Taylor Scheme

We have seen that the Euler scheme is the simplest Taylor scheme, from the point of view of both strong and weak approximations. Note that one can simplify the Euler scheme by using discrete random variables, as discussed in Sect. 11.2. When a higher accuracy is required, and thus, a scheme with higher order of convergence is sought, then it is important to distinguish between strong and weak schemes. Indeed, by adding to the Euler scheme the four multiple stochastic integrals appearing in the strong order 1.0 Taylor scheme presented in Sect. 6.2, we can improve the order of strong convergence from $\gamma = 0.5$ to $\gamma = 1.0$. However, it can be shown that the strong order 1.0 Taylor scheme has generally only the same order of weak convergence $\beta = 1.0$ as the Euler scheme. This indicates that the construction of efficient higher order weak schemes requires rules that are different from those applicable for strong schemes, as became evident in Sects. 11.1 and 11.2. At the end of this chapter we present a convergence theorem for weak Taylor approximations. It will provide a rule for selecting from the Wagner-Platen expansion those multiple stochastic integrals needed to achieve a given order of weak convergence $\beta = \{1.0, 2.0, \dots\}$. The rule is rather simple, one needs to include all the terms with multiple stochastic integrals of up to multiplicity β when constructing a weak order β scheme. In this way, we obtain the *weak order 2.0 Taylor scheme*.

by including all double integrals. In the one-dimensional case, $d = m = 1$, this scheme is given by

$$\begin{aligned}
Y_{n+1} = & Y_n + a\Delta_n + b\Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c(v) p_\varphi(dv, dz) \\
& + b b' \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1} dW_{z_2} \\
& + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b c'(v) dW_{z_1} p_\varphi(dv, dz_2) \\
& + \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ b(t_n, Y_n + c(v)) - b \right\} p_\varphi(dv, dz_1) dW_{z_2} \\
& + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ c(t_n, Y_n + c(v_1), v_2) - c(v_2) \right\} p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2) \\
& + \left(\frac{\partial a}{\partial t} + a a' + \frac{a''}{2} b^2 \right) \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dz_1 dz_2 + a' b \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1} dz_2 \\
& + \left(\frac{\partial b}{\partial t} + a b' + \frac{b''}{2} b^2 \right) \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dz_1 dW_{z_2} \\
& + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \left(\frac{\partial c(v)}{\partial t} + a c'(v) + \frac{c''(v)}{2} b^2 \right) dz_1 p_\varphi(dv, dz_2) \\
& + \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ a(t_n, Y_n + c(v)) - a \right\} p_\varphi(dv, dz_1) dz_2. \tag{12.1.3}
\end{aligned}$$

Here we have used the abbreviated notation (6.1.12)–(6.1.14), and have expressed the partial derivative with respect to time as $\frac{\partial}{\partial t}$. This scheme will be shown to achieve weak order $\beta = 2.0$. We emphasize, to achieve second order weak convergence, we have added all double stochastic integrals to the Euler scheme. The weak order 2.0 Taylor scheme was presented in Liu & Li (2000). It generalizes the weak order 2.0 Taylor scheme (11.2.6) for pure diffusion SDEs.

The scheme (12.1.3) is rather complex, as it involves all nine possible double stochastic integrals with respect to time, Wiener process and the Poisson random measure. By Itô's formula, the integration by parts formula and using (12.1.3), we can rewrite the weak order 2.0 Taylor scheme in the autonomous case as

$$\begin{aligned}
Y_{n+1} = & Y_n + a \Delta_n + b \Delta W_n + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c(\xi_i) + \frac{bb'}{2} \left((\Delta W_n)^2 - \Delta_n \right) \\
& + b \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} c'(\xi_i) (W_{\tau_i} - W_{t_n}) \\
& + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} \left\{ b(Y_n + c(\xi_i)) - b \right\} (W_{t_{n+1}} - W_{\tau_i}) \\
& + \sum_{j=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(\tau_j)} \left\{ c(Y_n + c(\xi_i), \xi_j) - c(\xi_j) \right\} \\
& + \frac{1}{2} \left(\frac{\partial a}{\partial t} + a a' + \frac{a''}{2} b^2 \right) (\Delta_n)^2 + a' b \Delta Z_n \\
& + \left(\frac{\partial b}{\partial t} + a b' + \frac{b''}{2} b^2 \right) (\Delta W_n \Delta_n - \Delta Z_n) \\
& + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} \left(\frac{\partial c(\xi_i)}{\partial t} + a c'(\xi_i) + \frac{c''(\xi_i)}{2} b^2 \right) (\tau_i - t_n) \\
& + \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} \left\{ a(Y_n + c(\xi_i)) - a \right\} (t_{n+1} - \tau_i). \tag{12.1.4}
\end{aligned}$$

This scheme is readily applicable for weak approximation and, thus, for Monte Carlo simulation. The correlated Gaussian random variables $\Delta W_n = W_{t_{n+1}} - W_{t_n}$ and

$$\Delta Z_n = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} dW_{s_1} ds_2 \tag{12.1.5}$$

can be generated as in (5.3.37). The order 2.0 Taylor scheme for the SDE (1.8.5) of the Merton model is given by

$$\begin{aligned}
Y_{n+1} = & Y_n + \mu Y_n \Delta_n + \sigma Y_n \Delta W_n + Y_n \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} (\xi_i - 1) \\
& + \frac{\sigma^2}{2} Y_n ((\Delta W_n)^2 - \Delta_n) + \sigma Y_n \Delta W_n \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} (\xi_i - 1)
\end{aligned}$$

$$\begin{aligned}
& + Y_n \sum_{j=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(\tau_j)} (\xi_i - 1)(\xi_j - 1) \\
& + \frac{\mu^2}{2} Y_n (\Delta_n)^2 + \mu \sigma Y_n \Delta W_n \Delta_n + \mu Y_n \Delta_n \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} (\xi_i - 1). \quad (12.1.6)
\end{aligned}$$

Mark-independent 2.0 Weak Taylor Scheme

In the special case of a mark-independent jump coefficient $c(t, x, v) = c(t, x)$, the weak order 2.0 Taylor scheme reduces to

$$\begin{aligned}
Y_{n+1} = & Y_n + a \Delta + b \Delta W_n + c \Delta p_n + b b' I_{(1,1)} + b c' I_{(1,-1)} \\
& + \{b(t_n, Y_n + c) - b\} I_{(-1,1)} + \{c(t_n, Y_n + c) - c\} I_{(-1,-1)} \\
& + \frac{1}{2} \left(\frac{\partial a}{\partial t} + a a' + \frac{a''}{2} b^2 \right) I_{(0,0)} + a' b I_{(1,0)} + \left(\frac{\partial b}{\partial t} + a b' + \frac{b''}{2} b^2 \right) I_{(0,1)} \\
& + \left(\frac{\partial c}{\partial t} + a c' + \frac{c''}{2} b^2 \right) I_{(0,-1)} + \{a(t_n, Y_n + c) - a\} I_{(-1,0)}, \quad (12.1.7)
\end{aligned}$$

where the multiple stochastic integrals $I_{(1,1)}$, $I_{(1,-1)}$, $I_{(-1,1)}$ and $I_{(-1,-1)}$ are defined in (6.2.5) taking the form

$$\begin{aligned}
I_{(0,0)} &= \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} ds_1 ds_2 = \frac{\Delta_n^2}{2} \\
I_{(1,0)} &= \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} dW_{s_1} ds_2 = \Delta Z_n \\
I_{(0,1)} &= \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} ds_1 dW_{s_2} = \Delta_n \Delta W_n - I_{(1,0)} \\
I_{(0,-1)} &= \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} ds_1 p_\varphi(\mathcal{E}, ds_2) = \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} \tau_i - \Delta p_n t_n \\
I_{(-1,0)} &= \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} p_\varphi(\mathcal{E}, ds_1) ds_2 = \Delta_n \Delta p_n - I_{(0,-1)}. \quad (12.1.8)
\end{aligned}$$

Even in the case of mark-independent jump sizes, the computational complexity of the weak order 2.0 Taylor scheme depends on the intensity of the Poisson jump measure. Indeed, as already discussed in Sect. (6.2), for the generation of double stochastic integrals involving the Poisson measure the random variables ΔW_n and Δp_n are not sufficient. In this case one needs also the location of jump times in every time interval $(t_n, t_{n+1}]$, with $n \in \{0, \dots, n_T - 1\}$.

Multi-dimensional Weak Order 2.0 Taylor Scheme

Let us consider the autonomous multi-dimensional SDE (12.1.2) with scalar Wiener process, $m = 1$. The k th component of the *weak order 2.0 Taylor scheme* is here given by

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + a^k \Delta_n + b^k \Delta W_n + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c^k(v) p_\varphi(dv, dz) \\
&\quad + \sum_{l=1}^d \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} b^l \frac{\partial b^k}{\partial x^l} dW_{z_1} dW_{z_2} \\
&\quad + \sum_{l=1}^d \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} b^l \frac{\partial c^k(v)}{\partial x^l} dW_{z_1} p_\varphi(dv, dz_2) \\
&\quad + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \left\{ b^k(t_n, Y_n + c(v)) - b^k \right\} p_\varphi(dv, dz_1) dW_{z_2} \\
&\quad + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ c^k(t_n, Y_n + c(v_1), v_2) - c^k(v_2) \right\} \\
&\quad \quad \times p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2) \\
&\quad + \left(\sum_{l=1}^d a^l \frac{\partial a^k}{\partial x^l} + \sum_{i,l=1}^d \frac{\partial^2 a^k}{\partial x^l \partial x^i} \frac{b^l b^i}{2} \right) \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dz_1 dz_2 \\
&\quad + \sum_{l=1}^d b^l \frac{\partial a^k}{\partial x^l} \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1} dz_2 \\
&\quad + \left(\sum_{l=1}^d a^l \frac{\partial b^k}{\partial x^l} + \sum_{i,l=1}^d \frac{\partial^2 b^k}{\partial x^l \partial x^i} \frac{b^l b^i}{2} \right) \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dz_1 dW_{z_2} \\
&\quad + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \left(\sum_{l=1}^d a^l \frac{\partial c^k(v)}{\partial x^l} + \sum_{i,l=1}^d \frac{\partial^2 c^k(v)}{\partial x^l \partial x^i} \frac{b^l b^i}{2} \right) dz_1 p_\varphi(dv, dz_2) \\
&\quad + \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} \left\{ a^k(t_n, Y_n + c(v)) - a^k \right\} p_\varphi(dv, dz_1) dz_2, \quad (12.1.9)
\end{aligned}$$

for $k \in \{1, 2, \dots, d\}$, where a^k , b^k , and c^k are the k th components of the drift, diffusion and jump coefficients, respectively. Note that the multiple stochastic integrals can be generated similarly as in the one-dimensional case.

In the general multi-dimensional case, the k th component of the *weak order 2.0 Taylor scheme* is given by

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} c^k(v) p_\varphi(dv, dz) \\
&\quad + \sum_{j_1, j_2=1}^m L^{j_1} b^{k, j_2} \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1}^{j_1} dW_{z_2}^{j_2} \\
&\quad + \sum_{j_1=1}^m \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} L^{j_1} c^k(v) dW_{z_1}^{j_1} p_\varphi(dv, dz_2) \\
&\quad + \sum_{j_1=1}^m \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} L_v^{-1} b^{k, j_1} p_\varphi(dv, dz_2) dW_{z_2}^{j_1} \\
&\quad + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} \int_{\mathcal{E}} L_{v_1}^{-1} c^k(v_2) p_\varphi(dv_1, dz_1) p_\varphi(dv_2, dz_2) \\
&\quad + \sum_{j=1}^m L^j a^k \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dW_{z_1}^j dz_2 + \sum_{j=1}^m L^0 b^{k,j} \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dz_1 dW_{z_2}^j \\
&\quad + L^0 a^k \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} dz_1 dz_2 + \int_{t_n}^{t_{n+1}} \int_{\mathcal{E}} \int_{t_n}^{z_2} L^0 c^k(v) dz_1 p_\varphi(dv, dz_2) \\
&\quad + \int_{t_n}^{t_{n+1}} \int_{t_n}^{z_2} \int_{\mathcal{E}} L_v^{-1} a^k p_\varphi(dv, dz_1) dz_2,
\end{aligned} \tag{12.1.10}$$

for $k \in \{1, 2, \dots, d\}$, where the operators L^j , with $j \in \{-1, \dots, m\}$, are defined in (4.3.4)–(4.3.6).

In the multi-dimensional case with mark-independent jump size, the k th component of the *weak order 2.0 Taylor scheme* simplifies to the algorithm

$$\begin{aligned}
Y_{n+1}^k &= Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + c^k \Delta p_n \\
&\quad + \sum_{j_1, j_2=1}^m L^{j_1} b^{k, j_2} I_{(j_1, j_2)} + \sum_{j_1=1}^m L^{j_1} c^k I_{(j_1, -1)} \\
&\quad + \sum_{j_1=1}^m \left\{ b^{k, j_1}(t_n, \mathbf{Y}_n + \mathbf{c}) - b^{k, j_1} \right\} I_{(-1, j_1)} + \left\{ c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k \right\} I_{(-1, -1)} \\
&\quad + \sum_{j=1}^m L^j a^k I_{(j, 0)} + \sum_{j=1}^m L^0 b^{k,j} I_{(0, j)} + L^0 a^k I_{(0, 0)} \\
&\quad + L^0 c^k I_{(0, -1)} + \left\{ a^k(t_n, \mathbf{Y}_n + \mathbf{c}) - a^k \right\} I_{(-1, 0)},
\end{aligned} \tag{12.1.11}$$

for $k \in \{1, 2, \dots, d\}$. All multiple stochastic integrals that do not involve Wiener processes can be generated as in the one-dimensional case (12.1.7). For those involving one Wiener process, we can use the relations

$$\begin{aligned} I_{(j,-1)} &= \sum_{i=p_\varphi(t_n)+1}^{p_\varphi(t_{n+1})} W_{\tau_i}^j - \Delta p_n W_{t_n}^j \quad I_{(-1,j)} = \Delta p_n \Delta W_n^j - I_{(j,-1)} \\ I_{(j,0)} &= \Delta Z_n^j \quad I_{(0,j)} = \Delta W_n^j \Delta_n - \Delta Z_n^j, \end{aligned} \quad (12.1.12)$$

for $j \in \{1, 2, \dots, m\}$ and $n \in \{0, 1, \dots, n_T - 1\}$. Recall that, for every $j \in \{1, 2, \dots, m\}$, the random variable ΔZ_n^j has a Gaussian distribution with mean $E(\Delta Z_n^j) = 0$, variance $E((\Delta Z_n^j)^2) = \frac{1}{3} (\Delta_n)^3$ and covariance $E(\Delta Z_n^j \Delta W_n^j) = \frac{1}{2} (\Delta_n)^2$. Therefore, with $2m$ independent $N(0, 1)$ distributed standard Gaussian random variables $U_{1,j}$ and $U_{2,j}$, for $j \in \{1, 2, \dots, m\}$, we obtain the required random variables by the transformation

$$\Delta W_n^j = U_{1,j} \sqrt{\Delta_n} \quad \text{and} \quad \Delta Z_n^j = \frac{1}{2} (\Delta_n)^{\frac{3}{2}} \left(U_{1,j} + \frac{1}{\sqrt{3}} U_{2,j} \right). \quad (12.1.13)$$

Finally, the generation of the multiple stochastic integrals $I_{(j_1, j_2)}$, with $j_1, j_2 \in \{1, 2, \dots, m\}$, generally requires an approximation such as those proposed in Chap. 6. As described in Sect. 11.2 for the diffusion case, simplified random variables can be used to match certain moments of the multiple stochastic integrals from the Wagner-Platen expansions.

12.2 Commutativity Conditions

As discussed previously, the generation of multiple stochastic integrals required in the weak order 2.0 Taylor scheme is computationally demanding. Let us now discuss some commutativity conditions under which the complexity of the weak order 2.0 Taylor scheme can be reduced. It is often in the hands of the modeler to construct or select a model that satisfies some of the following commutativity conditions. An appropriate choice of model can provide considerable computational advantages.

Mark-independent Jump Commutativity Conditions

The computational complexity of the weak order 2.0 Taylor scheme generally depends on the intensity of the Poisson measure. Indeed, as previously discussed, the generation of the double stochastic integrals $I_{(1,-1)}$, $I_{(-1,1)}$, $I_{(0,-1)}$ and $I_{(-1,0)}$ requires the knowledge of the exact location of the jump times in the time interval $[t_n, t_{n+1}]$. However, the sum of the above first two double integrals is given by

$$I_{(1,-1)} + I_{(-1,1)} = \Delta p_n \Delta W_n, \quad (12.2.1)$$

which yields an expression that is independent of the particular values of the jump times. Additionally, by (12.1.8) the sum of the above last two double integrals is obtained as

$$I_{(0,-1)} + I_{(-1,0)} = \Delta p_n \Delta_n, \quad (12.2.2)$$

which is also independent of the jump times. Therefore, in the one-dimensional case with mark-independent jump coefficient $c(t, x, v) = c(t, x)$, we can formulate the *first jump commutativity condition*

$$b(t, x) \frac{\partial c(t, x)}{\partial x} = b(t, x + c(t, x)) - b(t, x) \quad (12.2.3)$$

and the *second jump commutativity condition*

$$\frac{\partial c(t, x)}{\partial t} + a(t, x) \frac{\partial c(t, x)}{\partial x} + \frac{b^2(t, x)}{2} \frac{\partial^2 c(t, x)}{\partial x^2} = a(t, x + c(t, x)) - a(t, x), \quad (12.2.4)$$

for all $t \in [0, T]$ and $x \in \Re$.

The first jump commutativity condition (12.2.3) was discussed in Chap. 6. We refer to Sect. 6.3 for selected diffusion and corresponding jump coefficients, satisfying the jump commutativity condition (12.2.3). Note that in this case the implementation of the weak order 2.0 Taylor scheme does not require sampling the Wiener process at all jump times. The second jump commutativity condition (12.2.4) expresses the relationship in the form of a partial differential equation (PDE) involving the drift, diffusion and jump coefficients. If this condition is satisfied, then one needs only to sample the Gaussian random variable ΔW_n and the Poisson random variable Δp_n at each time step. In this special case the above algorithm is very convenient also for SDEs driven by a Poisson measure with high intensity.

Multi-dimensional Jump Commutativity Conditions

In the multi-dimensional case with mark-independent jump size, we also need to generate the multiple stochastic integrals $I_{(j,-1)}$ and $I_{(-1,j)}$, for $j \in \{1, 2, \dots, m\}$. As discussed in Sect. 6.3, the sum of two multiple stochastic integrals with respect to the j th component of the Wiener process and the Poisson measure is given by

$$I_{(j,-1)} + I_{(-1,j)} = \Delta p_n \Delta W_n^j. \quad (12.2.5)$$

This sum is independent of the particular jump times. Therefore, we obtain the *first jump commutativity condition*

$$\sum_{l=1}^d b^{l,j}(t, \mathbf{x}) \frac{\partial c^k(t, \mathbf{x})}{\partial x^l} = b^{k,j}(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x})) - b^{k,j}(t, \mathbf{x}) \quad (12.2.6)$$

and the *second jump commutativity condition*

$$L^0 c^k(t, \mathbf{x}) = a^k(t, \mathbf{x} + \mathbf{c}(t, \mathbf{x})) - a^k(t, \mathbf{x}), \quad (12.2.7)$$

for $j \in \{1, 2, \dots, m\}$, $k \in \{1, 2, \dots, d\}$, $t \in [0, T]$ and $\mathbf{x} \in \Re^d$. Here the differential operator L^0 is as defined in (4.3.4). Note that the above commutativity conditions consist of two systems of $d \times m$ equations each. Therefore, even for simple drift and diffusion coefficients, there may not exist any jump coefficients satisfying (12.2.6) or (12.2.7).

To simplify also the generation of the double Wiener integrals $I_{(j_1, j_2)}$ for $j_1, j_2 \in \{1, 2, \dots, m\}$, one should check if the SDE under consideration satisfies the *diffusion commutativity condition*

$$L^{j_1} b^{k, j_2}(t, \mathbf{x}) = L^{j_2} b^{k, j_1}(t, \mathbf{x}) \quad (12.2.8)$$

for $j_1, j_2 \in \{1, 2, \dots, m\}$, $k \in \{1, 2, \dots, d\}$, $t \in [0, T]$ and $\mathbf{x} \in \Re^d$, see also (6.3.14). In this situation, as discussed in Sect. 6.3, the double Wiener integrals can be expressed in terms of the increments $\Delta W_n^{j_1}$ and $\Delta W_n^{j_2}$ of the Wiener processes, that is

$$I_{(j_1, j_2)} + I_{(j_2, j_1)} = \Delta W_n^{j_1} \Delta W_n^{j_2}. \quad (12.2.9)$$

For a multi-dimensional SDE satisfying the commutativity conditions (12.2.6)–(12.2.8) and having mark-independent jump size, the weak order 2.0 Taylor scheme reduces to

$$\begin{aligned} Y_{n+1}^k &= Y_n^k + a^k \Delta_n + \sum_{j=1}^m b^{k,j} \Delta W_n^j + c^k \Delta p_n \\ &\quad + \frac{1}{2} \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{i, j_1} \frac{\partial b^{k, j_2}}{\partial x^i} \{ \Delta W_n^{j_1} \Delta W_n^{j_2} - \Delta_n \} \\ &\quad + \sum_{j_1=1}^m \{ b^{k, j_1}(t_n, \mathbf{Y}_n + \mathbf{c}) - b^{k, j_1} \} (\Delta p_n \Delta W_n^{j_1}) \\ &\quad + \frac{1}{2} \{ c^k(t_n, \mathbf{Y}_n + \mathbf{c}) - c^k \} ((\Delta p_n)^2 - \Delta p_n) + \sum_{j=1}^m \sum_{l=1}^d b^{l,j} \frac{\partial a^k}{\partial x^l} \Delta Z_n^j \\ &\quad + \sum_{j=1}^m \left(\frac{\partial b^{k,j}}{\partial t} + \sum_{l=1}^d a^l \frac{\partial b^{k,j}}{\partial x^l} + \sum_{i,l=1}^d \sum_{j_1=1}^m \frac{\partial^2 b^{k,j}}{\partial x^i \partial x^l} \frac{b^{i, j_1} b^{l, j_1}}{2} \right) \{ \Delta W_n^j \Delta_n - \Delta Z_n^j \} \\ &\quad + \left(\frac{\partial a^k}{\partial t} + \sum_{l=1}^d a^l \frac{\partial a^k}{\partial x^l} + \sum_{i,l=1}^d \sum_{j_1=1}^m \frac{\partial^2 a^k}{\partial x^i \partial x^l} \frac{b^{i, j_1} b^{l, j_1}}{2} \right) \frac{(\Delta_n)^2}{2} \\ &\quad + \{ a^k(t_n, \mathbf{Y}_n + \mathbf{c}) - a^k \} \Delta p_n \Delta_n, \end{aligned} \quad (12.2.10)$$

for $k \in \{1, 2, \dots, d\}$. The second weak order of convergence of the above scheme will be shown below.

12.3 Convergence Results

First, let us prepare some results that will be used to establish the order of weak convergence of regular weak Taylor approximations.

Consider the Itô process driven by a compensated jump measure

$$\mathbf{X}_t^{z,y} = \mathbf{y} + \int_z^t \tilde{\mathbf{a}}(\mathbf{X}_u^{z,y}) du + \int_z^t \mathbf{b}(\mathbf{X}_u^{z,y}) d\mathbf{W}_u + \int_z^t \int_{\mathcal{E}} \mathbf{c}(\mathbf{X}_{u-}^{z,y}, v) \tilde{p}_{\varphi}(dv, du) \quad (12.3.1)$$

for $z \leq t \leq T$, which starts at time $z \in [0, T]$ in $\mathbf{y} \in \mathbb{R}^d$. This process has the same solution as the Itô process $\mathbf{X} = \{\mathbf{X}_t, t \in [0, T]\}$, which solves the SDE (1.8.2).

For a given $\beta \in \{1, 2, \dots\}$ and function $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ define the functional

$$u(z, \mathbf{y}) = E(g(\mathbf{X}_T^{z,y})) \quad (12.3.2)$$

for $(z, \mathbf{y}) \in [0, T] \times \mathbb{R}^d$. Then we have

$$u(0, \mathbf{X}_0) = E(g(\mathbf{X}_T^{0,X_0})) = E(g(\mathbf{X}_T)). \quad (12.3.3)$$

We will need the following lemma, see [Mikulevicius & Platen \(1988\)](#) and [Mikulevicius \(1983\)](#), involving the operator

$$\tilde{L}^0 f(z, \mathbf{y}) = L^0 f(z, \mathbf{y}) + \int_{\mathcal{E}} L_v^{-1} f(z, \mathbf{y}) \phi(dv) \quad (12.3.4)$$

for a sufficiently smooth function f , see (4.3.7).

Lemma 12.3.1 *Let us assume that the drift, diffusion and jump coefficients of the SDE (1.8.2) have components $a^k, b^{k,j}, c^k \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ for all $k \in \{1, 2, \dots, d\}$ and $j \in \{1, 2, \dots, m\}$ with uniformly bounded derivatives. Then the functional u , defined in (12.3.2), is the unique solution of the Kolmogorov backward partial integro differential equation (PIDE)*

$$\tilde{L}^0 u(z, \mathbf{y}) = 0 \quad (12.3.5)$$

for all $(z, \mathbf{y}) \in (0, T) \times \mathbb{R}^d$ with terminal condition

$$u(T, \mathbf{x}) = g(\mathbf{x}) \quad (12.3.6)$$

for $\mathbf{x} \in \mathbb{R}^d$, with \tilde{L}^0 defined in (4.3.7). Moreover, we have

$$u(z, \cdot) \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R}) \quad (12.3.7)$$

for each $z \in [0, T]$.

Proof: Mikulevicius & Platen (1988) showed that $u(z, \cdot) \in \mathcal{C}_P^{2(\beta+1)}(\mathfrak{R}^d, \mathfrak{R})$ for each $z \in [0, T]$. Moreover, note that by the Markov property of \mathbf{X} we obtain that $u(z) = u(z, \mathbf{X}_z^{0, X_0})$ equals $E(g(\mathbf{X}_T) | \mathcal{A}_z)$ for $z \in [0, T]$. Therefore, one can show that $u(z)$ forms a martingale. By application of Itô's formula to $u(t, \mathbf{X}_t^{s, y})$, for $0 \leq s < t \leq T$ we obtain

$$\begin{aligned} u(t, \mathbf{X}_t^{s, y}) &= u(s, \mathbf{y}) + \int_s^t L^0 u(z, \mathbf{X}_z^{s, y}) dz \\ &\quad + \sum_{j=1}^m \int_s^t L^j u(z, \mathbf{X}_z^{s, y}) dW_z^j + \int_s^t \int_{\mathcal{E}} L_v^{-1} u(z, \mathbf{X}_z^{s, y}) p_{\varphi}(dv, dz), \end{aligned}$$

where the operators L^0, L^j with $j \in \{1, 2, \dots, m\}$ and L_v^{-1} are defined in (4.3.4), (4.3.5) and (4.3.6), respectively.

By the martingale property of $u(t) = u(t, \mathbf{X}_t^{0, X_0})$, the function u satisfies the PIDE $L^0 u(z, \mathbf{y}) + \int_{\mathcal{E}} L_v^{-1} u(z, \mathbf{y}) \varphi(dv) = \tilde{L}^0 u(z, \mathbf{y}) = 0$ for all $(z, \mathbf{y}) \in (0, T) \times \mathfrak{R}^d$. \square

Remark 12.3.2. For simplicity, we have assumed that the Itô process (12.3.1) can reach any point in \mathfrak{R}^d . If instead the Itô process (12.3.1) can take values only in a subset of \mathfrak{R}^d , then relations (12.3.5) and (12.3.6) of Lemma 12.3.1 hold only in this subset of \mathfrak{R}^d . This is sufficient since in the convergence theorems to be presented we will need the relations (12.3.5) and (12.3.6) only for values of $(t, \mathbf{x}) \in (0, T) \times \mathfrak{R}^d$ that can be reached by the Itô process (12.3.1).

By an application of Itô's formula, we obtain the following result.

Lemma 12.3.3 For all $n \in \{1, 2, \dots, n_T\}$ and $\mathbf{y} \in \mathfrak{R}^d$, we have

$$E\left(u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, y}) - u(t_{n-1}, \mathbf{y}) \mid \mathcal{A}_{t_{n-1}}\right) = 0. \quad (12.3.8)$$

Proof: By Itô's formula we obtain

$$\begin{aligned} u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, y}) &= u(t_{n-1}, \mathbf{y}) + \int_{t_{n-1}}^{t_n} L^0 u(z, \mathbf{X}_z^{t_{n-1}, y}) dz \\ &\quad + \sum_{j=1}^m \int_{t_{n-1}}^{t_n} L^j u(z, \mathbf{X}_z^{t_{n-1}, y}) dW_z^j \\ &\quad + \int_{t_{n-1}}^{t_n} \int_{\mathcal{E}} L_v^{-1} u(z, \mathbf{X}_z^{t_{n-1}, y}) p_{\varphi}(dv, dz). \end{aligned}$$

Then applying the expected value to both sides of relations (4.3.4)–(4.3.7), and using (12.3.5) together with the properties of the Itô integral, we obtain (12.3.8). \square

By using the notation introduced in Chap. 4, we define for every $\beta \in \{1, 2, \dots\}$ the hierarchical set

$$\Gamma_\beta = \{\alpha \in \mathcal{M} : l(\alpha) \leq \beta\}, \quad (12.3.9)$$

which will give us the rule for the construction of regular weak Taylor approximations of weak order β .

Given a regular time discretization $(t)_\Delta$, with maximum step size $\Delta \in (0, 1)$, we define the *weak order β Taylor scheme* by the vector equation

$$\mathbf{Y}_{n+1}^\Delta = \mathbf{Y}_n^\Delta + \sum_{\alpha \in \Gamma_\beta \setminus \{v\}} I_\alpha \left[f_\alpha(t_n, \mathbf{Y}_n^\Delta) \right]_{t_n, t_{n+1}} = \sum_{\alpha \in \Gamma_\beta} I_\alpha \left[f_\alpha(t_n, \mathbf{Y}_n^\Delta) \right]_{t_n, t_{n+1}}, \quad (12.3.10)$$

for $n \in \{0, 1, \dots, n_T - 1\}$ with $f(t, \mathbf{x}) = \mathbf{x}$, as defined in (4.3.8). Similarly, we define the *weak order β compensated Taylor scheme* by the vector equation

$$\mathbf{Y}_{n+1}^\Delta = \mathbf{Y}_n^\Delta + \sum_{\alpha \in \Gamma_\beta \setminus \{v\}} \tilde{I}_\alpha \left[\tilde{f}_\alpha(t_n, Y_n^\Delta) \right]_{t_n, t_{n+1}} = \sum_{\alpha \in \Gamma_\beta} \tilde{I}_\alpha \left[\tilde{f}_\alpha(t_n, Y_n^\Delta) \right]_{t_n, t_{n+1}}, \quad (12.3.11)$$

for $n \in \{0, 1, \dots, n_T - 1\}$ with $f(t, \mathbf{x}) = \mathbf{x}$ in (4.3.9).

For simplicity, for the next and subsequent theorems we will assume an autonomous multi-dimensional jump diffusion SDE. This formulation is not restrictive, since we can always rewrite an SDE with time-dependent coefficients as being autonomous, by modeling the time t as the first component of the process \mathbf{X} . However, the resulting conditions on the time component can be relaxed in a direct proof for the non-autonomous case.

We now present the following convergence theorem which states that, under suitable conditions, for any given $\beta \in \{1, 2, \dots\}$ the corresponding weak order β compensated Taylor scheme (12.3.11) achieves the weak order of convergence β .

Theorem 12.3.4. *For given $\beta \in \{1, 2, \dots\}$, let $\mathbf{Y}^\Delta = \{\mathbf{Y}_n^\Delta, n \in \{0, 1, \dots, n_T\}\}$ be the weak order β compensated Taylor approximation defined in (12.3.11) corresponding to a regular time discretization with maximum time step size $\Delta \in (0, 1)$.*

We assume that $E(|\mathbf{X}_0|^q) < \infty$, for $q \in \{1, 2, \dots\}$, and for any $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ there exists a positive constant C , independent of Δ , such that

$$|E(g(\mathbf{X}_0)) - E(g(\mathbf{Y}_0^\Delta))| \leq C \Delta^\beta. \quad (12.3.12)$$

Moreover, suppose that the drift, diffusion and jump coefficients are Lipschitz continuous with components $a^k, b^{k,j}, c^k \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ for all $k \in \{1, 2, \dots, d\}$ and $j \in \{1, 2, \dots, m\}$ and that the coefficients \tilde{f}_α , with $f(t, \mathbf{x}) = \mathbf{x}$, satisfy the linear growth condition

$$|\tilde{f}_\alpha(t, \mathbf{x})| \leq K(1 + |\mathbf{x}|), \quad (12.3.13)$$

with $K < \infty$, for all $t \in [0, T]$, $\mathbf{x} \in \mathbb{R}^d$ and $\alpha \in \Gamma_\beta \cup \mathcal{B}(\Gamma_\beta)$.

Then for any function $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ there exists a positive constant C , independent of Δ , such that

$$|E(g(\mathbf{X}_T)) - E(g(\mathbf{Y}_{n_T}^\Delta))| \leq C \Delta^\beta. \quad (12.3.14)$$

Remark 12.3.5. Note that the linear growth condition (12.3.13) on the coefficient functions \tilde{f}_α is satisfied if, for instance, the drift, diffusion and jump coefficients are uniformly bounded.

Remark 12.3.6. By replacing the conditions on the compensated Itô coefficient functions \tilde{f}_α with equivalent conditions on the Itô coefficient functions f_α , one can show for given $\beta \in \{1, 2, \dots\}$ that the weak order β Taylor scheme (12.3.10) also attains a weak order of convergence β .

Theorem 12.3.4 is an extension of the weak convergence theorem for diffusion SDEs presented in Kloeden & Platen (1999). The following proof of Theorem 12.3.4 has similarities to the one given in Liu & Li (2000).

Proof: For ease of notation, so no misunderstanding is possible, we write \mathbf{Y} for \mathbf{Y}^Δ . By (12.3.3) and the terminal condition of the Kolmogorov backward equation (12.3.6) we can write

$$\begin{aligned} H &= \left| E(g(\mathbf{Y}_{n_T})) - E(g(\mathbf{X}_T)) \right| \\ &= \left| E(u(T, \mathbf{Y}_{n_T}) - u(0, \mathbf{X}_0)) \right|. \end{aligned} \quad (12.3.15)$$

Moreover, by (12.3.12), (12.3.8), (12.3.7) and the deterministic Taylor expansion we obtain

$$\begin{aligned} H &\leq \left| E \left(\sum_{n=1}^{n_T} (u(t_n, \mathbf{Y}_n) - u(t_{n-1}, \mathbf{Y}_{n-1})) \right) \right| + K \Delta^\beta \\ &= \left| E \left(\sum_{n=1}^{n_T} (u(t_n, \mathbf{Y}_n) - u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}})) \right) \right| + K \Delta^\beta \\ &\leq H_1 + H_2 + K \Delta^\beta, \end{aligned} \quad (12.3.16)$$

where

$$H_1 = \left| E \left(\sum_{n=1}^{n_T} \left\{ \sum_{i=1}^d \left(\frac{\partial}{\partial y^i} u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}}) \right) (Y_n^i - X_{t_n}^{i; t_{n-1}, Y_{n-1}}) \right\} \right) \right| \quad (12.3.17)$$

and

$$H_2 = \left| E \left(\sum_{n=1}^{n_T} \left\{ \sum_{i,j=1}^d \frac{1}{2} \left(\frac{\partial^2}{\partial y^i \partial y^j} u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}} + \theta_n(\mathbf{Y}_n - \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}})) \right) \right. \right. \right. \\ \times (Y_n^i - X_{t_n}^{i; t_{n-1}, Y_{n-1}})(Y_n^j - X_{t_n}^{j; t_{n-1}, Y_{n-1}}) \left. \right\} \right) \right|. \quad (12.3.18)$$

Here, according to our notation we have used a superscript to denote vector components of \mathbf{Y} and \mathbf{X} . Moreover, θ_n is a $d \times d$ diagonal matrix with

$$\theta_n^{k,k} \in (0, 1) \quad (12.3.19)$$

for $k \in \{1, \dots, d\}$.

Note that by Theorem 4.4.1 we have

$$\mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}} - \mathbf{Y}_n = \sum_{\alpha \in \mathcal{B}(\Gamma_\beta)} \tilde{I}_\alpha [\tilde{f}_\alpha(\mathbf{X}_{\cdot}^{t_{n-1}, Y_{n-1}})]_{t_{n-1}, t_n}, \quad (12.3.20)$$

where $\mathcal{B}(\Gamma_\beta) = \{\alpha : l(\alpha) = \beta + 1\}$. Moreover, by the linear growth condition (12.3.13) on \tilde{f}_α one can show that for every $p \in \{1, 2, \dots\}$ there exist constants K and r such that for every $q \in \{1, \dots, p\}$

$$E \left(\max_{0 \leq n \leq n_T} |\mathbf{Y}_n|^{2q} \right) \leq K(1 + |\mathbf{Y}_0|^{2r}).$$

Therefore, by (12.3.13) for every $\alpha \in \Gamma_\beta \cup \mathcal{B}(\Gamma_\beta)$ and $p \in \{1, 2, \dots\}$ there exist constants K and r such that

$$E(|\tilde{f}_\alpha(\mathbf{X}_z^{t_{n-1}, Y_{n-1}})|^{2q}) \leq K(1 + |\mathbf{Y}_0|^{2r}), \quad (12.3.21)$$

for every $n \in \{1, \dots, n_T\}$ and $z \in [t_{n-1}, t_n]$.

We can now apply (12.3.20), and since (12.3.21) holds, then by Lemma 4.5.4, we obtain

$$H_1 \leq E \left(\sum_{n=1}^{n_T} \sum_{i=1}^d \sum_{\{\alpha: l(\alpha)=\beta+1\}} \right. \\ \left| E \left(\left(\frac{\partial}{\partial y^i} u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}}) \right) \tilde{I}_\alpha (\tilde{f}_\alpha^i(\mathbf{X}_{\cdot}^{t_{n-1}, Y_{n-1}}))_{t_{n-1}, t_n} | \mathcal{A}_{t_{n-1}} \right) \right| \\ \leq KE \left(\sum_{n=1}^{n_T} \sum_{i=1}^d \sum_{\{\alpha: l(\alpha)=\beta+1\}} (t_n - t_{n-1})^{\beta+1} \right) \\ \leq K\Delta^\beta. \quad (12.3.22)$$

To estimate H_2 we can apply (12.3.20), and since (12.3.21) holds, then by Lemma 4.5.5, we obtain

$$\begin{aligned}
H_2 &\leq \left| E \left(\sum_{n=1}^{n_T} \left\{ \sum_{i,j=1}^d \frac{1}{4} \left(\frac{\partial^2}{\partial y^i \partial y^j} u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}} + \boldsymbol{\theta}_n(\mathbf{Y}_n - \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}})) \right) \right. \right. \right. \\
&\quad \times \left[(Y_n^i - X_{t_n}^{i; t_{n-1}, Y_{n-1}})^2 + (Y_n^j - X_{t_n}^{j; t_{n-1}, Y_{n-1}})^2 \right] \left. \right\} \right) \Bigg| \\
&= \left| E \left(\sum_{n=1}^{n_T} \left\{ \sum_{i,j=1}^d \frac{1}{2} \left(\frac{\partial^2}{\partial y^i \partial y^j} u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}} + \boldsymbol{\theta}_n(\mathbf{Y}_n - \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}})) \right) \right. \right. \right. \\
&\quad \times (Y_n^i - X_{t_n}^{i; t_{n-1}, Y_{n-1}})^2 \left. \right\} \right) \Bigg| \\
&\leq K E \left(\sum_{n=1}^{n_T} \left\{ \sum_{i,j=1}^d \sum_{\{\alpha: l(\alpha)=\beta+1\}} \right. \right. \\
&\quad \times E \left(\left| \frac{\partial^2}{\partial y^i \partial y^j} u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}} + \boldsymbol{\theta}_n(\mathbf{Y}_n - \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}})) \right| \right. \\
&\quad \times \left| \tilde{I}_\alpha \left(\tilde{f}_\alpha^i(\mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}}) \right)_{t_{n-1}, t_n} \right|^2 | \mathcal{A}_{t_{n-1}} \right) \left. \right\} \right) \leq K \Delta^\beta. \quad (12.3.23)
\end{aligned}$$

Finally, by combining (12.3.16), (12.3.22) and (12.3.23) we complete the proof of Theorem 12.3.4. \square

Remark 12.3.7. Note that the proof of Theorem 12.3.4 holds also for the wider class of test functions $g \in C_P^{2k+1}(\mathfrak{R}^d, \mathfrak{R})$ with $k = \beta + 1 - [\frac{\beta+1}{2}]$, see Lemma 4.5.4.

We remark that there also exists a theorem on the leading error coefficients for regular schemes derived in Liu & Li (2000). This establishes for the case with jumps the weak convergence of regular extrapolation methods similar to those presented in Sect. 11.4.

12.4 Exercises

12.1. Consider the Merton type jump diffusion model, see (1.7.36), with SDE

$$dS_t = S_t - (a dt + \sigma dW_t + c dN_t)$$

for $t \geq 0$ and $S_0 > 0$, where $W = \{W_t, t \geq 0\}$ is a standard Wiener process and $N = \{N_t, t \geq 0\}$ a Poisson process with intensity $\lambda \in (0, \infty)$. For an equidistant time discretization provide the weak order 1.0 Taylor scheme.

12.2. For the Merton SDE in Exercise 12.1 provide the weak order 2.0 Taylor scheme.

Jump-Adapted Weak Approximations

In this chapter we consider weak approximations constructed on jump-adapted time discretizations similar to those presented in Chap. 8. Since a jump-adapted discretization includes the jump times of the Poisson measure, we can use various approximations for the pure diffusion part between discretization points. Higher order jump-adapted weak schemes avoid multiple stochastic integrals that involve the Poisson random measure. Only multiple stochastic integrals with respect to time and Wiener processes, or their equivalents, are required. This leads to easily implementable schemes. It needs to be emphasized that jump-adapted weak approximations become computationally demanding when the intensity of the Poisson measure is high.

13.1 Jump-Adapted Weak Schemes

Jump-Adapted Time Discretization

The weak jump-adapted schemes to be presented are constructed on jump-adapted time discretizations. We recall from Chap. 8 that a jump-adapted time discretization

$$(t)_\Delta = \{0 = t_0 < t_1 < \dots < t_{n_T} = T\}, \quad (13.1.1)$$

with maximum step size Δ , includes all the jump times $\{\tau_1, \tau_2, \dots, \tau_{p_\varphi(T)}\}$ of the Poisson random measure p_φ . Moreover, as discussed in Chap. 8, we assume that such a jump-adapted time discretization satisfies the following conditions:

$$P(t_{n+1} - t_n \leq \Delta) = 1, \quad (13.1.2)$$

and

$$t_{n+1} \text{ is } \mathcal{A}_{t_n} - \text{measurable}, \quad (13.1.3)$$

for $n \in \{0, 1, \dots, n_T - 1\}$, if it is not a jump time.

Since jump-adapted weak schemes are very convenient and can be applied often, we present them in what follows with some detail. Similarities with parts of Chap. 11, where we treated the case without jumps, naturally arise and are deliberate.

Weak Order 1.0 Taylor Scheme

The simplest jump-adapted weak Taylor scheme is the *jump-adapted Euler scheme*, presented in Chap. 8. We recall that in the general multi-dimensional case the k th component is given by

$$Y_{t_{n+1}-}^k = Y_{t_n}^k + a^k \Delta_{t_n} + \sum_{j=1}^m b^{k,j} \Delta W_{t_n}^j \quad (13.1.4)$$

and

$$Y_{t_{n+1}}^k = Y_{t_{n+1}-}^k + \int_{\mathcal{E}} c^k(t_{n+1}, \mathbf{Y}_{t_{n+1}-}, v) p_{\varphi}(dv, \{t_{n+1}\}), \quad (13.1.5)$$

where $t_n \in [0, T]$, $\Delta_{t_n} = t_{n+1} - t_n$ and the increments $\Delta W_{t_n}^j = W_{t_{n+1}}^j - W_{t_n}^j$ of the Wiener processes are, $N(0, \Delta_{t_n})$ Gaussian distributed with $j \in \{1, 2, \dots, m\}$ and $n \in \{0, 1, \dots, n_T - 1\}$.

The impact of jumps is generated by equation (13.1.5). If t_{n+1} is a jump time, then $\int_{\mathcal{E}} p_{\varphi}(dv, \{t_{n+1}\}) = 1$ and

$$\int_{\mathcal{E}} c(t_{n+1}, \mathbf{Y}_{t_{n+1}-}, v) p_{\varphi}(dv, \{t_{n+1}\}) = c(t_{n+1}, \mathbf{Y}_{t_{n+1}-}, \xi_{p_{\varphi}(t_{n+1})}),$$

while, if t_{n+1} is not a jump time one has $\mathbf{Y}_{t_{n+1}} = \mathbf{Y}_{t_{n+1}-}$, as $\int_{\mathcal{E}} p_{\varphi}(dv, \{t_{n+1}\}) = 0$. Therefore, the weak order of convergence of the jump-adapted Euler scheme is $\beta = 1.0$, and thus, equals the weak order of the Euler scheme used in (13.1.4) for the diffusive component.

As was discussed in Sect. 11.2 and will be exploited below, for weak convergence it is possible to replace the Gaussian distributed random variables by simpler multi-point distributed random variables that satisfy certain moment conditions. For instance, in the Euler scheme we can replace the random variables $\Delta W_{t_n}^j$ by simpler random variables $\Delta \hat{W}_{t_n}^j$ that satisfy the moment condition (11.2.3). In this case the order of weak convergence of the resulting simplified Euler scheme is still $\beta = 1.0$. It is popular to replace the random variables $\Delta W_{t_n}^j$ by the simpler two-point distributed random variables $\Delta \hat{W}_{2,t_n}^j$, with

$$P\left(\Delta \hat{W}_{2,t_n}^j = \pm \sqrt{\Delta_{t_n}}\right) = \frac{1}{2}, \quad (13.1.6)$$

for $j \in \{1, 2, \dots, m\}$. This yields the *jump-adapted simplified Euler scheme*, whose k th component is given by

$$Y_{t_{n+1}-}^k = Y_{t_n}^k + a^k \Delta_{t_n} + \sum_{j=1}^m b^{k,j} \Delta \hat{W}_{2,t_n}^j \quad (13.1.7)$$

together with (13.1.5). This scheme achieves weak order of convergence $\beta = 1.0$. In this way, we have obtained a scheme with the same order of weak convergence of the Euler scheme, requiring only the generation of simple two-point distributed random variables.

Weak Order 2.0 Taylor Scheme

By using a weak order 2.0 Taylor scheme for the diffusive part, we can derive the *jump-adapted weak order 2.0 Taylor scheme*. In the one-dimensional case, $d = m = 1$, it is given by

$$\begin{aligned} Y_{t_{n+1}-} &= Y_{t_n} + a \Delta_{t_n} + b \Delta W_{t_n} + \frac{bb'}{2} ((\Delta W_{t_n})^2 - \Delta_{t_n}) + a'b \Delta Z_{t_n} \quad (13.1.8) \\ &+ \frac{1}{2} \left(\frac{\partial a}{\partial t} + aa' + \frac{a''}{2} b^2 \right) \Delta_{t_n}^2 + \left(\frac{\partial b}{\partial t} + ab' + \frac{b''}{2} b^2 \right) \{ \Delta W_{t_n} \Delta_{t_n} - \Delta Z_{t_n} \} \end{aligned}$$

and

$$Y_{t_{n+1}} = Y_{t_{n+1}-} + \int_{\mathcal{E}} c(t_{n+1}, Y_{t_{n+1}-}, v) p_{\varphi}(dv, \{t_{n+1}\}), \quad (13.1.9)$$

and achieves weak order $\beta = 2.0$. Here ΔZ_{t_n} represents the double Itô integral defined in (8.2.23). The jump-adapted weak order 2.0 Taylor scheme was first presented in Mikulevicius & Platen (1988).

If we compare the regular weak order 2.0 Taylor scheme (12.1.3) with the jump-adapted weak order 2.0 Taylor scheme (13.1.8)–(13.1.9), then we notice that the latter is much simpler as it avoids multiple stochastic integrals with respect to the Poisson measure.

Also in this case, since we are constructing weak schemes, we have some freedom in the choice of the simplifying random variables. A *jump-adapted simplified weak order 2.0 scheme* is given by

$$\begin{aligned} Y_{t_{n+1}-} &= Y_{t_n} + a \Delta_{t_n} + b \Delta \hat{W}_{t_n} + \frac{bb'}{2} \{ (\Delta \hat{W}_{t_n})^2 - \Delta_{t_n} \} \quad (13.1.10) \\ &+ \frac{1}{2} \left(\frac{\partial a}{\partial t} + aa' + \frac{a''}{2} b^2 \right) \Delta_{t_n}^2 + \frac{1}{2} \left(\frac{\partial b}{\partial t} + a'b + ab' + \frac{b''}{2} b^2 \right) \Delta \hat{W}_{t_n} \Delta_{t_n} \end{aligned}$$

and (13.1.9). To obtain weak order 2.0 convergence, the random variable $\Delta \hat{W}_{t_n}$ should be $\mathcal{A}_{t_{n+1}}$ -measurable and satisfy the moment condition (11.2.8). Note that if we choose independent random variables $\Delta \hat{W}_{t_n}$, $n \in \{0, \dots, n_T - 1\}$, then we automatically obtain the required $\mathcal{A}_{t_{n+1}}$ -measurability. Recall that the moment condition (11.2.8) is satisfied, for instance, by a three-point distributed random variable $\Delta \hat{W}_{3,t_n}$, where

$$P(\Delta \hat{W}_{3,t_n} = \pm \sqrt{3\Delta_{t_n}}) = \frac{1}{6}, \quad P(\Delta \hat{W}_{3,t_n} = 0) = \frac{2}{3}. \quad (13.1.11)$$

Note that the jump-adapted simplified weak order 2.0 scheme (13.1.10) requires only one random variable to be generated at each time step. Therefore, it is computationally more efficient than the jump-adapted weak order 2.0 Taylor scheme (13.1.8).

In the general multi-dimensional case, the k th component of the *jump-adapted weak order 2.0 Taylor scheme* is of the form

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + a^k \Delta_{t_n} + \frac{L^0 a^k}{2} \Delta_{t_n}^2 \\ &\quad + \sum_{j=1}^m (b^{k,j} \Delta W_{t_n}^j + L^0 b^{k,j} I_{(0,j)} + L^j a^k I_{(j,0)}) + \sum_{j_1, j_2=1}^m L^{j_1} b^{k,j_2} I_{(j_1, j_2)} \end{aligned} \quad (13.1.12)$$

and (13.1.5), where the operators L^0 and L^j , with $j \in \{1, 2, \dots, m\}$, are defined in (4.3.4) and (4.3.5). The multiple stochastic integrals $I_{(0,j)}$ and $I_{(j,0)}$, for $j \in \{1, 2, \dots, m\}$, can be generated as in (12.1.12). However, the generation of the multiple stochastic integrals involving two components of the Wiener process, $I_{(j_1, j_2)}$, with $j_1, j_2 \in \{1, 2, \dots, m\}$, is not straightforward. In the special case of the diffusion commutativity condition (6.3.14), one can express these multiple stochastic integrals in terms of the Gaussian increments of the Wiener processes $\Delta W_{t_n}^j$. In general, since we are interested in a weak approximation, we can replace the multiple stochastic integrals by the corresponding simple multi-point distributed random variables satisfying appropriate moment conditions. In this way, we obtain the *jump-adapted simplified weak order 2.0 scheme* with k th component of the form

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + a^k \Delta_{t_n} + \frac{L^0 a^k}{2} \Delta_{t_n}^2 + \sum_{j=1}^m \left\{ b^{k,j} + \frac{\Delta_{t_n}}{2} (L^0 b^{k,j} + L^j a^k) \right\} \Delta \hat{W}_{t_n}^j \\ &\quad + \frac{1}{2} \sum_{j_1, j_2=1}^m L^{j_1} b^{k,j_2} (\Delta \hat{W}_{t_n}^{j_1} \Delta \hat{W}_{t_n}^{j_2} + V_{t_n}^{j_1, j_2}) \end{aligned} \quad (13.1.13)$$

and (13.1.5), where $\Delta \hat{W}_{t_n}^j$, with $j \in \{1, 2, \dots, m\}$, are independent random variables such as (13.1.11), satisfying the moment conditions (11.2.8). Additionally as in Sect. 11.2, the quantities $V_{t_n}^{j_1, j_2}$ are independent two-point distributed random variables given by

$$P(V_{t_n}^{j_1, j_2} = \pm \Delta_{t_n}) = \frac{1}{2}, \quad (13.1.14)$$

for $j_2 \in \{1, \dots, j_1 - 1\}$

$$V_{t_n}^{j_1, j_1} = -\Delta_{t_n} \quad (13.1.15)$$

and

$$V_{t_n}^{j_1, j_2} = -V_{t_n}^{j_2, j_1} \quad (13.1.16)$$

for $j_2 \in \{j_1 + 1, \dots, m\}$ and $j_1 \in \{1, 2, \dots, m\}$.

Weak Order 3.0 Taylor Scheme

By including in the diffusive component all multiple stochastic integrals of multiplicity three with respect to time and Wiener processes, one obtains the *jump-adapted weak order 3.0 Taylor scheme*. In the general multi-dimensional case, its k th component is given by

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + a^k \Delta_{t_n} + \frac{L^0 a^k}{2} \Delta_{t_n}^2 \\ &\quad + \sum_{j=1}^m (b^{k,j} \Delta W_{t_n}^j + L^0 b^{k,j} I_{(0,j)} + L^j a^k I_{(j,0)}) + \sum_{j_1, j_2=1}^m L^{j_1} b^{k,j_2} I_{(j_1, j_2)} \\ &\quad + \sum_{j_1, j_2=0}^m L^{j_1} L^{j_2} a^k I_{(j_1, j_2, 0)} + \sum_{j_1, j_2=0}^m \sum_{j_3=1}^m L^{j_1} L^{j_2} b^{k,j_3} I_{(j_1, j_2, j_3)} \end{aligned} \quad (13.1.17)$$

and (13.1.5). Because of the difficulties in the generation of multiple stochastic integrals involving different components of the Wiener process, this scheme is usually too complex for practical implementation. However, in the following we will consider some special cases that lead to conveniently implementable schemes with third order weak convergence. These schemes generalize those derived in Platen (1984) for pure diffusion SDEs, see also Sect. 11.2.

In the one-dimensional case, $d = m = 1$, by approximating the multiple stochastic integrals with Gaussian random variables, we obtain a *jump-adapted weak order 3.0 scheme* given by

$$\begin{aligned} Y_{n+1} &= Y_n + a \Delta + b \Delta W_{t_n} + \frac{1}{2} L^1 b \left\{ (\Delta W_{t_n})^2 - \Delta_{t_n} \right\} \\ &\quad + L^1 a \Delta Z_{t_n} + \frac{1}{2} L^0 a (\Delta_{t_n})^2 + L^0 b \{ \Delta W_{t_n} \Delta_{t_n} - \Delta Z_{t_n} \} \\ &\quad + \frac{1}{6} (L^0 L^0 b + L^0 L^1 a + L^1 L^0 a) \{ \Delta W_{t_n} (\Delta_{t_n})^2 \} \\ &\quad + \frac{1}{6} (L^1 L^1 a + L^1 L^0 b + L^0 L^1 b) \{ (\Delta W_{t_n})^2 - \Delta_{t_n} \} \Delta_{t_n} \\ &\quad + \frac{1}{6} L^0 L^0 a (\Delta_{t_n})^3 + \frac{1}{6} L^1 L^1 b \{ (\Delta W_{t_n})^2 - 3 \Delta_{t_n} \} \Delta W_{t_n}, \end{aligned} \quad (13.1.18)$$

and equation (13.1.5). Here ΔW_{t_n} and ΔZ_{t_n} are the correlated Gaussian random variables of the type defined in (5.3.37) or (11.2.18)–(11.2.19).

To construct a third order simplified method, the required simplified random variables $\Delta \hat{W}_n$ need to satisfy the following moment condition:

$$\begin{aligned} &|E(\Delta \hat{W}_{t_n})| + |E((\Delta \hat{W}_{t_n})^3)| + |E((\Delta \hat{W}_{t_n})^5)| + |E((\Delta \hat{W}_{t_n})^7)| \\ &\quad + |E((\Delta \hat{W}_{t_n})^2) - \Delta_{t_n}| + |E((\Delta \hat{W}_{t_n})^4) - 3 \Delta_{t_n}^2| \\ &\quad + |E((\Delta \hat{W}_{t_n})^6) - 15 \Delta_{t_n}^3| \leq K \Delta^4. \end{aligned} \quad (13.1.19)$$

In [Hofmann \(1994\)](#) a four-point distributed random variable $\Delta\hat{W}_{4,t_n}$ that satisfies condition [\(13.1.19\)](#) was proposed, where

$$\begin{aligned} P\left(\Delta\hat{W}_{4,t_n} = \pm\sqrt{3+\sqrt{6}} \sqrt{\Delta_{t_n}}\right) &= \frac{1}{12+4\sqrt{6}}, \\ P\left(\Delta\hat{W}_{4,t_n} = \pm\sqrt{3-\sqrt{6}} \sqrt{\Delta_{t_n}}\right) &= \frac{1}{12-4\sqrt{6}}, \end{aligned} \quad (13.1.20)$$

see [\(11.2.21\)](#). However, since the probabilities in [\(13.1.20\)](#) are not rational numbers, the corresponding four-point distributed random variable cannot be efficiently implemented by random bit generation, see [Bruti-Liberati & Platen \(2004\)](#) and [Bruti-Liberati, Martini, Piccardi & Platen \(2008\)](#). Instead, we present an alternative five-point distributed random variable $\Delta\hat{W}_{5,t_n}$, with

$$\begin{aligned} P\left(\Delta\hat{W}_{5,t_n} = \pm\sqrt{6\Delta_{t_n}}\right) &= \frac{1}{30}, & P\left(\Delta\hat{W}_{5,t_n} = \pm\sqrt{\Delta_{t_n}}\right) &= \frac{9}{30}, \\ P\left(\Delta\hat{W}_{5,t_n} = 0\right) &= \frac{1}{3}, \end{aligned} \quad (13.1.21)$$

which still satisfies condition [\(13.1.19\)](#) and is a highly efficient implementation based on random bit generators, see [Bruti-Liberati et al. \(2008\)](#). This five-point distributed random variable has also been suggested in [Milstein & Tretjakov \(2004\)](#). The corresponding *jump-adapted simplified weak order 3.0 scheme* is then given by

$$\begin{aligned} Y_{n+1} = Y_n + a\Delta + b\Delta\hat{W}_{5,t_n} + \frac{1}{2} L^1 b \left\{ (\Delta\hat{W}_{5,t_n})^2 - \Delta_{t_n} \right\} + \frac{1}{2} L^0 a \Delta_{t_n}^2 \\ + \frac{1}{2} L^1 a \left\{ \Delta\hat{W}_{5,t_n} + \frac{1}{\sqrt{3}} \Delta\hat{W}_{2,t_n} \right\} \Delta_{t_n} \\ + \frac{1}{2} L^0 b \left\{ \Delta\hat{W}_{5,t_n} - \frac{1}{\sqrt{3}} \Delta\hat{W}_{2,t_n} \right\} \Delta_{t_n} \\ + \frac{1}{6} (L^0 L^0 b + L^0 L^1 a + L^1 L^0 a) \Delta\hat{W}_{5,t_n} \Delta_{t_n}^2 \\ + \frac{1}{6} (L^1 L^1 a + L^1 L^0 b + L^0 L^1 b) \left\{ (\Delta\hat{W}_{5,t_n})^2 - \Delta_{t_n} \right\} \Delta_{t_n} \\ + \frac{1}{6} L^0 L^0 a \Delta_{t_n}^3 + \frac{1}{6} L^1 L^1 b \left\{ (\Delta\hat{W}_{5,t_n})^2 - 3 \Delta_{t_n} \right\} \Delta\hat{W}_{5,t_n}, \end{aligned} \quad (13.1.22)$$

and relation [\(13.1.5\)](#). It involves the five-point distributed random variables $\Delta\hat{W}_{5,t_n}$ and the two-point distributed random variables $\Delta\hat{W}_{2,t_n}$, see [\(13.1.6\)](#). This scheme achieves an order of weak convergence $\beta = 3.0$.

We remark that in [Kloeden & Platen \(1999\)](#) there exist multi-dimensional versions of weak order $\beta = 3.0$ schemes for the case with additive noise using Gaussian random variables. Furthermore, there is an alternative weak order 3.0 scheme with multi-point distributed random variables given in [Hofmann \(1994\)](#).

13.2 Derivative-Free Schemes

In the previous section we explained how to construct jump-adapted weak Taylor schemes with higher order of weak convergence β . However, the jump-adapted order 2.0 and weak order 3.0 Taylor schemes require the computation of derivatives of the drift and diffusion coefficients. In this section we present higher order weak schemes that avoid the computation of those derivatives.

In the one-dimensional, autonomous case, $d = m = 1$, the *jump-adapted order weak 2.0 derivative-free scheme* is given by

$$\begin{aligned} Y_{t_{n+1}-} &= Y_{t_n} + \frac{1}{2} \left(a(\bar{Y}_{t_n}) + a \right) \Delta_{t_n} + \frac{1}{4} \left(b(\bar{Y}_{t_n}^+) + b(\bar{Y}_{t_n}^-) + 2b \right) \Delta W_{t_n} \\ &\quad + \frac{1}{4} \left(b(\bar{Y}_{t_n}^+) - b(\bar{Y}_{t_n}^-) \right) \left((\Delta W_{t_n})^2 - \Delta_{t_n} \right) (\Delta_{t_n})^{-\frac{1}{2}} \end{aligned}$$

and

$$Y_{t_{n+1}} = Y_{t_{n+1}-} + \int_{\mathcal{E}} c(t_{n+1}, Y_{t_{n+1}-}, v) p_\varphi(dv, \{t_{n+1}\}), \quad (13.2.1)$$

with supporting values

$$\bar{Y}_{t_n} = Y_{t_n} + a \Delta_{t_n} + b \Delta W_{t_n},$$

and

$$\bar{Y}_{t_n}^\pm = Y_{t_n} + a \Delta_{t_n} \pm b \sqrt{\Delta_{t_n}}.$$

This scheme attains weak order $\beta = 2.0$ and generalizes a scheme for pure diffusions presented in Platen (1984), see also Sect. 11.3.

If we replace the Gaussian random variables ΔW_{t_n} by the three-point distributed random variables $\Delta \hat{W}_{3,t_n}$ defined in (13.1.11), then we obtain the *jump-adapted simplified weak order 2.0 derivative-free scheme* that still achieves weak order $\beta = 2.0$.

In the autonomous, multi-dimensional case, the k th component of a *jump-adapted weak order 2.0 derivative-free scheme* is given by

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + \frac{1}{2} \left(a^k(\bar{Y}_{t_n}) + a^k \right) \Delta_{t_n} \\ &\quad + \frac{1}{4} \sum_{j=1}^m \left\{ \left(b^{k,j}(\bar{R}_{t_n}^{+,j}) + b^{k,j}(\bar{R}_{t_n}^{-,j}) + 2b^{k,j} \right) \Delta \hat{W}_{t_n}^j \right. \\ &\quad \left. + \sum_{\substack{r=1 \\ r \neq j}}^m \left(b^{k,j}(\bar{U}_{t_n}^{+,r}) + b^{k,j}(\bar{U}_{t_n}^{-,r}) + 2b^{k,j} \right) \Delta \hat{W}_{t_n}^j (\Delta_{t_n})^{-\frac{1}{2}} \right\} \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{4} \sum_{j=1}^m \left\{ (b^{k,j}(\bar{\mathbf{R}}_{t_n}^{+,j}) - b^{k,j}(\bar{\mathbf{R}}_{t_n}^{-,j})) (\left(\Delta \hat{W}_{t_n}^j \right)^2 - \Delta_{t_n}) \right. \\
& \left. + \sum_{\substack{r=1 \\ r \neq j}}^m (b^{k,j}(\bar{\mathbf{U}}_{t_n}^{+,r}) - b^{k,j}(\bar{\mathbf{U}}_{t_n}^{-,r})) (\Delta \hat{W}_{t_n}^j \Delta \hat{W}_{t_n}^r + V_{t_n}^{r,j}) \right\}
\end{aligned} \quad (13.2.2)$$

and

$$Y_{t_{n+1}}^k = Y_{t_{n+1}-}^k + \int_{\mathcal{E}} c^k(t_{n+1}, \mathbf{Y}_{t_{n+1}-}, v) p_\varphi(dv, \{t_{n+1}\}), \quad (13.2.3)$$

for $k \in \{1, 2, \dots, d\}$. The supporting values being

$$\bar{\mathbf{Y}}_{t_n} = \mathbf{Y}_{t_n} + \mathbf{a} \Delta_{t_n} + \sum_{j=1}^m \mathbf{b}^j \Delta \hat{W}_{t_n}^j,$$

$$\bar{\mathbf{R}}_{t_n}^\pm = \mathbf{Y}_{t_n} + \mathbf{a} \Delta_{t_n} \pm \mathbf{b}^j \sqrt{\Delta_{t_n}},$$

and

$$\bar{\mathbf{U}}_{t_n}^\pm = \mathbf{Y}_{t_n} \pm \mathbf{b}^j \sqrt{\Delta_{t_n}}.$$

Moreover, $\Delta \hat{W}_{t_n}^j$ has to be an independent $\mathcal{A}_{t_{n+1}}$ -measurable, random variable satisfying the moment condition (11.2.8). For instance, one can choose independent Gaussian $N(0, \Delta_{t_n})$ distributed random variables or three-point distributed random variables of the form (13.1.11). Furthermore, $V_{t_n}^{r,j}$ are the previously introduced two-point distributed random variables defined in (11.2.13)–(11.2.15) or (13.1.14)–(13.1.16).

Additional weak order 2.0 methods for approximating the diffusion part can be found in Chap. 11 or Kloeden & Platen (1999). This includes implicit methods, which can also be used in jump-adapted weak approximations.

13.3 Predictor-Corrector Schemes

As discussed in Chap. 7 and will be the topic of Chap. 14, it is important to develop schemes with good numerical stability properties and computationally convenient structures. Predictor-corrector schemes have reasonable numerical stability, and are computationally efficient. Since the difference between the predicted and the corrected values provides an indication of the local error, one can use this information to design advanced schemes with step size control or for selecting a suitable general step size, see Hofmann, Müller-Gronbach & Ritter (2000a).

Again, as in Sect. 11.5, the diffusive component of jump-adapted weak predictor-corrector schemes can be derived in the following way: The corrector equation is obtained using the Wagner-Platen expansion for pure diffusions in implicit form, see Kloeden & Platen (1999). This implicit scheme is then

made explicit by using a predictor equation to generate the next value. The predictor component of the scheme can be derived by the explicit Wagner-Platen expansion for pure diffusions.

Order 1.0 Predictor-Corrector Scheme

In the one-dimensional case, $d = m = 1$, the *jump-adapted predictor-corrector Euler scheme* is given by the corrector

$$Y_{t_{n+1}-} = Y_{t_n} + \frac{1}{2} \left\{ a(\bar{Y}_{t_{n+1}-}) + a \right\} \Delta_{t_n} + b \Delta W_{t_n}, \quad (13.3.1)$$

the predictor

$$\bar{Y}_{t_{n+1}-} = Y_{t_n} + a \Delta_{t_n} + b \Delta W_{t_n}, \quad (13.3.2)$$

and the jump condition

$$Y_{t_{n+1}} = Y_{t_{n+1}-} + \int_{\mathcal{E}} c(t_{n+1}, Y_{t_{n+1}-}, v) p_{\varphi}(dv, \{t_{n+1}\}). \quad (13.3.3)$$

By introducing some effects of implicitness in the diffusion part of the corrector, we obtain, for $\theta, \eta \in [0, 1]$, a *family of jump-adapted weak order 1.0 predictor-corrector schemes* with corrector

$$Y_{t_{n+1}-} = Y_{t_n} + \left\{ \theta \bar{a}(\bar{Y}_{t_{n+1}-}) + (1 - \theta) \bar{a} \right\} \Delta_{t_n} + \left\{ \eta b(\bar{Y}_{t_{n+1}-}) + (1 - \eta) b \right\} \Delta W_{t_n},$$

where $\bar{a} = a - \eta b b'$, and predictor

$$\bar{Y}_{t_{n+1}-} = Y_{t_n} + a \Delta_{t_n} + b \Delta W_{t_n},$$

again using (13.3.3). Note that this family of Euler schemes coincides with the family of jump-adapted predictor-corrector Euler schemes (8.5.1)–(8.5.2), which achieve strong order $\gamma = 0.5$. However, the weak order convergence of this family of schemes is $\beta = 1.0$. Moreover, if we replace the Gaussian random variables ΔW_{t_n} , by the two-point distributed random variables $\Delta \hat{W}_{2,t_n}$, as defined in (13.1.6), then we obtain a family of *jump-adapted simplified weak order 1.0 predictor-corrector schemes* that still attain weak order $\beta = 1.0$. The resulting family of weak order 1.0 predictor-corrector schemes generalizes the family of schemes presented in Platen (1995) for pure diffusion SDEs, see also (11.5.6).

In the general multi-dimensional case, we can construct a *family of jump-adapted weak order 1.0 predictor-corrector schemes*. Its k th component is given by the corrector

$$\begin{aligned} Y_{t_{n+1}-}^k &= Y_{t_n}^k + \left\{ \theta \bar{a}_{\eta}^k(t_{n+1}, \bar{Y}_{t_{n+1}-}) + (1 - \theta) \bar{a}_{\eta}^k \right\} \Delta_{t_n} \\ &+ \sum_{j=1}^m \left\{ \eta b^{k,j}(t_{n+1}, \bar{Y}_{t_{n+1}-}) + (1 - \eta) b^{k,j} \right\} \Delta \hat{W}_{t_n}^j, \end{aligned} \quad (13.3.4)$$

for $\theta, \eta \in [0, 1]$, where

$$\bar{a}_\eta = a - \eta \sum_{j_1, j_2=1}^m \sum_{i=1}^d b^{k, j_1} \frac{\partial b^{k, j_2}}{\partial x^i}. \quad (13.3.5)$$

Here we use the predictor

$$\bar{Y}_{t_{n+1}-}^k = Y_{t_n}^k + a^k \Delta_{t_n} + \sum_{j=1}^m b^{k, j} \Delta \hat{W}_{t_n}^j, \quad (13.3.6)$$

and the jump condition

$$Y_{t_{n+1}}^k = Y_{t_{n+1}-}^k + \int_{\mathcal{E}} c^k(t_{n+1}, \mathbf{Y}_{t_{n+1}-}, v) p_\varphi(dv, \{t_{n+1}\}). \quad (13.3.7)$$

The random variables $\Delta \hat{W}_{t_n}^j$, with $j \in \{1, 2, \dots, m\}$ and $n \in \{0, 1, \dots, n_T - 1\}$, can be chosen, for instance, as independent Gaussian $N(0, \Delta_{t_n})$ distributed random variables or as two-point distributed random variables in the form (13.1.6).

Weak Order 2.0 Predictor-Corrector Scheme

By using higher order weak schemes for the diffusive component of the predictor and corrector algorithms, we obtain predictor-corrector schemes of higher weak order.

In the general multi-dimensional case, the k th component of the *jump-adapted weak order 2.0 predictor-corrector scheme* has corrector

$$Y_{t_{n+1}-}^k = Y_{t_n}^k + \frac{1}{2} \{ a^k(t_{n+1}, \bar{\mathbf{Y}}_{t_{n+1}-}) + a^k \} \Delta_{t_n} + \Psi_{t_n}^k,$$

with

$$\begin{aligned} \Psi_{t_n}^k &= \sum_{j=1}^m \left\{ b^{k, j} + \frac{1}{2} L^0 b^{k, j} \Delta_{t_n} \right\} \Delta \hat{W}_{t_n}^j \\ &\quad + \frac{1}{2} \sum_{j_1, j_2=1}^m L^{j_1} b^{k, j_2} \left(\Delta \hat{W}_{t_n}^{j_1} \Delta \hat{W}_{t_n}^{j_2} + V_{t_n}^{j_1, j_2} \right), \end{aligned}$$

predictor

$$\bar{Y}_{t_{n+1}-}^k = Y_{t_n}^k + a^k \Delta_{t_n} + \Psi_{t_n}^k + \frac{1}{2} L^0 a^k (\Delta_{t_n})^2 + \frac{1}{2} \sum_{j=1}^m L^j a^k \Delta \hat{W}_{t_n}^j \Delta_{t_n}$$

and using (13.3.7). The random variables $\Delta \hat{W}_{t_n}^j$, with $j \in \{1, 2, \dots, m\}$ and $n \in \{0, 1, \dots, n_T - 1\}$, can be chosen as independent Gaussian $N(0, \Delta_{t_n})$ distributed random variables or three-point distributed random variables, see (13.1.11). The two-point distributed random variables $V_{t_n}^{j_1, j_2}$, $j_1, j_2 \in \{1, 2, \dots, m\}$, are as defined in (13.1.14)–(13.1.16).

13.4 Some Jump-Adapted Exact Weak Schemes

SDEs with Jumps

As a generalization of Chap. 2 we now discuss a special class of SDEs for which it is possible to develop jump-adapted schemes that do not generate any discretization error. Similar to Sect. 8.6, we introduce a jump-adapted time discretization, obtained by a superposition of jump times generated by the Poisson measure and times at which we are interested in sampling the simulated values of the solution \mathbf{X} . Between jump times the SDE is assumed to have an explicit transition density.

Let us consider the d -dimensional SDE with jumps

$$d\mathbf{X}_t = \mathbf{a}(t, \mathbf{X}_t) dt + \mathbf{b}(t, \mathbf{X}_t) d\mathbf{W}_t + \int_{\mathcal{E}} \mathbf{c}(t, \mathbf{X}_{t-}, v) p_{\varphi}(dv, dt) \quad (13.4.1)$$

for $t \in [0, T]$ and $\mathbf{X}_0 \in \mathbb{R}^d$, that we aim to solve. We have already discussed in Chap. 2 and Sect. 8.6 cases where the corresponding diffusion SDE

$$d\mathbf{Z}_t = \mathbf{a}(t, \mathbf{Z}_t) dt + \mathbf{b}(t, \mathbf{Z}_t) d\mathbf{W}_t, \quad (13.4.2)$$

admits an explicit solution. Now, in this section we assume less, as we require only that the transition density of \mathbf{Z}_t is explicitly known. In Chap. 2 we describe several cases of solutions of SDEs with explicitly known transition densities.

An Example for an Exact Weak Scheme

Since we are interested in weak approximations, we can construct exact or almost exact jump-adapted schemes for a wider class of jump diffusion SDEs than in Sect. 8.6. Let us present an illustrative example: Consider the SDE with jumps given by

$$dX_t = \alpha(b - X_t) dt + \sigma\sqrt{X_t} dW_t + \int_{\mathcal{E}} c(t, X_{t-}, v) p_{\varphi}(dv, dt), \quad (13.4.3)$$

where $\alpha, b, \sigma \in \mathbb{R}$ and $2\alpha b > \sigma^2$. In this case the corresponding diffusion, given by

$$dZ_t = \alpha(b - Z_t) dt + \sigma\sqrt{Z_t} dW_t, \quad (13.4.4)$$

describes a square root process, see Cox et al. (1985) and Chap. 2. The transition density of Z is known in closed form. The distribution of Z_t given Z_s , with $s < t$, is a non-central chi-square distribution, see Platen & Heath (2006) and Chap. 2. Therefore, we can construct a jump-adapted weak scheme given by

$$Y_{t_{n+1}-} = \frac{\sigma^2(1 - e^{-\alpha\Delta_{t_n}})}{4\alpha} \chi^2(\delta, l_{t_n}), \quad (13.4.5)$$

and

$$Y_{t_{n+1}} = Y_{t_{n+1}-} + \int_{\mathcal{E}} c(t_{n+1}, Y_{t_{n+1}-}, v) p_\varphi(dv, \{t_{n+1}\}). \quad (13.4.6)$$

Here the quantities $\chi^2(\delta, l_{t_n})$, with $n \in \{0, \dots, n_T - 1\}$, denote independent non-central chi-square distributed random variables with

$$\delta = \frac{4ab}{\sigma^2}$$

degrees of freedom and non-centrality parameter

$$l_{t_n} = \frac{4\alpha e^{-\alpha \Delta_{t_n}}}{\sigma^2(1 - e^{-\alpha \Delta_{t_n}})} Y_{t_n}.$$

In this way, the jump-adapted weak scheme (13.4.5)–(13.4.6) is exact in the sense that it does not generate any weak error. In fact, we obtain the result that at the discretization points the distribution of the numerical approximation Y coincides with that of the solution X of (13.4.3).

It is clear from Chap. 2 that this approach can be generalized to the case of d -dimensional jump diffusion SDEs of the type (13.4.1). In some cases one may only be able to construct an almost exact weak approximation, as discussed in Chap. 2. However, this is still very valuable in practice since the systematic error becomes negligible. In modeling, one should always check whether the diffusion part of the dynamics under consideration could be modeled by an SDE that belongs to the special subclass of SDEs whose solution admits a closed form transition density. This can bring significant computational advantages.

13.5 Convergence of Jump-Adapted Weak Taylor Schemes

Jump-Adapted Time Discretization

In this section we present a convergence theorem for jump-adapted weak Taylor approximations of any weak order of convergence $\beta \in \{1, 2, \dots\}$. This theorem also covers the convergence of the schemes presented in Sect. 13.2. The results of this section resemble those in Mikulevicius & Platen (1988). They will also be employed in the next section, where we will prove weak convergence of general jump-adapted approximations. Similar results regarding the order of weak convergence of the jump-adapted Euler scheme under weaker conditions on the coefficients of SDEs were obtained in Kubilius & Platen (2002) and Glasserman & Merener (2003b).

As suggested in Platen (1982a), we define a *jump-adapted time discretization* $0 = t_0 < t_1 < \dots < t_{n_T} = T$ with maximum step size $\Delta \in (0, 1)$,

with n_t defined in (6.1.11). The term *jump-adapted* indicates again that all jump times $\{\tau_1, \tau_2, \dots, \tau_{p_\varphi(T)}\}$ of the Poisson measure p_φ are included in the time discretization. Moreover, we require a maximum step size $\Delta \in (0, 1)$, which means that $P(t_{n+1} - t_n \leq \Delta) = 1$ for every $n \in \{0, 1, 2, \dots, n_T - 1\}$, and if the discretization time t_{n+1} is not a jump time, then t_{n+1} should be \mathcal{A}_{t_n} -measurable. Let us introduce an additional filtration

$$\tilde{\mathcal{A}}_{t_n} = \sigma(\mathbf{1}_{p_\varphi(\mathcal{E}, \{t_{n+1}\}) \neq 0}) \vee \mathcal{A}_{t_n} \quad (13.5.1)$$

for every $n \in \{0, 1, \dots, n_T\}$. We assume that t_{n+1} is $\tilde{\mathcal{A}}_{t_n}$ -measurable. We also assume a finite number of time discretization points, which means that $n_t < \infty$ a.s. for $t \in [0, T]$. The superposition of the jump times with an equidistant time discretization, as discussed in Chap. 6, provides an example of such jump-adapted time discretization.

We recall that for $m \in \mathcal{N}$ we denote the set of all multi-indices α that do not include components equal to -1 by

$$\widehat{\mathcal{M}}_m = \{(j_1, \dots, j_l) : j_i \in \{0, 1, \dots, m\}, i \in \{1, 2, \dots, l\} \text{ for } l \in \mathcal{N}\} \cup \{v\}$$

where v is the multi-index of length zero, see Chap. 4 and Sect. 8.7.

Given a set $\mathcal{A} \subset \widehat{\mathcal{M}}_m$, the *remainder* set $\hat{\mathcal{B}}(\mathcal{A})$ of \mathcal{A} is defined by

$$\hat{\mathcal{B}}(\mathcal{A}) = \{\alpha \in \widehat{\mathcal{M}}_m \setminus \mathcal{A} : -\alpha \in \mathcal{A}\}.$$

Moreover, for every $\beta \in \{1, 2, \dots\}$ we define the hierarchical set

$$\hat{\Gamma}_\beta = \{\alpha \in \widehat{\mathcal{M}}_m : l(\alpha) \leq \beta\}.$$

For a jump-adapted time discretization $(t)_\Delta$, with maximum time step size $\Delta \in (0, 1)$, we define the *jump-adapted weak order β Taylor scheme* by

$$\mathbf{Y}_{t_{n+1}-} = \mathbf{Y}_{t_n} + \sum_{\alpha \in \hat{\Gamma}_\beta \setminus \{v\}} f_\alpha(t_n, \mathbf{Y}_{t_n}) I_\alpha \quad (13.5.2)$$

and

$$\mathbf{Y}_{t_{n+1}} = \mathbf{Y}_{t_{n+1}-} + \int_{\mathcal{E}} c(t_n, \mathbf{Y}_{t_{n+1}-}, v) p_\varphi(dv, \{t_{n+1}\}), \quad (13.5.3)$$

where I_α is the multiple stochastic integral of the multi-index α over the time period $(t_n, t_{n+1}]$, $n \in \{0, 1, \dots, n_T - 1\}$, and f_α is the corresponding Itô coefficient function defined in (4.3.8) with $f(t, \mathbf{x}) = \mathbf{x}$.

Convergence Theorem

Now we can formulate a convergence theorem for jump-adapted schemes similar to that in Mikulevicius & Platen (1988).

Theorem 13.5.1. For a given $\beta \in \{1, 2, \dots\}$, let $\mathbf{Y}^\Delta = \{\mathbf{Y}_{t_n}^\Delta, n \in \{0, 1, \dots, n_T\}\}$ be the weak order β jump-adapted Taylor scheme (13.5.2)–(13.5.3) corresponding to a jump-adapted time discretization with maximum step size $\Delta \in (0, 1)$. We assume that $E(|\mathbf{X}_0|^i) < \infty$, for $i \in \{1, 2, \dots\}$, and \mathbf{Y}_0^Δ converges weakly to \mathbf{X}_0 with order β . Moreover, suppose that the drift, diffusion and jump coefficients have components $a^k, b^{k,j}, c^k \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ for all $k \in \{1, 2, \dots, d\}$ and $j \in \{1, 2, \dots, m\}$ and the coefficients f_α , with $f(t, \mathbf{x}) = \mathbf{x}$, satisfy the linear growth condition $|f_\alpha(t, \mathbf{y})| \leq K(1 + |\mathbf{y}|)$, with $K < \infty$, for all $t \in [0, T]$, $\mathbf{y} \in \mathbb{R}^d$ and $\alpha \in \hat{I}_\beta$, see also Remark 12.3.5.

Then for any function $g \in \mathcal{C}_P^{2(\beta+1)}$ there exists a positive constant C , independent of Δ , such that

$$|E(g(\mathbf{X}_T)) - E(g(\mathbf{Y}_{t_{n_T}}^\Delta))| \leq C\Delta^\beta.$$

First, we present some results in preparation of the proof of Theorem 13.5.1 along the lines of Mikulevicius & Platen (1988).

Some Auxiliary Results

Lemma 13.5.2 For all $n \in \{1, \dots, n_T\}$ and $\mathbf{y} \in \mathbb{R}^d$, we have

$$E\left(u(t_n, \mathbf{X}_{t_{n-1}}^{t_{n-1}, \mathbf{y}}) - u(t_{n-1}, \mathbf{y}) + \int_{t_{n-1}}^{t_n} \int_{\mathcal{E}} L_v^{-1} u(z, \mathbf{X}_z^{t_{n-1}, \mathbf{y}}) \varphi(dv) dz \mid \mathcal{A}_{t_{n-1}}\right) = 0, \quad (13.5.4)$$

where the process $\mathbf{X}^{s, \mathbf{y}} = \{\mathbf{X}_t^{s, \mathbf{y}}, t \in [0, T]\}$, for $(s, \mathbf{y}) \in [0, t] \times \mathbb{R}^d$, is defined in (12.3.1), and

$$u(s, \mathbf{y}) = E(g(\mathbf{X}_T^{s, \mathbf{y}}) \mid \mathcal{A}_s). \quad (13.5.5)$$

Proof: Note that since all jump times are included in the time discretization, $X_z^{t_{n-1}, \mathbf{y}}$ evolves as a diffusion in the time interval (t_{n-1}, t_n) for every $n \in \{1, \dots, n_T\}$. Therefore, by Itô's formula we obtain

$$\begin{aligned} u(t_n, \mathbf{X}_{t_{n-1}}^{t_{n-1}, \mathbf{y}}) &= u(t_{n-1}, \mathbf{y}) + \int_{t_{n-1}}^{t_n} L^0 u(z, \mathbf{X}_z^{t_{n-1}, \mathbf{y}}) dz \\ &\quad + \sum_{j=1}^m \int_{t_{n-1}}^{t_n} L^j u(z, \mathbf{X}_z^{t_{n-1}, \mathbf{y}}) dW_z^j \\ &= u(t_{n-1}, \mathbf{y}) + \int_{t_{n-1}}^{t_n} \tilde{L}^0 u(z, \mathbf{X}_z^{t_{n-1}, \mathbf{y}}) dz \\ &\quad + \sum_{j=1}^m \int_{t_{n-1}}^{t_n} L^j u(z, \mathbf{X}_z^{t_{n-1}, \mathbf{y}}) dW_z^j \\ &\quad - \int_{t_{n-1}}^{t_n} \int_{\mathcal{E}} L_v^{-1} u(z, \mathbf{X}_z^{t_{n-1}, \mathbf{y}}) \varphi(dv) dz, \end{aligned}$$

where the operators L^0 , L^j , $L_v^{(-1)}$ and \tilde{L}^0 are defined in (4.3.4)–(4.3.7).

We complete the proof of the lemma by applying the expected value, using the result (12.3.5) and the properties of Itô integrals. \square

Lemma 13.5.3 *For each $p \in \{1, 2, \dots\}$ there exists a finite constant K such that*

$$E\left(\left|\mathbf{X}_{t_{n-}}^{t_{n-1}, y} - \mathbf{y}\right|^{2q} \middle| \tilde{\mathcal{A}}_{t_{n-1}}\right) \leq K(1 + |\mathbf{y}|^{2q})(t_n - t_{n-1})^q \quad (13.5.6)$$

for all $q \in \{1, \dots, p\}$ and $n \in \{1, \dots, n_T\}$, where $\tilde{\mathcal{A}}_{t_{n-1}}$ is defined in (13.5.1).

Proof: Since there are no jumps between discretization points, the proof of (13.5.6) follows from that of a similar lemma for the pure diffusion case in Kloeden & Platen (1999).

Let us also define the process $\boldsymbol{\eta}_{ja} = \{\boldsymbol{\eta}_{ja}(t), t \in [0, T]\}$ by

$$\begin{aligned} \boldsymbol{\eta}_{ja}(t) &= \boldsymbol{\eta}_{ja}(t_n) + \sum_{\alpha \in \hat{\Gamma}_\beta \setminus \{v\}} I_\alpha(f_\alpha(t_n, \boldsymbol{\eta}_{ja}(t_n)))_{t_n, t} \\ &\quad + \int_{(t_n, t]} \int_{\mathcal{E}} \mathbf{c}(z, \boldsymbol{\eta}_{ja}(z-), v) p_\varphi(dv, dz), \end{aligned} \quad (13.5.7)$$

for $n \in \{0, \dots, n_T - 1\}$ and $t \in (t_n, t_{n+1}]$, with $\boldsymbol{\eta}_{ja}(0) = \mathbf{Y}_0$. Note that

$$\boldsymbol{\eta}_{ja}(t_n) = \mathbf{Y}_{t_n} \quad (13.5.8)$$

for every $n \in \{0, \dots, n_T\}$. \square

The following result is shown in Mikulevicius & Platen (1988).

Lemma 13.5.4 *For each $p \in \{1, 2, \dots\}$ there exists a finite constant K such that for every $q \in \{1, \dots, p\}$*

$$E\left(\sup_{0 \leq t \leq T} |\boldsymbol{\eta}_{ja}(t)|^{2q}\right) \leq K(1 + |\mathbf{Y}_0|^{2q}). \quad (13.5.9)$$

We shall also write

$$P_l = \{1, 2, \dots, d\}^l$$

and

$$F_{\mathbf{p}}(\mathbf{y}) = \prod_{h=1}^l y^{p_h} \quad (13.5.10)$$

for all $\mathbf{y} = (y^1, \dots, y^d)^\top \in \Re^d$ and $\mathbf{p} = (p_1, \dots, p_l) \in P_l$ where $l \in \{1, 2, \dots\}$.

Lemma 13.5.5 For each $p \in \{1, 2, \dots\}$ there exist finite constants K and $r \in \{1, 2, \dots\}$ such that

$$\begin{aligned} & \left| E \left(\left| F_{\mathbf{p}}(\boldsymbol{\eta}_{ja}(z) - \mathbf{Y}_{t_{n_z}}^{\Delta}) \right|^{2q} + \left| F_{\mathbf{p}}(\mathbf{X}_z^{t_{n_z}, Y_{t_{n_z}}^{\Delta}} - \mathbf{Y}_{t_{n_z}}^{\Delta}) \right|^{2q} \middle| \tilde{\mathcal{A}}_{t_{n_z}} \right) \right| \\ & \leq K (1 + |\mathbf{Y}_{t_{n_z}}^{\Delta}|^{2r}) (t_{n_z+1} - t_{n_z})^{ql} \end{aligned} \quad (13.5.11)$$

for each $l \in \{1, \dots, 2(\beta + 1)\}$, $q \in \{1, \dots, p\}$, $\mathbf{p} \in P_l$ and $z \in [0, T]$, where $F_{\mathbf{p}}$ is defined in (13.5.10).

Proof: Note that at discretization times the estimate (13.5.11) is trivial since $\boldsymbol{\eta}_{ja}(z) = \mathbf{Y}_{t_{n_z}}^{\Delta}$ and $\mathbf{X}_z^{t_{n_z}, Y_{t_{n_z}}^{\Delta}} = \mathbf{Y}_{t_{n_z}}^{\Delta}$ for $z \in \{t_0, t_1, \dots, t_{n_T}\}$. Moreover, since jump times arise only at discretization times, we obtain the estimate (13.5.11) by Itô's formula as in the case of pure diffusion SDEs, see Kloeden & Platen (1999). \square

Lemma 13.5.6 For each $\mathbf{p} \in P_l$, $l \in \{1, \dots, 2\beta + 1\}$, $n \in \{1, 2, \dots, n_T\}$ and $z \in [t_{n-1}, t_n]$ there exist two finite constants K and $r \in \{1, 2, \dots\}$ such that

$$\begin{aligned} & \left| E \left(F_{\mathbf{p}}(\boldsymbol{\eta}_{ja}(z) - \mathbf{Y}_{t_{n-1}}^{\Delta}) - F_{\mathbf{p}}(\mathbf{X}_z^{t_{n-1}, Y_{t_{n-1}}^{\Delta}} - \mathbf{Y}_{t_{n-1}}^{\Delta}) \middle| \tilde{\mathcal{A}}_{t_{n-1}} \right) \right| \\ & \leq K (1 + |\mathbf{Y}_{t_{n-1}}^{\Delta}|^r) \Delta^{\beta} (z - t_{n-1}). \end{aligned} \quad (13.5.12)$$

Proof: Since the time discretization includes all jump times, the proof of the lemma follows from that for pure diffusion SDEs, see Kloeden & Platen (1999). \square

Proof of the Convergence Theorem

Proof of Theorem 13.5.1: By (12.3.3) and the terminal condition of the Kolmogorov backward equation (12.3.6) we obtain

$$\begin{aligned} H &= \left| E(g(\mathbf{Y}_{t_{n_T}}^{\Delta})) - E(g(\mathbf{X}_T)) \right| \\ &= \left| E(u(T, \mathbf{Y}_{t_{n_T}}^{\Delta}) - u(0, \mathbf{X}_0)) \right|. \end{aligned} \quad (13.5.13)$$

Note that for the ease of notation we will write \mathbf{Y} for \mathbf{Y}^{Δ} when no misunderstanding is possible. Since \mathbf{Y}_0 converges weakly to \mathbf{X}_0 with order β we obtain

$$\begin{aligned} H &\leq \left| E \left(\sum_{n=1}^{n_T} (u(t_n, \mathbf{Y}_{t_n}) - u(t_n, \mathbf{Y}_{t_n-}) + u(t_n, \mathbf{Y}_{t_n-}) - u(t_{n-1}, \mathbf{Y}_{t_{n-1}})) \right) \right| \\ &\quad + K \Delta^{\beta}. \end{aligned} \quad (13.5.14)$$

By (13.5.4) we can write

$$\begin{aligned} H \leq & \left| E \left(\sum_{n=1}^{n_T} \left[\{u(t_n, \mathbf{Y}_{t_n}) - u(t_n, \mathbf{Y}_{t_n-}) + u(t_n, \mathbf{Y}_{t_n-}) - u(t_{n-1}, \mathbf{Y}_{t_{n-1}})\} \right. \right. \right. \\ & - \{u(t_n, \mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}}) - u(t_{n-1}, \mathbf{Y}_{t_{n-1}}) \\ & \left. \left. \left. + \int_{t_{n-1}}^{t_n} \int_{\mathcal{E}} L_v^{-1} u(z, \mathbf{X}_z^{t_{n-1}, Y_{t_{n-1}}}) \varphi(dv) dz \} \right] \right) \right| + K \Delta^\beta. \end{aligned}$$

Note that by (13.5.8) and the properties of the stochastic integral with respect to the Poisson measure, we have

$$\begin{aligned} E \left(\sum_{n=1}^{n_T} \{u(t_n, \mathbf{Y}_{t_n}) - u(t_n, \mathbf{Y}_{t_n-})\} \right) &= E \left(\int_0^T \int_{\mathcal{E}} L_v^{-1} u(z, \boldsymbol{\eta}_{ja}(z-)) p_\varphi(dv, dz) \right) \\ &= E \left(\int_0^T \int_{\mathcal{E}} L_v^{-1} u(z, \boldsymbol{\eta}_{ja}(z)) \varphi(dv) dz \right). \end{aligned}$$

Therefore, we obtain

$$H \leq H_1 + H_2 + K \Delta^\beta, \quad (13.5.15)$$

where

$$\begin{aligned} H_1 = & \left| E \left(\sum_{n=1}^{n_T} \left[\left(u(t_n, \mathbf{Y}_{t_n-}) - u(t_n, \mathbf{Y}_{t_{n-1}}) \right) \right. \right. \right. \\ & \left. \left. \left. - \left(u(t_n, \mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}}) - u(t_n, \mathbf{Y}_{t_{n-1}}) \right) \right] \right) \right| \quad (13.5.16) \end{aligned}$$

and

$$\begin{aligned} H_2 = & \left| E \left(\int_0^T \int_{\mathcal{E}} \left[\left(L_v^{-1} u(z, \boldsymbol{\eta}_{ja}(z)) - L_v^{-1} u(z, \mathbf{Y}_{t_n}) \right) \right. \right. \right. \\ & \left. \left. \left. - \left(L_v^{-1} u(z, \mathbf{X}_z^{t_n, Y_{t_n}}) - L_v^{-1} u(z, \mathbf{Y}_{t_n}) \right) \right] \varphi(dv) dz \right) \right|. \quad (13.5.17) \end{aligned}$$

1. Let us note that by (12.3.7) the function u is smooth enough to apply the deterministic Taylor expansion. Therefore, by expanding the increments of u in H_1 we obtain

$$\begin{aligned}
H_1 = & \left| E \left(\sum_{n=1}^{n_T} \left\{ \left[\sum_{l=1}^{2\beta+1} \frac{1}{l!} \sum_{\mathbf{p} \in P_l} (\partial_y^{\mathbf{p}} u(t_n, \mathbf{Y}_{t_{n-1}})) F_{\mathbf{p}}(\mathbf{Y}_{t_n-} - \mathbf{Y}_{t_{n-1}}) + R_n(\mathbf{Y}_{t_n-}) \right] \right. \right. \right. \\
& - \left[\sum_{l=1}^{2\beta+1} \frac{1}{l!} \sum_{\mathbf{p} \in P_l} (\partial_y^{\mathbf{p}} u(t_n, \mathbf{Y}_{t_{n-1}})) F_{\mathbf{p}}(\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}} - \mathbf{Y}_{t_{n-1}}) \right. \\
& \left. \left. \left. + R_n(\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}}) \right] \right\} \right) \right|, \tag{13.5.18}
\end{aligned}$$

where the remainders are

$$\begin{aligned}
R_n(\mathbf{Z}) = & \frac{1}{2(\beta+1)!} \sum_{\mathbf{p} \in P_{2(\beta+1)}} \partial_y^{\mathbf{p}} u(t_n, \mathbf{Y}_{t_{n-1}} + \boldsymbol{\theta}_{\mathbf{p},n}(\mathbf{Z})(\mathbf{Z} - \mathbf{Y}_{t_{n-1}})) \\
& \times F_{\mathbf{p}}(\mathbf{Z} - \mathbf{Y}_{t_{n-1}}) \tag{13.5.19}
\end{aligned}$$

for $\mathbf{Z} = \mathbf{Y}_{t_n-}$ and $\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}}$, respectively. Here $\boldsymbol{\theta}_{\mathbf{p},n}(\mathbf{Z})$ is a $d \times d$ diagonal matrix with

$$\theta_{\mathbf{p},n}^{k,k}(\mathbf{Z}) \in (0, 1) \tag{13.5.20}$$

for $k \in \{1, 2, \dots, d\}$.

Therefore, we have

$$\begin{aligned}
H_1 \leq & E \left(\sum_{n=1}^{n_T} \left\{ \sum_{l=1}^{2\beta+1} \frac{1}{l!} \sum_{\mathbf{p} \in P_l} |\partial_y^{\mathbf{p}} u(t_n, \mathbf{Y}_{t_{n-1}})| \right. \right. \\
& \times \left| E(F_{\mathbf{p}}(\mathbf{Y}_{t_n-} - \mathbf{Y}_{t_{n-1}}) - F_{\mathbf{p}}(\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}} - \mathbf{Y}_{t_{n-1}})) \middle| \tilde{\mathcal{A}}_{t_{n-1}} \right| \\
& \left. \left. + E(|R_n(\mathbf{Y}_{t_n-})| \middle| \tilde{\mathcal{A}}_{t_{n-1}}) + E(|R_n(\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}})| \middle| \tilde{\mathcal{A}}_{t_{n-1}}) \right| \right).
\end{aligned}$$

By (13.5.19), the Cauchy-Schwarz inequality, (12.3.7), (13.5.20) and (13.5.11) we obtain

$$\begin{aligned}
& E(|R_n(\mathbf{Y}_{t_n-})| \middle| \tilde{\mathcal{A}}_{t_{n-1}}) \\
& \leq K \sum_{\mathbf{p} \in P_{2(\beta+1)}} \left[E(|\partial_y^{\mathbf{p}} u(t_n, \mathbf{Y}_{t_{n-1}} + \boldsymbol{\theta}_{\mathbf{p},n}(\mathbf{Y}_{t_n-})(\mathbf{Y}_{t_n-} - \mathbf{Y}_{t_{n-1}}))|^2 \middle| \tilde{\mathcal{A}}_{t_{n-1}}) \right]^{\frac{1}{2}} \\
& \quad \times \left[E(|F_{\mathbf{p}}(\mathbf{Y}_{t_n-} - \mathbf{Y}_{t_{n-1}})|^2 \middle| \tilde{\mathcal{A}}_{t_{n-1}}) \right]^{\frac{1}{2}}
\end{aligned}$$

$$\begin{aligned}
&\leq K \left[E \left(1 + |\mathbf{Y}_{t_{n-1}}|^{2r} + |\mathbf{Y}_{t_n-} - \mathbf{Y}_{t_{n-1}}|^{2r} \middle| \tilde{\mathcal{A}}_{t_{n-1}} \right) \right]^{\frac{1}{2}} \\
&\quad \times \left[E \left(|F_{\mathbf{p}}(\mathbf{Y}_{t_n-} - \mathbf{Y}_{t_{n-1}})|^2 \middle| \tilde{\mathcal{A}}_{t_{n-1}} \right) \right]^{\frac{1}{2}} \\
&\leq K (1 + |\mathbf{Y}_{t_{n-1}}|^{2r}) (t_n - t_{n-1})^{\beta+1},
\end{aligned} \tag{13.5.21}$$

using the estimate

$$E(|\mathbf{Y}_{t_n-}|^{2r} | \mathcal{A}_{t_{n-1}}) \leq K (1 + |\mathbf{Y}_{t_{n-1}}|^{2r}),$$

for every $n \in \{1, 2, \dots, n_T\}$.

In a similar way, by (13.5.19), the Cauchy-Schwarz inequality, (12.3.7), (13.5.20), (13.5.10) and (13.5.6) we have

$$\begin{aligned}
&E(|R_n(\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}})| | \tilde{\mathcal{A}}_{t_{n-1}}) \\
&\leq K \sum_{\mathbf{p} \in P_{2(\beta+1)}} \left[E \left(|\partial_y^{\mathbf{p}} u(t_n, \mathbf{Y}_{t_{n-1}} + \theta_{\mathbf{p}, n}(\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}})(\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}} - \mathbf{Y}_{t_{n-1}}))|^2 \middle| \tilde{\mathcal{A}}_{t_{n-1}} \right) \right]^{\frac{1}{2}} \\
&\quad \times \left[E \left(|F_{\mathbf{p}}(\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}} - \mathbf{Y}_{t_{n-1}})|^2 \middle| \tilde{\mathcal{A}}_{t_{n-1}} \right) \right]^{\frac{1}{2}} \\
&\leq K \left[E \left(1 + |\mathbf{Y}_{t_{n-1}}|^{2r} + |\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}} - \mathbf{Y}_{t_{n-1}}|^{2r} \middle| \tilde{\mathcal{A}}_{t_{n-1}} \right) \right]^{\frac{1}{2}} \\
&\quad \times \left[E \left(|\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}} - \mathbf{Y}_{t_{n-1}}|^{4(\beta+1)} \middle| \tilde{\mathcal{A}}_{t_{n-1}} \right) \right]^{\frac{1}{2}} \\
&\leq K (1 + |\mathbf{Y}_{t_{n-1}}|^{2r}) (t_n - t_{n-1})^{\beta+1}
\end{aligned} \tag{13.5.22}$$

for every $n \in \{1, \dots, n_T\}$. Here we have used the estimates

$$E(|\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}}|^{2r} | \mathcal{A}_{t_{n-1}}) \leq K (1 + |\mathbf{Y}_{t_{n-1}}|^{2r}),$$

and

$$\left| F_{\mathbf{p}}(\mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}} - \mathbf{Y}_{t_{n-1}}) \right|^2 \leq K \left| \mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}} - \mathbf{Y}_{t_{n-1}} \right|^{4(\beta+1)},$$

for every $n \in \{1, \dots, n_T\}$, with $\mathbf{p} \in P_{2(\beta+1)}$.

Finally, by applying the Cauchy-Schwarz inequality, (12.3.7), (13.5.8), (13.5.12), (13.5.21), (13.5.22) and (13.5.9) we obtain

$$\begin{aligned}
H_1 &\leq E \left(K \sum_{n=1}^{n_T} (1 + |\mathbf{Y}_{t_{n-1}}|^{2r}) \Delta^\beta (t_n - t_{n-1}) \right) \\
&\leq K \Delta^\beta \left(1 + E \left(\max_{0 \leq n \leq n_T} |\mathbf{Y}_{t_n}|^{2r} \right) \right) \\
&\leq K \Delta^\beta (1 + |\mathbf{Y}_0|^{2r}) \leq K \Delta^\beta.
\end{aligned} \tag{13.5.23}$$

2. Let us now estimate the term H_2 in (13.5.15). By the smoothness of the function u , see (12.3.7), and of the jump coefficient c , we can apply the deterministic Taylor formula and obtain

$$\begin{aligned}
H_2 &= \left| E \left(\int_0^T \int_{\mathcal{E}} \left\{ \left[\sum_{l=1}^{2\beta+1} \frac{1}{l!} \sum_{\mathbf{p} \in P_l} (\partial_y^{\mathbf{p}} L_v^{-1} u(z, \mathbf{Y}_{t_{n_z}})) F_{\mathbf{p}}(\boldsymbol{\eta}_{ja}(z) - \mathbf{Y}_{t_{n_z}}) \right. \right. \right. \right. \\
&\quad \left. \left. \left. \left. + R_{n_z}(\boldsymbol{\eta}_{ja}(z)) \right] \right. \right. \\
&\quad \left. \left. \left. \left. - \left[\sum_{l=1}^{2\beta+1} \frac{1}{l!} \sum_{\mathbf{p} \in P_l} (\partial_y^{\mathbf{p}} L_v^{-1} u(z, \mathbf{Y}_{t_{n_z}})) F_{\mathbf{p}}(\mathbf{X}_z^{t_{n_z}, Y_{t_{n_z}}} - \mathbf{Y}_{t_{n_z}}) \right. \right. \right. \right. \\
&\quad \left. \left. \left. \left. + R_{n_z}(\mathbf{X}_z^{t_{n_z}, Y_{t_{n_z}}}) \right] \right\} \varphi(dv) dz \right) \right| \\
&\leq \int_0^T \int_{\mathcal{E}} E \left(\sum_{l=1}^{2\beta+1} \frac{1}{l!} \sum_{\mathbf{p} \in P_l} \left| \partial_y^{\mathbf{p}} L_v^{-1} u(z, \mathbf{Y}_{t_{n_z}}) \right| \right. \\
&\quad \times \left| E \left(F_{\mathbf{p}}(\boldsymbol{\eta}_{ja}(z) - \mathbf{Y}_{t_{n_z}}) - F_{\mathbf{p}}(\mathbf{X}_z^{t_{n_z}, Y_{t_{n_z}}} - \mathbf{Y}_{t_{n_z}}) \middle| \tilde{\mathcal{A}}_{t_{n_z}} \right) \right| \tag{13.5.24} \\
&\quad \left. + E \left(\left| R_{n_z}(\boldsymbol{\eta}_{ja}(z)) \right| \middle| \tilde{\mathcal{A}}_{t_{n_z}} \right) + E \left(\left| R_{n_z}(\mathbf{X}_z^{t_{n_z}, Y_{t_{n_z}}}) \right| \middle| \tilde{\mathcal{A}}_{t_{n_z}} \right) \right) \varphi(dv) dz.
\end{aligned}$$

We can estimate the remainders, as in (13.5.21) and (13.5.22), for every $z \in [0, T]$ by

$$E \left(\left| R_{n_z}(\boldsymbol{\eta}_{ja}(z)) \right| \middle| \tilde{\mathcal{A}}_{t_{n_z}} \right) \leq K (1 + |\mathbf{Y}_{t_{n_z}}|^{2r}) (z - t_{n_z})^{\beta+1} \tag{13.5.25}$$

and

$$E \left(\left| R_{n_z}(\mathbf{X}_z^{t_{n_z}, Y_{t_{n_z}}}) \right| \middle| \tilde{\mathcal{A}}_{t_{n_z}} \right) \leq K (1 + |\mathbf{Y}_{t_{n_z}}|^{2r}) (z - t_{n_z})^{\beta+1}. \tag{13.5.26}$$

Then, by applying the Cauchy-Schwarz inequality, the polynomial growth conditions on $u(t, \cdot)$, see (12.3.7), and on the jump coefficient c , (13.5.20), (13.5.12), (13.5.25), (13.5.26) and (13.5.9) to the estimate (13.5.24), we obtain

$$\begin{aligned}
H_2 &\leq K \int_0^T \int_{\mathcal{E}} E\left(1 + |\mathbf{Y}_{t_{n_z}}|^{2r}\right) \Delta^\beta(z - t_{n_z}) \varphi(dv) dz \\
&\leq K \Delta^\beta \int_0^T E\left(1 + \max_{0 \leq n \leq n_T} |\mathbf{Y}_{t_n}|^{2r}\right) (z - t_{n_z}) dz \\
&\leq K \Delta^\beta.
\end{aligned} \tag{13.5.27}$$

3. Finally, by (13.5.15), (13.5.23) and (13.5.27) we can summarize to obtain

$$\left| E(g(\mathbf{Y}_{t_{n_T}})) - E(g(\mathbf{X}_T)) \right| \leq K \Delta^\beta. \quad \square$$

13.6 Convergence of Jump-Adapted Weak Schemes

Jump-Adapted Weak Approximations

In this section we present a convergence theorem for general jump-adapted weak approximations of any weak order of convergence $\beta \in \{1, 2, \dots\}$. This theorem requires certain conditions to be satisfied by the increments of a given discrete-time approximation in order to obtain a jump-adapted approximation of weak order $\beta \in \{1, 2, \dots\}$. It covers the weak convergence of the schemes presented previously in this chapter.

Let us consider a jump-adapted discrete-time approximation $\mathbf{Y}^\Delta = \{\mathbf{Y}_{t_n}^\Delta, n \in \{0, 1, \dots, n_T\}\}$ corresponding to a jump-adapted time discretization with maximum step size $\Delta \in (0, 1)$. We simulate the jump impact as before by

$$\mathbf{Y}_{t_{n+1}}^\Delta = \mathbf{Y}_{t_{n+1}-}^\Delta + \int_{\mathcal{E}} \mathbf{c}(t_n, \mathbf{Y}_{t_{n+1}-}^\Delta, v) p_\varphi(dv, \{t_{n+1}\}), \tag{13.6.1}$$

for $n \in \{0, 1, \dots, n_T - 1\}$. Note that we use here the same notation \mathbf{Y}^Δ as applied in (13.5.2)–(13.5.3) for the jump-adapted weak Taylor scheme. However, the jump-adapted approximation considered here is more general since we do not specify its evolution between discretization points. The scheme (13.5.2)–(13.5.3) is a special case of the approximation considered here. The theorem below will state the conditions for obtaining weak order of convergence $\beta \in \{1, 2, \dots\}$.

First, let us formulate an important condition on the evolution of \mathbf{Y}^Δ . Define a stochastic process $\boldsymbol{\eta}_{ja} = \{\boldsymbol{\eta}_{ja}(t), t \in [0, T]\}$ such that for every $n \in \{0, \dots, n_T\}$

$$\boldsymbol{\eta}_{ja}(t_n) = \mathbf{Y}_{t_n} \tag{13.6.2}$$

and

$$\boldsymbol{\eta}_{ja}(t_n-) = \mathbf{Y}_{t_n-}. \tag{13.6.3}$$

Assume that the process $\boldsymbol{\eta}_{ja} = \{\boldsymbol{\eta}_{ja}(t), t \in [0, T]\}$ satisfies Lemmas 13.5.4 with respect to the initial value \mathbf{Y}_0 of the general jump-adapted approximation. Moreover, we require that the process $\boldsymbol{\eta}_{ja}$ satisfies also Lemma 13.5.5 and Lemma 13.5.6, where \mathbf{Y} is again the general jump-adapted approximation under consideration.

Weak Convergence Theorem

Now we can formulate the following weak convergence theorem:

Theorem 13.6.1. *Let us assume that $E(|\mathbf{X}_0|^i) < \infty$ for $i \in \{1, 2, \dots\}$, and that \mathbf{Y}_0^Δ converges weakly to \mathbf{X}_0 with order $\beta \in \{1, 2, \dots\}$. Suppose that the drift, diffusion and jump coefficients $a^k, b^{k,j}, c^k$, respectively, belong to the space $C_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$, for all $k \in \{1, 2, \dots, d\}$ and $j \in \{1, 2, \dots, m\}$, and the coefficients f_α , with $f(t, \mathbf{x}) = \mathbf{x}$, satisfy the linear growth condition $|f_\alpha(t, \mathbf{y})| \leq K(1 + |\mathbf{y}|)$, with $K < \infty$, for all $t \in [0, T]$, $\mathbf{y} \in \mathbb{R}^d$, and $\alpha \in \hat{\Gamma}_\beta$.*

Moreover, assume that for each $p \in \{1, 2, \dots\}$ there exist constants $K < \infty$ and $r \in \{1, 2, \dots\}$, which do not depend on Δ , such that for each $q \in \{1, 2, \dots, p\}$

$$E \left(\max_{0 \leq n \leq n_T} |\mathbf{Y}_{t_n-}^\Delta|^{2q} | \mathcal{A}_0 \right) \leq K(1 + |\mathbf{Y}_0^\Delta|^{2r}), \quad (13.6.4)$$

$$E \left(|\mathbf{Y}_{t_{n+1}-}^\Delta - \mathbf{Y}_{t_n}^\Delta|^{2q} | \tilde{\mathcal{A}}_{t_n} \right) \leq K \left(1 + \max_{0 \leq k \leq n} |\mathbf{Y}_{t_k}^\Delta|^{2r} \right) (t_{n+1} - t_n)^q \quad (13.6.5)$$

for $n \in \{0, 1, \dots, n_T - 1\}$, and

$$\begin{aligned} & \left| E \left(\prod_{h=1}^l (Y_{t_{n+1}}^{\Delta, p_h} - Y_{t_n}^{\Delta, p_h}) - \prod_{h=1}^l \left(\sum_{\alpha \in \Gamma_\beta \setminus \{v\}} I_\alpha \left(f_\alpha^{p_h}(t_n, \mathbf{Y}_{t_n}^\Delta) \right)_{t_n, t_{n+1}} \right) \middle| \tilde{\mathcal{A}}_{t_n} \right) \right| \\ & \leq K(1 + \max_{0 \leq k \leq n_T} |\mathbf{Y}_{t_k}^\Delta|^{2r}) \Delta^\beta (t_{n+1} - t_n) \end{aligned} \quad (13.6.6)$$

for all $n \in \{0, 1, \dots, n_T - 1\}$ and $(p_1, \dots, p_l) \in \{1, 2, \dots, d\}^l$, where $l \in \{1, \dots, 2\beta + 1\}$ and Y^{Δ, p_h} denotes the p_h th component of \mathbf{Y}^Δ .

Then for any function $g \in C_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ there exists a positive constant C , independent of Δ , such that

$$|E(g(\mathbf{X}_T)) - E(g(\mathbf{Y}_{t_{n_T}}^\Delta))| \leq C \Delta^\beta. \quad (13.6.7)$$

Theorem 13.6.1 formulates sufficient conditions on a discrete-time approximation so that it approximates the diffusion component of a jump-adapted scheme. The most important condition is the estimate (13.6.6). It requires the first $2\beta + 1$ conditional moments of the increments of the diffusion approximation to be sufficiently close to those of the truncated Wagner-Platen expansion for pure diffusions.

Proof of Weak Convergence

To prove Theorem 13.6.1, let us for $n \in \{1, \dots, n_T\}$ and $\mathbf{y} \in \mathbb{R}^d$ define

$$\eta_{n,\beta}^y = \mathbf{y} + \sum_{\alpha \in \hat{\Gamma}_\beta \setminus \{v\}} I_\alpha \left(f_\alpha(t_n, \mathbf{y}) \right)_{t_{n-1}, t_n}. \quad (13.6.8)$$

Proof of Theorem 13.6.1: Note that in the following we will write \mathbf{Y} for \mathbf{Y}^Δ . By (12.3.3) and (12.3.6) we obtain

$$\begin{aligned} H &= \left| E(g(\mathbf{Y}_{t_{n_T}})) - E(g(\mathbf{X}_T)) \right| \\ &= \left| E(u(T, \mathbf{Y}_{t_{n_T}}) - u(0, \mathbf{X}_0)) \right|. \end{aligned} \quad (13.6.9)$$

Moreover, since \mathbf{Y}_0 converges weakly with order β to \mathbf{X}_0 we obtain

$$\begin{aligned} H &\leq \left| E \left(\sum_{n=1}^{n_T} (u(t_n, \mathbf{Y}_{t_n}) - u(t_n, \mathbf{Y}_{t_n-}) + u(t_n, \mathbf{Y}_{t_n-}) - u(t_{n-1}, \mathbf{Y}_{t_{n-1}})) \right) \right| \\ &\quad + K \Delta^\beta. \end{aligned} \quad (13.6.10)$$

By (13.5.4), the definition of $\boldsymbol{\eta}_{ja}$ and (13.6.2)–(13.6.3), we obtain

$$\begin{aligned} H &\leq \left| E \left(\sum_{n=1}^{n_T} \left[\{u(t_n, \mathbf{Y}_{t_n}) - u(t_n, \mathbf{Y}_{t_n-}) + u(t_n, \mathbf{Y}_{t_n-}) - u(t_{n-1}, \mathbf{Y}_{t_{n-1}})\} \right. \right. \right. \\ &\quad \left. \left. \left. - \{u(t_n, \mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}}) - u(t_{n-1}, \mathbf{Y}_{t_{n-1}}) \right. \right. \right. \\ &\quad \left. \left. \left. + \int_{t_{n-1}}^{t_n} \int_{\mathcal{E}} L_v^{-1} u(z, \mathbf{X}_z^{t_{n-1}, Y_{t_{n-1}}}) \varphi(dv) dz \} \right] \right) \right| + K \Delta^\beta \\ &\leq \left| E \left(\int_0^T \int_{\mathcal{E}} \left[L_v^{-1} u(z, \boldsymbol{\eta}_{ja}(z)) - L_v^{-1} u(z, \mathbf{X}_z^{t_{n_z}, Y_{t_{n_z}}}) \right] \varphi(dv) dz \right. \right. \\ &\quad \left. \left. + \sum_{n=1}^{n_T} \left[(u(t_n, \mathbf{Y}_{t_n-}) - u(t_n, \mathbf{Y}_{t_{n-1}})) - (u(t_n, \mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}}) - u(t_n, \mathbf{Y}_{t_{n-1}})) \right] \right) \right| \\ &\quad + K \Delta^\beta \\ &\leq H_1 + H_2 + H_3 + K \Delta^\beta, \end{aligned} \quad (13.6.11)$$

where

$$\begin{aligned} H_1 &= \left| E \left(\sum_{n=1}^{n_T} \left[(u(t_n, \mathbf{Y}_{t_n-}) - u(t_n, \mathbf{Y}_{t_{n-1}})) \right. \right. \right. \\ &\quad \left. \left. \left. - (u(t_n, \boldsymbol{\eta}_{n,\beta}^{Y_{t_{n-1}}}) - u(t_n, \mathbf{Y}_{t_{n-1}})) \right] \right) \right|, \end{aligned} \quad (13.6.12)$$

$$H_2 = \left| E \left(\sum_{n=1}^{n_T} \left[\left(u(t_n, \boldsymbol{\eta}_{n,\beta}^{Y_{t_{n-1}}}) - u(t_n, \mathbf{Y}_{t_{n-1}}) \right) \right. \right. \right. \\ \left. \left. \left. - \left(u(t_n, \mathbf{X}_{t_n-}^{t_{n-1}, Y_{t_{n-1}}}) - u(t_n, \mathbf{Y}_{t_{n-1}}) \right) \right] \right) \right| \quad (13.6.13)$$

and

$$H_3 = \left| E \left(\int_0^T \int_{\mathcal{E}} \left[\left(L_v^{-1} u(z, \boldsymbol{\eta}_{ja}(z)) - L_v^{-1} u(z, \mathbf{Y}_{t_n z}) \right) \right. \right. \right. \\ \left. \left. \left. - \left(L_v^{-1} u(z, \mathbf{X}_z^{t_n z, Y_{t_n z}}) - L_v^{-1} u(z, \mathbf{Y}_{t_n z}) \right) \right] \varphi(dv) dz \right) \right|. \quad (13.6.14)$$

By applying the deterministic Taylor expansion to the increments in H_1 and H_2 , we obtain the following estimate

$$H_i \leq K \Delta^\beta \quad (13.6.15)$$

for $i \in \{1, 2\}$, see Kloeden & Platen (1999).

The estimate of H_3 follows as in the estimate (13.5.24) of Theorem 13.5.1, since $\boldsymbol{\eta}_{ja}$ satisfies Lemmas 13.5.4, 13.5.5 and 13.5.6. This completes the proof of Theorem 13.6.1. \square

Simplified and Predictor-Corrector Schemes

Theorem 13.6.1 can be used to establish the order of weak convergence of the jump-adapted approximations presented in the current chapter. To do so, we should first construct a stochastic process η_{ja} which satisfies properties (13.6.2)–(13.6.3) and Lemmas 13.5.4, 13.5.5 and 13.5.6. Then, we have to show that the increments of the approximation between discretization points satisfy conditions (13.6.5) and (13.6.6). Finally, we need to check the regularity of the approximation with condition (13.6.4).

Let us consider, for example, the jump-adapted predictor-corrector Euler scheme (13.3.1)–(13.3.2). We can define the following stochastic process $\eta_{ja} = \{\eta_{ja}(t), t \in [0, T]\}$ by

$$\eta_{ja}(t) = \eta_{ja}(t_n) + \frac{1}{2} \left\{ a(\xi_{ja}(t)) + a(\eta_{ja}(t_n)) \right\} (t - t_n) + b(\eta_{ja}(t_n)) (W_t - W_{t_n}) \\ + \int_{(t_n, t]} \int_{\mathcal{E}} c(z, \eta_{ja}(z-), v) p_\varphi(dv, dz), \quad (13.6.16)$$

and

$$\xi_{ja}(t) = \xi_{ja}(t_n) + a(\eta_{ja}(t_n)) (t - t_n) + b(\eta_{ja}(t_n)) (W_t - W_{t_n}) \quad (13.6.17)$$

for $n \in \{0, \dots, n_T - 1\}$ and $t \in (t_n, t_{n+1}]$, with $\eta_{ja}(0) = Y_0$. By construction, η_{ja} satisfies conditions (13.6.2)–(13.6.3). Conditions (13.6.4) and (13.6.5) can be obtained using the linear growth conditions (1.9.3) on the coefficients a and b , the Wagner-Platen expansion and the properties of the increments of Wiener processes. Furthermore, condition (13.6.6) holds for $\beta = 1.0$, as in the case of pure diffusion SDEs, see Kloeden & Platen (1999). Finally, one has to show that the process η_{ja} defined in (13.6.16) satisfies Lemmas 13.5.4, 13.5.5 and 13.5.6 with $\beta = 1.0$. In doing so, the weak order of convergence $\beta = 1.0$ of the jump-adapted predictor-corrector Euler scheme is established.

The order of weak convergence of other jump-adapted predictor-corrector and derivative-free schemes, as presented in this chapter, can be established in a similar fashion. Also the general multi-dimensional case can be handled similarly.

To show the order of weak convergence of simplified jump-adapted schemes, based on multi-point distributed random variables, one has to construct a process η_{ja} satisfying conditions (13.6.2)–(13.6.3). For example, when considering the jump-adapted simplified Euler scheme (13.1.7) in the one-dimensional case, we can define

$$\begin{aligned} \eta_{ja}(t) &= \eta_{ja}(t_n) + a(\eta_{ja}(t_n))(t - t_n) + b(\eta_{ja}(t_n))\sqrt{t - t_n} \operatorname{sgn}(W_t - W_{t_n}) \\ &\quad + \int_{(t_n, t]} \int_{\mathcal{E}} c(z, \eta_{ja}(z-), v) p_\varphi(dv, dz), \end{aligned} \tag{13.6.18}$$

for $n \in \{0, \dots, n_T - 1\}$ and $t \in (t_n, t_{n+1}]$, with $\eta_{ja}(0) = Y_0$, where we have used the sign function

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ -1 & \text{for } x < 0. \end{cases} \tag{13.6.19}$$

Then the process η_{ja} satisfies conditions (13.6.2)–(13.6.3) and the Lemmata 13.5.4, 13.5.5 and 13.5.6 with $\beta = 1$.

To establish the order of weak convergence of the jump-adapted simplified weak order 2.0 scheme (13.1.10) in the one-dimensional case, we define

$$\begin{aligned} \eta_{ja}(t) &= \eta_{ja}(t_n) + a(\eta_{ja}(t_n))(t - t_n) + b(\eta_{ja}(t_n))\omega_n(t) \\ &\quad + \frac{b(\eta_{ja}(t_n))b'(\eta_{ja}(t_n))}{2} \left((\omega_n(t))^2 - (t - t_n) \right) \\ &\quad + \frac{1}{2} \left(a(\eta_{ja}(t_n))a'(\eta_{ja}(t_n)) + \frac{1}{2}a''(\eta_{ja}(t_n))(b(\eta_{ja}(t_n)))^2 \right) (t - t_n)^2 \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{2} \left(a'(\eta_{ja}(t_n)) b(\eta_{ja}(t_n)) + a(\eta_{ja}(t_n)) b'(\eta_{ja}(t_n)) \right. \\
& \quad \left. + \frac{1}{2} b''(\eta_{ja}(t_n)) (b(\eta_{ja}(t_n)))^2 \right) \omega_n(t)(t - t_n) \\
& + \int_{(t_n, t]} \int_{\mathcal{E}} c(z, \eta_{ja}(z-), v) p_\varphi(dv, dz),
\end{aligned} \tag{13.6.20}$$

where

$$\omega_n(t) = \sqrt{t - t_n} \left(\mathbf{1}_{\{W_t - W_{t_n} < N^{-1}(\frac{1}{6}) \sqrt{t - t_n}\}} - \mathbf{1}_{\{W_t - W_{t_n} > N^{-1}(\frac{5}{6}) \sqrt{t - t_n}\}} \right) \tag{13.6.21}$$

for $n \in \{0, \dots, n_T - 1\}$ and $t \in (t_n, t_{n+1}]$, with $\eta_{ja}(0) = Y_0$. In (13.6.21) we have denoted by $\mathbf{1}$ the indicator function defined in (1.8.8) and by $N^{-1}(x)$ the inverse function of the distribution function of a standard Gaussian random variable. Therefore, the process η_{ja} satisfies conditions (13.6.2)–(13.6.3) and the Lemmata 13.5.4, 13.5.5 and 13.5.6 with $\beta = 2.0$. In a similar way one can use Theorem 13.6.1 to construct simplified weak schemes with a higher order of weak convergence including the general multi-dimensional case, as for the third order weak schemes presented in this chapter.

Finally, we remark that for a jump-adapted weak Euler scheme for an SDE with jumps where the drift and diffusion coefficients are only Hölder continuous, one can still derive some weak order of convergence. As shown in Kubilius & Platen (2002) the corresponding weak order β depends on the degree of Hölder continuity and is then typically below one.

13.7 Numerical Results on Weak Schemes

Computing Time

Now, we present some numerical results in the diffusion case for the Euler and the weak order 2.0 Taylor schemes, as well as for their simplified versions. The case with jumps via jump-adapted schemes is, in principle, identical. As a test case we choose an SDE with multiplicative noise, where

$$dX_t = a X_t dt + \sigma X_t dW_t \tag{13.7.1}$$

for $t \in [0, T]$ and $X_0 \in \mathfrak{R}$. Recall from Chap. 1 that the SDE admits the closed form solution

$$X_T = X_0 \exp \left\{ \left(a - \frac{\sigma^2}{2} \right) T + \sigma W_T \right\}. \tag{13.7.2}$$

When we analyze the central processor unit (CPU) time needed to run a Monte Carlo simulation that computes the payoff of a call option with 400000 paths and 128 time steps on a standard desktop machine, we report a CPU

time of 22 and 23.5 seconds for the Euler scheme and the weak order 2.0 Taylor scheme, respectively. The corresponding simplified versions only require 0.75 and 4.86 seconds, respectively. Thus, for the Euler method the simplified version is about 29 times faster than the Gaussian one. The simplified weak order 2.0 scheme is nearly five times faster than the weak order 2.0 Taylor scheme.

Accuracy for Smooth Payoff

In the following we will discuss the accuracy of the Euler scheme, the weak order 2.0 Taylor scheme and their simplified versions when applied to the test SDE (13.7.1). As in Chap. 11, we show several log-log plots with the logarithm $\log_2(\mu(\Delta))$ of the weak error, as defined in (11.1.2), versus the logarithm $\log_2(\Delta)$ of the maximum time step size Δ .

We first study the computation of the functional $E(g(X_T))$, where g is a smooth function and X_T is the solution of the SDE (13.7.1). Note that the estimation of the expected value of a smooth function of the solution X arises, for instance, when computing expected utilities. Another important application is the calculation of Value at Risk, via the simulation of moments, as applied in Edgeworth expansions and saddle point methods in Studer (2001). Therefore, we consider here the estimation of the k th moment $E((X_T)^k)$ of X_T at time T , for $k \in \mathcal{N}$.

By (13.7.2) we obtain the theoretical k th moment of X_T in closed form as

$$E\left((X_T)^k\right) = (X_0)^k \exp\left\{\left(kT(a + \frac{\sigma^2}{2}(k-1))\right)\right\} \quad (13.7.3)$$

for $k \in \mathcal{N}$, which we can use for comparison.

The particular structure of the SDE (13.7.1) allows us to obtain a closed form solution also for the estimator of the k th moment provided by the Euler and the weak order 2.0 Taylor schemes, and by their simplified versions.

Let us rewrite the Euler scheme (11.2.1) for the SDE (13.7.1) as

$$Y_T = Y_0 \prod_{n=0}^{N-1} \left(1 + \frac{aT}{N} + \sigma \Delta W_n\right), \quad (13.7.4)$$

where ΔW_n are i.i.d Gaussian $N(0, \frac{T}{N})$ random variables and $N = \frac{T}{\Delta}$. Note that

$$E\left((\Delta W_n)^k\right) = \begin{cases} \frac{k!}{(k/2)!2^{k/2}} \left(\frac{T}{N}\right)^{k/2} & \text{for } k \text{ even} \\ 0 & \text{for } k \text{ odd.} \end{cases} \quad (13.7.5)$$

By the independence of ΔW_n , for $n \in \{0, 1, \dots, N-1\}$ and (13.7.5) we obtain

$$\begin{aligned}
E((Y_T)^k) &= (Y_0)^k E\left(\prod_{n=0}^{N-1} \left(1 + \frac{aT}{N} + \sigma \Delta W_n\right)^k\right) \\
&= (Y_0)^k \prod_{n=0}^{N-1} E\left(\left(1 + \frac{aT}{N} + \sigma \Delta W_n\right)^k\right) \\
&= (Y_0)^k \prod_{n=0}^{N-1} \sum_{i=0}^k \binom{k}{i} \left(1 + \frac{aT}{N}\right)^{k-i} \sigma^i E\left((\Delta W_n)^i\right) \\
&= (Y_0)^k \left(\sum_{q=0}^{[k/2]} \binom{k}{2q} \left(1 + \frac{aT}{N}\right)^{k-2q} \frac{(2q)!}{q!} \left(\frac{\sigma^2 T}{2N}\right)^q \right)^N. \quad (13.7.6)
\end{aligned}$$

We recall that by $[z]$ we denote the integer part of $z \in \mathbb{R}$ and by $\binom{i}{l}$, for $i \geq l$, the combinatorial coefficient, see (4.1.24).

Similarly, for the simplified Euler scheme (11.2.2) we obtain

$$\begin{aligned}
E((Y_T)^k) &= (Y_0)^k E\left(\prod_{n=0}^{N-1} \left(1 + \frac{aT}{N} + \sigma \Delta \hat{W}_{2,n}\right)^k\right) \\
&= (Y_0)^k \left(\sum_{q=0}^{[k/2]} \binom{k}{2q} \left(1 + \frac{aT}{N}\right)^{k-2q} \left(\frac{\sigma^2 T}{N}\right)^q \right)^N, \quad (13.7.7)
\end{aligned}$$

where $\Delta \hat{W}_{2,n}$ is the two-point distributed random variable given in (11.2.4) or (13.1.6). Here we have used the result

$$E\left(\left(\Delta \hat{W}_{2,n}\right)^k\right) = \begin{cases} \left(\frac{T}{N}\right)^{k/2} & \text{for } k \text{ even} \\ 0 & \text{for } k \text{ odd.} \end{cases} \quad (13.7.8)$$

By comparing (13.7.6) to (13.7.7), we notice that the Euler scheme and the simplified Euler scheme give the same estimator for the expectation $E((Y_T)^k)$ with $k \in \{1, 2, 3\}$.

Let us now rewrite the weak order 2.0 Taylor scheme (11.2.6) for the SDE (13.7.1) as

$$Y_T = Y_0 \prod_{n=0}^{N-1} \left(h_1 + h_2 \Delta W_n + h_3 (\Delta W_n)^2 \right), \quad (13.7.9)$$

with

$$h_1 = 1 + \left(a - \frac{\sigma^2}{2}\right) \frac{T}{N} + \frac{a^2}{2} \frac{T^2}{N^2}, \quad h_2 = \sigma + a\sigma \frac{T}{N} \quad \text{and} \quad h_3 = \frac{\sigma^2}{2}. \quad (13.7.10)$$

By the independence of the random increments ΔW_n , for $n \in \{0, 1, \dots, N-1\}$, and (13.7.5) we obtain

$$\begin{aligned}
E \left((Y_T)^k \right) &= (Y_0)^k \prod_{n=0}^{N-1} E \left(\left(h_1 + h_2 \Delta W_n + h_3 (\Delta W_n)^2 \right)^k \right) \\
&= (Y_0)^k \prod_{n=0}^{N-1} \sum_{i=0}^k \binom{k}{i} h_1^{k-i} E \left(\left(h_2 \Delta W_n + h_3 (\Delta W_n)^2 \right)^i \right) \\
&= (Y_0)^k \prod_{n=0}^{N-1} \sum_{i=0}^k \binom{k}{i} h_1^{k-i} \sum_{j=0}^i \binom{i}{j} h_2^{i-j} h_3^j E \left((\Delta W_n)^{i+j} \right) \\
&= (Y_0)^k \prod_{n=0}^{N-1} \left(\sum_{q=0}^{[k/2]} \binom{k}{2q} h_1^{k-2q} \sum_{l=0}^q \binom{2q}{2l} h_2^{2(q-l)} h_3^{2l} E \left((\Delta W_n)^{2(q+l)} \right) \right. \\
&\quad \left. + \sum_{q=0}^{[(k-1)/2]} \binom{k}{2q+1} h_1^{k-(2q+1)} \sum_{l=0}^q \binom{2q+1}{2l+1} h_2^{2(q-l)} h_3^{2l+1} \right. \\
&\quad \left. \times E \left((\Delta W_n)^{2(q+l+1)} \right) \right) \\
&= (Y_0)^k \left(\sum_{q=0}^{[k/2]} \binom{k}{2q} h_1^{k-2q} \sum_{l=0}^q \binom{2q}{2l} h_2^{2(q-l)} h_3^{2l} \frac{(2(q+l))!}{(q+l)!} \left(\frac{T}{2N} \right)^{q+l} \right. \\
&\quad \left. + \sum_{q=0}^{[(k-1)/2]} \binom{k}{2q+1} h_1^{k-(2q+1)} \sum_{l=0}^q \binom{2q+1}{2l+1} h_2^{2(q-l)} h_3^{2l+1} \right. \\
&\quad \left. \times \frac{(2(q+l+1))!}{(q+l+1)!} \left(\frac{T}{2N} \right)^{q+l+1} \right)^N. \tag{13.7.11}
\end{aligned}$$

For the simplified weak order 2.0 Taylor scheme (11.2.7), with the three-point distributed random variable $\Delta \hat{W}_{3,n}$, see (11.2.9) or (13.1.11), we obtain

$$\begin{aligned}
E \left((Y_T)^k \right) &= (Y_0)^k \prod_{n=0}^{N-1} E \left(\left(h_1 + h_2 \Delta \hat{W}_{3,n} + h_3 (\Delta \hat{W}_{3,n})^2 \right)^k \right) \\
&= (Y_0)^k \left(h_1^k + \sum_{q=1}^{[k/2]} \binom{k}{2q} h_1^{k-2q} \sum_{l=0}^q \binom{2q}{2l} h_2^{2(q-l)} h_3^{2l} \frac{1}{3} \left(\frac{3T}{N} \right)^{q+l} \right)
\end{aligned}$$

$$\begin{aligned}
& + \sum_{q=0}^{[(k-1)/2]} \binom{k}{2q+1} h_1^{k-(2q+1)} \sum_{l=0}^q \binom{2q+1}{2l+1} h_2^{2(q-l)} h_3^{2l+1} \\
& \times \left(\frac{1}{3} \left(\frac{3T}{N} \right)^{q+l+1} \right)^N,
\end{aligned} \tag{13.7.12}$$

where we have used the result

$$E \left(\left(\Delta \hat{W}_{3,n} \right)^k \right) = \begin{cases} \frac{1}{3} \left(\frac{3T}{N} \right)^{k/2} & \text{for } k \text{ even} \\ 0 & \text{for } k \text{ odd,} \end{cases} \tag{13.7.13}$$

for $k \in \{1, 2, \dots\}$. Therefore, the weak order 2.0 Taylor scheme and the simplified weak order 2.0 Taylor scheme provide the same estimate for $E((Y_T)^k)$ with $k \in \{1, 2\}$.

In the following, we consider the estimation of the fifth moment, that is $E((Y_T)^5)$. We choose the following parameters: $X_0 = 1$, $a = 0.1$, $\sigma = 0.15$, $T = 1$. Note that by comparing the closed form solution (13.7.3) with the estimators (13.7.6), (13.7.11), (13.7.7) and (13.7.12), we can compute explicitly the weak error $\mu(\Delta)$ for the Euler and weak order 2.0 Taylor schemes, as well as for their simplified versions.

In Fig. 13.7.1 we show the logarithm $\log_2(\mu(\Delta))$ of the weak error for the Euler and the weak order 2.0 Taylor schemes together with their simplified versions versus the logarithm $\log_2(\Delta)$ of the time step size. Note that the slope of the log-log plot for each scheme corresponds to the theoretically predicted weak order, being $\beta = 1.0$ for both the Euler and simplified Euler, and $\beta = 2.0$ for both the weak order 2.0 Taylor and simplified weak order 2.0 Taylor schemes, as expected. The errors generated by the simplified schemes are very similar to those of the corresponding weak Taylor schemes.

We know that simplified schemes achieve the same order of weak convergence as their counterparts that are based on Gaussian random variables. However, it is useful to check whether there is a significant loss in accuracy when the time step size Δ is relatively large.

Relative Weak Error

We will now present some plots of the relative weak error, that is

$$\hat{\mu}(\Delta) = \left| \frac{E(g(X_T)) - E(g(Y_T))}{E(g(X_T))} \right|, \tag{13.7.14}$$

for fixed time step size Δ .

More precisely, in the following we analyze the relative weak error when estimating the fifth moment in our example. This means that we consider the quantity

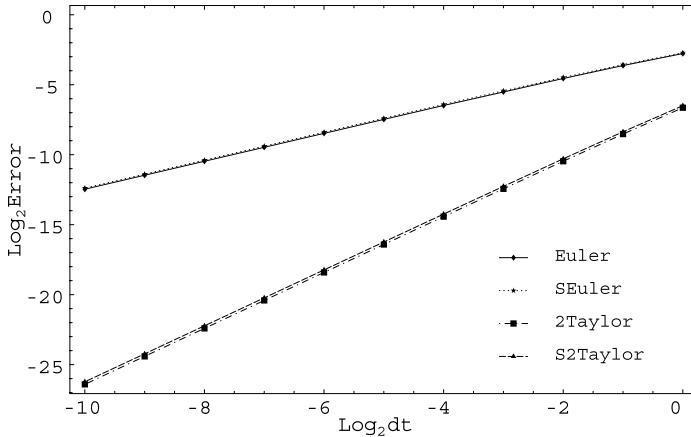


Fig. 13.7.1. Log-log plot of the weak error for the Euler, the simplified Euler, the weak order 2.0 Taylor and the simplified weak order 2.0 Taylor schemes

$$\left| \frac{E((X_T)^5) - E((Y_T)^5)}{E((X_T)^5)} \right|, \quad (13.7.15)$$

where the parameters are set as before with $T \in [1/365, 3]$ and $\sigma \in [0.05, 0.5]$. Moreover, we use only one time step, which means that the time step size Δ used for the discrete-time approximation Y_T equals T . In Fig. 13.7.2 we report the relative error generated by the Euler scheme. For small values of the time to maturity T and volatility σ , the Euler scheme is very precise, even when using only one time step as considered here. For instance, when the maturity time is set to 2/12 and the volatility to 0.1 we obtain a relative error of 0.13%. When the time to maturity and the volatility increase, the accuracy of the Euler scheme is not satisfactory. For instance, for $T = 3$ and $\sigma = 0.5$ the relative weak error amounts to 99.6%. In Fig. 13.7.3 we report the relative weak error generated by the simplified Euler scheme for the same parameter values as above. The results are similar to those obtained in Fig. 13.7.2 for the Euler scheme based on Gaussian random variables. Finally, Fig. 13.7.4 reports the difference between the relative errors generated by the Euler and simplified Euler schemes. Here we notice that the loss in accuracy due to the use of simplified schemes does not exceed 4%.

The relative errors generated by the weak order 2.0 Taylor scheme and its simplified version are significantly smaller than those reported for the Euler and simplified Euler schemes. However, the qualitative behavior with respect to the values of the time to maturity T , and volatility σ is similar. In Fig. 13.7.5 we report the relative error of the simplified weak order 2.0 scheme minus the relative error of the weak order 2.0 Taylor scheme for the same parameters considered in the previous plots. Also in this case the loss in accuracy generated by the multi-point distributed random variables is limited for all parameter values tested.

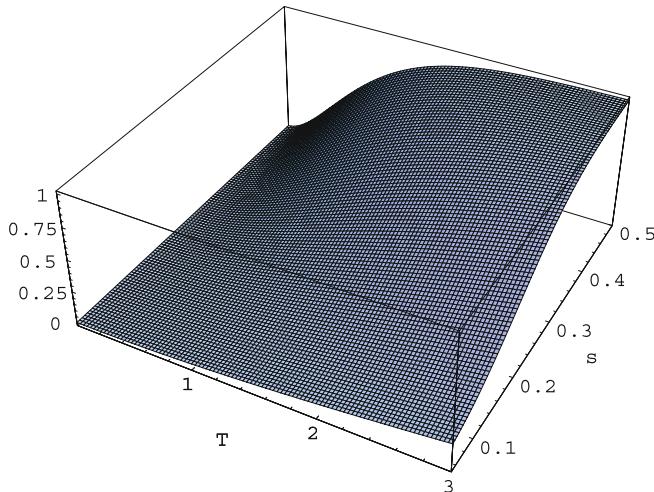


Fig. 13.7.2. Relative error for Euler scheme with $\Delta = T$

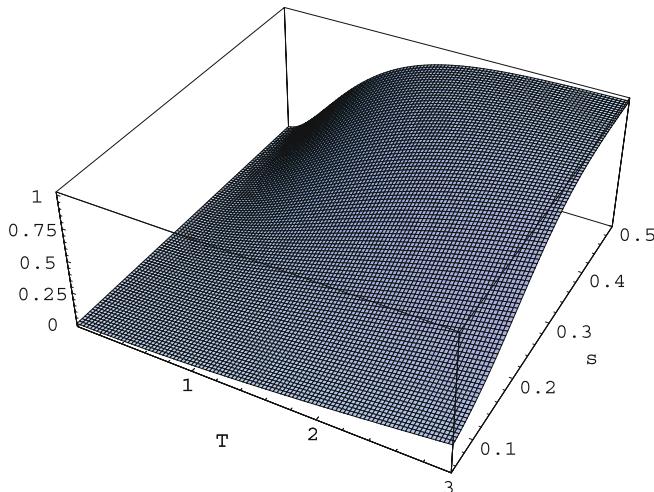


Fig. 13.7.3. Relative error for simplified Euler scheme with $\Delta = T$

We also obtained similar results when estimating higher moments. For instance, in Fig. 13.7.6 we plot the difference in the relative errors of the Euler and simplified Euler schemes for the tenth moment.

What really matters in practice is the time needed to reach a given level of accuracy. In Fig. 13.7.7 we plot the logarithm of the CPU time versus the negative of the logarithm, of the weak error observed for the two methods described above, together with those for the fully implicit Euler scheme and simplified versions.

Since the accuracy for a given time step size is almost identical for the schemes of the same order, the increase in efficiency simply reflects the fact

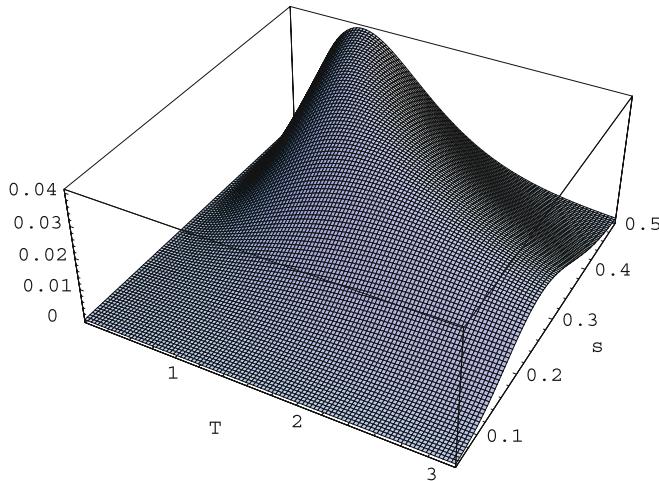


Fig. 13.7.4. Relative error of simplified Euler scheme minus relative error of Euler scheme

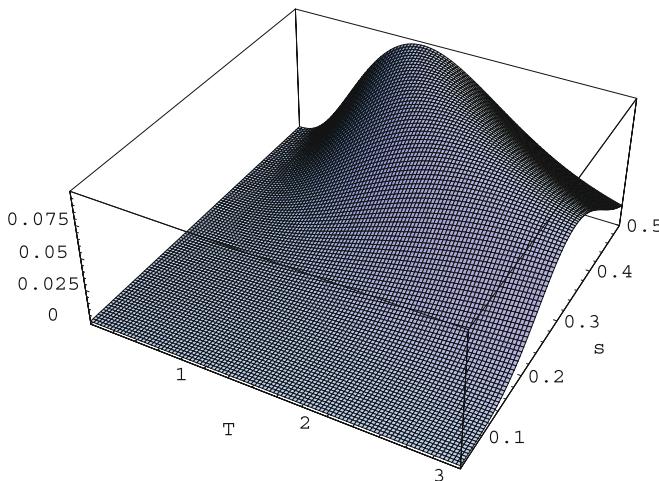


Fig. 13.7.5. Relative error of simplified weak order 2.0 scheme minus relative error of weak order 2.0 Taylor scheme

that the simplified schemes are computationally less expensive than their Gaussian counterparts. We recall that, for instance, the simplified Euler scheme in our example is 29 times faster than the Euler scheme.

By comparing all six methods, we conclude that the second order simplified scheme is significantly more efficient for the given example than any other scheme considered. This result is rather important since it directs us towards higher order simplified methods for efficient Monte Carlo simulation for smooth payoffs.

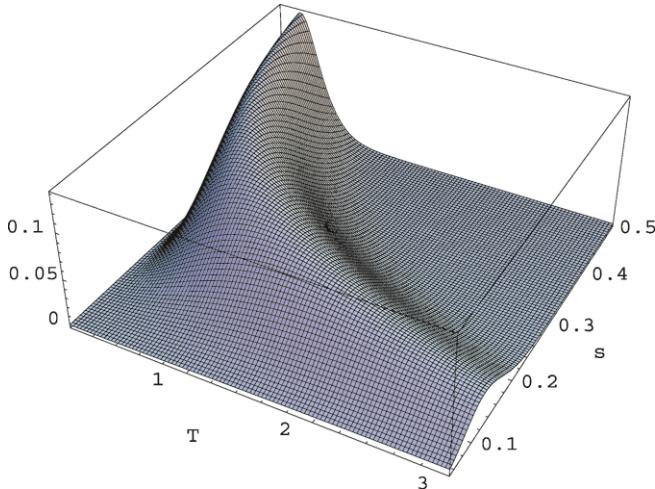


Fig. 13.7.6. Relative error of simplified Euler scheme minus relative error of Euler scheme for the tenth moment

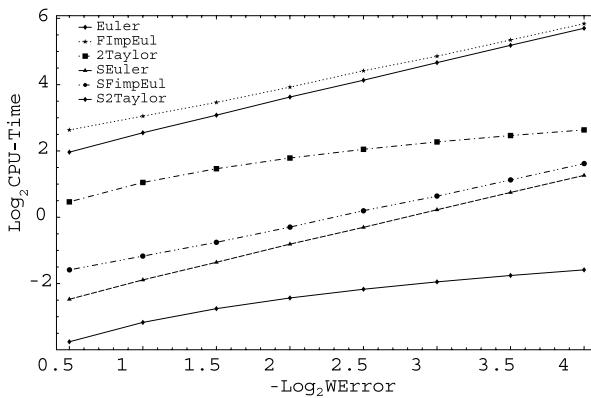


Fig. 13.7.7. Log-log plot of CPU time versus weak error

Non-Smooth Payoffs

In option pricing we are confronted with the computation of expectations of non-smooth payoffs. Note that weak convergence theorems for discrete-time approximations require typically smooth payoff functions. We refer to [Bally & Talay \(1996a, 1996b\)](#) and [Guyon \(2006\)](#) for weak convergence theorems for the Euler scheme in the case of non-smooth payoff functions.

To give a simple example, let us compute the price of a European call option. In this case we assume that the Black-Scholes dynamics of the SDE (13.7.1) is specified under the risk neutral measure, so that the drift coefficient a equals the risk-free rate r . In this case the price of the European call option is given by the expected value of the continuous, but only piecewise

differentiable payoff $e^{-rT}(X_T - K)^+ = e^{-rT} \max(X_T - K, 0)$ with strike price K and maturity T . By (13.7.2), we obtain the European call price from the well-known Black-Scholes formula

$$c_{T,K}(X_0) = E(e^{-rT}(X_T - K)^+) = X_0 N(d_1) - K e^{-rT} N(d_2), \quad (13.7.16)$$

where $d_1 = \frac{\ln(\frac{X_0}{K}) + (r + \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}$ and $d_2 = d_1 - \sigma\sqrt{T}$, see (2.7.56).

For this particular example the estimator of the simplified Euler scheme and that of the simplified weak order 2.0 Taylor scheme can be obtained in closed form. For the simplified Euler scheme we obtain

$$\begin{aligned} c_{T,K}(Y_0) &= \frac{e^{-rT}}{2^N} \sum_{i=0}^N \binom{N}{i} \left(Y_0 \left(1 + \frac{rT}{N} + \sqrt{\frac{T}{N}} \right)^{N-i} \right. \\ &\quad \times \left. \left(1 + \frac{rT}{N} - \sqrt{\frac{T}{N}} \right)^i - K \right)^+. \end{aligned} \quad (13.7.17)$$

For the simplified weak order 2.0 scheme we have

$$\begin{aligned} c_{T,K}(Y_0) &= \frac{e^{-rT}}{6^N} \sum_{j=0}^N \binom{N}{j} 4^{N-j} \sum_{i=0}^j \binom{j}{i} \left(Y_0 h_1^{N-j} \left(h_1 + h_2 \sqrt{\frac{3T}{N}} + h_3 \frac{3T}{N} \right)^{j-i} \right. \\ &\quad \times \left. \left(h_1 - h_2 \sqrt{\frac{3T}{N}} + h_3 \frac{3T}{N} \right)^i - K \right)^+, \end{aligned} \quad (13.7.18)$$

where h_1 , h_2 and h_3 are defined in (13.7.10).

For the Euler scheme and the order 2.0 Taylor scheme, we resort to Monte Carlo simulation to obtain an estimate of the expected value $E(e^{-rT}(Y_T - K)^+)$. Note that the number of generated sample paths used in the following numerical experiments is made large enough to ensure that the statistical error is negligible when compared to the systematic error. In the special case of one time step, that is $\Delta = T$, the estimator of the call price for the Euler scheme and for the weak order 2.0 Taylor scheme can be obtained in closed form. For the Euler scheme with one time step we have

$$c_{T,K}(Y_0) = e^{-rT} \left((Y_0(1 + rT) - K) N(b) + Y_0 \sigma \sqrt{T} N'(b) \right),$$

with

$$b = \frac{K - Y_0(1 + rT)}{Y_0 \sigma \sqrt{T}} \quad \text{and} \quad N'(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}},$$

for $x \in \mathbb{R}$.

For the weak order 2.0 Taylor scheme with one time step, if

$$Y_0(Y_0(h_2)^2 - 4h_3(Y_0h_1 - K)) > 0,$$

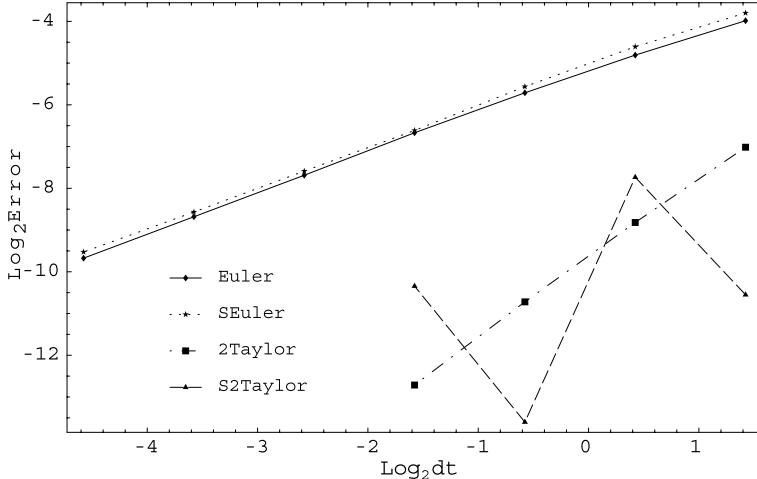


Fig. 13.7.8. Log-log plot of the weak errors for the Euler, the simplified Euler, the weak order 2.0 Taylor and the simplified weak order 2.0 Taylor schemes

then we obtain

$$\begin{aligned} c_{T,K}(Y_0) = & e^{-rT} \left((Y_0(h_1 + h_3) - K)(N(b_-) + N(b_+)) \right. \\ & \left. + Y_0 \left((N'(b_+) - N'(b_-))(h_2 + h_3(b_+ - b_-)) \right) \right), \end{aligned}$$

where

$$b_{\pm} = \frac{-Y_0 h_2 \pm \sqrt{(Y_0 h_2)^2 - 4X_0 h_3(Y_0 h_1 - K)}}{2X_0 h_3},$$

and h_1 , h_2 and h_3 are defined in (13.7.10). If instead

$$Y_0(Y_0(h_2)^2 - 4h_3(Y_0 h_1 - K)) \leq 0,$$

then we have

$$c_{T,K}(Y) = Y_0(h_1 + h_3) - K.$$

These closed form solutions allow us to show the weak error generated by the respective schemes when $\Delta = T$. Here we avoid any statistical error that would arise from a Monte Carlo simulation. Note that for some range of parameters used in our study the weak error is very small and it would be unfeasible to use Monte Carlo simulation to obtain a reliable estimate. The above methodology allows us to also study the theoretical value of the weak error in such cases.

In Fig. 13.7.8 we show the log-log weak error plot of an at-the-money European call option with strike $K = X_0$. The Euler and the simplified Euler schemes have very similar accuracy and generate weak order $\beta = 1.0$, with

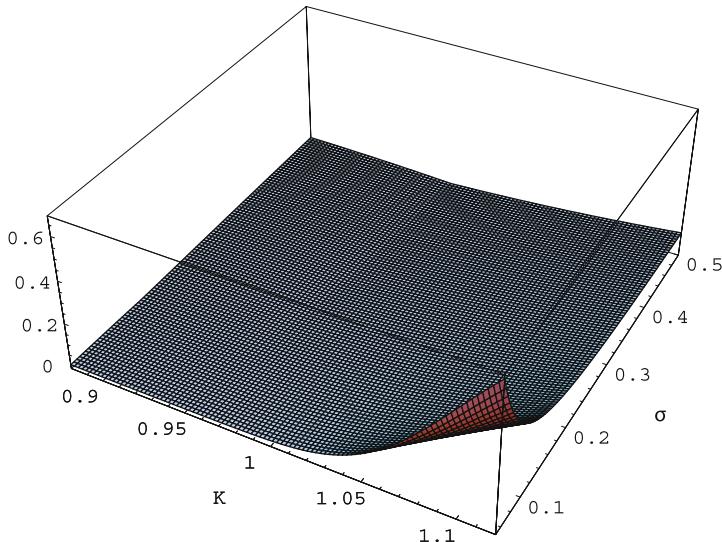


Fig. 13.7.9. Relative error for Euler scheme with $\Delta = T$

the log-error forming a perfect line dependent on the log-time step size. The weak order 2.0 Taylor scheme is more accurate and achieves an order of weak convergence of about $\beta = 2.0$. The accuracy of the simplified weak order 2.0 scheme is comparable to that of the weak order 2.0 Taylor scheme, but its convergence is more erratic.

We report that when testing these schemes with different sets of parameters, we noticed that simplified schemes have an accuracy similar to that of corresponding Taylor schemes, but their error exhibits oscillations depending on the step size. This effect seems to be more pronounced in the simplified weak order 2.0 Taylor scheme, but it can also be present, when using the simplified Euler scheme for certain parameters.

As we will see in Sect. 17.2, the simplified Euler scheme and the simplified weak order 2.0 schemes are related to binomial and trinomial trees, respectively. The erratic behavior of the weak error that we observed appears to be due to the discrete nature of the presence of multi-point distributed random variables used in the simplified schemes. This effect appears to be related to that which was noticed for tree methods in [Boyle & Lau \(1994\)](#), and which we will also discuss in Chap. 17.

As in the previous section, we now compare for a wide range of parameters the accuracy of weak Taylor schemes with that of simplified weak Taylor schemes when using only one time step. We consider the relative weak error,

$$\left| \frac{c_{T,K}(X) - c_{T,K}(Y)}{c_{T,K}(X)} \right|, \quad (13.7.19)$$

with $X_0 = Y_0 = 1$, $r = 0.1$, $\sigma \in [0.05, 0.5]$, $K \in [0.88, 1.12]$ and $T = 3/12$. In Fig. 13.7.9 we report the relative error generated by the Euler scheme. We

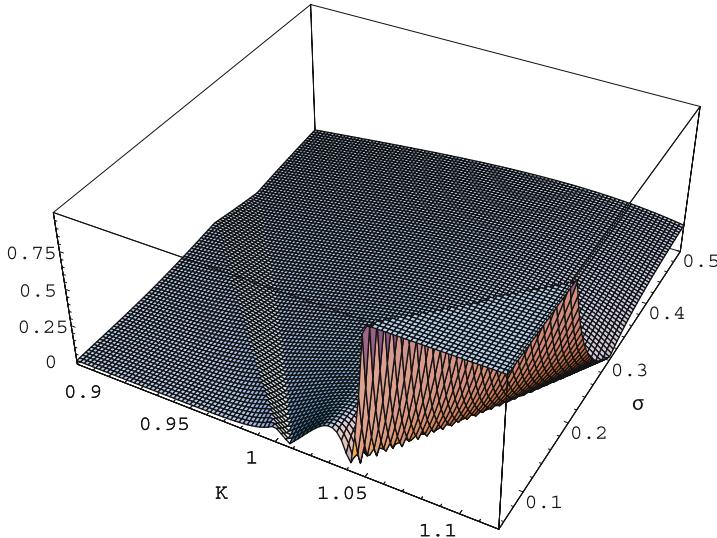


Fig. 13.7.10. Relative error for simplified Euler scheme with $\Delta = T$

notice that for small values of the volatility and large values of the strike, the relative error increases as much as 70%. Fig. 13.7.10 shows the relative error for the simplified Euler scheme, which is extreme for large strikes and small volatilities. Also in this case the loss in accuracy due to the use of two-point distributed random variables (11.2.4) is rather small outside this region of high strike prices and low volatilities. Note that since we are using only one time step, for every value of $K \geq Y_0(1 + rT + \sigma\sqrt{T})$ all sample paths of the simplified Euler scheme finish out-of-the-money and, thus, the expected call payoff is valued close to zero. On the other hand, some of the paths generated by the Euler scheme end up in-the-money, providing a positive value for the expected call payoff, which is closer to the exact solution. Note however that in this region the exact value of the expected payoff is very small. For instance, for $K = 1.05$ and $\sigma = 0.05$ the exact value equals about 0.002. Thus, the absolute error generated by the simplified Euler scheme remains under control.

In Figs. 13.7.11 and 13.7.12 we report the difference between the relative error of the simplified Euler scheme and that of the Euler scheme with a medium maturity $T = 1$ and a long maturity $T = 4$, respectively. Again we can see that the accuracy of schemes based on simplified random variables is similar to that of schemes based on Taylor schemes. Furthermore, we notice that for certain sets of parameters the simplified Euler scheme is even more accurate than the Euler scheme.

Finally, we conducted similar numerical experiments comparing the accuracy of the weak order 2.0 Taylor scheme to that of the simplified weak order 2.0 scheme. We report also that in this case the loss in accuracy due to the use of the three-point distributed random variable (13.1.11) is quite small.

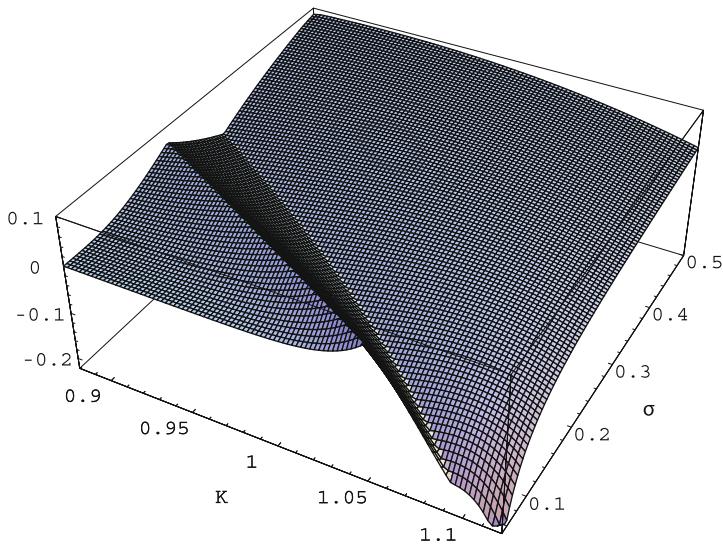


Fig. 13.7.11. Relative error of simplified Euler scheme minus relative error of Euler scheme for $T = 1$

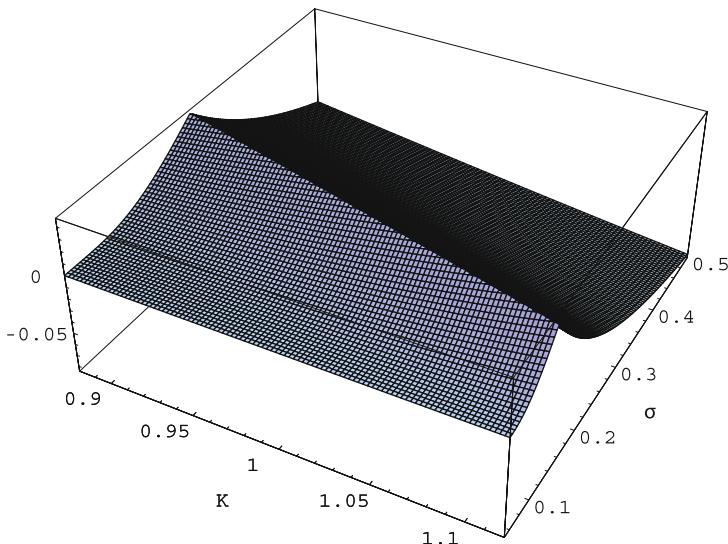


Fig. 13.7.12. Relative error of simplified Euler scheme minus relative error of Euler scheme for $T = 4$

In summary, our experimental results suggest that the accuracy of weak schemes based on multi-point distributed random variables is similar to that of weak Taylor schemes. The loss in accuracy, which is due to the use of multi-point random variables, is for typical parameters below 10% when measured in terms of relative errors. We have also found that in some particular cases

the accuracy of simplified weak schemes can be superior to that of weak Taylor schemes.

Smooth Payoffs under the Merton Model

We study the weak approximation of the SDE (1.8.5), describing the Merton model, which is of the form

$$dX_t = X_{t-} \left(a dt + \sigma dW_t + \int_{\mathcal{E}} (v - 1) p_\varphi(dv, dt) \right), \quad (13.7.20)$$

for $t \in [0, T]$ and $X_0 > 0$. We recall the explicit solution

$$X_t = X_0 e^{(a - \frac{1}{2}\sigma^2)t + \sigma W_t} \prod_{i=1}^{p_\varphi(t)} \xi_i, \quad (13.7.21)$$

where the marks ξ_i are distributed according to a given probability measure $F(dv) = \frac{\varphi(dv)}{\lambda}$ and $p_\varphi = \{p_\varphi(t), t \in [0, T]\}$ denotes a Poisson process with intensity $\lambda = \varphi(\mathcal{E}) < \infty$.

In this subsection we consider several schemes with weak order of convergence $\beta \in \{1.0, 2.0\}$. The schemes with weak order $\beta = 1.0$ are the following: the regular Euler, jump-adapted Euler and the jump-adapted predictor-corrector Euler schemes. Moreover, we consider the following schemes with weak order $\beta = 2.0$: the regular order 2.0 Taylor, jump-adapted order 2.0 Taylor and the jump-adapted order 2.0 predictor-corrector schemes.

Let us consider the estimation of moments of the solution X at a final time T . For the SDE (13.7.20), we obtain, via its closed form solution (13.7.21), the expression

$$E((X_T)^k) = (X_0)^k \exp \left\{ kT \left(a + \frac{\sigma^2}{2}(k-1) \right) + \lambda T \left(E((\xi)^k) - 1 \right) \right\} \quad (13.7.22)$$

for $k \in \mathcal{N}$, so that the weak error $\mu(\Delta)$ can be estimated. In the following example we choose the smooth function $g(x) = x^4$, so that we estimate the fourth moment of the solution X at a given maturity date T . We select the following default parameters: $a = 0.05$, $\sigma = 0.15$, $\lambda = 0.05$, $X_0 = 1$ and $T = 4$.

First, we consider the case of a mark-independent jump coefficient $c(t, x) = x(\psi - 1)$, which is equivalent to the case of constant marks $\xi_i = \psi > 0$, and setting $\psi = 0.85$. Therefore, at each jump time the value of X drops by 15%. In Fig. 13.7.13 we report the results for the regular Euler, jump-adapted Euler, and jump-adapted predictor-corrector Euler schemes. Here and in the remainder of this section, the implicitness parameters have been set to $\theta = 0.5$ and $\eta = 0$. The slopes of the lines in Fig. 13.7.13 are about one, which means that these schemes achieve their theoretical order of weak convergence $\beta = 1.0$. Furthermore, the accuracies of the regular and the jump-adapted

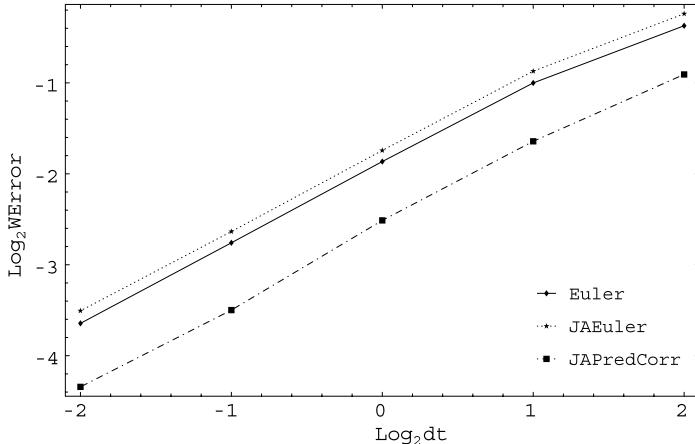


Fig. 13.7.13. Log-log plot of weak error versus time step size

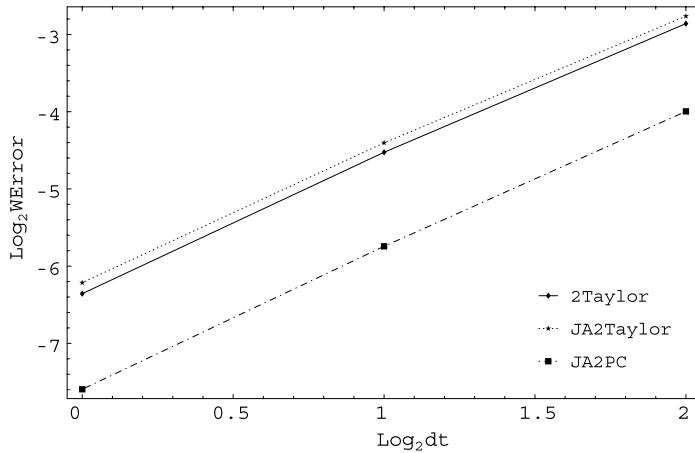


Fig. 13.7.14. Log-log plot of weak error versus time step size

Euler schemes are both very similar. The jump-adapted predictor-corrector Euler scheme is slightly more accurate.

In Fig. 13.7.14 we show the errors generated by the regular order 2.0 Taylor, jump-adapted order 2.0 Taylor and jump-adapted order 2.0 predictor-corrector schemes. Here we see that the order of convergence achieved is about $\beta = 2.0$, which has been suggested by the convergence theorems of previous chapters. Also in this case, the predictor-corrector scheme is more accurate than the corresponding explicit scheme. Finally, in Fig. 13.7.15 we plot the weak second order schemes together with the regular Euler and the jump-adapted predictor-corrector Euler schemes. This figure highlights the higher accuracy of schemes with second order of weak convergence. Furthermore, we report that the jump-adapted weak order 2.0 predictor-corrector scheme is the most accurate among all schemes implemented.

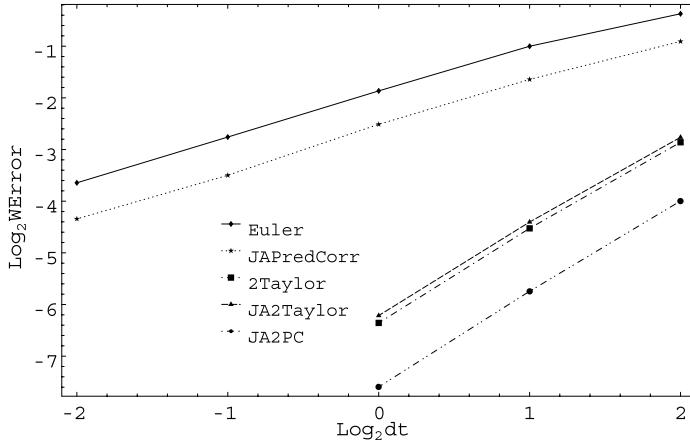


Fig. 13.7.15. Log-log plot of weak error versus time step size

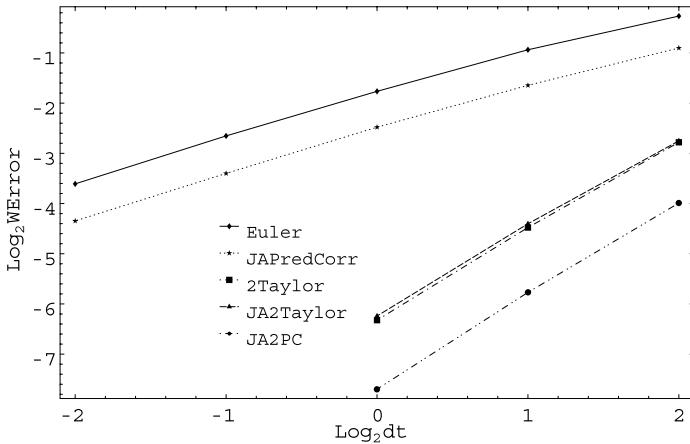


Fig. 13.7.16. Log-log plot of weak error versus time step size

Let us now consider the case of lognormally distributed marks, where the logarithm of the mark $\zeta_i = \ln(\xi_i)$ is an independent Gaussian random variable, $\zeta_i \sim N(\varrho, \varsigma)$, with mean $\varrho = -0.1738$ and standard deviation $\sqrt{\varsigma} = 0.15$. This implies that at a jump time the value of X drops on average by 15%, since $E(\xi) - 1 = -0.15$. In Fig. 13.7.16 we report the results for the regular Euler, jump-adapted predictor-corrector Euler, regular and jump-adapted order 2.0 Taylor, and the jump-adapted order 2.0 predictor-corrector schemes. The accuracy of these schemes in the case of lognormally distributed jump sizes is almost the same as that achieved in the case of a mark-independent jump coefficient. Here all schemes achieve the theoretical orders of weak convergence prescribed by the convergence theorems in the previous chapters.

Non-Smooth Payoff under the Merton Model

We now consider the case of some non-smooth payoff function g when computing the price of a European call option. The convergence theorems presented in Chaps. 12 and 13 do not cover the case of a non-differentiable payoff. It seems that there does not exist in the literature a general convergence theorem for weak Taylor approximations in the case of non-smooth payoff functions when considering SDEs with jumps. The existing results are limited to the Euler scheme for pure diffusion SDEs, see [Bally & Talay \(1996a, 1996b\)](#) and [Guyon \(2006\)](#), and to the Euler scheme for pure jump SDEs, see [Hausenblas \(2002\)](#). It is, however, interesting to obtain some numerical results that could indicate the theoretical performance of the weak schemes presented, when applied to the non-smooth European call option payoff $g(X_T) = e^{-rT}(X_T - K)^+ = e^{-rT} \max(X_T - K, 0)$. Here r is the risk-free rate, K the strike price and T the maturity. Often, the presence of jumps causes market incompleteness, and thus, precludes the possibility of perfect hedging. We refer to [Cont & Tankov \(2004\)](#) for a discussion on pricing and hedging in incomplete markets with jumps, see also Sect. 10.5 and Chap. 15. Here we assume for simplicity, similar to [Merton \(1976\)](#), that the jump risk is non-systematic, and thus, diversifiable. In our example the price of a European call option is then given by

$$c_{T,X} = E_\theta(e^{-rT}(X_T - K)^+), \quad (13.7.23)$$

where under the above risk neutral pricing measure the drift a of the process X has to be set to $r - q - \lambda(E_\theta(\xi) - 1)$, where q is the continuous dividend yield provided by the security X .

In the case of the SDE (13.7.20), where the logarithm of the mark $\zeta_i = \ln(\xi_i)$ is a Gaussian random variable $\zeta_i \sim N(\varrho, \varsigma)$ with mean ϱ and variance ς , we obtain a closed form solution, see [Merton \(1976\)](#), given by

$$c_{T,K}(X) = \sum_{j=0}^{\infty} \frac{e^{-\lambda' T} (\lambda' T)^j}{j!} f_j,$$

where $\lambda' = \lambda E_\theta(\xi)$. Here

$$f_j = X_0 N(d_{1,j}) - e^{-a_j T} K N(d_{2,j}),$$

is the Black-Scholes price of a call option with the parameters specified as

$$d_{1,j} = \frac{\ln(\frac{X_0}{K}) + (a_j + \frac{\sigma_j^2}{2}) T}{\sigma_j \sqrt{T}},$$

$d_{2,j} = d_{1,j} - \sigma_j \sqrt{T}$, $a_j = r - q - \lambda(E_\theta(\xi) - 1) + \frac{j \ln E_\theta(\xi)}{T}$ and $\sigma_j^2 = \sigma^2 + \frac{j \varsigma}{T}$. Recall that we denote by $N(\cdot)$ the probability distribution of a standard Gaussian random variable.

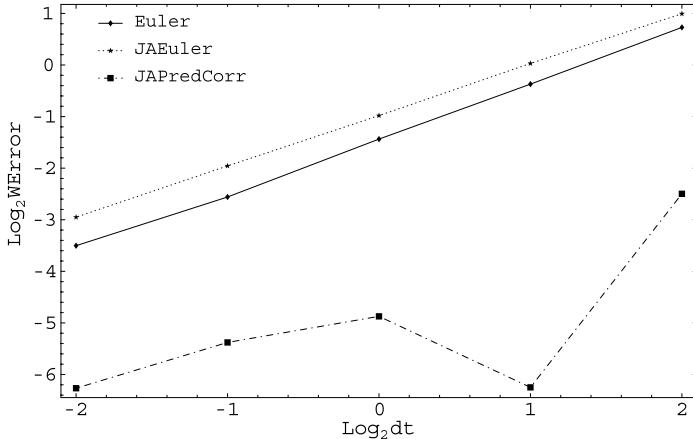


Fig. 13.7.17. Log-log plot of weak error versus time step size

Let us first consider the case of constant marks $\xi_i = \psi > 0$, with $\psi = 0.85$ as in the previous section. Moreover, we set $r = 0.055$, $q = 0.01125$, $\lambda = 0.05$, $\sigma = 0.15$, $X_0 = 100$ and $K = 100$. Note that this implies the risk neutral drift $a = r - q - \lambda(\psi - 1) = 0.05$.

In Fig. 13.7.17 we consider the case of a mark-independent jump coefficient for the regular and jump-adapted Euler schemes, and the jump-adapted predictor-corrector Euler scheme. All these schemes achieve, in our study, an order of weak convergence of about one. The regular and the jump-adapted Euler schemes achieve experimentally first order weak convergence, with the regular scheme being slightly more accurate. The jump-adapted predictor-corrector Euler scheme is far more accurate than the explicit schemes considered here. Its order of weak convergence is about one, with an oscillatory behavior for large time step sizes. We will give in the next chapter some explanation for this excellent performance.

Fig. 13.7.18 shows the accuracy of the regular order 2.0 Taylor, jump-adapted order 2.0 Taylor and jump-adapted order 2.0 predictor-corrector schemes. All schemes attain numerically an order of weak convergence equal to $\beta = 2.0$. In this example the jump-adapted order 2.0 Taylor scheme is the most accurate. Finally, in Fig. 13.7.20 we report the results for these second weak order schemes together with those of the regular Euler and jump-adapted predictor-corrector Euler schemes. Notice the difference in accuracy between first order and second order schemes. However, we emphasize again the excellent performance of the jump-adapted predictor-corrector Euler scheme already for large step sizes. This scheme reaches an accuracy similar to that of some second order schemes. The jump-adapted weak order 2.0 Taylor scheme is the most accurate for all time step sizes considered.

We have also tested the schemes mentioned above in the case of lognormal marks, where $\ln(\xi_i) \sim N(\varrho, \varsigma)$, with mean $\varrho = -0.1738$ and standard deviation $\sqrt{\varsigma} = 0.15$. In Fig. 13.7.19 one notices that the numerical results are

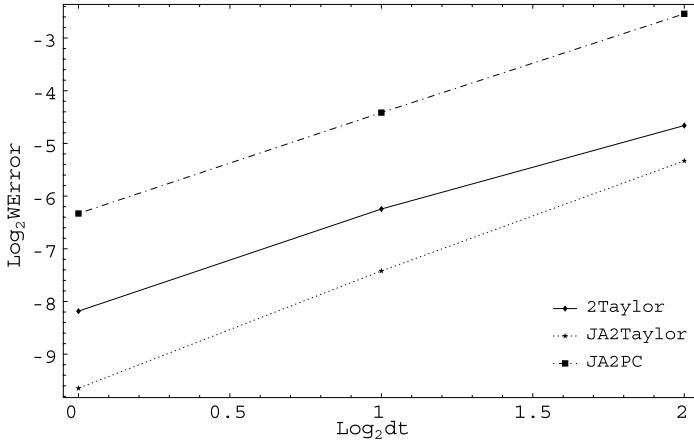


Fig. 13.7.18. Log-log plot of weak error versus time step size

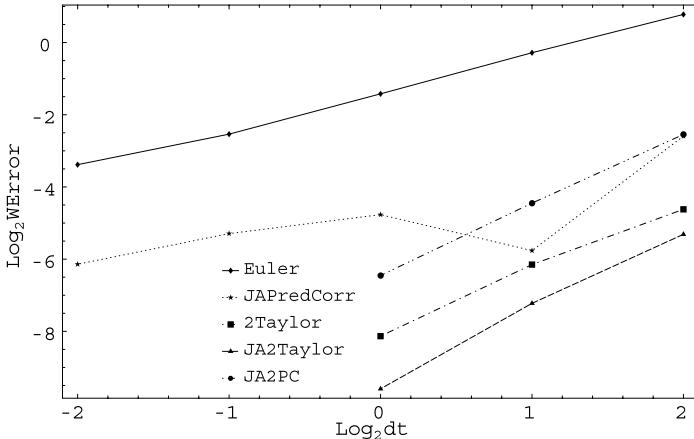


Fig. 13.7.19. Log-log plot of weak error versus time step size, lognormal marks

practically identical to those reported in Fig. 13.7.20 for the case of constant marks.

Let us now test the weak schemes proposed when approximating the SDE (13.7.20), for the case of lognormal marks, by using parameter values reasonably consistent with market data. We use here the parameters reported in Andersen & Andreasen (2000a) from European call options on the S&P 500 in April 1999. The risk free rate and the dividend yield are given by $r = 0.0559$ and $q = 0.01114$, respectively. With a least-squares fit of the Merton model to the implied midpoint Black-Scholes volatilities of the S&P 500 options in April 1999, Andersen & Andreasen (2000a) obtained the following parameters: $\sigma = 0.1765$, $\lambda = 0.089$, $\varrho = -0.8898$ and $\sqrt{\varsigma} = 0.4505$. The last two parameters being the mean and standard deviation of the logarithm of the marks, $\ln(\xi_i)$.

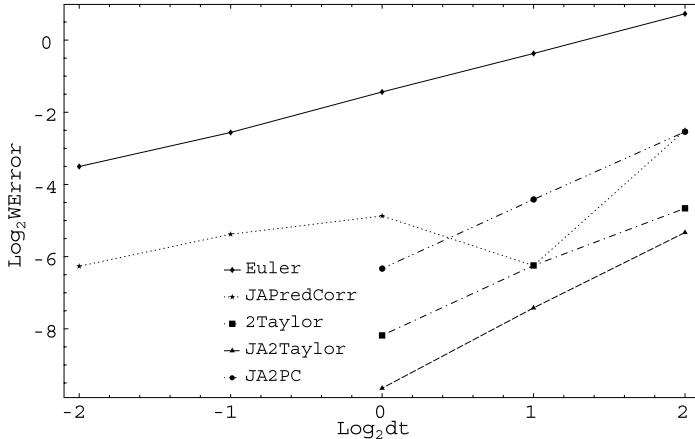


Fig. 13.7.20. Log-log plot of weak error versus time step size, constant marks

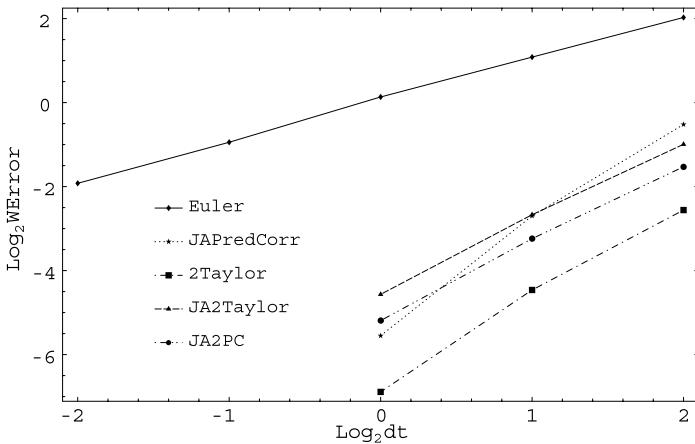


Fig. 13.7.21. Log-log plot of weak error versus time step size, lognormal marks

In Fig. 13.7.21 we show the results for the regular Euler, jump-adapted predictor-corrector Euler, regular and jump-adapted order 2.0 Taylor, and the jump-adapted order 2.0 predictor-corrector schemes, when pricing an at-the-money European call option with maturity time $T = 4$. We report that the regular Euler scheme and the jump-adapted Euler scheme, which is not shown in the figure, achieve first order of weak convergence, with the jump-adapted Euler scheme being slightly less accurate than its regular counterpart. The regular and jump-adapted order 2.0 Taylor and the jump-adapted order 2.0 predictor-corrector schemes are significantly more accurate and achieve an order of weak convergence of approximately $\beta = 2.0$. We emphasize that the jump-adapted predictor-corrector Euler scheme achieves in this example a remarkable accuracy. For large time step sizes it is as good as some second order schemes.

In summary, the numerical results obtained indicate that predictor-corrector schemes for the given underlying dynamics remain accurate, even when using large time step sizes. Also, [Hunter, Jäckel & Joshi \(2001\)](#) report that the predictor-corrector Euler scheme is very accurate when pricing interest rate options under the diffusion LIBOR market model. Finally, we remark that in this section we have analyzed only the accuracy and not the CPU time needed to run the algorithms. This is, of course, strongly dependent on the specific implementation and the problem at hand. We emphasize that predictor-corrector schemes are only slightly computationally more expensive than corresponding explicit schemes, as they use the same random variables.

13.8 Exercises

- 13.1.** For the Merton type SDE given in Exercise [12.1](#) provide a jump-adapted simplified weak order 2.0 scheme.
- 13.2.** For the SDE in Exercise [12.1](#) formulate a jump-adapted fully implicit weak order 1.0 predictor-corrector scheme.

Numerical Stability

When simulating discrete-time approximations of solutions of SDEs, in particular martingales, numerical stability is clearly more important than higher order of convergence. The stability criterion presented is designed to handle both scenario and Monte Carlo simulation, that is, both strong and weak approximation methods. Stability regions for various schemes are visualized. The result being that schemes, which have implicitness in both the drift and the diffusion terms, exhibit the largest stability regions. Refining the time step size in a simulation can lead to numerical instabilities, which is not what one experiences in deterministic numerical analysis. This chapter follows closely [Platen & Shi \(2008\)](#).

14.1 Asymptotic p -Stability

Numerical Stability of Discrete-Time Methods

One of the common challenges in numerical analysis is to choose algorithms which are *numerically stable* for the task at hand. The precise definition of numerical stability depends on the context but is typically related to the long term accuracy of an algorithm when applied to a given dynamic. In the context of solving SDEs via simulation, one can use different discrete-time approximations to simulate the paths of an approximating stochastic process. However, numerical errors such as roundoff and truncation errors are often unavoidable during practical simulations. Therefore, it is important to choose a numerical scheme which does not propagate uncontrolled approximation errors. In comparison with the already discussed concepts of strong and weak higher order of convergence, the basic question on numerical stability should be answered first when deciding which numerical scheme to use. Higher order of convergence is a secondary issue.

Various concepts of numerical stability have been introduced in the literature for numerical schemes approximating solutions of SDEs. These con-

cepts typically make use of specifically designed test equations, which have explicit solutions available, see, for instance, Milstein (1995a), Kloeden & Platen (1999), Hernandez & Spigler (1992, 1993), Saito & Mitsui (1993a, 1993b, 1996), Hofmann & Platen (1994, 1996) and Higham (2000).

Stability regions have been identified for the parameter sets of test equations where the propagation of errors is under control in a well-defined sense. In some cases authors use solutions of complex valued linear SDEs with additive noise as test dynamics, see Hernandez & Spigler (1992, 1993), Milstein (1995a), Kloeden & Platen (1999) and Higham (2000). However, in financial applications, these test equations are potentially not realistic enough to capture numerical instabilities, typically arising when simulating asset prices denominated in units of a certain numéraire. The resulting SDEs have no drift and their diffusion coefficients are often level dependent in a multiplicative manner. To study the stability of the corresponding numerical schemes, test SDEs with multiplicative noise have been suggested in real valued and complex valued form, see Saito & Mitsui (1993a, 1993b, 1996), Hofmann & Platen (1994, 1996) and Higham (2000). Recently, Bruti-Liberati & Platen (2008) analyzed the stability regions of a family of strong predictor-corrector Euler schemes under a criterion which captures a rather weak form of numerical stability, known as asymptotic stability.

In the spirit of Bruti-Liberati & Platen (2008), we will use a more general stability criterion. It unifies, in some sense, criteria that identify asymptotic stability, mean square stability and the stability of absolute moments. The corresponding stability regions allow us to visualize the stability properties for a range of simulation schemes. The interpretation of these plots will turn out to be rather informative. The methodology can be employed for the study of numerical stability properties of both strong and weak schemes.

Test Dynamics

Similar to Hofmann & Platen (1994) and Bruti-Liberati & Platen (2008) we use as *linear test SDE* the linear Itô SDE with multiplicative noise

$$dX_t = \left(1 - \frac{3}{2}\alpha\right)\lambda X_t dt + \sqrt{\alpha|\lambda|} X_t dW_t \quad (14.1.1)$$

for $t \geq 0$, where $X_0 > 0$, $\lambda < 0$, $\alpha \in [0, 1]$. Here we call α the degree of stochasticity and λ the growth parameter. The corresponding Stratonovich SDE has the form

$$dX_t = (1 - \alpha)\lambda X_t dt + \sqrt{\alpha|\lambda|} X_t \circ dW_t, \quad (14.1.2)$$

where “ \circ ” denotes the Stratonovich stochastic differential, see Sect. 1.4 and Kloeden & Platen (1999).

For the two equivalent real valued SDEs (14.1.1) and (14.1.2), the parameter $\alpha \in [0, 1]$ describes the degree of stochasticity of the test dynamics. When $\alpha = 0$, the SDEs (14.1.1) and (14.1.2) become deterministic. In the

case $\alpha = \frac{2}{3}$, the Itô SDE (14.1.1) has no drift, and X is a martingale. This case models a typical Black-Scholes asset price dynamic when the price is expressed in units of a numéraire under the corresponding pricing measure. When the numéraire is the savings account, then the pricing measure is, under appropriate assumptions, the equivalent risk neutral probability measure. In Chap. 3 we have shown that the pricing measure is simply the real world probability measure when the numéraire is the growth optimal portfolio. In the case $\alpha = 1$, the Stratonovich SDE (14.1.2) has no drift.

Stability Criterion

Now, let us introduce the following notion of stability:

Definition 14.1.1. *For $p > 0$ a process $Y = \{Y_t, t > 0\}$ is called asymptotically p -stable if*

$$\lim_{t \rightarrow \infty} E(|Y_t|^p) = 0. \quad (14.1.3)$$

Consequently, a process is asymptotically p -stable if in the long run its p th moment vanishes. By the Lyapunov inequality it follows that if a process is asymptotically p_1 -stable, for $p_1 > p_2 > 0$, then it is also asymptotically p_2 stable.

The explicit solution of (14.1.1) is of the form

$$X_t = X_0 \exp \left\{ (1 - \alpha) \lambda t + \sqrt{\alpha |\lambda|} W_t \right\} \quad (14.1.4)$$

for $t \geq 0$, $\lambda < 0$ and $\alpha \geq 0$. For $p > 0$ and $\lambda < 0$ it follows from (14.1.4), by application of Itô's formula and by taking expectation, that

$$\lim_{t \rightarrow \infty} E(|X_t|^p) = 0 \quad (14.1.5)$$

if and only if $\alpha < \frac{1}{1+\frac{p}{2}}$. This means, in the case when $\lambda < 0$ and $\alpha \in [0, \frac{1}{1+p/2})$, that X is asymptotically p -stable. Furthermore, note that for $\lambda < 0$ by the law of iterated logarithm and the law of large numbers, one has

$$P \left(\lim_{t \rightarrow \infty} X_t = 0 \right) = 1 \quad (14.1.6)$$

if and only if $\alpha \in [0, 1)$, see for instance Protter (2005).

There is only limited potential for identifying numerically stable schemes for unstable test dynamics. Therefore, we will consider the family of test dynamics given by (14.1.1) only for some negative growth parameter $\lambda < 0$ and degree of stochasticity $\alpha \in [0, 1)$. This choice of parameter range will allow us to detect also situations where a scheme is asymptotically p -stable for unstable test dynamics.

Numerically, one would prefer the discrete-time approximation Y and the original continuous process X to have similar stability properties. Ideally, for

given $p > 0$ the approximation Y should generate asymptotically p -stable paths when X was asymptotically p -stable. Any impact from errors should decline over time. However, in our subsequent analysis we will see that Y and X have typically different stability properties for different combinations of values of λ , Δ , α and p .

Stability Region

To explore the properties discussed above, we introduce for a given discrete-time approximation the notion of a *stability region*, which will permit the convenient visualization of asymptotic p -stability properties.

Definition 14.1.2. *The stability region Γ is determined by those triplets $(\lambda\Delta, \alpha, p) \in (-\infty, 0) \times [0, 1] \times (0, \infty)$ for which the discrete-time approximation Y with time step size Δ , is asymptotically p -stable when applied to the test equation (14.1.2).*

In Monte Carlo simulation, where expectations of functionals are approximated, some moments of a discrete-time approximation need typically to exhibit certain asymptotic stability properties. Otherwise, errors in the expectations of functionals may be uncontrollably propagated. For correctly approximating the p th moment of a stochastic process X the numerical approximation Y needs to be asymptotically p -stable.

Note that the asymptotic p -stability criterion (14.1.3) can be related to several concepts of numerical stability previously introduced in the literature, such as the popular concept of mean-square stability, see [Saito & Mitsui \(1996\)](#), [Higham \(2000\)](#), [Higham & Kloeden \(2005\)](#) and [Alcock & Burrage \(2006\)](#). For some schemes this interesting concept leads to explicit characterizations of the resulting region of mean-square stability. However in general, stability regions need to be calculated numerically, and the choice $p = 2$ represents only a special case among many that are of potential interest. Important information is gained by studying the stability region as a function of p . When p increases, the cross section of the stability region of a numerical scheme shrinks, as can be expected due to the Lyapunov inequality, see (1.2.35). It is interesting to analyze for a given scheme how fast this happens. Recall that for $\lambda < 0$ and given $p \in (0, \infty)$ the test dynamics X is p -stable when $\alpha < \frac{1}{1+\frac{p}{2}}$. We will see that it requires typically an implicit scheme to obtain, for instance, for $p = 2$ a stability region that reaches beyond $\alpha = 0.5$.

Transfer Function

We will concentrate on the concept of asymptotic p -stability for both strong and weak discrete-time approximations of the test dynamics (14.1.2) for $\lambda < 0$ and $\alpha \in [0, 1]$. For p close to 0 this approach provides the widest stability region, which will turn out to be the one that refers to the kind of asymptotic

stability proposed in [Bruti-Liberati & Platen \(2008\)](#) for assessing scenario simulations. Our aim is now to study for given strong and weak discrete-time schemes the corresponding stability regions. This analysis will provide guidance for the choice of particular schemes and time step sizes.

We will see for many discrete-time approximations Y with time step size $\Delta > 0$, when applied to the test equation (14.1.1) with a given degree of stochasticity $\alpha \in [0, 1]$, that the ratio

$$\left| \frac{Y_{n+1}}{Y_n} \right| = G_{n+1}(\lambda \Delta, \alpha) \quad (14.1.7)$$

is of crucial interest, $n \in \{0, 1, \dots\}$, $Y_n > 0$, $\lambda < 0$. We call the random variable $G_{n+1}(\lambda \Delta, \alpha)$ the *transfer function* of the approximation Y at time t_n . It transfers the previous approximate value Y_n to the approximate value Y_{n+1} of the next time step.

Furthermore, let us assume for a given scheme and $\lambda < 0$, $\Delta \in (0, \infty)$ and $\alpha \in [0, 1)$ that the random variables $G_{n+1}(\lambda \Delta, \alpha)$ are for $n \in \{0, 1, \dots\}$ independent and identically distributed with $E((\ln(G_{n+1}(\lambda \Delta, \alpha)))^2) < \infty$. This assumption is satisfied for a wide range of schemes. It will allow us to obtain the corresponding stability regions for all schemes that we consider by employing the following result:

Lemma 14.1.3 *A discrete-time approximation for a given $\lambda \Delta < 0$, $\alpha \in [0, 1)$ and $p > 0$, is asymptotically p -stable if and only if*

$$E((G_{n+1}(\lambda \Delta, \alpha))^p) < 1. \quad (14.1.8)$$

The proof of Lemma 14.1.3 is straightforward, since by (14.1.7) the p th moment of the approximation vanishes as $n \rightarrow \infty$, if and only if $E((G_{n+1}(\lambda \Delta, \alpha))^p) < 1$, almost surely, for all $n \in \{0, 1, \dots\}$.

Note that for $p > 0$, the condition (14.1.8) yields by the use of the inequality $\ln(x) \leq x - 1$, the relation

$$E(\ln(G_{n+1}(\lambda \Delta, \alpha))) \leq \frac{1}{p} E((G_{n+1}(\lambda \Delta, \alpha))^p - 1) < 0 \quad (14.1.9)$$

for all $n \in \{0, 1, \dots\}$. This provides the already indicated link to the kind of asymptotic stability studied in [Bruti-Liberati & Platen \(2008\)](#) when p vanishes. In this case a scheme is called *asymptotically stable* for a given $\lambda \Delta$ and α if $E(\ln(G_{n+1}(\lambda \Delta, \alpha))) < 0$ for all $n \in \{0, 1, \dots\}$.

In the following we visualize the stability regions of various schemes by identifying the set of triplets $(\lambda \Delta, \alpha, p)$ for which the inequality (14.1.8) is satisfied. More precisely, we will display the stability regions for $\lambda \Delta \in [-10, 0)$, $\alpha \in [0, 1)$ and $p \in (0, 2]$. Note that when p approaches zero, we obtain the stability region identified under the asymptotic stability criterion, see (14.1.9). We deliberately truncate in our visualization the stability regions at $p = 2$, since the area shown on the top of the graph then provides information about

the popular mean-square stability of the scheme. For higher values of p the stability region shrinks further, and for some schemes it may even become an empty set. Asymptotically, for very large p the resulting stability region relates to the worst case stability concept studied in [Hofmann & Platen \(1994, 1996\)](#). As such the employed asymptotic p -stability covers a relatively wide range of stability criteria.

14.2 Stability of Predictor-Corrector Methods

Within this section we study the numerical stability of the class of strong predictor-corrector Euler schemes introduced in Sect. [7.4](#).

Predictor-Corrector Methods

The Euler scheme is the simplest and most popular scheme for the discrete-time approximation of SDEs, see [\(5.2.2\)](#). It simply keeps the drift and diffusion coefficients constant over the discretization interval. Furthermore, it is an explicit scheme and known to exhibit numerical instabilities for particular SDEs and large time step sizes. For instance, in [Milstein, Platen & Schurz \(1998\)](#) it has been demonstrated that for a martingale, modeled by geometric Brownian motion, the simulation of the solution of the corresponding SDE can be rather erratic if the time step size is not small enough, see also Sect. [7.3](#).

For simplicity, in the schemes we consider below we will study only the one-dimensional time homogeneous case. Let us consider a time discretization $0 = t_0 < t_1 < \dots$ with equal time step size $\Delta = t_{n+1} - t_n$, $n \in \{0, 1, \dots\}$. The following strong predictor-corrector schemes generalize the Euler scheme and can provide some improved numerical stability while avoiding to solve an algebraic equation in each time step, as is required by implicit methods. At the n th time step, first the predictor is constructed using an explicit Euler scheme which predicts a value \bar{Y}_{n+1} . Then, the corrector is applied, which is in its structure similar to an implicit Euler scheme and corrects the predicted value. We emphasize that not only the predictor step is explicit, but that the corrector step is also explicit since it only uses some predicted value. We will see later that with such a predictor-corrector procedure one can introduce some stabilizing effect in the simulation and may avoid the propagation of errors. The family of strong predictor-corrector Euler schemes we consider is given by the corrector

$$\begin{aligned} Y_{n+1} = Y_n + & \left\{ \theta \bar{a}_\eta(\bar{Y}_{n+1}) + (1 - \theta) \bar{a}_\eta(Y_n) \right\} \Delta \\ & + \left\{ \eta b(\bar{Y}_{n+1}) + (1 - \eta) b(Y_n) \right\} \Delta W_n, \end{aligned} \quad (14.2.1)$$

and the predictor

$$\bar{Y}_{n+1} = Y_n + a(Y_n) \Delta + b(Y_n) \Delta W_n, \quad (14.2.2)$$

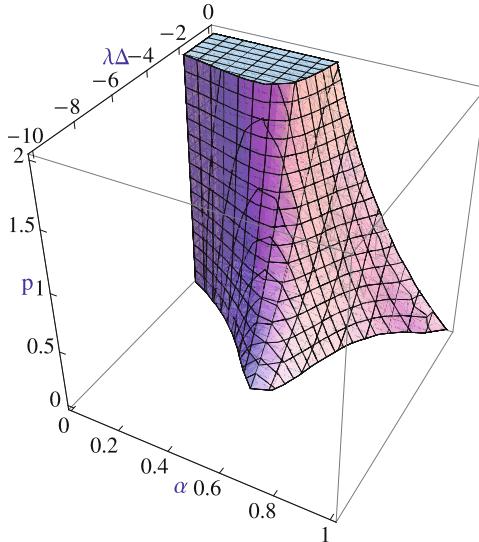


Fig. 14.2.1. Stability region for the Euler scheme

see (7.4.1) and (7.4.2). Note that we need to use in (14.2.1) the adjusted drift function $\bar{a}_\eta = a - \eta b b'$. The parameters $\theta, \eta \in [0, 1]$ are the degrees of implicitness in the drift and the diffusion coefficient, respectively. For the case $\eta = \theta = 0$ one recovers the Euler scheme. A major advantage of the above family of schemes is that they have flexible degrees of implicitness parameters η and θ . For each possible combination of these two parameters the scheme is of strong order $r = 0.5$. This allows us to compare for a given problem simulated paths for different degrees of implicitness. If these paths differ significantly from each other, then some numerical stability problem is likely to be present and one needs to make an effort in providing extra numerical stability. The above strong predictor-corrector schemes potentially offer the required numerical stability. However, other schemes, in particular, implicit schemes may have to be employed to secure sufficient numerical stability for a given simulation task.

Stability of the Euler Scheme

For the Euler scheme, which results for $\theta = \eta = 0$ in the algorithm (14.2.1), it follows by (14.1.7) that

$$G_{n+1}(\lambda \Delta, \alpha) = \left| 1 + \left(1 - \frac{3}{2} \alpha \right) \lambda \Delta + \sqrt{-\alpha \lambda} \Delta W_n \right|, \quad (14.2.3)$$

where $\Delta W_n \sim \mathcal{N}(0, \Delta)$ is a Gaussian distributed random variable with mean zero and variance Δ . The transfer function (14.2.3) yields the stability region

shown in Fig. 14.2.1. It is the region where $E((G_{n+1}(\lambda\Delta, \alpha))^p) < 1$. This stability region has been obtained numerically by identifying for each pair $(\lambda\Delta, \alpha) \in [-10, 0] \times [0, 1]$ those values $p \in (0, 2]$ for which the inequality (14.1.8) holds. One notes that for the purely deterministic dynamics of the test SDE, that is $\alpha = 0$, the stability region covers the area $(-2, 0) \times (0, 2)$. Also in later figures of this type we note that for the deterministic case $\alpha = 0$ we do not see any dependence on p . This is explained by the fact that in this case there is no stochasticity and the relation $\lambda\Delta \in (-2, 0)$ reflects the classical interval of numerical stability for the Euler scheme, as obtained under various criteria, see Hernandez & Spigler (1992). For a stochasticity parameter of about $\alpha \approx 0.5$, the stability region has in Fig. 14.2.1 the largest cross section in the direction of $\lambda\Delta$ for most values of p . The stability region covers an interval $\lambda\Delta$ of increasing length up to about $[-6.5, 0]$ when p is close to zero, and about $[-3, 0]$ when p increases to 2. For increased stochasticity parameter α beyond 0.5 the stability region declines in Fig. 14.2.1. We observe that if the Euler scheme is stable for a certain step size, then it is also stable for smaller step sizes. We see later that this is not the case for some other schemes.

The top of Fig. 14.2.1 shows the mean-square stability region of the Euler scheme, where we have $p = 2$. As the order p of the moments increases, the cross section in the direction of α and $\lambda\Delta$ will be further reduced. The intuition is that as the stability requirements in terms of the order p of the moment becomes stronger, there are less pairs $(\lambda\Delta, \alpha)$ for which the simulated approximation's p th moment tends to zero.

Semi-Drift-Implicit Method

Let us now consider the semi-drift-implicit predictor-corrector Euler method with $\theta = \frac{1}{2}$ and $\eta = 0$ in (14.2.1). The transfer function for this method equals

$$G_{n+1}(\lambda\Delta, \alpha) = \left| 1 + \lambda\Delta \left(1 - \frac{3}{2}\alpha \right) \left\{ 1 + \frac{1}{2} \left(\lambda\Delta \left(1 - \frac{3}{2}\alpha \right) + \sqrt{-\alpha\lambda} \Delta W_n \right) \right\} + \sqrt{-\alpha\lambda} \Delta W_n \right|.$$

Its stability region is displayed in Fig. 14.2.2, showing that this scheme has good numerical stability around $\alpha \approx 0.6$, where it is stable for almost all $\lambda\Delta \in [-10, 0]$. When p is close to 2, its stability region begins to show a gouge around $\lambda\Delta \approx -9$. This means, when we consider the numerical stability for the second moment $p = 2$, instability could arise when using a step size close to $\lambda\Delta \approx -9$. Unfortunately, for the stochasticity parameter value $\alpha = \frac{2}{3}$, that is when X forms a martingale, the stability region is considerably reduced when compared to the region for $\alpha = 0.6$. Near the value of $\alpha = 0.6$ the semi-drift-implicit predictor-corrector scheme outperforms the Euler scheme in terms of stability for most values of p .

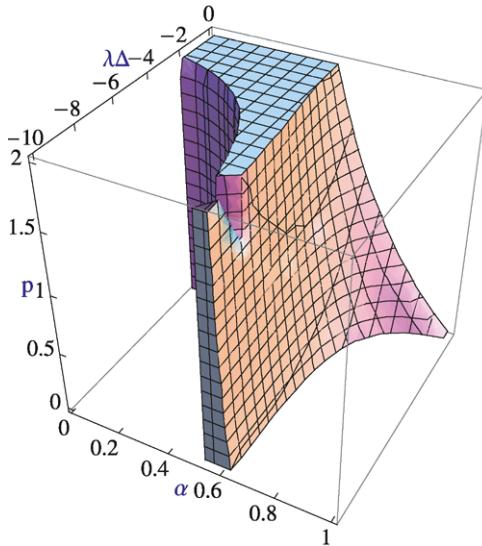


Fig. 14.2.2. Stability region for semi-drift-implicit predictor-corrector Euler method

Drift-Implicit Predictor-Corrector Method

Similarly, we obtain for the drift-implicit predictor-corrector Euler method, derived from (14.2.1) with $\theta = 1$ and $\eta = 0$, the transfer function

$$G_{n+1}(\lambda \Delta, \alpha) = \left| 1 + \lambda \Delta \left(1 - \frac{3}{2} \alpha \right) \left\{ 1 + \lambda \Delta \left(1 - \frac{3}{2} \alpha \right) + \sqrt{-\alpha \lambda} \Delta W_n \right\} + \sqrt{-\alpha \lambda} \Delta W_n \right|.$$

The corresponding stability region is plotted in Fig. 14.2.3. It has similar features as the stability region of the semi-drift-implicit predictor-corrector scheme. However, the stability region appears reduced. It no longer extends as far in the negative direction of $\lambda \Delta$ as in Fig. 14.2.2. In particular, this is visible around a value of $\alpha \approx 0.6$. When p is close to 2.0, the gouge observed in Fig. 14.2.2 near $\lambda \Delta \approx -9$ now occurs at $\lambda \Delta \approx -4$. It appears as though the stability region decreased by making the drift fully “implicit”. A symmetric choice in (14.2.1) with $\theta = 1/2$ seems to be more appropriate than the extreme choice of full drift implicitness with $\theta = 1$.

Semi-Implicit Diffusion Term

Now, let us study the impact of making the diffusion term “implicit” in the predictor-corrector Euler method (14.2.1). First, we consider a predictor-

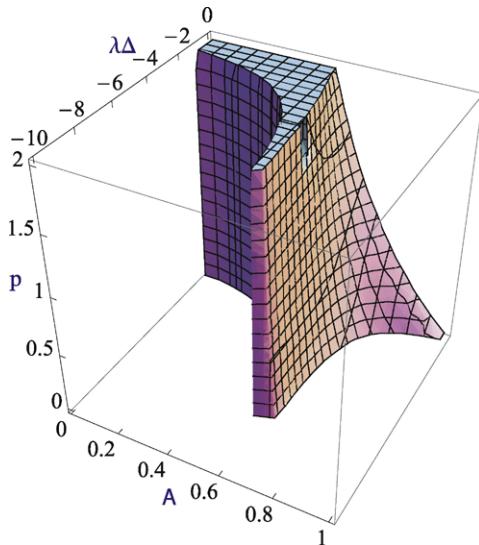


Fig. 14.2.3. Stability region for drift-implicit predictor-corrector Euler method

corrector Euler method with semi-implicit diffusion term where $\theta = 0$ and $\eta = \frac{1}{2}$. Its transfer function has the form

$$G_{n+1}(\lambda \Delta, \alpha) = \left| 1 + \lambda \Delta (1 - \alpha) + \sqrt{-\alpha \lambda} \Delta W_n \right. \\ \left. \times \left\{ 1 + \frac{1}{2} \left(\lambda \Delta \left(1 - \frac{3}{2} \alpha \right) + \sqrt{-\alpha \lambda} \Delta W_n \right) \right\} \right|.$$

The corresponding stability region is shown in Fig. 14.2.4. Although it appears to be rather restricted, when compared with the stability region for the semi-drift-implicit predictor-corrector method, it is important to note that for the martingale case, that is $\alpha = \frac{2}{3}$, the stability region is wider than those of the previous schemes. This observation could become relevant for applications in finance. In some sense, this should be expected since when simulating martingale dynamics, one only gains more numerical stability by making the diffusion term “implicit” rather than the drift term, which is simply zero for a martingale.

Furthermore, although it may seem counter-intuitive to the common experience in deterministic numerical analysis, this scheme can lose its numerical stability by reducing the time step-size. For example, for p close to 0.01 and α near $\frac{2}{3}$, we observe that for about $\lambda \Delta \approx -3$ the method is asymptotically p -stable, whereas for smaller step sizes yielding a $\lambda \Delta$ above -2 , the method is no longer asymptotically p -stable. This is a phenomenon, which is not common

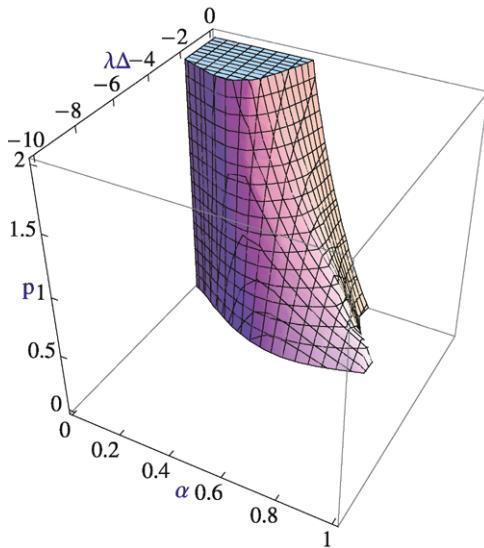


Fig. 14.2.4. Stability region for the predictor-corrector Euler method with $\theta = 0$ and $\eta = \frac{1}{2}$

in deterministic numerical analysis. In stochastic numerics it arises occasionally, as we will also see for some other schemes. In Fig. 14.2.4 it gradually disappears as the value of p increases.

Symmetric Predictor-Corrector Method

An interesting scheme is the symmetric predictor-corrector Euler method, obtained from (14.2.1) for $\theta = \eta = \frac{1}{2}$. It has transfer function

$$G_{n+1}(\lambda \Delta, \alpha) = \left| 1 + \lambda \Delta (1 - \alpha) \left\{ 1 + \frac{1}{2} \left(\lambda \Delta \left(1 - \frac{3}{2} \alpha \right) + \sqrt{-\alpha \lambda} \Delta W_n \right) \right\} \right. \\ \left. + \sqrt{-\alpha \lambda} \Delta W_n \left\{ 1 + \frac{1}{2} \left(\lambda \Delta \left(1 - \frac{3}{2} \alpha \right) + \sqrt{-\alpha \lambda} \Delta W_n \right) \right\} \right|.$$

Its stability region is shown in Fig. 14.2.5. In particular, for the martingale case $\alpha = \frac{2}{3}$ it has a rather large region of stability when compared to that of the Euler scheme and those of previously discussed schemes. For p near zero, α close to one and $\lambda \Delta$ close to -1 , there exists a small area where this scheme is not asymptotically p -stable. Again, this is an area where a decrease in the step size can make a simulation numerically unstable. Also here this phenomenon disappears as the value of p increases.

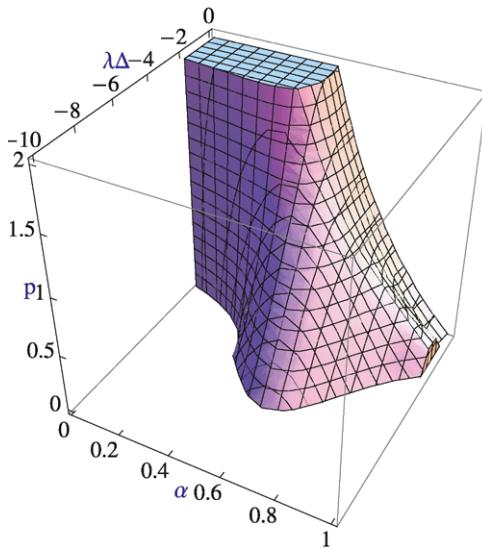


Fig. 14.2.5. Stability region for the symmetric predictor-corrector Euler method

Fully Implicit Predictor-Corrector Method

Next we check the stability region of a fully implicit predictor-corrector Euler method, which is obtained when setting in (14.2.1) both degrees of implicitness to one, that is $\theta = \eta = 1$. The corresponding transfer function is of the form

$$\begin{aligned} G_{n+1}(\lambda \Delta, \alpha) &= \left| 1 + \lambda \Delta \left(1 - \frac{1}{2} \alpha \right) \left\{ 1 + \lambda \Delta \left(1 - \frac{3}{2} \alpha \right) + \sqrt{-\alpha \lambda} \Delta W_n \right\} \right. \\ &\quad \left. + \sqrt{-\alpha \lambda} \Delta W_n \left\{ 1 + \lambda \Delta \left(1 - \frac{3}{2} \alpha \right) + \sqrt{-\alpha \lambda} \Delta W_n \right\} \right| \\ &= \left| 1 + \left\{ 1 + \lambda \Delta \left(1 - \frac{3}{2} \alpha \right) + \sqrt{-\alpha \lambda} \Delta W_n \right\} \right. \\ &\quad \left. \times \left\{ \lambda \Delta \left(1 - \frac{1}{2} \alpha \right) + \sqrt{-\alpha \lambda} \Delta W_n \right\} \right|. \end{aligned}$$

The resulting stability region is shown in Fig. 14.2.6. We notice here that this stability region is considerably smaller than that of the symmetric predictor-corrector Euler scheme. It seems that asymptotic p -stability does not always increase monotonically with the degrees of implicitness in both the drift and the diffusion term. Symmetric levels of implicitness tend to achieve large stability regions.

In summary, in the case of predictor-corrector Euler methods, the stability regions displayed suggest that the symmetric predictor-corrector Euler

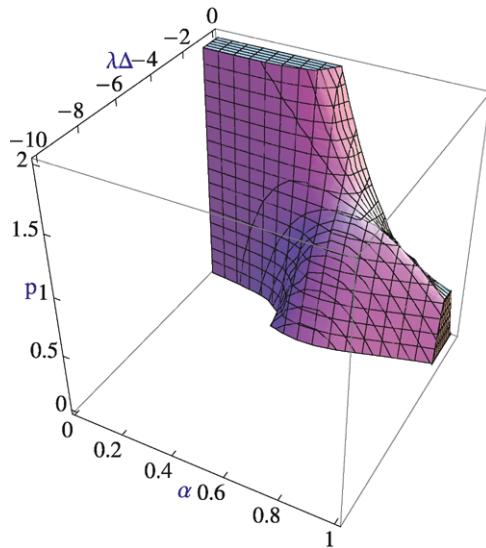


Fig. 14.2.6. Stability region for fully implicit predictor-corrector Euler method

method, see Fig. 14.2.5, is the most suitable of those visualized when simulating martingale dynamics with multiplicative noise. The symmetry between the drift and diffusion terms of the method balances well its numerical stability properties, making it an appropriate choice for a range of simulation tasks.

14.3 Stability of Some Implicit Methods

After having studied the numerical stability properties of the family of predictor-corrector Euler methods, it is worthwhile to compare the observed stability regions with those of some implicit methods, which from the deterministic case, see Sect. 5.1, are known to show good numerical stability properties.

Semi-Drift-Implicit Euler Method

Let us first consider the semi-drift-implicit Euler scheme in the form

$$Y_{n+1} = Y_n + \frac{1}{2}(a(Y_{n+1}) + a(Y_n))\Delta + b(Y_n)\Delta W_n,$$

see (7.2.1). It has the transfer function

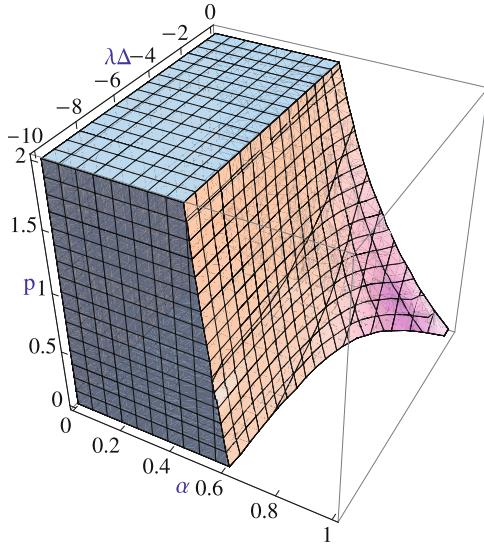


Fig. 14.3.1. Stability region for semi-drift-implicit Euler method

$$G_{n+1}(\lambda \Delta, \alpha) = \left| \frac{1 + \frac{1}{2} \left(1 - \frac{3}{2} \alpha\right) \lambda \Delta + \sqrt{-\alpha \lambda} \Delta W_n}{1 - \frac{1}{2} \left(1 - \frac{3}{2} \alpha\right) \lambda \Delta} \right|. \quad (14.3.1)$$

Its stability region is shown in Fig. 14.3.1. We can see from (14.3.1) that the transfer function involves in this case some division by α and $\lambda\Delta$. This kind of implicit scheme requires, in general, solving an algebraic equation at each time step in order to obtain the approximate solution. This extra computational effort makes the stability region significantly wider compared to those obtained for the previous predictor-corrector Euler schemes. For instance, for all considered values of p and $\lambda\Delta$, one obtains in Fig. 14.3.1 always asymptotic p -stability for a degree of stochasticity of up to $\alpha = 0.5$. Unfortunately, there is no such stability for $\alpha = 2/3$ for any value of p when $\lambda\Delta < -3$. This means that this numerical scheme may not be well suited for simulating martingale dynamics when the step size needs to remain relatively large. The symmetric predictor-corrector Euler method, with stability region shown in Fig. 14.2.5, appears to be better prepared for such a task.

Full-Drift-Implicit Euler Method

Now, let us also consider the full-drift-implicit Euler scheme

$$Y_{n+1} = Y_n + a(Y_{n+1})\Delta + b(Y_n)\Delta W_n,$$

see Sect. 7.2, with transfer function

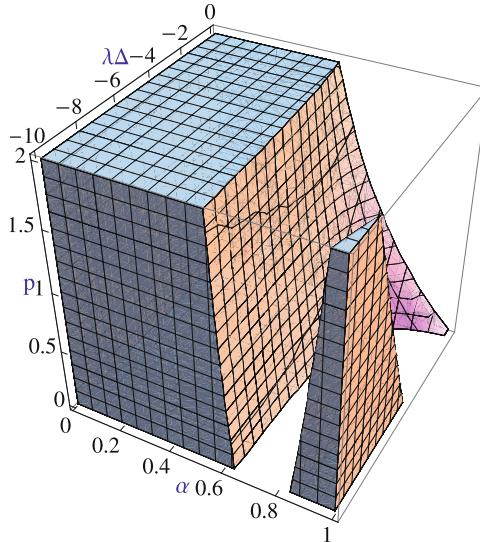


Fig. 14.3.2. Stability region for full-drift-implicit Euler method

$$G_{n+1}(\lambda \Delta, \alpha) = \left| \frac{1 + \sqrt{-\alpha \lambda} \Delta W_n}{1 - (1 - \frac{3}{2} \alpha) \lambda \Delta} \right|.$$

Fig. 14.3.2 shows its stability region which looks similar to the region obtained for the semi-drift-implicit Euler scheme. However, it has an additional area of stability for α close to one and $\lambda \Delta$ near -10. Additionally, this region diminishes as p becomes larger and does not cover the martingale case $\alpha = 2/3$. It seems that the region of stability is likely to increase with the degree of implicitness in the drift term of an implicit Euler scheme. However, this increase of asymptotic p -stability appears to weaken as p becomes large.

Balanced Implicit Euler Method

Finally, we mention in this section the balanced implicit Euler method, proposed in Milstein et al. (1998), see Sect. 7.3. We study this scheme in the particular form

$$Y_{n+1} = Y_n + \left(1 - \frac{3}{2} \alpha\right) \lambda Y_n \Delta + \sqrt{\alpha |\lambda|} Y_n \Delta W_n + c |\Delta W_n| (Y_n - Y_{n+1}).$$

When c is chosen to be $\sqrt{-\alpha \lambda}$, then its transfer function equals

$$G_{n+1}(\lambda \Delta, \alpha) = \left| \frac{1 + (1 - \frac{3}{2} \alpha) \lambda \Delta + \sqrt{-\alpha \lambda} (\Delta W_n + |\Delta W_n|)}{1 + \sqrt{-\alpha \lambda} |\Delta W_n|} \right|.$$

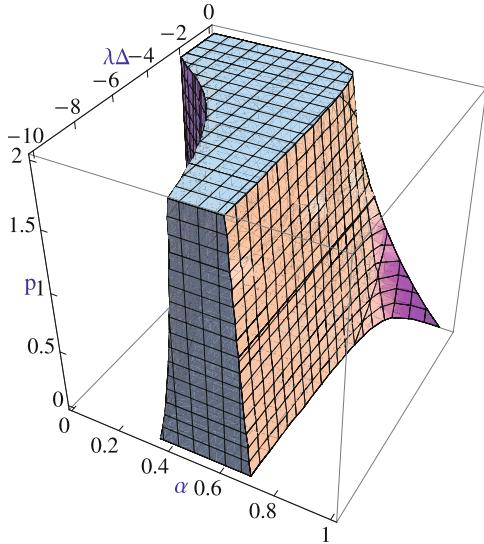


Fig. 14.3.3. Stability region for a balanced implicit Euler method

The corresponding stability region is shown in Fig. 14.3.3.

It turns out that this is the only type of method we have considered so far that provides in the martingale case $\alpha = 2/3$ for small p , asymptotic p -stability for all $\lambda\Delta$ we analyze, this means, also for large step sizes. Consequently, the balanced implicit Euler method is well suited for the scenario simulation of martingales, for which it was originally designed, see [Milstein et al. \(1998\)](#). Note that one has flexibility in constructing a range of balanced implicit methods, which potentially allows one to influence the stability region.

14.4 Stability of Simplified Schemes

Simplified Euler Method

All the schemes considered so far were originally designed for scenario simulation, see [Kloeden & Platen \(1999\)](#). However, Monte Carlo simulation is extremely important in finance and requires only weak approximations. As we discussed earlier in Sect. 11.2, the Wiener process increments ΔW_n , appearing in a Taylor scheme, can be typically replaced in a weak scheme by simpler multi-point random variables. These only need to satisfy certain moment matching conditions for the required weak order of convergence. In most of the previously analyzed schemes in this chapter, to obtain a weak order 1.0 method the Gaussian random variable ΔW_n may be replaced by a simpler two-point distributed random variable $\hat{\Delta W}_n$, which has the probabilities

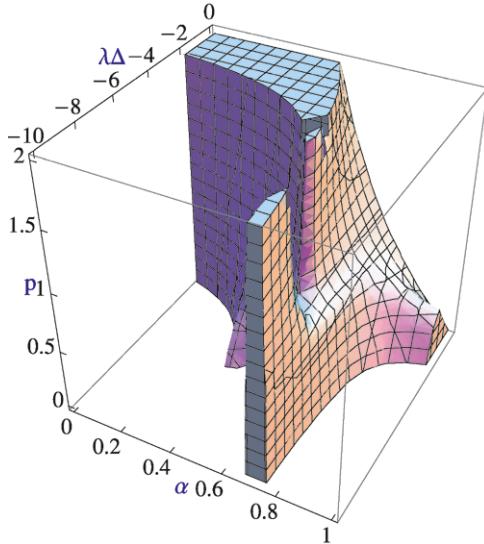


Fig. 14.4.1. Stability region for the simplified symmetric Euler method

$$P(\Delta \hat{W}_n = \pm \sqrt{\Delta}) = \frac{1}{2}. \quad (14.4.1)$$

Platen & Shi (2008) have investigated the impact of such simplification of the random variables in simplified weak schemes on corresponding stability regions. The resulting stability regions look rather similar to those of the corresponding previously studied schemes with Gaussian random variables. In most cases, the stability region increases slightly. To provide an example, Fig. 14.4.1 shows the stability region of the simplified symmetric predictor-corrector Euler scheme. It is, of course, different from the one displayed in Fig. 14.2.5. When comparing these two figures, the simplified scheme shows an increased stability region, in particular near the critical level of $\alpha \approx 2/3$, which is critical for the Monte Carlo simulation of martingales. In general, one can say that simplified schemes usually increase numerical stability. The stability region of the simplified balanced implicit Euler method shows a stability region similar to the one already presented in Fig. 14.3.3, and is therefore, not displayed here.

Simplified Implicit Euler Scheme

As pointed out in Sect. 11.5, terms that approximate the diffusion coefficient can be made implicit in a simplified weak scheme. The reason being that in such a scheme instead of an unbounded random variable ΔW_n one uses a bounded random variable $\Delta \hat{W}_n$. Let us now consider the family of simplified implicit Euler schemes, see Sect. 11.5, given in the form

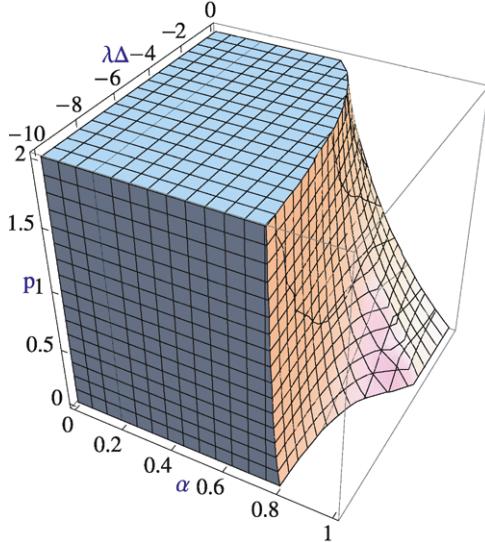


Fig. 14.4.2. Stability region for the simplified symmetric implicit Euler scheme

$$\begin{aligned} Y_{n+1} = Y_n + & \{\theta \bar{a}_\eta(Y_{n+1}) + (1 - \theta) \bar{a}_\eta(Y_n)\} \Delta \\ & + \{\eta b(Y_{n+1}) + (1 - \eta) b(Y_n)\} \Delta \hat{W}_n, \end{aligned} \quad (14.4.2)$$

for $\theta, \eta \in [0, 1]$ and $\bar{a}_\eta = a - \eta bb'$. When comparing this scheme with the strong predictor-corrector Euler scheme (14.2.1), one notices a resemblance to its corrector part. For sufficiently small step size Δ , the implicitness in the diffusion term makes sense for this scheme. Its transfer function is of the form

$$G_{n+1}(\lambda \Delta, \alpha) = \frac{1 + (1 + (\eta - \frac{3}{2})\alpha)\lambda(1 - \theta)\Delta + (1 - \eta)\sqrt{\alpha|\lambda|}\Delta \hat{W}_n}{1 - (1 + (\eta - \frac{3}{2})\alpha)\lambda\theta\Delta - \eta\sqrt{\alpha|\lambda|}\Delta \hat{W}_n}. \quad (14.4.3)$$

Platen & Shi (2008) have studied the resulting stability regions for this family of schemes. For instance, the stability region for the simplified semi-drift-implicit Euler scheme resembles that in Fig. 14.3.1, and that of the simplified full-drift-implicit Euler method shown in Fig. 14.3.2. The simplified symmetric implicit Euler scheme with $\theta = \eta = 0.5$ in (14.4.2) has its stability region displayed in Fig. 14.4.2 and that of the simplified fully implicit Euler method is shown in Fig. 14.4.3. We observe that both stability regions fill almost the entire area covered. For this reason, a simplified symmetric or fully implicit Euler scheme could be able to overcome potential numerical instabilities in a Monte Carlo simulation that are otherwise difficult to master. Note that we achieve here asymptotic p -stability even in some situations where the test dynamics are unstable, which is interesting to know.

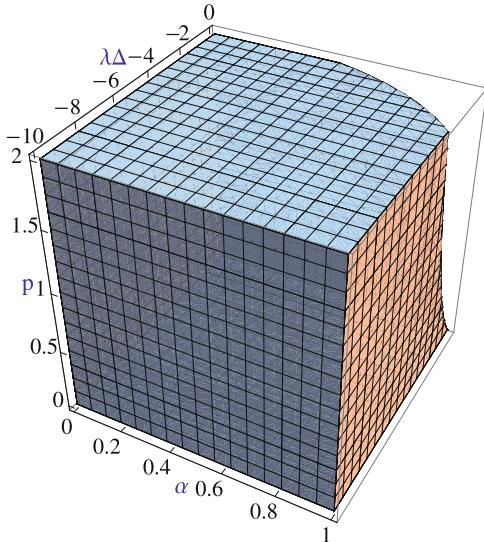


Fig. 14.4.3. Stability region for the simplified fully implicit Euler scheme

One has to be aware of the fact that for the simplified symmetric implicit Euler scheme, see Fig. 14.4.2, problems may arise for small step sizes. For instance, for the martingale case, that is $\alpha = 2/3$, a problem occurs at $p = 2$, the stability region is decreased by the use of a time step size that is too small. A straightforward calculation shows by using (14.4.3) for $\alpha = \eta = \theta = 1$ that the fully implicit Euler scheme is not asymptotically p -stable for $p = 2$ when the step size is below $\frac{2}{|\lambda|}$, as can be also visually confirmed from Fig. 14.4.3, see Exercise 14.2. This means, for an implementation, which may have been successfully working for some given large time step size in a derivative pricing tool, by decreasing the time step size, often expecting a more accurate valuation, one may in fact create severe numerical stability problems. This type of effect has already been observed for several other schemes. We emphasize that refining the time step size in a Monte Carlo simulation can lead to numerical stability problems.

Testing Second Moments

Finally, let us perform some Monte Carlo simulations with the simplified Euler scheme and the simplified semi-drift-implicit Euler scheme, where we estimate the second moment of X_T for $T \in [3, 729]$, $\lambda = -1$, $\alpha = 0.49$ and $\Delta = 3$. We simulated $N = 1,000,000$ sample paths to obtain extremely small confidence intervals, which are subsequently neglected in Table 14.4.1.

Table 14.4.1 shows the second moment exact solution, simplified Euler scheme estimate \tilde{Y}_T , and simplified semi-drift-implicit Euler scheme estimate

T	3	9	27	81	243	729
$E(X_T^2)$	0.89	0.7	0.34	0.039	0.00006	2.17×10^{-13}
$E(\tilde{Y}_T^2)$	1.51	3.47	41.3	69796	3.47×10^{14}	9.83×10^{42}
$E(\hat{Y}_T^2)$	0.94	0.83	0.57	0.069	5.07×10^{-8}	3.0×10^{-30}

Table 14.4.1. Second moment of the test dynamics calculated using the exact solution X_T ; the simplified Euler scheme \tilde{Y}_T ; and simplified semi-drift-implicit Euler scheme \hat{Y}_T

\hat{Y}_T in that order. One can observe that the Monte Carlo simulation with the simplified Euler scheme diverges and is not satisfactory, whereas the simplified semi-drift-implicit Euler scheme estimate is much more realistic for a time horizon that is not too far out. The better performance is explained by the fact that this scheme remains for $p = 2$ in its stability region, whereas the simplified Euler scheme is for the given parameter set in a region where there is no asymptotic p -stability for $p = 2$.

14.5 Exercises

14.1. Identify for the Euler scheme and asymptotic p -stability with $p = 2$ the boundary of its stability region for values $\lambda\Delta$ as a function of α .

14.2. Describe the step sizes for which the fully implicit simplified Euler scheme for the test SDE with $\alpha = 1$ is asymptotically p -stable for $p = 1$.

Martingale Representations and Hedge Ratios

The calculation of hedge ratios is fundamental to both the valuation of derivative securities and also the risk management procedures needed to replicate these instruments. In Monte Carlo simulation the following results on martingale representations and hedge ratios will be highly relevant. In this chapter we follow closely [Heath \(1995\)](#) and consider the problem of finding explicit Itô integral representations of the payoff structure of derivative securities. If such a representation can be found, then the corresponding hedge ratio can be identified and numerically calculated. For simplicity, we focus here on the case without jumps. The case with jumps is very similar.

15.1 General Contingent Claim Pricing

In this section we introduce a general setting which expresses the price dynamics of a derivative security as the conditional expectation of the payoff structure of the security under a suitable pricing measure.

Financial Market Model

Let $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \geq 0\}$ be an m -dimensional Brownian motion defined on the probability space $(\Omega, \mathcal{A}_T, \mathcal{A}, P)$. We assume that the filtration $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0, T]}$ is the P -augmentation of the natural filtration of \mathbf{W} , where we work with a bounded time horizon $T \in [0, \infty)$. These conditions ensure that $\underline{\mathcal{A}}$ satisfies the usual conditions, see Sect. [1.2](#).

Let $\mathbf{X}^{t_0, \underline{x}} = \{\mathbf{X}_t^{t_0, \underline{x}} = (X_t^{1, t_0, \underline{x}}, \dots, X_t^{d, t_0, \underline{x}})^\top, t \in [t_0, T]\}$ be a d -dimensional diffusion process that describes the factors in our financial market model and whose components satisfy the SDE

$$dX_t^{i, t_0, \underline{x}} = a^i(t, \mathbf{X}_t^{t_0, \underline{x}}) dt + \sum_{j=1}^m b^{i,j}(t, \mathbf{X}_t^{t_0, \underline{x}}) dW_t^j \quad (15.1.1)$$

for $t \in [t_0, T]$, $i \in \{1, 2, \dots, d\}$. Here $\mathbf{X}^{t_0, \underline{x}}$ starts at time t_0 with initial value $\underline{x} = (\underline{x}_1, \dots, \underline{x}_d)^\top \in \Re^d$. We assume that appropriate growth and Lipschitz conditions apply for the drift a^i and diffusion coefficients $b^{i,j}$ so that (15.1.1) admits a unique strong solution and is Markovian, see Chap. 1.

In order to model the chosen numéraire in the market model we may select the first component $X^{1,t_0,\underline{x}} = \beta = \{\beta_t, t \in [t_0, T]\}$ to model its price movements. Under real world pricing, β will be the growth optimal portfolio (GOP). However, under risk neutral pricing we use the savings account as numéraire. Both cases are covered by the following analysis.

The vector process $\mathbf{X}^{t_0, \underline{x}}$ could model several risky assets or factors that drive the securities, as well as, components that provide additional specifications or features of the model such as stochastic volatility, market activity, inflation or averages of risky assets for Asian options, etc.

General Contingent Claims

In order to build a setting that will support American option pricing and certain exotic option valuations, as well as hedging, we consider a stopping time formulation as follows:

Let $\Gamma_0 \subset [t_0, T] \times \Re^d$ be some region with $\Gamma_0 \cap (t_0, T] \times \Re^d$ an open set and define a stopping time $\tau : \Omega \rightarrow \Re$ by

$$\tau = \inf\{t > t_0 : (t, \mathbf{X}_t^{t_0, \underline{x}}) \notin \Gamma_0\}. \quad (15.1.2)$$

Using the stopping time τ we define the region

$$\Gamma_1 = \left\{ \left(\tau(\omega), \mathbf{X}_{\tau(\omega)}^{t_0, \underline{x}}(\omega) \right) \in [t_0, T] \times \Re^d : \omega \in \Omega \right\}.$$

Γ_1 contains all points of the boundary of Γ_0 which can be reached by the process $\mathbf{X}^{t_0, \underline{x}}$. We now consider contingent claims with payoff structures of the form

$$H_\tau = h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}),$$

where $h : \Gamma_1 \rightarrow \Re$ is some payoff function.

Using a terminology that is applied often for American option pricing, we call the set Γ_0 the *continuation region* and Γ_1 the *exercise boundary*, which forms part of the stopping region. For a process $\mathbf{X}^{t_0, \underline{x}}$ with continuous sample paths, an option is considered ‘alive’ at time $s \in [t_0, T]$, if $(s, \mathbf{X}_s^{t_0, \underline{x}}) \in \Gamma_0$. On the other hand, it is ‘exercised’ or ‘stopped’ at the first time $s \in [t_0, T]$ that $(s, \mathbf{X}_s^{t_0, \underline{x}})$ touches the exercise boundary Γ_1 . It is assumed that $(t_0, \underline{x}) \in \Gamma_0$, since otherwise the derivative security would be immediately ‘exercised’.

For example, if we take $\Gamma_0 = [t_0, T] \times \Re^d$, which implies $\tau = T$ and payoff structures of the form $h(T, \mathbf{X}_T^{t_0, \underline{x}})$, then this formulation reduces to the case of a multi-dimensional European style contingent claim. More generally, in the case of a process with jumps an option is ‘exercised’ or ‘stopped’ at the first time $s \in [t_0, T]$ that $(s, \mathbf{X}_s^{t_0, \underline{x}})$ leaves the continuation region.

Valuation Function

When applying the Markov property, the option pricing or *valuation function* $u : \Gamma_0 \cup \Gamma_1 \rightarrow \mathbb{R}$ corresponding to the payoff structure $h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})$ is given by

$$u(t, \mathbf{x}) = \tilde{E} \left(\frac{\beta_t}{\beta_T} h(\tau, \mathbf{X}_\tau^{t, \mathbf{x}}) \right) \quad (15.1.3)$$

for $(t, \mathbf{x}) \in \Gamma_0 \cup \Gamma_1$. Here \tilde{E} denotes expectation under the appropriately defined pricing measure \tilde{P} . When β_t is chosen to be the savings account, then \tilde{P} has to be chosen to be an equivalent risk neutral probability measure. A good choice in this case is the, so called, *minimal equivalent martingale measure*, which can be used for incomplete markets, as described in Föllmer & Schweizer (1991), Schweizer (1991) and Hofmann, Platen & Schweizer (1992). When performing real world pricing in an incomplete market under the existence of an equivalent risk neutral probability measure, then real world pricing is equivalent to the choice of the equivalent risk neutral martingale measure. If the financial market model does not admit an equivalent risk neutral probability measure, then the real world pricing concept of the benchmark approach can still be applied and the expectation of benchmarked contingent claims is taken with respect to the real world probability measure P , see Chap. 3 and Sect. 10.5. In this case β_t is the numéraire portfolio. The analysis and numerical challenges remain in both cases very similar.

Define the discounted or benchmarked functions $\bar{h} : \Gamma_1 \rightarrow \mathbb{R}$ and $\bar{u} : \Gamma_0 \cup \Gamma_1 \rightarrow \mathbb{R}$ by

$$\begin{aligned} \bar{h}(s, \mathbf{y}) &= \frac{1}{y_1} h(s, \mathbf{y}) \\ \bar{u}(t, \mathbf{x}) &= \tilde{E} (\bar{h}(\tau, \mathbf{X}_\tau^{t, \mathbf{x}})), \end{aligned} \quad (15.1.4)$$

respectively, for $(s, \mathbf{y}) \in \Gamma_1$, $(t, \mathbf{x}) \in \Gamma_0 \cup \Gamma_1$ with $\mathbf{y} = (y_1, \dots, y_d)^\top \in \mathbb{R}^d$, where we recall that $X_t^{1, t_0, \underline{x}}$ models the price movements of the numéraire β .

Let $Z^{t_0, t_0} = \{Z_t^{t_0, t_0}, t \in [t_0, T]\}$ be the solution of the SDE

$$dZ_t^{t_0, t_0} = \mathbf{1}_{\{t < \tau\}} dt \quad (15.1.5)$$

for $t \in [t_0, T]$, starting at time t_0 with initial value t_0 . We can write the solution to (15.1.5) in the form

$$Z_t^{t_0, t_0} = t \wedge \tau, \quad (15.1.6)$$

for $t \in [t_0, T]$. This expression together with the uniqueness of the solutions of (15.1.1) and (15.1.5) shows that

$$\mathbf{X}_\tau^{t_0, \underline{x}} = \mathbf{X}_\tau^{t \wedge \tau, X_{t \wedge \tau}^{t_0, \underline{x}}} \quad (15.1.7)$$

and

$$\tau = Z_{\tau}^{t_0, t_0} = Z_{\tau}^{t \wedge \tau, Z_{t \wedge \tau}^{t_0, t_0}} \quad (15.1.8)$$

for $t \in [t_0, T]$. Using these equalities, the functional (15.1.3), the Markov property and the assignment $\beta_{\tau} = X_{\tau}^{1, t_0, \underline{x}}$ with $\underline{x}_1 = 1$, we have

$$\begin{aligned} u_t &= u(t \wedge \tau, \mathbf{X}_{t \wedge \tau}^{t_0, \underline{x}}) \\ &= \tilde{E} \left(\frac{\beta_{t \wedge \tau}}{\beta_{\tau}} h \left(Z_{\tau}^{t \wedge \tau, Z_{t \wedge \tau}^{t_0, t_0}}, \mathbf{X}_{\tau}^{t \wedge \tau, X_{t \wedge \tau}^{t_0, \underline{x}}} \right) \right) \\ &= \tilde{E} \left(\frac{\beta_{t \wedge \tau}}{\beta_{\tau}} h \left(Z_{\tau}^{t \wedge \tau, Z_{t \wedge \tau}^{t_0, t_0}}, \mathbf{X}_{\tau}^{t \wedge \tau, X_{t \wedge \tau}^{t_0, \underline{x}}} \right) \middle| \mathcal{A}_t \right) \\ &= \beta_{t \wedge \tau} \tilde{E} \left(\frac{1}{\beta_{\tau}} h \left(Z_{\tau}^{t \wedge \tau, Z_{t \wedge \tau}^{t_0, t_0}}, \mathbf{X}_{\tau}^{t \wedge \tau, X_{t \wedge \tau}^{t_0, \underline{x}}} \right) \middle| \mathcal{A}_t \right) \\ &= \beta_{t \wedge \tau} \tilde{E} \left(\frac{1}{\beta_{\tau}} h \left(Z_{\tau}^{t \wedge \tau, Z_{t \wedge \tau}^{t_0, t_0}}, \mathbf{X}_{\tau}^{t \wedge \tau, X_{t \wedge \tau}^{t_0, \underline{x}}} \right) \right) \\ &= \beta_{t \wedge \tau} \tilde{E} \left(\frac{1}{\beta_{\tau}} h(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}}) \right) \\ &= \beta_{t \wedge \tau} \tilde{E} (\bar{h}(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}})) \\ &= \beta_{t \wedge \tau} \tilde{E} \left(\bar{h} \left(\tau, \mathbf{X}_{\tau}^{t \wedge \tau, X_{t \wedge \tau}^{t_0, \underline{x}}} \right) \right) \\ &= \beta_{t \wedge \tau} \bar{u}(t \wedge \tau, \mathbf{X}_{t \wedge \tau}^{t_0, \underline{x}}) \end{aligned} \quad (15.1.9)$$

for $(t \wedge \tau, \mathbf{X}_{t \wedge \tau}^{t_0, \underline{x}}) \in \Gamma_0$.

Define the $(\underline{A}, \tilde{P})$ -martingale $M = \{M_t : t \in [t_0, T]\}$ by

$$M_t = \tilde{E} \left(\bar{h}(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}}) \middle| \mathcal{A}_t \right), \quad (15.1.10)$$

for $t \in [t_0, T]$. We assume that an appropriate growth condition applies for \bar{h} so that the conditional expectation in (15.1.10) is well-defined. Applying once again equation (15.1.7), the Markov property and the definitions of \bar{h} and \bar{u} we have

$$\begin{aligned} M_t &= \tilde{E} \left(\bar{h} \left(\tau, \mathbf{X}_{\tau}^{t \wedge \tau, X_{t \wedge \tau}^{t_0, \underline{x}}} \right) \middle| \mathcal{A}_t \right) \\ &= \tilde{E} \left(\bar{h} \left(\tau, \mathbf{X}_{\tau}^{t \wedge \tau, X_{t \wedge \tau}^{t_0, \underline{x}}} \right) \right) \\ &= \bar{u}(t \wedge \tau, \mathbf{X}_{t \wedge \tau}^{t_0, \underline{x}}) \end{aligned} \quad (15.1.11)$$

for $t \in [t_0, T]$.

Consequently, the discounted or benchmarked valuation process

$$\bar{u} = \left\{ \bar{u}_t = \bar{u} \left(t \wedge \tau, \mathbf{X}_{t \wedge \tau}^{t_0, x} \right), t \in [t_0, T] \right\}$$

is an $(\underline{\mathcal{A}}, \tilde{P})$ -martingale. Also, from (15.1.9) we can determine values for the random variable u_t if the corresponding values for $\beta_{t \wedge \tau}$ and \bar{u}_t are known. Usually it is more convenient to compute prices via the function \bar{u} rather than u . This is mainly because the $(\underline{\mathcal{A}}, \tilde{P})$ -martingale property associated with \bar{u} enables us to apply a number of powerful results from stochastic calculus. In particular, subject to certain integrability conditions, the price process corresponding to \bar{u} will admit a *martingale representation*. From this, hedging parameters can be determined either in an implicit or explicit form. Some specific examples of hedging strategies will be considered when dealing with particular applications, see also Sect. 10.5.

The above analysis, leading in particular to equations (15.1.9) and (15.1.11), demonstrates that the valuation of contingent claims, as given by (15.1.3), can be reduced to valuations of the form

$$\bar{u}_t = \bar{u} \left(t \wedge \tau, \mathbf{X}_{t \wedge \tau}^{t_0, x} \right) = \tilde{E} \left(\bar{h} \left(\tau, \mathbf{X}_{\tau}^{t_0, x} \right) \mid \mathcal{A}_t \right) \quad (15.1.12)$$

for some payoff function $\bar{h} : \Gamma_1 \rightarrow \mathfrak{R}$. Consequently, in the remaining part of this chapter we will assume this type of structure for our pricing and hedging problems. For simplicity, we omit then the previous hint of a potential measure change and simply consider the problem of computing the conditional expectation

$$u \left(t, \mathbf{X}_t^{t_0, x} \right) = E \left(h \left(\tau, \mathbf{X}_{\tau}^{t_0, x} \right) \mid \mathcal{A}_t \right) \quad (15.1.13)$$

for $t \in [t_0, \tau]$.

In the special case, where $\tau = T$ is the bounded deterministic time horizon and the payoff structure takes the form $h(\mathbf{X}_T^{t_0, x})$, we will refer to the corresponding equations for (15.1.3) and (15.1.9)–(15.1.12) as the time-independent formulations.

15.2 Hedge Ratios for One-dimensional Processes

This section considers a simple one-dimensional dynamic for the risky security and describes an explicit form for the hedge ratios of given contingent claims.

A One-dimensional Model

To allow for an easier and simpler exposition of the underlying ideas, we assume in this section that $W = \{W_t, t \in [t_0, T]\}$ is a one-dimensional Wiener process defined on the filtered probability space $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$. Consider the one-dimensional SDE

$$dX_t = a(t, X_t) dt + b(t, X_t) dW_t \quad (15.2.1)$$

for $t \in [t_0, T]$. Here $a, b : [t_0, T] \times \mathfrak{R} \rightarrow \mathfrak{R}$ are measurable functions which, including partial derivatives, have linear growth, and are Lipschitz continuous in x . We denote by $X^{t_0, \underline{x}} = \{X_s^{t_0, \underline{x}}, s \in [t_0, T]\}$ the solution of (15.2.1) starting at $\underline{x} \in \mathfrak{R}$ at time t_0 .

Martingale Representation

We consider a European style contingent claim with payoff structure of the form

$$h(X_T^{t_0, \underline{x}}) = h(T, X_T^{t_0, \underline{x}})$$

for $\underline{x} \in \mathfrak{R}$ and maturity T . We emphasize that the only uncertainty in the model under consideration is assumed to be given by the Wiener process W .

Subject to certain growth bounds for the function h , the above conditions for the coefficients a and b ensure, see Sect. 1.9 and Platen & Heath (2006), that

$$E((h(X_T^{t_0, \underline{x}}))^2) < \infty$$

and the process $M = \{M_t, t \in [t_0, T]\}$, defined by

$$M_t = E\left(h\left(X_T^{t_0, \underline{x}}\right) \mid \mathcal{A}_t\right) \quad (15.2.2)$$

for $t \in [t_0, T]$, is a square integrable martingale, and therefore, admits a *martingale representation* of the form

$$M_t = M_0 + \int_{t_0}^t \xi_s dW_s, \quad (15.2.3)$$

where ξ is an $\underline{\mathcal{A}}$ -predictable process with

$$E\left(\int_{t_0}^T \xi_s^2 ds\right) < \infty.$$

The process ξ is unique in the sense that if

$$M_t = M_0 + \int_{t_0}^t \tilde{\xi}_s dW_s$$

for some other $\underline{\mathcal{A}}$ -predictable process $\tilde{\xi}$, then

$$E\left(\left(\int_{t_0}^t (\xi_s - \tilde{\xi}_s) dW_s\right)^2\right) = \int_{t_0}^t E((\xi_s - \tilde{\xi}_s)^2) ds = 0.$$

A general statement concerning this uniqueness property can be found in Karatzas & Shreve (1991).

Integrand in a Martingale Representation

Finding explicit expressions for the *integrand* ξ in the martingale representation (15.2.3) is of considerable practical value, as it is closely related to the computation of hedge ratios in option pricing. Here we seek explicit characterizations based on an application of the Kolmogorov backward equation, see Sect. 2.7.

Define the scalar function $u : [t_0, T] \times \mathfrak{R} \rightarrow \mathfrak{R}$ by

$$u(t, x) = E(h(X_T^{t,x})), \quad (15.2.4)$$

for $(t, x) \in [t_0, T] \times \mathfrak{R}$. We assume that the function u is of class $C^{1,3}$, that is, continuously differentiable with respect to t and three times continuously differentiable in x . Expanding

$$u(T, X_T^{t_0, \underline{x}}) = h(X_T^{t_0, \underline{x}}) \quad (15.2.5)$$

by the Itô formula and applying the Kolmogorov backward equation yields

$$u\left(t, X_t^{t_0, \underline{x}}\right) = u(t_0, \underline{x}) + \int_{t_0}^t b\left(s, X_s^{t_0, \underline{x}}\right) \frac{\partial}{\partial x} u\left(s, X_s^{t_0, \underline{x}}\right) dW_s \quad (15.2.6)$$

for $t \in [t_0, T]$, see Sect. 2.7. Using the time-independent formulations of (15.1.10), (15.1.11) and (15.2.6) we have

$$\begin{aligned} M_t &= E\left(h\left(X_T^{t_0, \underline{x}}\right) \mid \mathcal{A}_t\right) \\ &= u(t, X_t^{t_0, \underline{x}}) \\ &= u(t_0, \underline{x}) + \int_{t_0}^t b\left(s, X_s^{t_0, \underline{x}}\right) \frac{\partial}{\partial x} u\left(s, X_s^{t_0, \underline{x}}\right) dW_s \end{aligned} \quad (15.2.7)$$

for $t \in [t_0, T]$. This result can also be obtained by applying Itô's formula to (15.2.5), taking the conditional expectation on both sides of the resulting equation, and using the relations (15.2.2) and (15.2.6). Consequently, we obtain

$$M_{t_0} = u(t_0, \underline{x})$$

and the *integrand* in the form

$$\xi_s = b\left(s, X_s^{t_0, \underline{x}}\right) \frac{\partial}{\partial x} u\left(s, X_s^{t_0, \underline{x}}\right).$$

We now have a martingale representation of the form (15.2.3). However, this expression for the integrand ξ requires the solution of the valuation equation (15.2.4) to be known.

Alternative Characterization of the Integrand

A substantial amount of literature has been developed pertaining to the question of determining the integrand in a martingale representation, which even exploits Malliavin calculus, see for instance [Malliavin \(1997\)](#), [Fournie, Lasry, Lebuchoux, Lions & Touzi \(1999, 2001\)](#), [Avellaneda & Gamba \(2002\)](#), [Bernis, Gobet & Kohatsu-Higa \(2003\)](#), [Gobet & Kohatsu-Higa \(2003\)](#), [El-Khatib & Privault \(2004\)](#), [Bally, Caramellino & Zanette \(2005\)](#) and [Davis & Johansson \(2006\)](#) who use direct numerical simulation. In practical applications, one often uses finite differences of option prices to approximate numerically the partial derivatives $\frac{\partial}{\partial x} u(s, X_s^{t_0, \underline{x}})$ and from these ξ_s . We will now describe a simple alternative characterization of the integrand ξ , which does not depend directly on the solution of (15.2.4), and goes back to [Heath \(1995\)](#). Let L^0 and $\frac{\partial}{\partial x} L^0$ be the operators

$$\begin{aligned} L^0 &= \frac{\partial}{\partial s} + a \frac{\partial}{\partial x} + \frac{1}{2} b^2 \frac{\partial^2}{\partial x^2} \\ \frac{\partial}{\partial x} L^0 &= \frac{\partial^2}{\partial s \partial x} + \frac{\partial a}{\partial x} \frac{\partial}{\partial x} + \left(a + b \frac{\partial b}{\partial x} \right) \frac{\partial^2}{\partial x^2} + \frac{1}{2} b^2 \frac{\partial^3}{\partial x^3}, \end{aligned} \quad (15.2.8)$$

where the operator $\frac{\partial}{\partial x} L^0$ is obtained by computing the partial derivative of L^0 with respect to x . The Kolmogorov backward equation can now be written in the form

$$L^0 u(s, x) = 0,$$

for $(s, x) \in (t_0, T) \times \mathfrak{R}$ with terminal condition

$$u(T, x) = h(x) \quad (15.2.9)$$

for $x \in \mathfrak{R}$, so that

$$\frac{\partial}{\partial x} L^0(u(s, x)) = 0 \quad (15.2.10)$$

for $(s, x) \in (t_0, T) \times \mathfrak{R}$, see Sect. 2.7.

Consider the *linearized stochastic differential equation*

$$dZ_t = Z_t \frac{\partial a(t, X_t)}{\partial x} dt + Z_t \frac{\partial b(t, X_t)}{\partial x} dW_t \quad (15.2.11)$$

for $t \in [t_0, T]$. Let $Z^{s, \underline{z}} = \{Z_t^{s, \underline{z}}, t \in [s, T]\}$ be the solution of (15.2.11) starting at $\underline{z} \in \mathfrak{R}$ at time s , $s \in [t_0, T]$.

Introducing the scalar function $v : [t_0, T] \times \mathfrak{R}^2 \rightarrow \mathfrak{R}$ defined by

$$v(t, x, z) = z \frac{\partial u(t, x)}{\partial x} \quad (15.2.12)$$

for $(t, x, z) \in [t_0, T] \times \Re^2$, and applying the multi-dimensional version of Itô's formula to $v(T, X_T^{t,x}, Z_T^{t,z})$ we have

$$\begin{aligned} v(T, X_T^{t,x}, Z_T^{t,z}) &= v(t, \underline{x}, \underline{z}) + \int_t^T \bar{L}^0 v(s, X_s^{t,x}, Z_s^{t,z}) ds \\ &\quad + \int_t^T \bar{L}^1 v(s, X_s^{t,x}, Z_s^{t,z}) dW_s \end{aligned} \quad (15.2.13)$$

for $t \in [t_0, T]$, where

$$\begin{aligned} \bar{L}^0 &= \frac{\partial}{\partial s} + a \frac{\partial}{\partial x} + \frac{\partial a}{\partial x} z \frac{\partial}{\partial z} \\ &\quad + \frac{1}{2} \left(b^2 \frac{\partial^2}{\partial x^2} + \left(\frac{\partial b}{\partial x} \right)^2 z^2 \frac{\partial^2}{\partial z^2} \right) + bz \frac{\partial b}{\partial x} \frac{\partial^2}{\partial x \partial z} \end{aligned}$$

and

$$\bar{L}^1 = b \frac{\partial}{\partial x} + z \frac{\partial b}{\partial x} \frac{\partial}{\partial z}.$$

Calculating the partial derivatives of the function v using (15.2.12) and applying (15.2.10) yields

$$\begin{aligned} \bar{L}^0 v(s, x, z) &= z \frac{\partial}{\partial x} L^0 u(s, x) \\ &= 0, \end{aligned} \quad (15.2.14)$$

for $(s, x, z) \in (t_0, T) \times \Re^2$.

We now assume that

$$E \left(\left| v(T, X_T^{t,x}, Z_T^{t,1}) \right| \right) < \infty$$

for all $(t, \underline{x}) \in [t_0, T] \times \Re$. Subject to certain growth bounds applying for the derivative $\frac{\partial h}{\partial x}$, this condition will be verified in Sect. 15.4.

Consequently, taking expectations on both sides of (15.2.13) and using (15.2.9) we have

$$\begin{aligned} \frac{\partial u(t, \underline{x})}{\partial \underline{x}} &= v(t, \underline{x}, 1) \\ &= E \left(v \left(T, X_T^{t,x}, Z_T^{t,1} \right) \right) \\ &= E \left(Z_T^{t,1} \frac{\partial}{\partial x} u \left(T, X_T^{t,x} \right) \right) \\ &= E \left(Z_T^{t,1} \frac{\partial}{\partial x} h \left(X_T^{t,x} \right) \right) \end{aligned} \quad (15.2.15)$$

for $(t, \underline{x}) \in (t_0, T) \times \mathfrak{X}$.

By substituting this result into (15.2.6) we obtain

$$u(T, X_T^{t_0, \underline{x}}) = u(t_0, \underline{x}) + \int_{t_0}^T E \left(Z_T^{s,1} \frac{\partial}{\partial x} h(X_T^{s, X_s^{t_0, \underline{x}}}) \right) b(s, X_s^{t_0, \underline{x}}) dW_s. \quad (15.2.16)$$

Applying the Markov property of X , see the remarks following (15.1.1), and (15.1.7) with $\tau = T$, this representation becomes

$$u(T, X_T^{t_0, \underline{x}}) = u(t_0, \underline{x}) + \int_{t_0}^T E \left(Z_T^{s,1} \frac{\partial}{\partial x} h(X_T^{t_0, \underline{x}}) \mid \mathcal{A}_s \right) b(s, X_s^{t_0, \underline{x}}) dW_s. \quad (15.2.17)$$

Using the relations

$$u(T, X_T^{t_0, \underline{x}}) = h(X_T^{t_0, \underline{x}})$$

and

$$u(t_0, \underline{x}) = E(h(X_T^{t_0, \underline{x}})),$$

with the latter following by taking expectations on both sides of (15.2.16), we can express (15.2.17) equivalently as an *alternative martingale representation* in the form

$$h(X_T^{t_0, \underline{x}}) = E \left(h(X_T^{t_0, \underline{x}}) \right) + \int_{t_0}^T E \left(Z_T^{s,1} \frac{\partial}{\partial x} h(X_T^{t_0, \underline{x}}) \mid \mathcal{A}_s \right) b(s, X_s^{t_0, \underline{x}}) dW_s. \quad (15.2.18)$$

Thus, we have obtained an alternative explicit characterization of the integrand ξ appearing in (15.2.3).

Unbiased Estimator

We see from (15.2.15) that the variate

$$Z_T^{t,1} \frac{\partial}{\partial x} h(X_T^{t, \underline{x}})$$

is an *unbiased estimator* of

$$\frac{\partial u(t, x)}{\partial x},$$

unlike finite difference approximations. Note also that if a Monte Carlo method is used to estimate the price functional u at the point $(s, X_s^{t_0, \underline{x}})$, $t \in [t_0, T]$, then the same simulation trajectories for $X^{t_0, \underline{x}}$ can be used, together with new ones for $Z^{s,1}$, to approximate $\frac{\partial u}{\partial x}$ at $(s, X_s^{t_0, \underline{x}})$. This procedure is usually more accurate since it is unbiased. It is also often more efficient, since only one simulation run is required, compared to at least two separate simulation runs needed for finite difference estimates.

Martingale representations of the type (15.2.16)–(15.2.18), under different conditions, have been obtained by Elliott & Kohlmann (1988) and Colwell, Elliott & Kopp (1991) who use the Markov property and the differentiability of solutions of Itô SDEs with respect to the initial conditions. Broadie & Glasserman (1997a) also derive explicit representations of hedge ratios in the case where the payoff structure is a standard European call option and the diffusion process $X^{t_0, \underline{x}}$ follows a one-dimensional geometric Brownian motion. The result presented relies on certain smoothness conditions and an application of the Kolmogorov backward equation. It has the advantage of being rather simple and straightforward. Furthermore, it can also be extended to include stopping time boundaries and multi-dimensional processes with jumps.

As noted by Colwell et al. (1991) similar results can be obtained by an application of the Haussmann (1978) integral representation theorem. In fact, Haussmann's integral representation theorem can be used for certain classes of path dependent securities where the payoff function h depends on whole trajectories of the diffusion process $X^{t_0, \underline{x}}$. A proof of Haussmann's theorem using Malliavin calculus techniques, which is related to the formula of Clark (1970), is given by Ocone (1984), see also Davis (1980) and Haussmann (1979).

The use of Haussmann's integral representation theorem has the disadvantage of being less direct and the conditions of the theorem are more difficult to establish. In fact, the results presented in this section are sufficient for many types of valuation and hedging problems that arise in practice. They can also be extended to a wide class of path dependent options and to the case with jumps.

15.3 Explicit Hedge Ratios

In practice, one is often confronted with the problem of computing hedge ratios for derivative securities associated with multi-component hedge portfolios. We demonstrate how one can successfully apply the ideas of the previous section.

Multi-dimensional Framework

To derive, in this section, explicit expressions for the hedge ratios in the multi-dimensional situation we are forced to use more complex notations and formulations. Let $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \geq 0\}$ be an m -dimensional Wiener process and $\mathbf{X}^{t_0, \underline{x}} = \{\mathbf{X}_t^{t_0, \underline{x}} = (X_t^{1, t_0, \underline{x}}, \dots, X_t^{d, t_0, \underline{x}})^\top, t \in [t_0, T]\}$ a d -dimensional process which satisfies the SDE (15.1.1).

Using the results obtained in Sect. 15.1 we let τ be a stopping time given by (15.1.2), corresponding to the continuation region Γ_0 and exercise boundary Γ_1 , with $(t_0, \underline{x}) \in \Gamma_0$. We assume that there is a payoff function $h : \Gamma_1 \rightarrow \mathfrak{R}$ and valuation function $u : \Gamma_0 \cup \Gamma_1 \rightarrow \mathfrak{R}$ with

$$u(t, \underline{x}) = E(h(\tau, \mathbf{X}_\tau^{t, \underline{x}})). \quad (15.3.1)$$

for $(t, \underline{x}) \in \Gamma_0 \cup \Gamma_1$. We will say that the function $f : \Gamma_0 \rightarrow \mathbb{R}$ is of class $C^{1,\ell}$, for integers $\ell \geq 1$, if f is continuously differentiable with respect to t and ℓ -times continuously differentiable with respect to the spatial variables x_1, \dots, x_d on the domain Γ_0 . In this section we assume that the valuation function u given by (15.3.1) is of the class $C^{1,3}$ for the continuation region Γ_0 .

As in the case for one-dimensional diffusion processes, see equation (15.2.6), we can apply a multi-dimensional version of the Itô formula and the Kolmogorov backward equation to obtain a martingale representation of the form

$$\begin{aligned} u(t \wedge \tau, \mathbf{X}_{t \wedge \tau}^{t_0, \underline{x}}) &= u(t_0, \underline{x}) + \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^{t \wedge \tau} b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) \frac{\partial}{\partial x_i} u(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s^j \\ &= u(t_0, \underline{x}) + \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^{t \wedge \tau} b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) \frac{\partial}{\partial x_i} u(s \wedge \tau, \mathbf{X}_{s \wedge \tau}^{t_0, \underline{x}}) dW_s^j \end{aligned} \quad (15.3.2)$$

for $t \in [t_0, T]$. Applying (15.1.10) and (15.1.11), with $h = \bar{h}$ and $u = \bar{u}$, we also have

$$\begin{aligned} M_t &= E(h(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}}) \mid \mathcal{A}_t) \\ &= u(t \wedge \tau, \mathbf{X}_{t \wedge \tau}^{t_0, \underline{x}}) \end{aligned} \quad (15.3.3)$$

so that $u(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}}) = h(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}})$.

Define the operators L^0 and $\frac{\partial}{\partial x_p} L^0$, $p \in \{1, 2, \dots, d\}$, by

$$\begin{aligned} L^0 &= \frac{\partial}{\partial s} + \sum_{i=1}^d a^i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,k=1}^d \sum_{j=1}^m b^{i,j} b^{k,j} \frac{\partial^2}{\partial x_i \partial x_k}, \\ \frac{\partial}{\partial x_p} L^0 &= \frac{\partial^2}{\partial x_p \partial s} + \sum_{i=1}^d \left(\frac{\partial a^i}{\partial x_p} \frac{\partial}{\partial x_i} + a^i \frac{\partial^2}{\partial x_p \partial x_i} \right) \\ &\quad + \frac{1}{2} \sum_{i,k=1}^d \sum_{j=1}^m \left(\frac{\partial b^{i,j}}{\partial x_p} b^{k,j} \frac{\partial^2}{\partial x_i \partial x_k} + b^{i,j} \frac{\partial b^{k,j}}{\partial x_p} \frac{\partial^2}{\partial x_i \partial x_k} + b^{i,j} b^{k,j} \frac{\partial^3}{\partial x_p \partial x_i \partial x_k} \right). \end{aligned}$$

Here the operators $\frac{\partial}{\partial x_p} L^0$, $p \in \{1, 2, \dots, d\}$, are obtained by calculating the partial derivatives of the operator L^0 with respect to x_p .

The Kolmogorov backward equation now takes the form

$$L^0 u(t, \mathbf{x}) = 0$$

for $(t, \mathbf{x}) \in \Gamma_0$ with boundary condition $u(\tau, \mathbf{x}) = h(\tau, \mathbf{x})$ for $(t, \mathbf{x}) \in \Gamma_1$. From this equation we also have

$$\frac{\partial}{\partial x_p} L^0 u(t, \mathbf{x}) = 0 \quad (15.3.4)$$

for $(t, \mathbf{x}) \in \Gamma_0$, $p \in \{1, 2, \dots, d\}$.

Linearized SDE

Let $\mathbf{Z}^{s, \underline{z}} = \{\mathbf{Z}_t^{s, \underline{z}} = (Z_t^{1,1,s,\underline{z}}, \dots, Z_t^{d,d,s,\underline{z}})^\top, t \in [s, T]\}$ be the solution of the d^2 -dimensional linearized SDE

$$dZ_t^{k,i} = \sum_{p=1}^d Z_t^{p,i,s,\underline{z}} \frac{\partial a^k(t, \mathbf{X}_t)}{\partial x_p} dt + \sum_{p=1}^d \sum_{j=1}^m Z_t^{p,i,s,\underline{z}} \frac{\partial b^{k,j}(t, \mathbf{X}_t)}{\partial x_p} dW_t^j \quad (15.3.5)$$

for $t \in [s, T]$, $k, i \in \{1, 2, \dots, d\}$. We assume that $\mathbf{Z}^{s, \underline{z}}$ starts at time $s \in [t_0, T]$ with initial value $\underline{z} = (z_{1,1}, \dots, z_{d,d})^\top \in \mathbb{R}^{d^2}$.

For $i \in \{1, 2, \dots, d\}$ we introduce the scalar functions $v_i : \Gamma_0 \cup \Gamma_1 \times \mathbb{R}^{d^2} \rightarrow \mathbb{R}$ defined by

$$v_i(t, \mathbf{x}, \mathbf{z}) = \sum_{p=1}^d z_{p,i} \frac{\partial u(t, \mathbf{x})}{\partial x_p} \quad (15.3.6)$$

for $(t, \mathbf{x}, \mathbf{z}) \in \Gamma_0 \times \mathbb{R}^{d^2}$ with $\mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$, $\mathbf{z} = (z_{1,1}, \dots, z_{d,d})^\top \in \mathbb{R}^{d^2}$.

Let i , with $i \in \{1, 2, \dots, d\}$ be fixed. Applying the Itô formula to v_i at time τ and $t \wedge \tau$ for $t \in [t_0, T]$, using the system of processes $\mathbf{X} = (X^1, \dots, X^d)^\top$ and $\mathbf{Z}^i = (Z^{1,i}, \dots, Z^{d,i})^\top$ and noting that $\frac{\partial}{\partial z_{p,i} \partial z_{q,i}} v_i(t, \mathbf{x}, \mathbf{z}) = 0$ for $p, q \in \{1, 2, \dots, d\}$ and $(t, \underline{\mathbf{x}}, \underline{\mathbf{z}})^\top \in \Gamma_0 \times \mathbb{R}^{d^2}$ yields

$$\begin{aligned} v_i(\tau, \mathbf{X}_\tau^{t,x}, \mathbf{Z}_\tau^{t,\underline{z}}) &= v_i(t, \underline{\mathbf{x}}, \underline{\mathbf{z}}) + \int_{t \wedge \tau}^\tau \bar{L}_i^0 v_i(s, \mathbf{X}_s^{t,x}, \mathbf{Z}_s^{t,\underline{z}}) ds \\ &\quad + \sum_{j=1}^m \int_{t \wedge \tau}^\tau \bar{L}_i^j v_i(s, \mathbf{X}_s^{t,x}, \mathbf{Z}_s^{t,\underline{z}}) dW_s^j, \end{aligned} \quad (15.3.7)$$

where

$$\begin{aligned} \bar{L}_i^0 &= \frac{\partial}{\partial s} + \sum_{\ell=1}^d a^\ell \frac{\partial}{\partial x_\ell} + \sum_{k=1}^d \left(\sum_{p=1}^d z_{p,i} \frac{\partial a^k}{\partial x_p} \right) \frac{\partial}{\partial z_{k,i}} \\ &\quad + \frac{1}{2} \sum_{\ell,k=1}^d \sum_{j=1}^m b^{\ell,j} b^{k,j} \frac{\partial^2}{\partial x_\ell \partial x_k} + \sum_{\ell,k=1}^d \sum_{j=1}^m b^{\ell,j} \left(\sum_{p=1}^d z_{p,i} \frac{\partial b^{k,j}}{\partial x_p} \right) \frac{\partial^2}{\partial x_\ell \partial z_{k,i}}, \\ \bar{L}_i^j &= \sum_{\ell=1}^d b^{\ell,j} \frac{\partial}{\partial x_\ell} + \sum_{k=1}^d \left(\sum_{p=1}^d z_{p,i} \frac{\partial b^{k,j}}{\partial x_p} \right) \frac{\partial}{\partial z_{k,i}} \end{aligned}$$

for $(t, \underline{x}, \underline{z})^\top \in \Gamma_0 \times \Re^{d^2}$.

Computing the partial derivatives of v_i from (15.3.6), and using (15.3.4) we have

$$\begin{aligned} \bar{L}_i^0 v_i(s, \mathbf{x}, \mathbf{z}) &= \sum_{p=1}^d z_{p,i} \frac{\partial^2 u(s, \mathbf{x})}{\partial x_p \partial s} + \sum_{\ell=1}^d a^\ell \sum_{p=1}^d z_{p,i} \frac{\partial^2 u(s, \mathbf{x})}{\partial x_\ell \partial x_p} \\ &\quad + \sum_{k=1}^d \sum_{p=1}^d z_{p,i} \frac{\partial a^k}{\partial x_p} \frac{\partial u(s, \mathbf{x})}{\partial x_k} \\ &\quad + \frac{1}{2} \sum_{\ell, k=1}^d \sum_{j=1}^m b^{\ell, j} b^{k, j} \sum_{p=1}^d z_{p,i} \frac{\partial^3 u(s, \mathbf{x})}{\partial x_p \partial x_\ell \partial x_k} \\ &\quad + \sum_{\ell, k=1}^d \sum_{j=1}^m b^{i, j} \left(\sum_{p=1}^d z_{p,i} \frac{\partial b^{k, j}}{\partial x_p} \right) \frac{\partial^2 u(s, \mathbf{x})}{\partial x_\ell \partial x_k} \\ &= \sum_{p=1}^d z_{p,i} \left(\frac{\partial}{\partial x_p} L^0 u(s, \mathbf{x}) \right) = 0 \end{aligned}$$

for $(s, \mathbf{x}, \mathbf{z})^\top \in \Gamma_0 \times \Re^{d^2}$.

Consequently, if we take the initial value $\delta = (\delta^{1,1}, \delta^{1,2}, \dots, \delta^{d,d})^\top \in \Re^{d^2}$, where $\delta^{p,i}$ is the Kronecker delta function given by

$$\delta^{p,i} = \begin{cases} 1 & : p = i \\ 0 & : p \neq i \end{cases}, \quad (15.3.8)$$

for $p, i \in \{1, 2, \dots, d\}$, then taking expectation on both sides of (15.3.7) yields

$$\begin{aligned} \frac{\partial}{\partial \underline{x}_i} u(t, \underline{x}) &= v_i(t, \underline{x}, \delta) \\ &= E \left(v_i \left(\tau, \mathbf{X}_{\tau}^{t, \underline{x}}, \mathbf{Z}_{\tau}^{t, \delta} \right) \right) \\ &= E \left(\sum_{p=1}^d Z_{\tau}^{p, i, t, \delta} \frac{\partial}{\partial x_p} u(\tau, \mathbf{X}_{\tau}^{t, \underline{x}}) \right) \\ &= E \left(\sum_{p=1}^d Z_{\tau}^{p, i, t, \delta} \frac{\partial}{\partial x_p} h(\tau, \mathbf{X}_{\tau}^{t, \underline{x}}) \right) \quad (15.3.9) \end{aligned}$$

for $(t, \underline{x}) \in \Gamma_0$ and $i \in \{1, 2, \dots, d\}$.

Martingale Representation

We assume

$$E(|v_i(\tau, \mathbf{X}_\tau^{t,x}, \mathbf{Z}_\tau^{t,\delta})|) < \infty$$

for $i \in \{1, 2, \dots, d\}$ so that the expectation in (15.3.9) is well-defined. As in the one-dimensional case and subject to certain growth bounds applying for the partial derivatives $\frac{\partial h}{\partial x_1}, \dots, \frac{\partial h}{\partial x_d}$, this condition will be verified in Sect. 15.4.

The above expression for $\frac{\partial}{\partial \underline{x}_i} u(t, \underline{x})$ can be substituted into (15.3.2) yielding

$$u(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) = u(t_0, \underline{x}) + \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^\tau \gamma_s^i b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s^j, \quad (15.3.10)$$

where

$$\gamma_s^i = E \left(\sum_{p=1}^d Z_\tau^{p,i,s,\delta} \frac{\partial}{\partial x_p} h(\tau, \mathbf{X}_\tau^{s \wedge \tau, X_{s \wedge \tau}^{t_0, \underline{x}}}) \right).$$

Using equation (15.1.7) and the Markov property we can write

$$E \left(\sum_{p=1}^d Z_\tau^{p,i,s,\delta} \frac{\partial}{\partial x_p} h(\tau, \mathbf{X}_\tau^{s \wedge \tau, X_{s \wedge \tau}^{t_0, \underline{x}}}) \right) = E \left(\sum_{p=1}^d Z_\tau^{p,i,s,\delta} \frac{\partial}{\partial x_p} h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) \mid \mathcal{A}_s \right).$$

Combining this with the representation (15.3.10) we obtain

$$u(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) = u(t_0, \underline{x}) + \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^\tau \gamma_s^i b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s^j, \quad (15.3.11)$$

where

$$\gamma_s^i = E \left(\sum_{p=1}^d Z_\tau^{p,i,s,\delta} \frac{\partial}{\partial x_p} h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) \mid \mathcal{A}_s \right).$$

Taking expectations on both sides of (15.3.11) and using the boundary condition

$$u(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) = h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})$$

we can infer that

$$u(t_0, \underline{x}) = E(h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})).$$

Consequently, this *martingale representation* becomes

$$h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) = E(h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})) + \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^\tau \gamma_s^i b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s^j, \quad (15.3.12)$$

where γ_s^i is as given in (15.3.10) or (15.3.11). The equations (15.3.11) and (15.3.12) should be compared to (15.2.17) and (15.2.18) for the one-dimensional case.

We remark that the formulas (15.3.11) and (15.3.12) can be expressed using matrix notation and appear in this form under stronger assumptions in Colwell et al. (1991) and Ocone (1984). We have used the component form as these components are involved explicitly in the proof of the result, and for practical applications all components are computed separately.

15.4 Martingale Representation for Non-Smooth Payoffs

In this section we consider the problem of extending the representation results obtained in the previous section to non-smooth payoff functions. We deliberately show much detail since in most of the literature this important problem is neglected or avoided.

Martingale Representation Theorem

The following extension to non-smooth payoff functions is required even for standard derivative securities, such as the well-known European call option.

As in the previous section we let $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \geq 0\}$ be an m -dimensional Brownian motion and $\mathbf{X}^{t_0, \underline{x}} = \{\mathbf{X}_t^{t_0, \underline{x}} = (X_t^{1, t_0, \underline{x}}, \dots, X_t^{d, t_0, \underline{x}})^\top, t \in [t_0, T]\}$ a d -dimensional process which satisfies the SDE (15.1.1).

We assume that the drift and diffusion coefficients of (15.1.1) have linear growth and are Lipschitz continuous so that, in particular,

$$E \left(\sup_{s \in [t_0, T]} |\mathbf{X}_s^{t_0, \underline{x}}|^2 \right) < K_1 < \infty \quad (15.4.1)$$

for some constant $K_1 \in \mathbb{R}^+$, see (1.9.7).

We also assume that the drift and diffusion coefficients of (15.3.5) have linear growth and are Lipschitz continuous, so that by using similar arguments that lead to (1.9.7), there is a constant $K_2(s) \in \mathbb{R}^+$, which may depend on s , with

$$E \left(\sup_{r \in [s, T]} |Z_r^{p, i, s, \delta}|^2 \right) < K_2(s) < \infty \quad (15.4.2)$$

for $p, i \in \{1, 2, \dots, d\}$.

Let τ be a stopping time given by (15.1.2) with continuation region Γ_0 and exercise boundary Γ_1 . Consider a valuation function $u : \Gamma_0 \cup \Gamma_1 \rightarrow \mathbb{R}$ of the form (15.3.1) with payoff function $h : \Gamma_1 \rightarrow \mathbb{R}$. The following conditions will be used in the statement of the martingale representation that will be presented in this section.

A1 There exists a sequence of functions $h_n : \Gamma_1 \rightarrow \mathbb{R}$, $n \in \mathcal{N}$ of class $C^{1, \ell}$, $\ell \geq 3$ such that

(i) for each $(t, \mathbf{x}) \in \Gamma_1$

$$\lim_{n \rightarrow \infty} h_n(t, \mathbf{x}) = h(t, \mathbf{x}),$$

(ii) and for each $(t, \mathbf{x}) \in \Gamma_1$, $i \in \{1, 2, \dots, d\}$

$$\lim_{n \rightarrow \infty} \frac{\partial h_n(t, \mathbf{x})}{\partial x_i} = g_i(t, \mathbf{x})$$

for some set of functions $g_i : \Gamma_1 \rightarrow \mathbb{R}$.

A2 (i) The functions h_n satisfy a uniform linear growth bound of the form

$$|h_n(t, \mathbf{x})|^2 \leq K_3^2 (1 + |\mathbf{x}|^2)$$

for all $\mathbf{x} \in \mathbb{R}^d$ and $n \in \mathcal{N}$, where $|\mathbf{x}|^2 = \sum_{i=1}^d x_i^2$ and $K_3 < \infty$,

(ii) and the functions $\frac{\partial h_n}{\partial x_i}$, $i \in \{1, 2, \dots, d\}$ satisfy a uniform linear growth bound of the form

$$\sum_{i=1}^d \left| \frac{\partial h_n(t, \mathbf{x})}{\partial x_i} \right|^2 \leq K_4^2 (1 + |\mathbf{x}|^2)$$

for all $x \in \mathbb{R}^d$, $n \in \mathcal{N}$, where $|\mathbf{x}|$ is as given in A2(i) above and $K_4 < \infty$.

The following *martingale representation theorem* has been derived in [Heath \(1995\)](#).

Theorem 15.4.1. *Suppose that the valuation function $u : [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is defined according to (15.3.1), conditions A1 and A2 hold for the payoff function h , and the random variables $h_n(\mathbf{X}_T^{t_0, \underline{x}})$, $n \in \mathcal{N}$, can be represented in the form*

$$h_n(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}}) = E(h_n(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}})) + \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^{\tau} \gamma_{n,s}^i b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s^j, \quad (15.4.3)$$

where

$$\gamma_{n,s}^i = E \left(\sum_{p=1}^d Z_{\tau}^{p,i,s,\delta} \frac{\partial}{\partial x_p} h_n(\mathbf{X}_{\tau}^{t_0, \underline{x}}) \mid \mathcal{A}_s \right).$$

Then for $(t_0, \underline{x}) \in \Gamma_0$, u admits the martingale representation

$$u(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}}) = u(t_0, \underline{x}) + \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^{\tau} \gamma_s^i b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s^j \quad (15.4.4)$$

or equivalently using A1(ii),

$$h(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}}) = E(h(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}})) + \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^{\tau} \gamma_s^i b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s^j, \quad (15.4.5)$$

where

$$\gamma_s^i = E \left(\sum_{p=1}^d Z_\tau^{p,i,s,\delta} g_p(\tau, \mathbf{X}_\tau^{t_0,x}) \mid \mathcal{A}_s \right)$$

and $Z_T^{p,i,s,\delta}$ is the unique solution of the SDE (15.3.5) with initial value δ at time $s \in [t_0, T]$, as given by (15.3.8).

The above representation theorem allows, under general conditions, for the payoff structure to be expressed as a stochastic integral with respect to the driving Wiener processes. Furthermore, it provides explicit functionals for the corresponding hedge ratios. This enables these hedge ratios, as well as prices, to be accurately computed.

A First Lemma

In the following, we will establish the representation theorem using two lemmas and some general results from measure theory and integration.

Lemma 15.4.2 *Suppose that the payoff function h satisfies conditions A1 and A2 and the random variables $h_n(\tau, \mathbf{X}_\tau^{t_0,x})$, $n \in \mathcal{N}$, can be represented in the form*

$$h_n(\tau, \mathbf{X}_\tau^{t_0,x}) = E(h_n(\tau, \mathbf{X}_\tau^{t_0,x})) + \sum_{j=1}^m \int_{t_0}^\tau \xi_{n,s}^j dW_s^j, \quad (15.4.6)$$

where $\xi_n = (\xi_n^1, \dots, \xi_n^m)^\top$ is a vector of $\underline{\mathcal{A}}$ -predictable processes for each $n \in \mathcal{N}$. Then

$$\lim_{n \rightarrow \infty} \left\| h(\tau, \mathbf{X}_\tau^{t_0,x}) - E(h(\tau, \mathbf{X}_\tau^{t_0,x})) - \sum_{j=1}^m \int_{t_0}^\tau \xi_{n,s}^j dW_s^j \right\|_2 = 0,$$

where $\|\cdot\|_2 = \sqrt{E(|\cdot|^2)}$ denotes the norm in the Banach space $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$.

Proof: Applying the uniform linear growth bound A2(i) we see that

$$|h_n(\tau, \mathbf{X}_\tau^{t_0,x})|^2 \leq K_3^2 \left(1 + |\mathbf{X}_\tau^{t_0,x}|^2 \right) \quad (15.4.7)$$

for all $n \in \mathcal{N}$. This shows that the random variable $|h_n(\tau, \mathbf{X}_\tau^{t_0,x})|^2$ is dominated by the variate $K_3^2(1 + |\mathbf{X}_\tau^{t_0,x}|^2)$ for all $n \in \mathcal{N}$. Now, from the growth bound (15.4.1) we have

$$E \left(K_3^2 \left(1 + |\mathbf{X}_\tau^{t_0,x}|^2 \right) \right) < K_3^2(1 + K_1) < \infty. \quad (15.4.8)$$

The pointwise convergence of h_n given by condition A1(i) means that

$$\lim_{n \rightarrow \infty} |h_n(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) - h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})|^2 \stackrel{\text{a.s.}}{=} 0. \quad (15.4.9)$$

In fact, the convergence here holds for all $\omega \in \Omega$ although we do not require this stronger result. Combining (15.4.7), (15.4.8) and (15.4.9) we can apply a version of the Dominated Convergence Theorem applicable to L^p spaces, $p > 0$, or the Corollary to Theorem 2.6.3 in Shiryaev (1984), to obtain

$$\lim_{n \rightarrow \infty} E(|h_n(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) - h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})|^2) = 0,$$

which can be expressed using the $\|\cdot\|_2$ norm of $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ as

$$\lim_{n \rightarrow \infty} \|h_n(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) - h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})\|_2 = 0. \quad (15.4.10)$$

Furthermore,

$$\begin{aligned} |E(h_n(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})) - E(h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}))| &\leq E(|h_n(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) - h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})|) \\ &= \|h_n(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) - h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})\|_1 \end{aligned} \quad (15.4.11)$$

for all $n \in \mathcal{N}$, where $\|\cdot\|_1 = E(|\cdot|)$ denotes the norm in the Banach space $L^1(\Omega, \mathcal{A}_T, P)$. Since by Hölder's inequality $\|f\|_1 \leq \|f\|_2$ for any $f \in L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ then from (15.4.10) and (15.4.11) we can infer that

$$\lim_{n \rightarrow \infty} |E(h_n(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})) - E(h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}))| = 0. \quad (15.4.12)$$

Using (15.4.10) and (15.4.12) together with the representation of $h_n(\mathbf{X}_\tau^{t_0, \underline{x}})$ stated in the hypothesis of the lemma gives the required result. \square

A Second Lemma

Let us prepare another result for the proof of the martingale representation theorem.

Lemma 15.4.3 *Let $(\xi_n^j = \{\xi_{n,s}^j, s \in [t_0, T]\})_{n \in \mathcal{N}}$ for each $j \in \{1, 2, \dots, m\}$ be a sequence of stochastic processes, adapted to the filtration $\underline{\mathcal{A}}$ and satisfying the conditions:*

(i) *The sequence of random variables*

$$\left(\sum_{j=1}^m \int_{t_0}^{\tau} \xi_{n,s}^j dW_s^j \right)_{n \in \mathcal{N}},$$

is well-defined and forms a Cauchy sequence in $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$.

(ii) For each $j \in \{1, 2, \dots, m\}$, there is a stochastic process $\xi^j = \{\xi_s^j, s \in [t_0, T]\}$ with

$$\lim_{n \rightarrow \infty} |\xi_{n,s}^j - \xi_s^j| \xrightarrow{a.s.} 0$$

for all $s \in [t_0, T]$.

Then the random variable

$$\sum_{j=1}^m \int_{t_0}^{\tau} \xi_s^j dW_s^j$$

is well-defined and

$$\lim_{n \rightarrow \infty} \left\| \sum_{j=1}^m \int_{t_0}^{\tau} \xi_{n,s}^j dW_s^j - \sum_{j=1}^m \int_{t_0}^{\tau} \xi_s^j dW_s^j \right\|_2 = 0.$$

Proof: In this proof we will use some general arguments from measure theory and integration using the Banach space $L^{2,*} = L^2([t_0, T] \times \Omega, \mathcal{L} \otimes \mathcal{A}_T, u_L \times P)$, where \mathcal{L} is the sigma-algebra of Lebesgue subsets of \mathbb{R} and u_L is the Lebesgue measure. We assume that this Banach space is equipped with the norm

$$\|f\|_2^* = \sqrt{\int_{[t_0, T] \times \Omega} |f|^2 du_L \times P}$$

for any $f \in L^{2,*}$.

Let us introduce the process $\mathbf{1}^\tau = \{\mathbf{1}_s^\tau = \mathbf{1}_{\{s \leq \tau\}}, s \in [t_0, T]\}$. Note that $\mathbf{1}^\tau$ is \mathcal{A} -adapted since τ is a stopping time. Also $\mathbf{1}^\tau$ is right continuous and hence is measurable with respect to the product sigma-algebra $\mathcal{L} \otimes \mathcal{A}_T$, see Proposition 1.13 in Karatzas & Shreve (1991).

In what follows we will consider the processes $\xi_n^j, j \in \{1, 2, \dots, m\}, n \in \mathcal{N}$ as functions defined on $[t_0, T] \times \Omega$. By hypothesis the Itô integrals

$$\int_{t_0}^{\tau} \xi_{n,s}^j dW_s^j = \int_{t_0}^T \mathbf{1}_s^\tau \xi_{n,s}^j dW_s^j,$$

are well-defined for $j \in \{1, 2, \dots, m\}$ and $n \in \mathcal{N}$. Consequently, by definition of the Itô integral as the $L^{2,*}$ limit of appropriately defined step functions, the integrands $\mathbf{1}^\tau \xi_n^j, j \in \{1, 2, \dots, m\}, n \in \mathcal{N}$, are also $\mathcal{L} \otimes \mathcal{A}_T$ -measurable. Using this result and Fubini's Theorem we have

$$\begin{aligned}
\|\mathbf{1}^\tau(\xi_{n_1}^j - \xi_{n_2}^j)\|_2^* &= \sqrt{E \left(\int_{t_0}^T \mathbf{1}_s^\tau (\xi_{n_1,s}^j - \xi_{n_2,s}^j)^2 ds \right)} \\
&= \sqrt{E \left(\int_{t_0}^\tau (\xi_{n_1,s}^j - \xi_{n_2,s}^j)^2 ds \right)} \\
&\leq \sqrt{E \left(\sum_{j=1}^m \int_{t_0}^\tau (\xi_{n_1,s}^j - \xi_{n_2,s}^j)^2 ds \right)} \\
&= \sqrt{E \left(\left(\sum_{j=1}^m \int_{t_0}^\tau (\xi_{n_1,s}^j - \xi_{n_2,s}^j) dW_s^j \right)^2 \right)} \\
&= \left\| \sum_{j=1}^m \int_{t_0}^\tau \xi_{n_1,s}^j dW_s^j - \sum_{j=1}^m \int_{t_0}^\tau \xi_{n_2,s}^j dW_s^j \right\|_2 \quad (15.4.13)
\end{aligned}$$

for any integers $n_1, n_2 \in \mathcal{N}$ and $j \in \{1, 2, \dots, m\}$, where $\|\cdot\|_2$ is the $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ norm previously defined in the statement of Lemma 15.4.2. By hypothesis the random variables

$$\left(\sum_{j=1}^m \int_{t_0}^\tau \xi_{n,s}^j dW_s^j \right)_{n \in \mathcal{N}}$$

form a Cauchy sequence in $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ and, therefore, from (15.4.13) the functions $(\mathbf{1}^\tau \xi_n^j)_{n \in \mathcal{N}}$, for fixed $j \in \{1, 2, \dots, m\}$ form a Cauchy sequence in $L^{2,*}$.

This means that there is a function $\bar{\xi}^j : [t_0, T] \times \Omega \rightarrow \mathfrak{R}$, $\bar{\xi}^j \in L^{2,*}$, for each $j \in \{1, 2, \dots, m\}$, with

$$\lim_{n \rightarrow \infty} \|\mathbf{1}^\tau \xi_n^j - \bar{\xi}^j\|_2^* = 0. \quad (15.4.14)$$

Let $D = \{(s, \omega) \in [t_0, T] \times \Omega : \mathbf{1}_s^\tau(\omega) = 1\}$ so that $D \in \mathcal{L} \otimes \mathcal{A}_T$, since $\mathbf{1}^\tau$ is measurable with respect to the product sigma-algebra $\mathcal{L} \otimes \mathcal{A}_T$. Using the definition of the $\|\cdot\|_2^*$ norm we have for $j \in \{1, 2, \dots, m\}$:

$$\begin{aligned}
\|\mathbf{1}^\tau \xi_n^j - \mathbf{1}^\tau \bar{\xi}^j\|_2^* &= \sqrt{\int_D |\xi_n^j - \bar{\xi}^j|^2 du_L \times P} \\
&= \sqrt{\int_D |\mathbf{1}^\tau \xi_n^j - \bar{\xi}^j|^2 du_L \times P} \\
&\leq \sqrt{\int_{[t_0, T] \times \Omega} |\mathbf{1}^\tau \xi_n^j - \bar{\xi}^j|^2 du_L \times P} \\
&= \|\mathbf{1}^\tau \xi_n^j - \bar{\xi}^j\|_2^*
\end{aligned}$$

so that from (15.4.14)

$$\lim_{n \rightarrow \infty} \|\mathbf{1}^\tau \xi_n^j - \mathbf{1}^\tau \bar{\xi}^j\|_2^* = 0. \quad (15.4.15)$$

Consequently, the $L^{2,*}$ limit of $\mathbf{1}^\tau \xi_n^j$, $n \in \mathcal{N}$, $j \in \{1, 2, \dots, m\}$ as $n \rightarrow \infty$, can be written in the form $\mathbf{1}^\tau \bar{\xi}^j$.

We know from measure theory and integration, see for example Widom (1969), that any Cauchy sequence in an L^2 Banach space will have a subsequence converging almost everywhere and this limit is the L^2 -limit. For a proof of this result in a probabilistic setting we refer to Theorems 2.10.2 and 2.10.5 in Shiryaev (1984). Applying this result and Fubini's Theorem we see that for each $j \in \{1, 2, \dots, m\}$

$$E \left(\left| \mathbf{1}_s^\tau \xi_{n_i, s}^j - \mathbf{1}_s^\tau \bar{\xi}_s^j \right|^2 \right) = 0 \quad (15.4.16)$$

for some subsequence $(n_i)_{i \in \mathcal{N}}$, of positive integers, for all $s \in A$. Here $A \subseteq [t_0, T]$ is some Lebesgue measurable set with $u_L([t_0, T] \setminus A) = 0$, and $[t_0, T] \setminus A = \{s \in [t_0, T] : s \notin A\}$.

Equation (15.4.16) shows that for fixed integer $j \in \{1, 2, \dots, m\}$, and $s \in A$, $(\xi_{n_i, s}^j)_{i \in \mathcal{N}}$ is a Cauchy sequence in $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$. This means that there is a subsequence of $(n_i)_{i \in \mathcal{N}}$, say $(n_{i_k}(s))_{k \in \mathcal{N}}$, which may depend on s , with

$$\lim_{k \rightarrow \infty} \left| \mathbf{1}_s^\tau \xi_{n_{i_k}(s), s}^j - \mathbf{1}_s^\tau \bar{\xi}_s^j \right|^2 \stackrel{\text{a.s.}}{\equiv} 0 \quad (15.4.17)$$

for all $s \in A$. Therefore, from condition (b) in the statement of Lemma 15.4.3 we see that for each $j \in \{1, 2, \dots, m\}$

$$\mathbf{1}^\tau \xi_s^j = \mathbf{1}^\tau \bar{\xi}_s^j$$

a.s. for all $s \in A$.

Applying this result, Fubini's Theorem and recalling that $\|\cdot\|_2$ denotes the norm of $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ we have for any integers $j \in \{1, 2, \dots, m\}$ and $n \in \mathcal{N}$ the relation

$$\begin{aligned}
\|\mathbf{1}^\tau \xi_n^j - \mathbf{1}^\tau \bar{\xi}^j\|_2^* &= \sqrt{E \left(\int_{t_0}^{T_0} \mathbf{1}_s^\tau \left(\xi_{n,s}^j - \bar{\xi}_s^j \right)^2 ds \right)} \\
&= \sqrt{E \left(\int_A \mathbf{1}_s^\tau \left(\xi_{n,s}^j - \bar{\xi}_s^j \right)^2 ds \right)} \\
&= \sqrt{\int_A E \left(\mathbf{1}_s^\tau \left(\xi_{n,s}^j - \bar{\xi}_s^j \right)^2 \right) ds} \\
&= \sqrt{E \left(\int_A \mathbf{1}_s^\tau \left(\xi_{n,s}^j - \bar{\xi}_s^j \right)^2 ds \right)} \\
&= \sqrt{E \left(\int_{t_0}^T \mathbf{1}_s^\tau \left(\xi_{n,s}^j - \bar{\xi}_s^j \right)^2 ds \right)} \\
&= \sqrt{E \left(\int_{t_0}^\tau \left(\xi_{n,s}^j - \bar{\xi}_s^j \right)^2 ds \right)} \\
&= \left\| \int_{t_0}^\tau \xi_{n,s}^j dW_s^j - \int_{t_0}^\tau \bar{\xi}_s^j dW_s^j \right\|_2. \tag{15.4.18}
\end{aligned}$$

Combining (15.4.15) and (15.4.18), which hold for each $j \in \{1, 2, \dots, m\}$, we can infer that

$$\lim_{n \rightarrow \infty} \left\| \sum_{j=1}^m \int_{t_0}^\tau \xi_{n,s}^j dW_s^j - \sum_{j=1}^m \int_{t_0}^\tau \bar{\xi}_s^j dW_s^j \right\|_2 = 0.$$

□

Proof of the Martingale Representation Theorem

We will now show that the conditions required for an application of Lemmas 15.4.2 and 15.4.3 can be satisfied for suitable choices of processes ξ_n , $n \in \mathcal{N}$, and ξ under the Assumptions A1 and A2 of Theorem 15.4.1.

For integers $j \in \{1, 2, \dots, m\}$, $n \in \mathcal{N}$ and $s \in [t_0, T]$ define

$$\xi_{n,s}^j = \sum_{i=1}^d \gamma_{n,s}^i b^{i,j}(s, \mathbf{X}_s^{t_0,x}), \quad (15.4.19)$$

$$\xi_s^j = \sum_{i=1}^d \gamma_s^i b^{i,j}(s, \mathbf{X}_s^{t_0,x}), \quad (15.4.20)$$

where $\gamma_{n,s}^i$ and γ_s^i , $i \in \{1, 2, \dots, d\}$, $n \in \mathcal{N}$ are as given in the representations (15.4.3) and (15.4.5), respectively.

Substituting (15.4.19) into (15.4.6) we obtain the representation (15.4.3), which is assumed to be true by the hypothesis of Theorem 15.4.1. Also, as given in the statement of Theorem 15.4.1 we assume that the payoff function h satisfies conditions A1 and A2. Consequently, by applying the results of Lemma 15.4.2 we see that the random variables

$$\sum_{j=1}^m \int_{t_0}^T \xi_{n,s}^j dW_s^j,$$

$n \in \mathcal{N}$, are well-defined and form a Cauchy sequence in $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$. This verifies condition (a) in the statement of Lemma 15.4.3.

To establish condition (b) of Lemma 15.4.3 we will use the uniform linear growth bound A2(ii) and the pointwise limit A1(ii). Let $p, i \in \{1, 2, \dots, d\}$ and $s \in [t_0, T]$ be fixed. For $n \in \mathcal{N}$ define

$$\psi_{n,p,i,s} = Z_\tau^{p,i,s,\delta} \frac{\partial}{\partial x_p} h_n(\tau, \mathbf{X}_\tau^{t_0,x}) \quad (15.4.21)$$

so that from condition A2(ii) we have the inequality

$$\begin{aligned} |\psi_{n,p,i,s}| &\leq \sqrt{K_4^2 \left(1 + |\mathbf{X}_\tau^{t_0,x}|^2\right)} |Z_\tau^{p,i,s,\delta}| \\ &\leq K_4 (1 + |\mathbf{X}_\tau^{t_0,x}|) |Z_\tau^{p,i,s,\delta}| \end{aligned}$$

for $s \in [t_0, T]$. If we let

$$\Psi_{p,i,s} = K_4 (1 + |\mathbf{X}_\tau^{t_0,x}|) |Z_\tau^{p,i,s,\delta}|,$$

then using Hölder's and Minkowski's inequalities, and the bounds (15.4.1) and (15.4.2) we have

$$\begin{aligned} E(\Psi_{p,i,s}) &\leq \|K_4 (1 + |\mathbf{X}_\tau^{t_0,x}|)\|_2 \|Z_\tau^{p,i,s,\delta}\|_2 \\ &\leq K_4 (1 + \|\mathbf{X}_\tau^{t_0,x}\|_2) \|Z_\tau^{p,i,s,\delta}\|_2 \\ &< \infty \end{aligned} \quad (15.4.22)$$

for $s \in [t_0, T]$. Also the pointwise limit condition A1(ii) shows that

$$\lim_{n \rightarrow \infty} \left| \frac{\partial}{\partial x_p} h_n(\mathbf{X}_\tau^{t_0, \underline{x}}) - g_p(\mathbf{X}_\tau^{t_0, \underline{x}}) \right| |Z_\tau^{p,i,s,\delta}| \stackrel{\text{a.s.}}{=} 0 \quad (15.4.23)$$

for $s \in [t_0, T]$. These results demonstrate that the random variable $\psi_{n,p,i,s}$ is dominated by $\Psi_{p,i,s}$, which has finite expectation by (15.4.22) and a P -almost sure limit, as given by (15.4.23).

We can now apply the Dominated Convergence Theorem for conditional expectations, see Shiryaev (1984) Theorem 2.7.2, yielding

$$\lim_{n \rightarrow \infty} E \left(\left(\frac{\partial}{\partial x_p} h_n(\mathbf{X}_\tau^{t_0, \underline{x}}) - g_p(\mathbf{X}_\tau^{t_0, \underline{x}}) \right) Z_\tau^{p,i,s,\delta} \mid \mathcal{A}_s \right) \stackrel{\text{a.s.}}{=} 0 \quad (15.4.24)$$

for $s \in [t_0, T]$ and $p, i \in \{1, 2, \dots, d\}$. From the definition of $\xi_{n,s}^j$ and ξ_s^j given by (15.4.19) and (15.4.20), respectively, and by applying (15.4.24) we can infer that

$$\lim_{n \rightarrow \infty} |\xi_{n,s}^j - \xi_s^j| \stackrel{\text{a.s.}}{=} 0 \quad (15.4.25)$$

for each $j \in \{1, 2, \dots, m\}$ and $s \in [t_0, T]$. This establishes condition (b) in the statement of Lemma 15.4.3.

Finally, note that since the random variables $\gamma_{n,s}^i$, $i \in \{1, 2, \dots, d\}$, $n \in \mathcal{N}$, are expressed as conditional expectations with respect to \mathcal{A}_s , as can be seen from (15.4.3), they are \mathcal{A}_s -measurable. Also $b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}})$, $i \in \{1, 2, \dots, d\}$, $j \in \{1, 2, \dots, m\}$ is \mathcal{A}_s -measurable since $\mathbf{X}_s^{t_0, \underline{x}}$ is \mathcal{A}_s -measurable. Therefore, ξ_n^j , $j \in \{1, 2, \dots, m\}$, $n \in \mathcal{N}$, as given by (15.4.19), will be adapted to the filtration \mathcal{A} .

Consequently, the conditions required for the application of Lemma 15.4.3 are satisfied with ξ_n^j and ξ^j given by (15.4.19) and (15.4.20), respectively. Combining the results of Lemma 15.4.2 and Lemma 15.4.3 yields

$$\left\| u(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) - u(t_0, \underline{x}) - \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^\tau \gamma_s^i b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s^j \right\|_2 = 0$$

or equivalently

$$u(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) = u(t_0, \underline{x}) + \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^\tau \gamma_s^i b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s^j$$

a.s., which completes the proof of Theorem 15.4.1. \square

We remark that the smoothness condition, where we assume that h_n is of class $C^{1,\ell}$, $\ell \geq 3$, as stated in Assumption A1, has not been explicitly used in the proof of Theorem 15.4.1. In practice, such a smoothness condition is required for the results in the previous sections to be applied so that the representation (15.4.3), used in the statement of Theorem 15.4.1, can be obtained. A method for constructing these approximating functions for one-dimensional processes

and a class of absolutely continuous payoff functions will be considered in the next section. Finally, note that the condition

$$E \left(\sum_{p=1}^d Z_{\tau}^{p,i,s,\delta} \frac{\partial}{\partial x_p} h(\tau, \mathbf{X}_{\tau}^{t,x}) \right) < \infty$$

for $t \in [t_0, T]$, required in the proof of (15.2.15) and (15.3.9), can also be obtained from the inequality (15.4.22), which depends on the conditions (15.4.1), (15.4.2) and A2(ii). The arguments employed in the proof of the martingale representation theorem suggest that one should be able to prove along the same lines a generalization for SDEs with jumps.

15.5 Absolutely Continuous Payoff Functions

In this and the two following sections we will provide examples of applications of Theorem 15.4.1. These examples demonstrate the wide applicability of the representation theorem.

One-dimensional Payoff

We will show that the conditions A1 and A2, required for an application of Theorem 15.4.1, are satisfied for a class of one-dimensional absolutely continuous functions. This result, together with Theorem 15.4.1, will then be applied to show that these functions admit an Itô integral representation of the form (15.2.17) or (15.2.18) for a wide class of one-dimensional processes $X^{t_0,x}$.

For a one-dimensional payoff function h with $\tau = T$ and h of the form $h(x) = h(T, x)$ for $x \in \mathbb{R}$, these conditions can be simplified as follows:

A1* There exists a sequence of functions $h_n : \mathbb{R} \rightarrow \mathbb{R}$, $n \in \mathcal{N}$, of class C^ℓ for some integer $\ell \geq 3$ such that

(i) for each $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} h_n(x) = h(x),$$

(ii) and there is a function $g : \mathbb{R} \rightarrow \mathbb{R}$ such that for each $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} h'_n(x) = g(x).$$

A2* The functions h_n and h'_n satisfy uniform linear growth bounds of the form

$$(i) \quad |h_n(x)|^2 \leq K_3^2 (1 + |x|^2)$$

$$(ii) \quad |h'_n(x)|^2 \leq K_4^2 (1 + |x|^2)$$

for all $x \in \mathbb{R}$ and $n \in \mathcal{N}$ with $K_3, K_4 < \infty$.

Let h be an absolutely continuous function of the form

$$h(x) = h(0) + \int_0^x g(s) ds \quad (15.5.1)$$

for $x \in \mathbb{R}$, where both h and g satisfy linear growth bounds of the form

$$|h(x)|^2 \leq A_0 (1 + |x|^2), \quad (15.5.2)$$

$$|g(x)|^2 \leq B_0 (1 + |x|^2) \quad (15.5.3)$$

for all $x \in \mathbb{R}$ with $A_0, B_0 < \infty$.

We assume that g is continuous except possibly at a finite number of points x_1, \dots, x_N with $x_1 < x_2 < \dots < x_N$ and that the right hand limits are well-defined and satisfy

$$g^+(x_i) = \lim_{\delta \rightarrow 0^+} g(x_i + \delta) = g(x_i)$$

for each $i \in \{1, 2, \dots, N\}$. These conditions and (15.5.1) show that $h'(x) = g(x)$ for all $x \in \mathbb{R} \setminus \{x_1, \dots, x_N\}$. Note that standard payoff functions such as European calls or puts, where $h(x) = (x - K)^+$ or $h(x) = (K - x)^+$, respectively, are absolutely continuous functions of the form (15.5.1).

Some Continuous Payoff Functions

For a continuous function $f : \mathbb{R} \rightarrow \mathbb{R}$ define the function $I_{n,m}(f) : \mathbb{R} \rightarrow \mathbb{R}$, for integers $n \geq 1$, $m \geq 0$ iteratively as follows:

$$\begin{aligned} I_{n,0}(f)(x) &= f(x) \\ I_{n,m+1}(f)(x) &= n \int_x^{x+\frac{1}{n}} I_{n,m}(f)(s) ds \end{aligned} \quad (15.5.4)$$

for $x \in \mathbb{R}$. Here $I_{n,m}$ can be interpreted as a smoothing operator defined on the set of real-valued continuous functions with domain \mathbb{R} .

We will show that h satisfies conditions A1* and A2* using the approximating function $I_{n,m}(h)$, $n \in \mathcal{N}$ for any integer $m \geq 3$:

1. The Lebesgue integral in (15.5.4) is used to ensure that the functions $I_{n,m+1}(f)$, $n \in \mathcal{N}$ are of class C^{m+1} if $I_{n,m}(f)$ is of class C^m for integers $m \geq 0$. In fact, computing the derivative of $I_{n,m}(f)$ using (15.5.4) we see that

$$(I_{n,m+1}(f))'(x) = n \left(I_{n,m}(f) \left(x + \frac{1}{n} \right) - I_{n,m}(f)(x) \right) \quad (15.5.5)$$

for all $x \in \mathbb{R}$. In particular, since the function h is absolutely continuous, then $I_{n,m}(h)$ will be of class C^m for all integers $n \geq 1$, $m \geq 1$.

We will now show that

$$|I_{n,m}(h)(x)| \leq A_m \sqrt{1 + |x|^2} \quad (15.5.6)$$

for all $x \in \Re$, and integers $n \geq 1$, $m \geq 0$, where $A_m < \infty$ is some constant which depends on m .

Applying the inequality

$$(a + \frac{1}{n})^2 \leq 2(a^2 + \frac{1}{n^2}) \leq 2(a^2 + 1)$$

for $a \in \Re$, $n \in \mathcal{N}$ and assuming (15.5.6) holds for some fixed integer $m \geq 0$, for all integers $n \geq 1$, we have

$$\begin{aligned} |I_{n,m+1}(h)(x)| &\leq n \int_x^{x+\frac{1}{n}} |I_{n,m}(h)(s)| ds \\ &\leq n \int_x^{x+\frac{1}{n}} A_m \sqrt{1 + |s|^2} ds \\ &\leq A_m \sqrt{1 + \left(|x| + \frac{1}{n}\right)^2} \\ &\leq A_m \sqrt{3 + 2|x|^2} \\ &\leq A_{m+1} \sqrt{1 + |x|^2}, \end{aligned} \quad (15.5.7)$$

where $A_{m+1} = \sqrt{3}A_m$. Since $I_{n,0}(h) = h$, condition (15.5.2) shows that (15.5.6) holds for $m = 0$. Consequently, by an induction argument using (15.5.7), we see that (15.5.6) is valid for all integers $n \geq 1$, $m \geq 0$. This shows that for any integer $m \geq 0$, the functions $I_{n,m}(h)$, $n \in \mathcal{N}$ satisfy A2*(i).

2. Since $I_{n,m}(h)$, for integers $n \geq 1$, $m \geq 1$ is of class C^m , as previously noted, and $I_{n,0}(h) = h$ is continuous, the function $I_{n,m}(h)$ is continuous for all integers $n \geq 1$, $m \geq 0$. From definition (15.5.4) and by the Mean Value Theorem we can show that for any $x \in \Re$ and integers $n \geq 1$, $m \geq 0$

$$I_{n,m+1}(h)(x) = I_{n,m}(h)(\eta),$$

for some $\eta \in [x, x + \frac{1}{n}]$. Applying this result to the functions $I_{n,m}(h), \dots, I_{n,1}(h)$ we can infer that

$$I_{n,m}(h)(x) = I_{n,0}(h)(\eta_1) = h(\eta_1)$$

for some $\eta_1 \in [x, x + \frac{m}{n}]$. Taking the limit as $n \rightarrow \infty$ yields

$$\lim_{n \rightarrow \infty} I_{n,m}(h)(x) = h(x) \quad (15.5.8)$$

for any integer $m \geq 1$ and $x \in \Re$. Consequently, for any integer $m \geq 1$ condition A1*(i) is satisfied with the approximating functions $I_{n,m}(h)$, $n \in \mathcal{N}$.

3. To verify condition A2*(ii) we will show that

$$|(I_{n,m}(h))'(x)| \leq B_m \sqrt{1 + |x|^2} \quad (15.5.9)$$

for all $x \in \Re$, and integers $n \geq 1$, $m \geq 0$, where $B_m < \infty$ is a constant which depends on m .

By (15.5.5) we can write

$$\begin{aligned} (I_{n,m+1}(h))'(x) &= n \left(I_{n,m}(h) \left(x + \frac{1}{n} \right) - I_{n,m}(h)(x) \right) \\ &= n \int_x^{x+\frac{1}{n}} (I_{n,m}(h))'(s) ds, \end{aligned} \quad (15.5.10)$$

for all $x \in \Re$ and integers $n \geq 1$, $m \geq 1$. Since $I_{n,0}(h) = h$ we also have from (15.5.5) and (15.5.1) the relation

$$\begin{aligned} (I_{n,1}(h))'(x) &= n \left(h \left(x + \frac{1}{n} \right) - h(x) \right) \\ &= n \int_x^{x+\frac{1}{n}} g(s) ds. \end{aligned} \quad (15.5.11)$$

If we assume that (15.5.9) holds for some integer $m \geq 1$ and all $n \in \mathcal{N}$, then by using the inequality $(a + \frac{1}{n})^2 \leq 2(a^2 + 1)$, we have from (15.5.10) and (15.5.11) the inequalities

$$\begin{aligned} |(I_{n,m+1}(h))'(x)| &\leq n \int_x^{x+\frac{1}{n}} B_m \sqrt{1 + |s|^2} ds \\ &\leq B_m \sqrt{1 + \left(|x| + \frac{1}{n} \right)^2} \\ &\leq B_m \sqrt{3 + 2|x|^2} \\ &\leq B_{m+1} \sqrt{1 + |x|^2} \end{aligned} \quad (15.5.12)$$

for all $x \in \Re$ and integers $n \geq 1$, where $B_{m+1} = \sqrt{3}B_m$. If we use equation (15.5.11) and apply the inequality (15.5.3), then a similar proof shows that

$$(I_{n,1}(h))'(x) \leq B_1 \sqrt{1 + |x|^2},$$

where $B_1 = \sqrt{3}B_0$. This means that (15.5.9) is valid for $m = 1$ and hence by an induction argument using (15.5.12) we see that (15.5.9) holds for all integers $m \geq 0$. This means that the functions $I_{n,m}(h)$, $n \in \mathcal{N}$, satisfy conditions A2*(ii) for any integer $m \geq 0$.

4. To find the pointwise limits of $(I_{n,m}(h))'$, as $n \rightarrow \infty$ for a fixed integer $m \geq 0$, suppose $x \notin \{x_1, \dots, x_N\}$. Let

$$\alpha = \min_{i \in [1, N]} \{|x - x_i|\}.$$

Then for all integers $n > \frac{m}{\alpha}$, g is continuous on the interval $[x, x + \frac{m}{n}]$. For integers $m \geq 1$, $(I_{n,m}(h))'$ is of class C^{m-1} , as previously noted, and hence is continuous on \mathfrak{R} . From this fact, (15.5.10) and by using the Mean Value Theorem we have for any integers $n \geq 1$, $m \geq 1$ the relation

$$(I_{n,m+1}(h))'(x) = (I_{n,m}(h))'(\eta)$$

for some $\eta \in [x, x + \frac{1}{n}]$. A similar argument and (15.5.11) show that

$$I_{n,1}(h)'(x) = g(\eta_1)$$

for some $\eta_1 \in [x, x + \frac{1}{n}]$. Applying these results to the functions $(I_{n,m}(h))'$, \dots , $(I_{n,1}(h))'$ we can infer that for any integer $n \geq 1$, $m \geq 1$,

$$(I_{n,m}(h))'(x) = g(\eta_2)$$

for some $\eta_2 \in [x, x + \frac{m}{n}]$. This shows that

$$\lim_{n \rightarrow \infty} (I_{n,m}(h))'(x) = g(x) \quad (15.5.13)$$

for all integers $m \geq 1$. If $x \in \{x_1, \dots, x_N\}$, then for all integers $n > \frac{m}{\beta}$, where $\beta = \min_{1 \leq i \leq N-1} \{|x_{i+1} - x_i|\}$, g is continuous on the interval $[x, x + \frac{m}{n}]$. As noted above, $(I_{n,m}(h))'$ is for $m \geq 1$ continuous on \mathfrak{R} , and therefore, the limit (15.5.13) also applies. It follows that the approximating functions $I_{n,m}(h)$, $n \in \mathcal{N}$, for any integer $m \geq 0$ satisfies condition A1*(ii).

5. Summarizing these results, we have shown that the conditions A1* and A2* hold for h using the approximating functions $I_{n,m}(h)$, $n \in \mathcal{N}$, for any integer $m \geq 3$.

Let $X^{t_0, \underline{x}}$ be a one-dimensional process satisfying (15.2.1) with drift a and diffusion b coefficients of class $C^{m,m}$, $m \geq 1$, with uniformly bounded derivatives. From Theorem 4.8.6 in Kloeden & Platen (1999) due to Mikulevicius (1983), we know that the valuation function $u_{n,m} : [t_0, T] \times \mathfrak{R} \rightarrow \mathfrak{R}$ given by

$$u_{n,m}(t, x) = E(I_{n,m}(h)(X_T^{t,x})) \quad (15.5.14)$$

for $(t, x) \in [t_0, T] \times \mathfrak{R}$ and $n \in \mathcal{N}$ is of class $C^{1,m}$ for all even integers $m \geq 2$. In particular, for $m \geq 4$, $u_{n,m}$ will be of class $C^{1,4}$ and hence, using the results in Sect. 15.2 will admit an Itô integral representation of the form (15.2.17) or (15.2.18). If (15.4.1) holds for the process $X^{t_0, \underline{x}}$ and the linearized process $Z^{s,1}$, $s \in [t_0, T]$, given by (15.2.11), satisfies

(15.4.2), then by applying Theorem 15.4.1 we see that the valuation function $u : [t_0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ given by (15.2.4) will also admit a representation of the form (15.2.17) or (15.2.18). Thus, we have found an Itô integral representation of the random variable $u(T, X_T^{t_0, \underline{x}}) = h(X_T^{t_0, \underline{x}})$ with explicit expressions for the integrand even in the case where the derivative of h is discontinuous at a finite number of points.

This result, which includes a wide class of non-smooth payoff functions, justifies the effort needed to deal with the technical difficulties encountered in the previous section.

15.6 Maximum of Several Assets

In this section we will apply the martingale representation from Sect. 15.4 to obtain explicit representations for non-smooth functionals of the maximum of several assets as, for example, are used in basket options. This is a challenging problem and will lead to some complex notations and formulations.

Vector Payoff Structures

Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a payoff function which satisfies conditions A1* and A2* given in Sect. 15.5. We assume that $I_{n,4}(h)$, for $n \in \mathcal{N}$, are approximating functions of the class $C^{1,4}$, as given by (15.5.4). We now consider a payoff structure of the form

$$h\left(\max\left(\mathbf{X}_T^{t_0, \underline{x}}\right)\right) = h\left(\max\left(X_T^{1,t_0,\underline{x}}, \dots, X_T^{d,t_0,\underline{x}}\right)\right), \quad (15.6.1)$$

where $\mathbf{X}^{t_0, \underline{x}} = \{\mathbf{X}_t^{t_0, \underline{x}}, t \in [t_0, T]\}$ is the solution of the d -dimensional SDE (15.1.1), starting at time t_0 with initial value $\underline{x} = (\underline{x}_1, \dots, \underline{x}_d)^\top \in \mathbb{R}^d$ and $X^{1,t_0,\underline{x}}, \dots, X^{d,t_0,\underline{x}}$ are the components of $\mathbf{X}^{t_0, \underline{x}}$.

In this section we assume that the corresponding valuation function $u : [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is given by the time-independent version of (15.1.4) such that

$$u(t, x_1, \dots, x_d) = E\left(h\left(\max\left(X_T^{1,t,x}, \dots, X_T^{d,t,x}\right)\right)\right)$$

for $\mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$ and $t \in [t_0, T]$.

Our aim will be to find an explicit representation of u of the form (15.4.4) or (15.4.5) using the martingale representation in Theorem 15.4.1. We do this by constructing an appropriate sequence of approximating functions for the payoff functional $h \circ \max : \mathbb{R}^d \rightarrow \mathbb{R}$ corresponding to u given by $h \circ \max(\mathbf{x}) = h(\max(\mathbf{x}))$ for $\mathbf{x} \in \mathbb{R}^d$.

Let $n \in \mathcal{N}$ and define $f_n : \mathbb{R}^d \rightarrow \mathbb{R}$ by

$$f_n(\mathbf{x}) = (5n)^{4d} \int_{C_n} \int_{C_n} \int_{C_n} \int_{C_n} \max \left(\mathbf{x} - \boldsymbol{\alpha}^{(n)} - \sum_{k=1}^4 \mathbf{y}^{(k)} \right) d\mathbf{y}^{(1)} \dots d\mathbf{y}^{(4)} \quad (15.6.2)$$

for $\mathbf{x} \in \mathbb{R}^d$, where $\boldsymbol{\alpha}^{(n)} = (\frac{1}{n}, \frac{2}{n}, \dots, \frac{d}{n})^\top \in \mathbb{R}^d$ and C_n is the d -dimensional cube of length $\frac{1}{5n}$ given by

$$C_n = \left\{ \mathbf{a} \in \mathbb{R}^d : \mathbf{a} = (a_1, \dots, a_d)^\top, a_i \in \left[0, \frac{1}{5n}\right] \text{ for } i \in \{1, 2, \dots, d\} \right\}.$$

If we take $\tau = T$, then we can remove the time parameter t from the formulation of conditions A1 and A2 given in Sect. 15.4. For example, condition A1(i) becomes

A1*(i) For each $\mathbf{x} \in \mathbb{R}^d$ we have

$$\lim_{n \rightarrow \infty} h_n(\mathbf{x}) = h(\mathbf{x}).$$

As already mentioned, we refer to this representation of conditions A1 and A2 as the time-independent formulation with $\tau = T$. Note that $\boldsymbol{\alpha}^{(n)} \in \mathbb{R}^d$ has components $\alpha_i^{(n)} = \frac{i}{n}$ for $i \in \{1, 2, \dots, d\}$. We will show that the time-independent formulation of conditions A1 and A2 with $\tau = T$ holds for the payoff function $h \circ \max$ using the approximating functions

$$h_n = I_{n,4}(h) \circ f_n,$$

$n \in \mathcal{N}$ given by

$$I_{n,4}(h) \circ f_n(\mathbf{x}) = I_{n,4}(h)(f_n(\mathbf{x}))$$

for $\mathbf{x} \in \mathbb{R}^d$.

Verifying Condition A1(i)

From the definition of the functions f_n , given by (15.6.2), we see that

$$\begin{aligned} |f_n(\mathbf{x})| &\leq (5n)^{4d} \int_{C_n} \int_{C_n} \int_{C_n} \int_{C_n} \left| \max \left(\mathbf{x} - \boldsymbol{\alpha}^{(n)} - \sum_{k=1}^4 \mathbf{y}^{(k)} \right) \right| d\mathbf{y}^{(1)} \dots d\mathbf{y}^{(4)} \\ &\leq (5n)^{4d} \int_{C_n} \int_{C_n} \int_{C_n} \int_{C_n} \left(|\max(\mathbf{x})| + \frac{d+1}{n} \right) d\mathbf{y}^{(1)} \dots d\mathbf{y}^{(4)} \\ &\leq |\mathbf{x}| + \frac{d+1}{n} \end{aligned} \quad (15.6.3)$$

for all $\mathbf{x} \in \mathbb{R}^d$ and $n \in \mathcal{N}$. Consequently, from the linear growth bound that applies for $I_{n,4}(h)$ given by condition A2*(i) and the inequality $(a+b)^2 \leq 2(a^2 + b^2)$ we have

$$\begin{aligned}
|I_{n,4}(h) \circ f_n(\mathbf{x})|^2 &= |I_{n,4}(h)(f_n(\mathbf{x}))|^2 \\
&\leq K_3^2 (1 + |f_n(\mathbf{x})|^2) \\
&\leq K_3^2 \left(1 + \left(|\mathbf{x}| + \frac{d+1}{n} \right)^2 \right) \\
&\leq K_3^2 \left(1 + 2 \left(\frac{d+1}{n} \right)^2 + 2|\mathbf{x}|^2 \right) \\
&\leq 2K_3^2 \left(1 + (d+1)^2 \right) (1 + |\mathbf{x}|^2) \\
&\leq K_5^2 (1 + |\mathbf{x}|^2)
\end{aligned} \tag{15.6.4}$$

for all $\mathbf{x} \in \mathbb{R}^d$, and $n \in \mathcal{N}$. Here $K_5^2 = 2K_3^2(1 + (d+1)^2)$ and $K_3 = A_4$ are given by (15.5.6). This result shows that condition A2(i) holds in the time-independent case with $\tau = T$ for the approximating functions $h_n = I_{n,4}(h) \circ f_n$, $n \in \mathcal{N}$.

From the Mean Value Theorem and the definition of f_n , given by (15.6.2), we know that for any $\mathbf{x} \in \mathbb{R}^d$ and $n \in \mathcal{N}$

$$f_n(\mathbf{x}) = \max \left(\mathbf{x} - \boldsymbol{\alpha}^{(n)} - 4\boldsymbol{\beta}^{(n)} \right)$$

for some $\boldsymbol{\beta}^{(n)} \in C_n$. This shows that

$$\lim_{n \rightarrow \infty} f_n(\mathbf{x}) = \max(\mathbf{x}) \tag{15.6.5}$$

for all $\mathbf{x} \in \mathbb{R}^d$. By condition A1*(i), see (15.5.8), this means that

$$\begin{aligned}
\lim_{n \rightarrow \infty} I_{n,4}(h) \circ f_n(\mathbf{x}) &= \lim_{n \rightarrow \infty} I_{n,4}(h)(f_n(\mathbf{x})) \\
&= h(\max(\mathbf{x})) \\
&= h \circ \max(\mathbf{x})
\end{aligned} \tag{15.6.6}$$

for all $\mathbf{x} \in \mathbb{R}^d$. This validates condition A1(i) in the time-independent case with $\tau = T$ for the payoff function $h \circ \max$ using the approximating functions

$$h_n = I_{n,4}(h) \circ f_n.$$

Verifying Condition A1(ii)

From definition (15.6.2) it is a straightforward but tedious task to show that the functions f_n , $n \in \mathcal{N}$, are of class C^4 . To calculate pointwise limits of the partial derivatives of f_n as $n \rightarrow \infty$ we define $\pi : \mathbb{R}^d \rightarrow \{1, 2, \dots, d\}$ by the equation

$$\pi(\mathbf{x}) = \min_{i \in [1, d]} \{i : x_i = \max(\mathbf{x})\}. \quad (15.6.7)$$

In addition, for $\zeta > 0$ and $i \in \{1, 2, \dots, d\}$ define $\boldsymbol{\zeta}^{(i)} = (\zeta_1^{(i)}, \dots, \zeta_d^{(i)})^\top \in \Re^d$ by the rule $\zeta_k^{(i)} = 0$ for $i \neq k$ and $\zeta_k^{(i)} = \zeta$ for $i = k$.

Let $i \in \{1, 2, \dots, d\}$, $n \in \mathcal{N}$ and $\mathbf{x} \in \Re^d$ be fixed with $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(4)} \in C_n$. If $i \neq \pi(\mathbf{x})$, then either $x_i < \max(\mathbf{x})$ or there is an integer j , with $j < i$ and $x_j = x_i = \max(\mathbf{x})$.

If $x_i < \max(\mathbf{x})$ we choose some integer j with $x_j = \max(\mathbf{x})$ and $x_i < x_j$. Then for $0 \leq \zeta \leq \frac{x_j - x_i}{2}$ and $n > \frac{2(d+1)}{x_j - x_i}$, so that $\frac{d+1}{n} < \frac{x_j - x_i}{2}$, one can show that

$$x_i + \zeta + \left(\alpha_j^{(n)} - \alpha_i^{(n)} \right) + \sum_{k=1}^4 \left(y_j^{(k)} - y_i^{(k)} \right) \leq x_j$$

or

$$x_i + \zeta - \alpha_i^{(n)} - \sum_{k=1}^4 y_i^{(k)} \leq x_j - \alpha_j^{(n)} - \sum_{k=1}^4 y_j^{(k)}$$

for all $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(4)} \in C_n$. This means that for $0 \leq \zeta \leq \frac{x_j - x_i}{2}$ and $n > \frac{2(d+1)}{x_j - x_i}$ one has the relation

$$\max \left(\mathbf{x} + \boldsymbol{\zeta}^{(i)} - \boldsymbol{\alpha}^{(n)} - \sum_{k=1}^4 \mathbf{y}^{(k)} \right) = \max \left(\mathbf{x} - \boldsymbol{\alpha}^{(n)} - \sum_{k=1}^4 \mathbf{y}^{(k)} \right) \quad (15.6.8)$$

for all $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(4)} \in C_n$.

If $x_j = x_i = \max(\mathbf{x})$, $j < i$, then for $0 < \zeta \leq \frac{1}{5n}$ and using the identity $\alpha_i^{(n)} - \alpha_j^{(n)} = \frac{i-j}{n}$ we can show

$$x_i + \zeta + \sum_{k=1}^4 \left(y_j^{(k)} - y_i^{(k)} \right) \leq x_j + \left(\alpha_i^{(n)} - \alpha_j^{(n)} \right)$$

or

$$x_i + \zeta - \alpha_i^{(n)} - \sum_{k=1}^4 y_i^{(k)} \leq x_j - \alpha_j^{(n)} - \sum_{k=1}^4 y_j^{(k)}$$

for all $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(4)} \in C_n$ and $n \in \mathcal{N}$. Again this means that (15.6.8) holds for all $0 < \zeta \leq \frac{1}{5n}$ and $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(4)} \in C_n$.

Combining these two results and the definition of f_n given by (15.6.2) we have, in the case where $i \neq \pi(\mathbf{x})$, and for sufficiently large n and sufficiently small ζ , the identity

$$f_n(\mathbf{x} + \boldsymbol{\zeta}^{(i)}) - f_n(\mathbf{x}) = 0.$$

Letting $\zeta \rightarrow 0$ we therefore obtain

$$\lim_{n \rightarrow \infty} \frac{\partial f_n(\mathbf{x})}{\partial x_i} = 0. \quad (15.6.9)$$

If $i = \pi(\mathbf{x})$, then $x_i = \max(\mathbf{x})$ and $x_j < x_i$ for $j < i$, $j \in \{1, 2, \dots, d\}$. Thus, for $j < i$ and $n > \frac{d+1}{\min_{j < i} |x_i - x_j|}$, so that $\frac{d+1}{n} < \min_{j < i} |x_i - x_j|$, we have the inequality

$$x_j + \left(\alpha_i^{(n)} - \alpha_j^{(n)} \right) + \sum_{k=1}^4 \left(y_i^{(k)} - y_j^{(k)} \right) \leq x_j + \frac{d+1}{n} \leq x_i,$$

for all $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(4)} \in C_n$.

If $j > i$, $j \in \{1, 2, \dots, d\}$ and $n > \frac{d+1}{\min_{j < i} |x_i - x_j|}$, then $\alpha_i^{(n)} - \alpha_j^{(n)} = \frac{i-j}{n} \leq -\frac{1}{n}$ and, therefore, we can infer that

$$x_j + \left(\alpha_i^{(n)} - \alpha_j^{(n)} \right) + \sum_{k=1}^4 \left(y_i^{(k)} - y_j^{(k)} \right) \leq x_j \leq x_i,$$

for all $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(4)} \in C_n$.

These inequalities show that if $i = \pi(\mathbf{x})$ and $j \neq i$, $j \in \{1, 2, \dots, d\}$, then for sufficiently large n

$$x_j - \alpha_j^{(n)} - \sum_{k=1}^4 y_j^{(k)} \leq x_i - \alpha_i^{(n)} - \sum_{k=1}^4 y_i^{(k)}$$

or

$$\max \left(\mathbf{x} - \boldsymbol{\alpha}^{(n)} - \sum_{k=1}^4 \mathbf{y}^{(k)} \right) = x_i - \alpha_i^{(n)} - \sum_{k=1}^4 y_i^{(k)},$$

for all $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(4)} \in C_n$ and hence for $\zeta > 0$ and sufficiently large n

$$\max \left(\mathbf{x} + \zeta^{(i)} - \boldsymbol{\alpha}^{(n)} - \sum_{k=1}^4 \mathbf{y}^{(k)} \right) = x_i + \zeta - \alpha_i^{(n)} - \sum_{k=1}^4 y_i^{(k)}.$$

Consequently, if $i = \pi(\mathbf{x})$, then from the definition of f_n given by (15.6.2) we see that for sufficiently large n that

$$f_n(\mathbf{x} + \zeta^{(i)}) - f_n(\mathbf{x}) = \zeta.$$

Taking the limit as $n \rightarrow \infty$ we therefore obtain

$$\lim_{n \rightarrow \infty} \frac{\partial f_n(\mathbf{x})}{\partial x_i} = 1. \quad (15.6.10)$$

For $i \in \{1, 2, \dots, d\}$ define $q_i, Q_i : \Re^d \rightarrow \Re$ by

$$q_i(\mathbf{x}) = \begin{cases} 1 & : i = \pi(\mathbf{x}) \\ 0 & : i \neq \pi(\mathbf{x}) \end{cases} \quad (15.6.11)$$

and

$$Q_i(\mathbf{x}) = g(\max(\mathbf{x})) q_i(\mathbf{x})$$

for $\mathbf{x} \in \Re^d$, where

$$\lim_{n \rightarrow \infty} (I_{n,4}(h))'(y) = g(y),$$

with $y \in \Re$ as given by condition A1*(ii), see (15.5.13).

Using this definition, (15.6.9) and (15.6.10) we have

$$\lim_{n \rightarrow \infty} \frac{\partial f_n(\mathbf{x})}{\partial x_i} = q_i(\mathbf{x})$$

for all $\mathbf{x} \in \Re^d$. Thus by the chain rule, condition A1*(ii) and (15.6.5) we see that

$$\begin{aligned} \lim_{n \rightarrow \infty} I_{n,4}(h) \left(\frac{\partial f_n(\mathbf{x})}{\partial x_i} \right) &= \lim_{n \rightarrow \infty} (I_{n,4}(h))'(f_n(\mathbf{x})) \frac{\partial f_n(\mathbf{x})}{\partial x_i} \\ &= g(\max(\mathbf{x})) q_i(\mathbf{x}) \\ &= Q_i(\mathbf{x}) \end{aligned} \quad (15.6.12)$$

for all $\mathbf{x} \in \Re^d$ and $i \in \{1, 2, \dots, d\}$. This proves that condition A1(ii) holds in the time-independent case with $\tau = T$ for the approximating functions $\frac{\partial}{\partial x_i} (I_{n,4}(h) \circ f_n)$, $n \in \mathcal{N}$, $i \in \{1, 2, \dots, d\}$ with pointwise limit functions Q_i .

Verifying Condition A2 (ii)

A straightforward calculation using the definition of f_n , shows that for any integers $n \in \mathcal{N}$, $i \in \{1, 2, \dots, d\}$, $\mathbf{x} \in \Re^d$ and $\zeta > 0$ we have the inequality

$$f_n(\mathbf{x} + \zeta^i) - f_n(\mathbf{x}) \leq \zeta$$

and hence

$$\left| \frac{\partial f_n(\mathbf{x})}{\partial x_i} \right| \leq 1.$$

Consequently, applying condition A2*(ii), see (15.5.9), which holds for $(I_{n,4}(h))'$, (15.6.3) and similar arguments used in the derivation of (15.6.4), we can infer that

$$\begin{aligned} \left| I_{n,4}(h) \left(\frac{\partial f_n(\mathbf{x})}{\partial x_i} \right) \right|^2 &\leq |(I_{n,4}(h))'(f_n(\mathbf{x}))|^2 \\ &\leq K_4^2 (1 + |f_n(\mathbf{x})|^2) \\ &\leq K_4^2 \left(1 + \left(|\mathbf{x}| + \frac{d+1}{n} \right)^2 \right) \\ &\leq K_6^2 (1 + |\mathbf{x}|^2) \end{aligned}$$

for all $\mathbf{x} \in \mathbb{R}^d$ and $i \in \{1, 2, \dots, d\}$, where $K_6^2 = 2K_4^2(1 + (d+1)^2)$. This verifies that condition A2(ii) holds in the time-independent case with $\tau = T$ for the approximating functions $\frac{\partial}{\partial x_i}(I_{n,4}(h) \circ f_n)$, $n \in \mathcal{N}$.

We will now assume that the valuation function $u_n : [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ given by

$$u_n(t, \mathbf{x}) = E \left(I_{n,4}(h) \left(f_n \left(X_T^{1,t_0,\underline{x}}, \dots, X_T^{d,t_0,\underline{x}} \right) \right) \right) \quad (15.6.13)$$

for $\mathbf{x} = (x_1, \dots, x_d)^T \in \mathbb{R}^d$, $t \in [t_0, T]$ is of class $C^{1,3}$. For example, using Theorem 4.8.6 in Kloeden & Platen (1999) we see that if the drift a^i and diffusion coefficients $b^{i,j}$ have uniformly bounded derivatives and are of class $C^{4,4}$, then u_n will be of class $C^{1,4}$. Thus, applying the results of the previous sections, the payoff function $I_{n,4}(h) \circ f_n(X_T^{1,t_0,\underline{x}}, \dots, X_T^{d,t_0,\underline{x}})$ will admit an Itô integral representation of the form (15.4.3). Since the payoff function $h \circ \max : \mathbb{R}^d \rightarrow \mathbb{R}$ with approximating functions $h_n = I_{n,4}(h) \circ f_n$, $n \in \mathcal{N}$, satisfies conditions A1 and A2 in Sect. 15.4 in the time-independent case with $\tau = T$, we can apply Theorem 15.4.1 with $\tau = T$ to the functional $h \circ \max$ and the diffusion process $\mathbf{X}_T^{t_0,\underline{x}}$ yielding

$$h \left(\max \left(\mathbf{X}_T^{t_0,\underline{x}} \right) \right) = E \left(h \left(\max \left(\mathbf{X}_T^{t_0,\underline{x}} \right) \right) \right) + \sum_{\ell=1}^d \sum_{j=1}^m \int_{t_0}^T \gamma_s^i b^{i,j}(s, \mathbf{X}_s^{t_0,\underline{x}}) dW_s^j, \quad (15.6.14)$$

where

$$\gamma_s^i = E \left(\sum_{p=1}^d Q_p \left(\mathbf{X}_T^{t_0,\underline{x}} \right) Z_T^{p,i,s,\delta} \mid \mathcal{A}_s \right)$$

for $i \in \{1, 2, \dots, d\}$ and $Z_T^{p,i,s,\delta}$ is the unique solution of the SDE (15.3.5) with initial value $\boldsymbol{\delta}$ at time $s \in [t_0, T]$ as given by (15.3.8).

Thus, we have found an explicit representation of the payoff structure (15.6.1). This is the case when h is an absolutely continuous function of the form (15.5.1) and the approximating valuation functions u_n , $n \in \mathcal{N}$, are sufficiently smooth.

15.7 Hedge Ratios for Lookback Options

In this section another important application of the martingale representation theorem in Sect. 15.4 will be presented which is related to lookback options.

Lookback Payoff

Let $X^{t_0,\underline{x}} = \{X_t^{t_0,\underline{x}}, t \in [t_0, T]\}$ be the solution of the one-dimensional SDE (15.2.1) starting at time t_0 with initial value $\underline{x} \in \mathbb{R}$. We assume that $h : \mathbb{R} \rightarrow \mathbb{R}$

is a payoff function for which conditions A1* and A2* given in Sect. 15.5 hold. We now consider, so called, *lookback options* with payoff structures of the form

$$h \left(\sup_{t \in [t_0, T]} X_t^{t_0, \underline{x}} \right). \quad (15.7.1)$$

To ensure that option prices for this payoff structure are well-defined, we also assume that the mean square bound (15.4.1) holds. Our aim will be to find an explicit representation for the payoff structure (15.7.1) of the form (15.4.5).

We remark that finding an explicit Itô integral representation of the payoff structure for a lookback option is generally considered as one of the more difficult problems concerning the computation of hedge ratios for path-dependent options. For example Föllmer (1991) applies Clark's formula to obtain an explicit representation in the simple case, where $X^{t_0, \underline{x}}$ is a geometric Brownian motion and h is the identity function.

Other path-dependent options, such as Asian options, can also be handled using the methods developed in this section. In fact, the analysis is often simpler because the state space may only need to be increased by a single extra variable.

For integers $m \in \mathcal{N}$ and $i \in \{1, \dots, d_m\}$, where $d_m = 2^m$, consider an equidistant time discretization of the interval $[t_0, T]$ of the form $t_0 = \tau_0 < \tau_1 < \dots < \tau_{d_m} = T$ with step size $\Delta_m = \frac{T-t_0}{d_m}$. We also denote by $\bar{\mathbf{X}}^{t_0, \underline{x}} = \{\bar{\mathbf{X}}_t^{t_0, \underline{x}} = (X_t^{1, t_0, \underline{x}}, \dots, X_t^{d_m, t_0, \underline{x}})^\top, t \in [t_0, T]\}$ the unique solution of the d_m -dimensional SDE with components

$$dX_t^{i, t_0, \underline{x}} = \mathbf{1}_{\{t < \tau_i\}} \left(a \left(t, X_t^{i, t_0, \underline{x}} \right) dt + b \left(t, X_t^{i, t_0, \underline{x}} \right) dW_t \right) \quad (15.7.2)$$

for $t \in [t_0, T]$, $i \in \{1, \dots, d_m\}$, starting at time t_0 with initial value $\underline{x} = (\underline{x}, \dots, \underline{x})^\top \in \mathbb{R}^{d_m}$. For simplicity, we will use the symbol \underline{x} for the initial value of the vector diffusion $\bar{\mathbf{X}}^{t_0, \underline{x}}$ at time t_0 and \underline{x} the initial value of the one-dimensional diffusion $X^{t_0, \underline{x}}$ at time t_0 .

For each integer $i \in \{1, \dots, d_m\}$, the evolution of the component process $X^{i, t_0, \underline{x}}$ is stopped at time τ_i and we can write

$$X_t^{i, t_0, \underline{x}} = X_{t \wedge \tau_i}^{t_0, \underline{x}} \quad (15.7.3)$$

for $t \in [t_0, T]$. For the time discretization with step size Δ_m define

$$M_T^{m, t_0, \underline{x}} = \max_{i \in \{1, \dots, d_m\}} X_T^{i, t_0, \underline{x}} \quad (15.7.4)$$

and

$$M_T^{\infty, t_0, \underline{x}} = \sup_{t \in [t_0, T]} X_t^{t_0, \underline{x}}.$$

Let $h_n = I_{n,4}(h)$, $n \in \mathcal{N}$, be the sequence of approximating functions for h , as given by (15.5.4), and consider the payoff structure

$$h_n \left(\max_{i \in \{1, \dots, d_m\}} X_T^{i, t_0, \underline{x}} \right) = h_n \left(M_T^{m, t_0, \underline{x}} \right) \quad (15.7.5)$$

for $n \in \mathcal{N}$. As in Sect. 15.5 we assume that the valuation function $u_{n,m}$ given by

$$u_{n,m}(t, x_1, \dots, x_{d_m}) = E(h_n(M_T^{m, t_0, x}))$$

for $\underline{x} = (x_1, \dots, x_{d_m})^\top \in \Re^{d_m}$ and $t \in [t_0, T]$ is of class $C^{1,3}$, and hence, equation (15.6.14) admits a representation of the form

$$h_n(M_T^{m, t_0, \underline{x}}) = E(h_n(M_T^{m, t_0, \underline{x}})) + \sum_{i=1}^{d_m} \int_{t_0}^T \gamma_s^i \mathbf{1}_{\{s < \tau_i\}} b(t, \bar{\mathbf{X}}_s^{t_0, \underline{x}}) dW_s, \quad (15.7.6)$$

where

$$\begin{aligned} \gamma_s^i &= E \left(\sum_{p=1}^{d_m} Q_p(\bar{\mathbf{X}}_T^{t_0, \underline{x}}) Z_T^{p, i, s, \delta} \mid \mathcal{A}_s \right) \\ &= E \left(\sum_{p=1}^{d_m} h'_n(M_T^{m, t_0, \underline{x}}) q_p(\bar{\mathbf{X}}_T^{t_0, \underline{x}}) Z_T^{p, i, s, \delta} \mid \mathcal{A}_s \right) \end{aligned}$$

for $n, m \in \mathcal{N}$, with the functions Q_p and q_p , $p \in \{1, \dots, d_m\}$, defined by (15.6.12) and (15.6.11), respectively. Here $Z^{s, \delta}$ is the unique solution of the d_m^2 -dimensional SDE (15.3.5) with initial value δ given by (15.3.8) at time $s \in [t_0, T]$. Note that an indicator function is included as part of the integrand in (15.7.6) because it forms part of the diffusion coefficient appearing in (15.7.2). Since this representation depends on an application of Theorem 15.4.1, we assume that the growth bound (15.4.2) holds for the process $Z^{s, \delta}$.

Linearized SDE

Using the specific structure that applies for the underlying vector process $\bar{\mathbf{X}}^{t_0, \underline{x}}$ given by (15.7.2) the linearized SDE for the component $Z^{p, i, s, \delta}$, $p, i \in \{1, \dots, d_m\}$, can be simplified and expressed in the form

$$dZ_t^{p, i, s, \delta} = \mathbf{1}_{\{t < \tau_p\}} Z_t^{p, i, s, \delta} \left(\frac{\partial}{\partial x} a(t, X_t^{p, t_0, \underline{x}}) dt + \frac{\partial}{\partial x} b(t, X_t^{p, t_0, \underline{x}}) dW_t \right) \quad (15.7.7)$$

for $t_0 \leq s \leq t \leq T$. The solution to (15.7.7) for each $p, i \in \{1, \dots, d_m\}$, has an exponential form, and therefore, by applying the initial condition $\delta^{p,i} = 0$, for $p \neq i$ given by (15.3.8), we obtain

$$Z_t^{p, i, s, \delta} = 0 \quad (15.7.8)$$

a.s. for $t_0 < s \leq t \leq T$. If we let $Z^{s, 1} = \{Z_t^{s, 1}, t \in [s, T]\}$ be the solution of the one-dimensional linearized SDE (15.2.11), then from (15.7.7) and the initial condition $\delta^{p,i} = 1$ for $p = i$ we can write

$$Z_t^{i,i,s,\delta} = Z_{t \wedge \tau_i}^{s,1}, \quad (15.7.9)$$

for $t \in [t_0, T]$, $i \in \{1, \dots, d_m\}$. By (15.7.8) the representation given by (15.7.6) can also be simplified and expressed as

$$\begin{aligned} h_n(M_T^{m,t_0,x}) &= E(h_n(M_T^{m,t_0,x})) \\ &+ \sum_{i=0}^{d_m} \int_{t_0}^T E(h'_n(M_T^{m,t_0,x}) q_i(\bar{\mathbf{X}}_T^{t_0,x}) Z_T^{i,i,s,\delta} | \mathcal{A}_s) \\ &\quad \times \mathbf{1}_{\{s < \tau_i\}} b(s, \mathbf{X}_s^{t_0,x}) dW_s \\ &= E(h_n(M_T^{m,t_0,x})) \\ &+ \int_{t_0}^T \sum_{i=0}^{d_m} E(h'_n(M_T^{m,t_0,x}) q_i(\bar{\mathbf{X}}_T^{t_0,x}) Z_T^{i,i,s,\delta} \mathbf{1}_{\{s < \tau_i\}} | \mathcal{A}_s) \\ &\quad \times b(s, \mathbf{X}_s^{t_0,x}) dW_s \end{aligned} \quad (15.7.10)$$

for $n, m \in \mathcal{N}$. For each $m \in \mathcal{N}$ and time discretization with step size Δ_m let $\pi_m = \pi(\bar{\mathbf{X}}_T^{t_0,x})$, where $\pi : \mathbb{R}^{d_m} \rightarrow \{1, \dots, d_m\}$ is given by (15.6.7). Using this function we define $\bar{\tau}_m : \Omega \rightarrow \mathbb{R}$ by

$$\bar{\tau}_m = \tau_{\pi_m}. \quad (15.7.11)$$

The variable $\bar{\tau}_m$ is a random time, which is \mathcal{A}_T -measurable but $t \wedge \bar{\tau}_m$ may not be \mathcal{A}_t -measurable, and consequently, $\bar{\tau}_m$ will not in general be a stopping time.

Approximate Representation for Smooth Payoffs

From the definition of the random time $\bar{\tau}_m$ and q_i , $i \in \{1, \dots, d_m\}$, together with (15.7.9) and (15.7.10) we obtain the approximate representation

$$\begin{aligned} h_n(M_T^{m,t_0,x}) &= E(h_n(M_T^{m,t_0,x})) \\ &+ \int_{t_0}^T E(h'_n(M_T^{m,t_0,x}) Z_{\bar{\tau}_m}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_m\}} | \mathcal{A}_s) b(s, \mathbf{X}_s^{t_0,x}) dW_s. \end{aligned} \quad (15.7.12)$$

By continuity of the sample paths of $\mathbf{X}^{t_0,x}$ we know that for a fixed integer $n \in \mathcal{N}$

$$\lim_{m \rightarrow \infty} h_n(M_T^{m,t_0,x}) \stackrel{\text{a.s.}}{=} h_n(M_T^{\infty,t_0,x}), \quad (15.7.13)$$

where $M_T^{\infty,t_0,x}$ is given in (15.7.4). Also, the linear growth of h_n , given by condition A2*(i), means that

$$\begin{aligned} \left| h_n \left(M_T^{m,t_0,\underline{x}} \right) \right|^2 &\leq K_3^2 \left(1 + \left| M_T^{m,t_0,\underline{x}} \right|^2 \right) \\ &\leq K_3^2 \left(1 + \sup_{t \in [t_0, T]} \left| \mathbf{X}_t^{t_0,\underline{x}} \right|^2 \right). \end{aligned} \quad (15.7.14)$$

Note that this bound applies uniformly for both integer variables n and m . In addition, from the mean square bound (15.4.1) we see that

$$E \left(K_1^2 \left(1 + \sup_{t \in [t_0, T]} \left| \mathbf{X}_t^{t_0,\underline{x}} \right|^2 \right) \right) < \infty.$$

This result together with (15.7.13) and (15.7.14) means that we can exploit a version of the Dominated Convergence Theorem applicable for L^p spaces, $p > 0$, see for example the Corollary to Theorem 2.6.3 in Shiryaev (1984), to obtain

$$\lim_{m \rightarrow \infty} \left\| h_n \left(M_T^{m,t_0,\underline{x}} \right) - h_n \left(M_T^{\infty,t_0,\underline{x}} \right) \right\|_2 = 0 \quad (15.7.15)$$

for all integers $n \in \mathcal{N}$. Furthermore, by Hölder's inequality

$$\begin{aligned} \left| E \left(h_n \left(M_T^{m,t_0,\underline{x}} \right) \right) - E \left(h_n \left(M_T^{\infty,t_0,\underline{x}} \right) \right) \right| &\leq E \left(\left| h_n \left(M_T^{m,t_0,\underline{x}} \right) - h_n \left(M_T^{\infty,t_0,\underline{x}} \right) \right| \right) \\ &= \left\| h_n \left(M_T^{m,t_0,\underline{x}} \right) - h_n \left(M_T^{\infty,t_0,\underline{x}} \right) \right\|_1 \\ &\leq \left\| h_n \left(M_T^{m,t_0,\underline{x}} \right) - h_n \left(M_T^{\infty,t_0,\underline{x}} \right) \right\|_2 \end{aligned}$$

so that from (15.7.15)

$$\lim_{m \rightarrow \infty} \left| E \left(h_n \left(M_T^{m,t_0,\underline{x}} \right) \right) - E \left(h_n \left(M_T^{\infty,t_0,\underline{x}} \right) \right) \right| = 0 \quad (15.7.16)$$

for all $n \in \mathcal{N}$.

Since $d_m = 2^m$ for $m \in \mathcal{N}$ any discretization point τ_i , $i \in \{1, \dots, d_m\}$, belonging to the time discretization with step size Δ_m will be an element of the set of discretization points for any time discretization with step size $\Delta_{m'}$ with $m' \in \mathcal{N}$ and $m' \geq m$. This means that $(\bar{\tau}_m)_{m \in \mathcal{N}}$, is a non-decreasing sequence of random times and thus $\lim_{m \rightarrow \infty} \bar{\tau}_m(\omega)$ exists for each $\omega \in \Omega$. Let

$$\lim_{m \rightarrow \infty} \bar{\tau}_m(\omega) = \bar{\tau}_\infty(\omega) \quad (15.7.17)$$

for some random time $\bar{\tau}_\infty : \Omega \rightarrow \mathbb{R}$ and let $s \in [t_0, T]$ be fixed. If $\omega \in \Omega$ and $s < \bar{\tau}_\infty(\omega)$, then by (15.7.17) there is an $M > 0$ such that for $m > M$ we have $s < \bar{\tau}_m(\omega)$. If $s \geq \bar{\tau}_\infty(\omega)$ then clearly $s \geq \bar{\tau}_m(\omega)$ for all $m \in \mathcal{N}$. Thus

$$\lim_{m \rightarrow \infty} \mathbf{1}_{\{s < \bar{\tau}_m\}} \stackrel{\text{a.s.}}{=} \mathbf{1}_{\{s < \bar{\tau}_\infty\}} \quad (15.7.18)$$

Combining this result, (15.7.13) and by continuity of the function h'_n and the sample paths of $Z^{s,1}$ we can infer that

$$\lim_{m \rightarrow \infty} h'_n(M_T^{m,t_0,x}) Z_{\bar{\tau}_m}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_m\}} \stackrel{\text{a.s.}}{=} h'_n(M_T^{\infty,t_0,x}) Z_{\bar{\tau}_\infty}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}} \quad (15.7.19)$$

for any $n \in \mathcal{N}$.

Applying the linear growth bound A2*(ii), we know that for integers $n, m \in \mathcal{N}$ and $s \in [t_0, T]$

$$\begin{aligned} \left| h'_n(M_T^{m,t_0,x}) Z_{\bar{\tau}_m}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_m\}} \right| &\leq K_4 \sqrt{\left(1 + |M_T^{m,t_0,x}|^2\right)} |Z_{\bar{\tau}_m}^{s,1}| \\ &\leq K_4 \sqrt{\left(1 + \sup_{t \in [t_0, T]} |\mathbf{X}_t^{t_0,x}|^2\right)} |Z_{\bar{\tau}_m}^{s,1}|. \end{aligned} \quad (15.7.20)$$

Also, applying the mean square bounds (15.4.1) and (15.4.2) together with Hölder's inequality we have

$$\begin{aligned} E\left(K_4 \sqrt{\left(1 + \sup_{t \in [t_0, T]} |\mathbf{X}_t^{t_0,x}|^2\right)} |Z_{\bar{\tau}_m}^{s,1}|\right) &\leq K_4 \left\| \sqrt{\left(1 + \sup_{t \in [t_0, T]} |\mathbf{X}_t^{t_0,x}|^2\right)} \right\|_2 \|Z_{\bar{\tau}_m}^{s,1}\|_2 \\ &< \infty \end{aligned} \quad (15.7.21)$$

for $t \in [t_0, T]$. This inequality combined with (15.7.19) and (15.7.20) means that we can apply the Dominated Convergence Theorem for conditional expectations, see for example [Shiryayev \(1984\)](#), Theorem 2.7.2, to obtain

$$\lim_{m \rightarrow \infty} E\left(h'_n(M_T^{m,t_0,x}) Z_{\bar{\tau}_m}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_m\}} \mid \mathcal{A}_s\right) \stackrel{\text{a.s.}}{=} E\left(h'_n(M_T^{\infty,t_0,x}) Z_{\bar{\tau}_\infty}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}} \mid \mathcal{A}_s\right) \quad (15.7.22)$$

for all $n \in \mathcal{N}$ and $s \in [t_0, T]$.

Now define the process ξ_m , $m \in \mathcal{N}$, and ξ_∞ by

$$\begin{aligned} \xi_m &= E\left(h'_n(M_T^{m,t_0,x}) Z_{\bar{\tau}_m}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_m\}} \mid \mathcal{A}_s\right) b(s, \mathbf{X}_s^{t_0,x}), \\ \xi_\infty &= E\left(h'_n(M_T^{\infty,t_0,x}) Z_{\bar{\tau}_\infty}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}} \mid \mathcal{A}_s\right) b(s, \mathbf{X}_s^{t_0,x}). \end{aligned}$$

Equation (15.7.22) shows that

$$\lim_{m \rightarrow \infty} \xi_m \stackrel{\text{a.s.}}{=} \xi_\infty$$

and therefore, with these choices for ξ_m , $m \in \mathcal{N}$ and ξ_∞ , condition (b) is satisfied in the statement of Lemma 15.4.3.

In addition, the limits (15.7.15) and (15.7.16) combined with the representation (15.7.12) show that the Itô integral in (15.7.12) forms a Cauchy

sequence in $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ for $m \in \{1, 2, \dots\}$ and fixed $n \in \mathcal{N}$. Moreover, using similar arguments to those given in Sect. 15.4, see the commentary following equation (15.4.24), it can be shown that the integrand in the Itô integral of (15.7.12) is $\underline{\mathcal{A}}$ -adapted with continuous sample paths.

Consequently, condition (a) is satisfied in the statement of Lemma 15.4.3. These results mean that Lemma 15.4.3 can be applied, which together with (15.7.15) and (15.7.16) shows that

$$\begin{aligned} h_n(M_T^{\infty, t_0, \underline{x}}) &= E\left(h_n(M_T^{\infty, t_0, \underline{x}})\right) \\ &\quad + \int_{t_0}^T E\left(h'_n(M_T^{\infty, t_0, \underline{x}}) Z_{\bar{\tau}_\infty}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}} \mid \mathcal{A}_s\right) b(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s. \end{aligned} \quad (15.7.23)$$

We have thus found an explicit representation for a lookback option for the smooth payoff function h_n , $n \in \mathcal{N}$.

Representation for Non-Smooth Payoffs

Fortunately, these methods also extend to the case of a non-smooth payoff function h . In fact, using conditions A1*, A2*(i) and similar arguments to those given for the proof of (15.7.15), (15.7.16), (15.7.19) and (15.7.20) we can show

$$\begin{aligned} \lim_{n \rightarrow \infty} \|h_n(M_T^{\infty, t_0, \underline{x}}) - h(M_T^{\infty, t_0, \underline{x}})\|_2 &\stackrel{\text{a.s.}}{=} 0, \\ \lim_{n \rightarrow \infty} |E\left(h_n(M_T^{\infty, t_0, \underline{x}})\right) - E\left(h(M_T^{\infty, t_0, \underline{x}})\right)| &\stackrel{\text{a.s.}}{=} 0, \\ \lim_{n \rightarrow \infty} h'_n(M_T^{\infty, t_0, \underline{x}}) Z_{\bar{\tau}_\infty}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}} &\stackrel{\text{a.s.}}{=} g(M_T^{\infty, t_0, \underline{x}}) Z_{\bar{\tau}_\infty}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}}, \\ |h'_n(M_T^{\infty, t_0, \underline{x}}) Z_{\bar{\tau}_\infty}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}}| &\leq K_4 \sqrt{\left(1 + \sup_{t \in [t_0, T]} |\mathbf{X}_t^{t_0, \underline{x}}|^2\right)} |Z_{\bar{\tau}_\infty}^{s,1}|. \end{aligned} \quad (15.7.24)$$

The last two relations and (15.7.21) show that

$$\lim_{n \rightarrow \infty} E\left(h'_n(M_T^{\infty, t_0, \underline{x}}) Z_{\bar{\tau}_\infty}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}} \mid \mathcal{A}_s\right) \stackrel{\text{a.s.}}{=} E\left(g(M_T^{\infty, t_0, \underline{x}}) Z_{\bar{\tau}_\infty}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}} \mid \mathcal{A}_s\right). \quad (15.7.25)$$

The first two equations in (15.7.24) and relation (15.7.23) prove that the Itô integral in (15.7.23) is a Cauchy sequence in $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ for $n \in \mathcal{N}$. Also from (15.7.25) we see that

$$\begin{aligned} \lim_{n \rightarrow \infty} E\left(h'_n(M_T^{\infty, t_0, \underline{x}}) Z_{\bar{\tau}_\infty}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}} \mid \mathcal{A}_s\right) b(s, \mathbf{X}_s^{t_0, \underline{x}}) \\ \stackrel{\text{a.s.}}{=} E\left(g(M_T^{\infty, t_0, \underline{x}}) Z_{\bar{\tau}_\infty}^{s,1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}} \mid \mathcal{A}_s\right) b(s, \mathbf{X}_s^{t_0, \underline{x}}). \end{aligned} \quad (15.7.26)$$

Furthermore, the random variables on both sides of the above equation are clearly \mathcal{A}_s -measurable, and consequently, by an application of Lemma 15.4.3 and using again the first two equations in (15.7.23) we can infer that

$$\begin{aligned} h(M_T^{\infty, t_0, \underline{x}}) &= E \left(h \left(M_T^{\infty, t_0, \underline{x}} \right) \right) \\ &\quad + \int_{t_0}^T E \left(g \left(M_T^{\infty, t_0, \underline{x}} \right) Z_{\bar{\tau}_\infty}^{s, 1} \mathbf{1}_{\{s < \bar{\tau}_\infty\}} \mid \mathcal{A}_s \right) b(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s. \end{aligned} \quad (15.7.27)$$

Thus, we have found a martingale representation of the payoff structure (15.7.1) of the form (15.4.5), with the non-smooth payoff function h , which satisfies conditions A1* and A2* given in Sect. 15.5.

The methods can also be adapted to the case, where the underlying model is continuous but trading activities or observations are restricted to certain times, say, $\tau_0 < \tau_1 < \dots < \tau_{d_m} = T$ with $d_m = 2^m$.

The payoff structure (15.7.5) with h replacing h_n , $n \in \mathcal{N}$, becomes

$$h \left(\max_{i \in \{1, \dots, d_m\}} X_T^{i, t_0, \underline{x}} \right) = h \left(M_T^{m, t_0, \underline{x}} \right).$$

Applying similar arguments to those used to obtain the representation (15.7.27), with $M_T^{m, t_0, \underline{x}}$ replacing $M_T^{\infty, t_0, \underline{x}}$, and $\bar{\tau}_m$ replacing $\bar{\tau}_\infty$, one can show that

$$\begin{aligned} h \left(M_T^{m, t_0, \underline{x}} \right) &= E \left(h \left(M_T^{m, t_0, \underline{x}} \right) \right) \\ &\quad + \int_{t_0}^T E \left(g \left(M_T^{m, t_0, \underline{x}} \right) Z_{\bar{\tau}_m}^{s, 1} \mathbf{1}_{\{s < \bar{\tau}_m\}} \mid \mathcal{A}_s \right) b(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s. \end{aligned} \quad (15.7.28)$$

In this case we have, therefore, found an explicit martingale representation of the payoff structure of a lookback option with discrete observations or fixings. We do not require here the limiting arguments needed in the derivation of (15.7.13), (15.7.15)–(15.7.19) and (15.7.22).

Although most sections of this chapter appear to be rather technical, they provide powerful and constructive methods which enables a wide class of payoff structures to be represented in the form of martingale stochastic integrals with explicit expressions for the integrands. Using these stochastic integrals one can find corresponding explicit formulas to compute the hedge ratios for the underlying derivative security. Most of the results presented in this chapter should carry over to the case with jumps.

Some incomplete markets can be handled by the above methodology using the same martingale representation theorem. One distinguishes in this case between traded and non-traded noise processes. Those Wiener processes that represent traded noise appear as driving noise in the hedgable martingale part of the martingale representation. The remaining martingale part represents the non-hedgable risk. The hedge ratios can be determined as previously described. The hedge is no longer perfect and the hedge error is driven by the Wiener processes that model non-traded noise.

15.8 Exercises

15.1. Consider a market with zero interest rate and a geometric Brownian motion $S = \{S_t, t \in [0, T]\}$ as risky security, satisfying the SDE

$$dS_t = S_t \sigma dW_t$$

for $t \in [0, T]$ with $S_0 > 0$. Here W is a Wiener process on a filtered probability space $(\Omega, \mathcal{A}_T, \mathcal{A}, \tilde{P})$ under the equivalent risk neutral martingale measure \tilde{P} . Identify for a payoff $H_T = (S_T)^q$ with $q \in (0, \infty)$ the martingale representation involving the corresponding pricing function $u(t, S_t)$ at time $t \in [0, T]$.

15.2. Calculate for the pricing problem in Exercise 15.1 the hedge ratio dependent on time t and asset price S_t .

15.3. Derive for the risky asset price given in Exercise 15.1 the martingale representation and hedge ratio for a European call option price $c_{T,K}(t, S_t)$ at time $t \in [0, T]$ when using the Black-Scholes formula.

Variance Reduction Techniques

The evaluation of the expectation of a given function of a solution of an SDE with jumps provides via the Feynman-Kac formula, see Sect. 2.7, the solution of a partial integro differential equation. In many applications it is of major interest to obtain numerically these expectations, in particular in multi-dimensional settings. Monte Carlo simulation appears to be a method that may be able to provide answers to this question under rather general circumstances. However, raw Monte Carlo estimates of the expectation of a payoff structure, for instance for derivative security prices, can be very expensive in terms of computer resource usage. In this chapter we investigate the problem of constructing variance reduced estimators for the expectation of functionals of solutions of SDEs that can speed up the simulation enormously. We follow again closely [Heath \(1995\)](#). As we will see, variance reduction is more of an art and can be applied in many ways. This chapter shall enable the reader to design her or his own variance reduction method for a given problem at hand.

16.1 Various Variance Reduction Methods

One can increase the efficiency in Monte Carlo simulation for SDEs considerably by using variance reduction techniques. These reduce primarily the variance of the random variable that has to be simulated. There exist many ways of achieving substantial variance reductions. Only some of them can be mentioned here to give examples. A variance reduction technique to be effective, needs to be adapted and engineered to the given specific problem.

Some general variance reduction techniques from classical Monte Carlo theory can often be applied and usually result in moderate improvements. Techniques that exploit to a high degree the stochastic analytic structure of the given functional of an SDE may yield substantial savings in computer time. Useful references on variance reduction techniques along more classical lines include [Hammersley & Handscomb \(1964\)](#), [Ermakov \(1975\)](#), [Boyle \(1977\)](#),

Maltz & Hitzl (1979), Rubinstein (1981), Ermakov & Mikhailov (1982), Ripley (1983), Kalos & Whitlock (1986), Bratley, Fox & Schrage (1987), Chang (1987), Law & Kelton (1991) and Ross (1990).

The application of Monte Carlo simulation to option pricing problems was first demonstrated in the seminal work of Boyle (1977). The basic Monte Carlo method has now become standard in quantitative finance but also in insurance, and the associated literature has rapidly expanded. Its variants can also handle high-dimensional problems, which is difficult to impossible by most other numerical techniques. Some publications, which extend the Monte Carlo method to include alternative variance reduced estimators, path dependent options or quasi random number sequences, include Niederreiter (1988a), Hofmann et al. (1992), Boyle, Broadie & Glasserman (1997), Broadie & Detemple (1997a), Broadie & Glasserman (1997b), Fu (1995), Grant, Vora & Weeks (1997), Joy, Boyle & Tan (1996), Jäckel (2002), Kohatsu-Higa & Peterson (2002), Glasserman (2004) and Kebaier (2005). Longstaff & Schwartz (2001) have proposed a least-squares Monte Carlo method that can be applied to path dependent securities including American style derivatives.

The use of measure transformations, to construct a range of variance reduced estimators, was first proposed by Milstein (1988a) and has subsequently been used and extended by Kloeden & Platen (1999), Hofmann et al. (1992), Newton (1994), Heath (1995), Fournie, Lasry & Touzi (1997) and others. A particular Itô integral representation method has been developed in Heath & Platen (2002d). It provides a continuous time extension of the martingale control variates described in Clewlow & Carverhill (1992) and will be detailed in Sect. 16.5.

In what follows, we first mention several more classical Monte Carlo variance reduction techniques. Later we will point to variance reduction methods that use martingale representations or measure transformations for functionals of SDEs and exploit the stochastic analytic structure of the problem.

Antithetic Variates

We describe at first a variance reduction technique that is based on the use of *antithetic variates*. This technique is standard and has been applied, for example, in Duffie & Glynn (1995), Hull & White (1987, 1988), Clewlow & Carverhill (1992, 1994) and Barraquand (1995). A description of its use in general Monte Carlo simulation is given by Ross (1990) and Law & Kelton (1991). Here we describe a version of antithetic variance reduction, which together with other variance reduction procedures will be used later.

The construction of antithetic variates can be illustrated using the probability space $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ where $\Omega = \mathcal{C}([t_0, T], \mathbb{R}^2)$ is the space of two-dimensional continuous functions on $[t_0, T]$. For $\omega \in \Omega$ we denote by $\omega_1(t)$ and $\omega_2(t)$ the two components of $\omega(t) \in \mathbb{R}^2$ so that $\omega(t) = (\omega_1(t), \omega_2(t))^\top \in \mathbb{R}^2$ for $t \in [t_0, T]$. We assume that P is the two-dimensional Wiener measure under which the coordinate mappings $W_t^1(\omega) = \omega_1(t)$ and $W_t^2(\omega) = \omega_2(t)$ define

a two-dimensional Wiener process on $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$. Let $X^{t_0, \underline{x}} = \{X_t^{t_0, \underline{x}}, t \in [t_0, T]\}$ be a one-dimensional diffusion process that satisfies the SDE

$$dX_t^{t_0, \underline{x}} = a(t, X_t^{t_0, \underline{x}}) dt + b(t, X_t^{t_0, \underline{x}}) dW_t \quad (16.1.1)$$

for $t \in [t_0, T]$ with $X_{t_0}^{t_0, \underline{x}} = \underline{x} \in \mathfrak{R}$. Here W is a standard Wiener process on $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$.

For $\omega \in \Omega$, define $\bar{\omega} \in \Omega$ by $\bar{\omega}(t) = (-\omega_1(t), -\omega_2(t))^\top$, $t \in [t_0, T]$ and the random variables $\bar{h}(X_T^{t_0, \underline{x}})$ by

$$\bar{h}\left(X_T^{t_0, \underline{x}}\right)(\omega) = h\left(X_T^{t_0, \underline{x}}\right)(\bar{\omega}) \quad (16.1.2)$$

for $\omega \in \Omega$.

The variate

$$\hat{h}\left(X_T^{t_0, \underline{x}}\right) = \frac{1}{2} \left(h\left(X_T^{t_0, \underline{x}}\right) + \bar{h}\left(X_T^{t_0, \underline{x}}\right) \right)$$

is an unbiased estimator for $E(h(X_T^{t_0, \underline{x}}))$ since $E(\bar{h}(X_T^{t_0, \underline{x}})) = E(h(X_T^{t_0, \underline{x}}))$. In addition

$$\begin{aligned} \text{Var}\left(\hat{h}\left(X_T^{t_0, \underline{x}}\right)\right) &= \frac{1}{4} \left(\text{Var}\left(h\left(X_T^{t_0, \underline{x}}\right)\right) + \text{Var}\left(\bar{h}\left(X_T^{t_0, \underline{x}}\right)\right) \right. \\ &\quad \left. + 2 \text{Cov}\left(h\left(X_T^{t_0, \underline{x}}\right), \bar{h}\left(X_T^{t_0, \underline{x}}\right)\right) \right) \end{aligned}$$

and, because the random variables $h(X_T^{t_0, \underline{x}})$ and $\bar{h}(X_T^{t_0, \underline{x}})$ will typically be negatively correlated, $\hat{h}(X_T^{t_0, \underline{x}})$ will then be a variance reduced estimator for $E(h(X_T^{t_0, \underline{x}}))$. We remark that other combinations are possible for $\bar{\omega}$, for example, reflection for only one component or partial reflections over time. These alternatives can then be assembled in various ways to produce new variance reduced unbiased estimators for $E(h(X_T^{t_0, \underline{x}}))$. The use of antithetic variates is a basic variance reduction method which can always be combined with other variance reduction techniques.

Variance Reduction by Conditioning

The classical *conditional variance formula*, see for example Ross (1990), takes the form

$$\text{Var}(Y) = E(\text{Var}(Y | Z)) + \text{Var}(E(Y | Z)) \quad (16.1.3)$$

for any two random variables defined on some sample space Ω' with probability measure P' .

To extend this result to stochastic processes we first use the concept of conditional variance with respect to sub-sigma-algebras of \mathcal{A}_T , see Sect. 1.2. Let X be an integrable random variable and $\mathcal{G} \subseteq \mathcal{A}_T$ a sub-sigma-algebra. The

conditional variance of X with respect to \mathcal{G} , denoted $\text{Var}(X \mid \mathcal{G})$, is defined by

$$\begin{aligned}\text{Var}(X \mid \mathcal{G}) &= E\left(\left(X - E(X \mid \mathcal{G})\right)^2 \mid \mathcal{G}\right) \\ &= E(X^2 \mid \mathcal{G}) - (E(X \mid \mathcal{G}))^2.\end{aligned}\quad (16.1.4)$$

Let $\mathcal{G}_1, \mathcal{G}_2$ be sub-sigma-algebras of \mathcal{A}_T with $\mathcal{G}_1 \subseteq \mathcal{G}_2$. Using the law of iterated conditional expectations, see (1.1.10), and the definition of conditional variance we have for any integrable random variable X the relations

$$\begin{aligned}\text{Var}(X \mid \mathcal{G}_1) &= E(X^2 \mid \mathcal{G}_1) - (E(X \mid \mathcal{G}_1))^2 \\ &= E\left(E(X^2 \mid \mathcal{G}_2) \mid \mathcal{G}_1\right) - \left(E\left(E(X \mid \mathcal{G}_2) \mid \mathcal{G}_1\right)\right)^2 \\ &= E\left(E(X^2 \mid \mathcal{G}_2) \mid \mathcal{G}_1\right) - E\left((E(X \mid \mathcal{G}_2))^2 \mid \mathcal{G}_1\right) \\ &\quad + E\left((E(X \mid \mathcal{G}_2))^2 \mid \mathcal{G}_1\right) - \left(E\left(E(X \mid \mathcal{G}_2) \mid \mathcal{G}_1\right)\right)^2 \\ &= E\left(\text{Var}(X \mid \mathcal{G}_2) \mid \mathcal{G}_1\right) + \text{Var}\left(E(X \mid \mathcal{G}_2) \mid \mathcal{G}_1\right).\end{aligned}\quad (16.1.5)$$

In particular, if we let $\mathcal{G}_1 = \{\phi, \Omega\}$ we obtain

$$\text{Var}(X) = E(\text{Var}(X \mid \mathcal{G}_2)) + \text{Var}(E(X \mid \mathcal{G}_2)) \quad (16.1.6)$$

for \mathcal{G}_2 a sub-sigma-algebra of \mathcal{A}_T . This is an analogue of the classical conditional variance formula.

For $t \in [t_0, T]$ consider the random variable $E(h(X_T^{t_0, x}) \mid \mathcal{A}_t)$, where $X^{t_0, x}$ is a stochastic process, for instance, as in (16.1.1). Since

$$E\left(E\left(h\left(X_T^{t_0, x}\right) \mid \mathcal{A}_t\right)\right) = E\left(h\left(X_T^{t_0, x}\right)\right)$$

this variate is an unbiased estimator for $E(h(X_T^{t_0, x}))$. By applying the conditional variance formula (16.1.6) we obtain the inequality

$$\text{Var}\left(E\left(h\left(X_T^{t_0, x}\right) \mid \mathcal{A}_t\right)\right) \leq \text{Var}\left(h\left(X_T^{t_0, x}\right)\right) \quad (16.1.7)$$

for $t \in [t_0, T]$. In general, the inequality in (16.1.7) will be strict. In this case $E(h(X_T^{t_0, x}) \mid \mathcal{A}_t)$ will provide a variance reduced unbiased estimator for $E(h(X_T^{t_0, x}))$. This classical type of error minimization is known in its standard formulation as *variance reduction by conditioning*, see Law & Kelton (1991).

Stratified Sampling

We will now turn our attention to the variance reduction technique of *stratified sampling*. This is an error minimization technique of long standing which has

been widely used in classical Monte Carlo simulation, see for example Ross (1990). Surprisingly, the method seems to have been underutilized for financial modeling problems. However, Curran (1994) has described a simple form of stratified sampling for Asian options and geometric Brownian motion. We will describe another version of stratified sampling which can be applied to a wide class of SDEs and valuation functionals.

Let $A_i \subseteq \mathcal{A}_T$, $i \in \{1, 2, \dots, N\}$, be a set of disjoint events satisfying

$$\bigcup_{i=1}^N A_i = \Omega, \quad A_i \cap A_j = \emptyset$$

for $i, j \in \{1, 2, \dots, N\}$. We assume $P(A_i) = \frac{1}{N}$ for all i , $i \in \{1, 2, \dots, N\}$. Let $\mathcal{A} = \sigma\{A_i, i \in \{1, 2, \dots, N\}\}$ be the sigma-algebra generated by finite unions of these sets.

Let $Z : \Omega \rightarrow \mathbb{R}$ be a random variable. We denote by $Z_{A_i} : A_i \rightarrow \mathbb{R}$ the restriction of Z to A_i given by $Z_{A_i}(\omega) = Z(\omega)$ for $\omega \in A_i$. Z_{A_i} is a random variable defined on the probability space $(A_i, \mathcal{A}_i, P_i)$, where $\mathcal{A}_i = \{A_i \cap F : F \in \mathcal{A}_T\}$ and $P_i : \mathcal{A}_i \rightarrow [0, 1]$ is given by

$$P_i(A_i \cap F) = \frac{P(A_i \cap F)}{P(A_i)}$$

for $F \in \mathcal{A}_T$.

Let

$$\bar{Z} = \frac{1}{N} \sum_{i=1}^N Z_{A_i},$$

where Z_{A_i} , $i \in \{1, 2, \dots, N\}$, are assumed to be independent. Since

$$E(\bar{Z}) = \frac{1}{N} \sum_{i=1}^N E(Z_{A_i}) = \sum_{i=1}^N \int_{A_i} Z dP = \int_{\Omega} Z dP = E(Z), \quad (16.1.8)$$

\bar{Z} will be an unbiased estimator for $E(Z)$. Furthermore, using the independence property of Z_{A_i} , $i \in \{1, 2, \dots, N\}$, we have from (16.1.6)

$$\begin{aligned} \text{Var}(\bar{Z}) &= \sum_{i=1}^N \frac{\text{Var}(Z_{A_i})}{N^2} \\ &= \frac{1}{N} E(\text{Var}(Z | \mathcal{A})) \\ &\leq \frac{1}{N} \text{Var}(Z). \end{aligned} \quad (16.1.9)$$

This inequality will be strict if $\text{Var}(E(Z | \mathcal{A})) > 0$. Consequently, if we set

$$Z = h(X_T^{t_0, x}),$$

then we obtain from (16.1.8) an unbiased estimator for $E(h(X_T^{t_0, \underline{x}}))$, which from (16.1.9) will usually be a variance reduced estimator.

As an example let Y^Δ be a simplified weak Euler approximation, see (11.2.2), of a given one-dimensional diffusion process, say, $X^{t_0, \underline{x}}$ as given in (16.1.1), where the $\Delta \hat{W}_k$, $k \in \{0, 1, \dots, N-1\}$, are two-point distributed random variables given by (11.2.4). Since we are using two-point variates with N time steps, we can replace for theoretical purposes the underlying sample space Ω with

$$\Omega_N = \{-1, 1\}^{\{0, 1, \dots, N-1\}}. \quad (16.1.10)$$

For each $\omega \in \Omega_N$ we denote by ω_k , $k \in \{0, 1, \dots, N-1\}$, the value of ω at the k th step. For $\omega \in \Omega_N$ the corresponding random walk $\hat{W} = \{\hat{W}_k, k \in \{0, 1, \dots, N-1\}\}$ with the increments $\hat{W}_{k+1}(\omega) - \hat{W}_k(\omega) = \omega_k \sqrt{\Delta}$ has probability $P_N(\omega) = \frac{1}{2^N}$. Although the number of states $\omega \in \Omega_N$ is finite, nevertheless, with current computing technology and values of N say greater than 30, usually only a tiny fraction of these paths can be sampled.

Let $a \in \Omega_{N'}$ with $N' \in \{1, 2, \dots, N\}$, $N' < N$ and define $A_a = \{\omega \in \Omega_N : \omega_k = a_k \text{ for } 0 \leq k < N'\}$. The sets A_a , $a \in \Omega_{N'}$ form a collection of $2^{N'}$ sets with probabilities $P(A_a) = \frac{1}{2^{N'}}$. It is apparent that $\cup_{a \in \Omega_{N'}} = \Omega_N$ and that for $a_1, a_2 \in \Omega_{N'}$ with $a_1 \neq a_2$, then $A_{a_1} \cap A_{a_2} = \emptyset$. Consequently, the partitioning rules required for the relations (16.1.8) and (16.1.9) are valid with $Z = h(Y_N^\Delta)$.

With this method our discrete sample space Ω_N forms a binary lattice or binomial tree, as will be described in Chap. 17, which is divided into sub-lattices A_a , $a \in \Omega_{N'}$ starting at time $t_{N'} < t_N = T$. A stratified Monte Carlo estimation of $E(h(Y_N^\Delta))$ using the sets A_a , $a \in \Omega_{N'}$ consists of exhausting all paths $\omega \in \Omega_N$ up to time $t_{N'}$ and then sampling randomly and independently within the sub-lattices A_a , $a \in \Omega_{N'}$. This method not only reduces the variance it also reduces the computational load because the raw Monte Carlo estimation procedure would involve, usually, many duplicate traversals of the early nodes of the lattice.

Measure Transformation Method

We consider now a method that is based on a purely stochastic analytic argument, which can be found in Milstein (1988a), Newton (1994) and Kloeden & Platen (1999). Given a d -dimensional diffusion process $\mathbf{X}^{s, \underline{x}} = \{\mathbf{X}_t^{s, \underline{x}}, t \in [s, T]\}$, starting at $\underline{x} \in \mathbb{R}^d$ at time $s \in [0, T]$, satisfying the SDE

$$d\mathbf{X}_t^{s, \underline{x}} = \mathbf{a}(t, \mathbf{X}_t^{s, \underline{x}}) dt + \sum_{j=1}^m \mathbf{b}^j(t, \mathbf{X}_t^{s, \underline{x}}) dW_t^j. \quad (16.1.11)$$

for $t \in [s, T]$ with $\mathbf{X}_s^{s, \underline{x}} = \underline{x}$. Here \mathbf{W} is a standard m -dimensional Wiener process on the filtered probability space $(\Omega, \mathcal{A}_T, \mathcal{A}, P)$, where \mathcal{A}_0 is the trivial sigma-algebra. Our aim is to approximate the functional

$$u(s, \mathbf{x}) = E(g(\mathbf{X}_T^{s,x}) | \mathcal{A}_s) \quad (16.1.12)$$

for a given real-valued function g and time $s = 0$.

If we assume that the function g and the drift and diffusion coefficients \mathbf{a} and \mathbf{b} are sufficiently smooth and satisfy appropriate conditions, then the function $u(\cdot, \cdot)$ defined by (16.1.12) satisfies by the Feynman-Kac formula, (2.7.3)–(2.7.5), the Kolmogorov backward equation. That is,

$$L^0 u(s, \mathbf{x}) = 0 \quad (16.1.13)$$

for $(s, \mathbf{x}) \in (0, T) \times \Re^d$ with

$$u(T, \mathbf{y}) = g(\mathbf{y}) \quad (16.1.14)$$

for all $\mathbf{y} \in \Re^d$, where L^0 is the operator

$$L^0 = \frac{\partial}{\partial s} + \sum_{k=1}^d a^k \frac{\partial}{\partial x^k} + \frac{1}{2} \sum_{k,\ell=1}^d \sum_{j=1}^m b^{k,j} b^{\ell,j} \frac{\partial^2}{\partial x^k \partial x^\ell}.$$

In Milstein (1988a) the use of the Girsanov transformation, see Sect. 2.7, was proposed to transform the underlying probability measure P such that the process \tilde{W} defined by

$$\tilde{W}_t^j = W_t^j - \int_0^t d^j(z, \tilde{\mathbf{X}}_z^{0,x}) dz \quad (16.1.15)$$

becomes a Wiener process with respect to a transformed probability measure \tilde{P} with Radon-Nikodym derivative

$$\frac{d\tilde{P}}{dP} = \frac{\Theta_t}{\Theta_0}. \quad (16.1.16)$$

Here the process $\tilde{\mathbf{X}}^{0,x}$ satisfies the SDE

$$\begin{aligned} d\tilde{\mathbf{X}}_t^{0,x} &= \mathbf{a}\left(t, \tilde{\mathbf{X}}_t^{0,x}\right) dt + \sum_{j=1}^m \mathbf{b}^j\left(t, \tilde{\mathbf{X}}_t^{0,x}\right) d\tilde{W}_t^j \\ &= \left(\mathbf{a}\left(t, \tilde{\mathbf{X}}_t^{0,x}\right) - \sum_{j=1}^m \mathbf{b}^j\left(t, \tilde{\mathbf{X}}_t^{0,x}\right) d^j\left(t, \tilde{\mathbf{X}}_t^{0,x}\right)\right) dt + \sum_{j=1}^m \mathbf{b}^j\left(t, \tilde{\mathbf{X}}_t^{0,x}\right) dW_t^j \end{aligned} \quad (16.1.17)$$

and the Radon-Nikodym derivative process Θ satisfies the SDE

$$\Theta_t = \Theta_0 + \sum_{j=1}^m \int_0^t \Theta_z d^j\left(z, \tilde{\mathbf{X}}_z^{0,x}\right) dW_z^j \quad (16.1.18)$$

with $\Theta_0 > 0$. The quantities $d^j(\cdot, \cdot)$ denote here real-valued functions for $j \in \{1, 2, \dots, m\}$. Note that $\tilde{\mathbf{X}}^{0,x}$ is d -dimensional, whereas Θ is only one-dimensional.

Obviously, the process $\tilde{\mathbf{X}}^{0,x}$ in (16.1.17) is a diffusion process with respect to \tilde{P} with the same drift and diffusion coefficients as the diffusion process $\mathbf{X}^{s,x}$ in (16.1.11). It then follows from (16.1.16) that

$$\begin{aligned} E\left(g\left(\mathbf{X}_T^{0,x}\right)\right) &= \int_{\Omega} g\left(\mathbf{X}_T^{0,x}\right) dP \\ &= \int_{\Omega} g\left(\tilde{\mathbf{X}}_T^{0,x}\right) d\tilde{P} \\ &= \int_{\Omega} g\left(\tilde{\mathbf{X}}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0} dP \\ &= E\left(g\left(\tilde{\mathbf{X}}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0}\right). \end{aligned} \quad (16.1.19)$$

Hence, we can estimate the expectation of the random variable

$$g\left(\tilde{\mathbf{X}}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0} \quad (16.1.20)$$

to evaluate the functional (16.1.12). So far, our analysis does not depend on the particular choice of the functions d^j , $j \in \{1, 2, \dots, m\}$. Therefore, we can use the process Θ as a control process to reduce the variance of the random variable (16.1.20).

Now, let us study the following idealized situation, which is interesting from a theoretical viewpoint. Assume that $u(t, \mathbf{x}) > 0$ everywhere and that the corresponding solutions of (16.1.17) and (16.1.18) exist. Furthermore, let us choose the parameter functions d^j in the form

$$d^j(t, \mathbf{x}) = -\frac{1}{u(t, \mathbf{x})} \sum_{k=1}^d b^{k,j}(t, \mathbf{x}) \frac{\partial u(t, \mathbf{x})}{\partial x^k} \quad (16.1.21)$$

for all $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d$ and $j \in \{1, 2, \dots, m\}$. Then, with the Itô formula and by using (16.1.17) and (16.1.18), (16.1.21) and (16.1.13), it can be shown that

$$u\left(t, \tilde{\mathbf{X}}_t^{0,x}\right) \Theta_t = u(0, \mathbf{x}) \Theta_0 \quad (16.1.22)$$

When combining now (16.1.14) and (16.1.22), we can conclude that

$$u(0, \mathbf{x}) = g\left(\tilde{\mathbf{X}}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0}. \quad (16.1.23)$$

Consequently, with the choice of the parameter functions in the form (16.1.21) the variable

$$g\left(\tilde{\mathbf{X}}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0} \quad (16.1.24)$$

is not random and the variance is even fully reduced to zero.

We would expect, in general, to obtain a small variance for the estimator if we would simulate the expression (16.1.24) approximately by using weak discrete-time approximations for the SDE (16.1.17) and (16.1.18).

Unfortunately, for the construction of the parameter functions in (16.1.21) one needs to know the solution $u(\cdot, \cdot)$ of the Kolmogorov backward equation, which is exactly what we are trying to determine by means of Monte Carlo simulation. The above discussion shows that it may be possible to obtain a substantial reduction in the variance of the estimator by an application of a measure transformation if one can exploit sufficient information about the pricing function.

In practice, when implementing a measure transformation method, one needs to find or guess a function \bar{u} which is sufficiently similar to the solution u of the Kolmogorov backward equation (16.1.13) and (16.1.14). One can then use \bar{u} instead of u to define the parameter functions in (16.1.21) by choosing the form

$$d^j(t, \mathbf{x}) = -\frac{1}{\bar{u}(t, \mathbf{x})} \sum_{k=1}^d b^{k,j}(t, \mathbf{x}) \frac{\partial \bar{u}(t, \mathbf{x})}{\partial x^k} \quad (16.1.25)$$

for all $(t, \mathbf{x}) \in [0, T] \times \Re^d$ and $j \in \{1, 2, \dots, m\}$. Then the quantity

$$g\left(\tilde{\mathbf{X}}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0}$$

is still a random variable, but in general, with small variance if \bar{u} is chosen sufficiently close to u .

The weak discrete-time approximations presented in Chap. 11 can be readily applied. They provide weak approximations of the SDEs governing $\tilde{\mathbf{X}}^{0,x}$ and Θ to estimate the functional

$$E\left(g\left(\tilde{\mathbf{X}}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0}\right) = E\left(g\left(\mathbf{X}_T^{0,x}\right)\right).$$

The measure transformation variance reduction technique is a rather flexible method with substantial potential, as will be indicated in the following section.

16.2 Measure Transformation Techniques

In this section we apply a measure transformation and the Itô formula to build variance reduced estimators that can be used for a wide class of derivative security valuation problems. The construction of these estimators follows Heath (1995) and requires only minor integrability conditions for the underlying payoff structure, and in this sense extends the results of Milstein (1988a), Hofmann et al. (1992) and Fournie, Lasry & Touzi (1997).

Unbiased Estimators

As in Chap. 15 let $\mathbf{W} = \{\mathbf{W}_t = (W^1, \dots, W^m)^\top, t \geq 0\}$ be an m -dimensional Wiener process defined on the filtered probability space $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$, where $\mathcal{A} = (\mathcal{A}_t)_{t \in [t_0, T]}$ denotes the P -augmentation of the natural filtration of \mathbf{W} . Here \mathcal{A}_{t_0} is assumed to be the trivial sigma-algebra. Let $M = \{M_t, t \in [t_0, T]\}$ be a square integrable $(\underline{\mathcal{A}}, P)$ -martingale with respect to the filtration $\underline{\mathcal{A}}$ and measure P with the Kunita-Watanabe representation, that is, the martingale representation

$$M_t = M_{t_0} + \sum_{j=1}^m \int_{t_0}^t \xi_s^j dW_s^j, \quad (16.2.1)$$

where $\boldsymbol{\xi} = (\xi^1, \dots, \xi^m)^\top$ is a vector of $\underline{\mathcal{A}}$ -predictable processes with

$$E \left(\int_{t_0}^T \boldsymbol{\xi}_s^\top \boldsymbol{\xi}_s ds \right) < \infty \quad (16.2.2)$$

for $t \in [t_0, T]$, see Chap. 15.

Let $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, \tilde{P})$ be another probability space with the same sample space Ω and filtration $\underline{\mathcal{A}}$ but a different probability measure \tilde{P} .

We will say that an adapted process $G = \{G_t, t \in [t_0, T]\}$ is an *unbiased estimator* under \tilde{P} for $E(M_T) = M_{t_0}$ at time t if

$$\tilde{E}(G_t) = E(M_T),$$

where \tilde{E} denotes expectation with respect to the measure \tilde{P} . Ideally we seek unbiased estimators of $E(M_T)$ since in these cases we only need to deal with the process G in order to approximate $E(M_T)$. If in addition, the inequality

$$\tilde{\text{Var}}(G_t) < \text{Var}(M_T)$$

holds, where $\tilde{\text{Var}}(G_t)$ denotes the variance of G_t under \tilde{P} , we will say that G is a *variance reduced unbiased estimator* under \tilde{P} for $E(M_T)$ at time t , $t \in [t_0, T]$. For many practical valuation problems we require variance reduced estimators of $E(M_T)$ because the statistical error associated with simulation estimates of $\tilde{E}(G_t)$, $t \in [t_0, T]$, depends on the variance under \tilde{P} of G_t . This error will be manageable if the variance is small. Clearly, these definitions extend to the case where we replace t with some stopping time τ and when we have jumps.

Measure Transformation

Let $\mathbf{d} = (d^1, \dots, d^m)^\top$ be a vector of adapted measurable processes satisfying the *Novikov condition*

$$E \left(\exp \left\{ \frac{1}{2} \int_{t_0}^T \mathbf{d}_s^\top \mathbf{d}_s ds \right\} \right) < \infty, \quad (16.2.3)$$

see Sect. 2.7 and Novikov (1972a). Also for $t \in [t_0, T]$, let P_t denote the restriction of the measure P to the sigma-algebra \mathcal{A}_t . Using the Girsanov transformation, see Sect. 2.7, we know that there is a corresponding measure \tilde{P}_t such that the process $\tilde{\mathbf{W}} = \{\tilde{\mathbf{W}}_t = (\tilde{W}^1, \dots, \tilde{W}^m)^\top, t \geq 0\}$ given by

$$\tilde{W}_t^j = W_t^j - \int_{t_0}^t d_s^j ds \quad (16.2.4)$$

for $t \in [t_0, T]$, $j \in \{1, 2, \dots, m\}$, is an m -dimensional Wiener process under \tilde{P}_t . Here \tilde{P}_t is defined using the Radon-Nikodym derivative

$$\frac{d\tilde{P}_t}{dP_t} = \theta_t = \exp \left\{ -\frac{1}{2} \int_{t_0}^t \mathbf{d}_s^\top \mathbf{d}_s ds + \sum_{j=1}^m \int_{t_0}^t d_s^j dW_s^j \right\}$$

and $\theta = \{\theta_t, t \in [t_0, T]\}$ is the process which is the solution of the integral equation

$$\theta_t = 1 + \sum_{j=1}^m \int_{t_0}^t d_s^j \theta_s dW_s^j \quad (16.2.5)$$

for $t \in [t_0, T]$ and $\theta_{t_0} = 1$.

From (16.2.3) it can be shown that θ is a square integrable martingale under the measure P_t with $E(\theta_t) = 1$, $t \in [t_0, T]$. Since the process θ is a martingale and $P_T = P$, then for any event $A \in \mathcal{A}_t$ we have

$$\tilde{P}_t(A) = \int_A \theta_t dP_t = \int_A \theta_t dP_T = \int_A E(\theta_T | \mathcal{A}_t) dP_T = \int_A \theta_T dP_T = \tilde{P}_T(A).$$

This means that \tilde{P}_t is the restriction of \tilde{P}_T to the sigma-algebra \mathcal{A}_t .

Constructing an Unbiased Estimator

Let $\tilde{M} = \{\tilde{M}_t, t \in [t_0, T]\}$ be a square integrable martingale with respect to the filtration $\underline{\mathcal{A}}$ and the measure \tilde{P} , with representation

$$\tilde{M}_t = \tilde{M}_{t_0} + \sum_{j=1}^m \int_{t_0}^t \tilde{\xi}_s^j d\tilde{W}_s^j, \quad (16.2.6)$$

where $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}^1, \dots, \tilde{\xi}^m)^\top$ is an $\underline{\mathcal{A}}$ -predictable process with

$$E \left(\int_{t_0}^T \tilde{\boldsymbol{\xi}}_s^\top \tilde{\boldsymbol{\xi}}_s ds \right) < \infty. \quad (16.2.7)$$

For the moment we will regard the process $\tilde{\boldsymbol{\xi}}$ as being unspecified, except for the requirement that $\tilde{\boldsymbol{\xi}}$ be $\underline{\mathcal{A}}$ -predictable and that the mean square integrability condition (16.2.7) holds. We assume that

$$\tilde{E}(\tilde{M}_T) = E(M_T). \quad (16.2.8)$$

This condition will be verified later for a wide class of contingent claims.

Equation (16.2.8) together with (16.2.1) and (16.2.6) imply the relations

$$E(M_t) = E(M_T) = \tilde{E}(\tilde{M}_T) = \tilde{E}(\tilde{M}_t) \quad (16.2.9)$$

for $t \in [t_0, T]$.

Since the restrictions of the measures $P = P_T$ and $\tilde{P} = \tilde{P}_T$ to \mathcal{A}_t are P_t and \tilde{P}_t , respectively, we have, using the Radon-Nikodym derivative $\theta_t = \frac{d\tilde{P}_t}{dP_t}$, the relations

$$\begin{aligned} \tilde{E}(\tilde{M}_t) &= \int_{\Omega} \tilde{M}_t d\tilde{P}_T \\ &= \int_{\Omega} \tilde{M}_t d\tilde{P}_t \\ &= \int_{\Omega} \tilde{M}_t \theta_t dP_t \\ &= \int_{\Omega} \tilde{M}_t \theta_t dP_T \\ &= E(\tilde{M}_t \theta_t) \end{aligned} \quad (16.2.10)$$

for $t \in [t_0, T]$. Combining (16.2.9) and (16.2.10) we see that the process $\tilde{M}\theta = \{\tilde{M}_t \theta_t, t \in [t_0, T]\}$ is an unbiased estimator under P for $E(M_T)$ at time $t \in [t_0, T]$.

Variance of the Estimator

We will now compute the variance of the estimator $\tilde{M}\theta$ under the measure P . To do this we first express the martingale \tilde{M} as a semimartingale using the Wiener process \mathbf{W} rather than $\tilde{\mathbf{W}}$.

Applying (16.2.4) and (16.2.6) we have

$$\tilde{M}_t = \tilde{M}_{t_0} - \sum_{j=1}^m \int_{t_0}^t d_s^j \tilde{\xi}_s^j ds + \sum_{j=1}^m \int_{t_0}^t \tilde{\xi}_s^j dW_s^j. \quad (16.2.11)$$

Expanding the estimator $\tilde{M}\theta$ by the Itô formula together with (16.2.5) and (16.2.11), the integral equation for $\tilde{M}_t \theta_t$ becomes

$$\tilde{M}_t \theta_t = \tilde{M}_{t_0} + \sum_{j=1}^m \int_{t_0}^t \theta_s (\tilde{\xi}_s^j + d_s^j \tilde{M}_s) dW_s^j \quad (16.2.12)$$

for $t \in [t_0, T]$.

This shows that $\tilde{M}\theta$ is an (\mathcal{A}, P) -martingale. From this formula we can also verify the previously noted fact that the process $\tilde{M}\theta$ is an unbiased estimator under P for $E(M_t)$ at time $t \in [t_0, T]$. The variance of the product $\tilde{M}_t \theta_t$ under P , denoted by $\text{Var}(\tilde{M}_t \theta_t)$, can now be computed using the equations

$$\begin{aligned} \text{Var}(\tilde{M}_t \theta_t) &= E \left(\left(\sum_{j=1}^m \int_{t_0}^t \theta_s (\tilde{\xi}_s^j + d_s^j \tilde{M}_s) dW_s^j \right)^2 \right) \\ &= \int_{t_0}^t E \left(\theta_s^2 \sum_{j=1}^m \left(\tilde{\xi}_s^j + d_s^j \tilde{M}_s \right)^2 \right) ds \end{aligned} \quad (16.2.13)$$

for $t \in [t_0, T]$.

Consequently, if the inequality $\tilde{M}_t > 0$ holds P -a.s. for $t \in [t_0, T]$ and we choose

$$d_s^j = -\frac{\tilde{\xi}_s^j}{\tilde{M}_s}$$

for all $j \in \{1, 2, \dots, m\}$ and $s \in [t_0, T]$, then the variance of the random variable $\tilde{M}_t \theta_t$ under P is reduced to zero for any $t \in [t_0, T]$. Of course, in this form we have assumed that the integrand process $\tilde{\xi}$ and the martingale \tilde{M} can somehow be determined. In practice, this condition is difficult to satisfy. They usually cannot be easily computed even in the case where an explicit form exists for the process $\tilde{\xi}$.

Alternative Variance Reduced Estimator

An alternative is to find approximations $\hat{\xi} = (\hat{\xi}^1, \dots, \hat{\xi}^m)^\top$ and \hat{M} for the processes $\tilde{\xi}$ and \tilde{M} , respectively, with $\hat{M}_t > 0$ P_T a.s., for all $t \in [t_0, T]$ and set

$$d_t^j = -\frac{\hat{\xi}_t^j}{\hat{M}_t}, \quad (16.2.14)$$

for $t \in [t_0, T]$ and $j \in \{1, 2, \dots, m\}$, so that from (16.2.13)

$$\begin{aligned} \text{Var}(\tilde{M}_t \theta_t) &= E \left(\left(\sum_{j=1}^m \int_{t_0}^t \theta_s \left(\tilde{\xi}_s^j - \hat{\xi}_s^j \frac{\tilde{M}_s}{\hat{M}_s} \right) dW_s^j \right)^2 \right) \\ &= \int_{t_0}^t E \left(\theta_s^2 \sum_{j=1}^m \left(\tilde{\xi}_s^j - \hat{\xi}_s^j \frac{\tilde{M}_s}{\hat{M}_s} \right)^2 \right) ds. \end{aligned} \quad (16.2.15)$$

This formula for the variance of $\tilde{M}_t \theta_t$ under P is an important practical result as it tells us exactly what factors need to be controlled to reduce the

variance of the estimator $\tilde{M}\theta$. In particular, it shows that the process θ^2 also contributes to the variance of $\tilde{M}_t\theta_t$. If θ^2 is allowed to explode, then $\text{Var}(\tilde{M}_t\theta_t)$ may become unacceptably large. Stability problems can also easily arise if $\tilde{M}_t\theta_t$ is being approximated using stochastic numerical methods and θ is not constrained.

To see how these problems can be controlled, suppose the process \hat{M} is bounded from below by the value $\alpha > 0$, that is,

$$\hat{M}_t(\omega) \geq \alpha \quad (16.2.16)$$

for all $(t, \omega) \in [t_0, T] \times \Omega$ and the approximation $\hat{\xi}$ satisfies a mean square integrability bound of the form (16.2.7). Then for a suitably large choice of α , the process $\mathbf{d} = (d^1, \dots, d^m)^\top$, given by (16.2.14) will be small, in a mean square sense, since

$$|d_t^j| = \left| \frac{\hat{\xi}_t^j}{\hat{M}_t} \right| \leq \frac{1}{\alpha} |\hat{\xi}_t^j|$$

for $t \in [t_0, T]$ and $j \in \{1, 2, \dots, m\}$. This will ensure that the exponential $\theta_t = \frac{d\hat{P}_t}{dP_t}$, $t \in [t_0, T]$ does not become unbounded but remains close to 1.

Unfortunately, for most derivative security valuation problems a lower bound condition of this type will usually not hold. However, if we apply the above analysis to the processes \tilde{M}^α , \hat{M}^α and $\hat{\xi}^{j,\alpha}$ given by

$$\begin{aligned} \tilde{M}_t^\alpha &= \alpha + \tilde{M}_t, \\ \hat{M}_t^\alpha &= \alpha + \hat{M}_t, \\ \hat{\xi}_t^{j,\alpha} &= \hat{\xi}_t^j, \end{aligned} \quad (16.2.17)$$

for $t \in [t_0, T]$, $j \in \{1, 2, \dots, m\}$, then using the assignment

$$d_t^j = -\frac{\hat{\xi}_t^{j,\alpha}}{\hat{M}_t^\alpha}$$

for $t \in [t_0, T]$, $j \in \{1, 2, \dots, m\}$, and the above arguments we can generally ensure that the corresponding process θ will be close to 1. Consequently, $\text{Var}(\tilde{M}_t\theta_t)$ will be small for the approximations \hat{M}^α and $\hat{\xi}^{j,\alpha}$, $j \in \{1, 2, \dots, m\}$, and α sufficiently large.

We remark that, in the above construction, we could define θ_t and \tilde{M}_t directly by equations (16.2.5) and (16.2.11). This would enable us to build estimators of the form (16.2.12) without measure transformation. However, we also need to verify condition (16.2.8) and have a means for building the process \tilde{M} , so that the estimator $\tilde{M}_t\theta_t$ based on (16.2.12) can be computed. In practice, for most types of financial evaluation problems this can be achieved via a measure transformation. Also the machinery developed in this section is needed in the next section when we construct other variance reduced estimators.

Note that the above variance reduction procedure can be applied iteratively because the estimator $\tilde{M}\theta$ is a martingale under P with an Itô integral representation of the form (16.2.12). We assume, of course, that the integrands in (16.2.12) satisfy mean square integrability conditions of the form (16.2.7).

Thus, if we let $M^{(1)} = \tilde{M}\theta$, then we can apply the above procedure to $M^{(1)}$ rather than the original M . The new estimator will be of the form $M^{(2)} = \tilde{M}^{(1)}\theta^{(1)}$ for a suitably chosen process $\theta^{(1)}$. In a similar fashion, additional estimators $M^{(3)}, \dots, M^{(k)}$ for some positive integer k can also be built.

For a large class of evaluation problems in quantitative finance and insurance we can find good analytic approximations for the valuation process M and integrands ξ^j with $j \in \{1, 2, \dots, m\}$, see Sects. 3.5 and 3.6. Using these one can construct the estimator $M^{(1)}$. To construct $M^{(2)}$ we require approximations of the integrands $\theta(\tilde{\xi}^j + d^j M^{(1)})$, $j \in \{1, 2, \dots, m\}$ appearing in (16.2.12). Analytic approximations are usually more difficult to find for this type of integrand.

Variance of Alternative Estimator

Let us now compute the variance of the estimator $M\theta^{-1}$ under the measure \tilde{P} . From the relation $\theta_t^{-1} = \frac{dP_t}{d\tilde{P}_t}$ we have

$$\begin{aligned} E(M_t) &= \int_{\Omega} M_t dP_T \\ &= \int_{\Omega} M_t dP_t \\ &= \int_{\Omega} M_t \theta_t^{-1} d\tilde{P}_t \\ &= \int_{\Omega} M_t \theta_t^{-1} d\tilde{P}_T \\ &= \tilde{E}(M_t \theta_t^{-1}) \end{aligned} \tag{16.2.18}$$

for $t \in [t_0, T]$. This expression can be compared to (16.2.10) for the estimator $\tilde{M}\theta$. Since $E(M_t) = E(M_T)$ the process $M\theta^{-1} = \{M_t \theta_t^{-1}, t \in [t_0, T]\}$ is, thus, an unbiased estimator under the measure $\tilde{P} = \tilde{P}_T$ for $E(M_T)$ at time t , $t \in [t_0, T]$. This result does not use the (\mathcal{A}, \tilde{P}) -martingale \tilde{M} , and in particular, does not require condition (16.2.8) to be verified.

We now express the random variables M_t and θ_t as semimartingales using the Wiener process \tilde{W} under \tilde{P} rather than W . Using (16.2.1), (16.2.4) and (16.2.5) we have

$$M_t = M_{t_0} + \sum_{j=1}^m \int_{t_0}^t \xi_s^j d_s^j ds + \sum_{j=1}^m \int_{t_0}^t \xi_s^j d\tilde{W}_s^j, \quad (16.2.19)$$

$$\theta_t = 1 + \sum_{j=1}^m \int_{t_0}^t (d_s^j)^2 \theta_s ds + \sum_{j=1}^m \int_{t_0}^t d_s^j \theta_s d\tilde{W}_s^j. \quad (16.2.20)$$

Applying the above expressions and Itô's formula, the integral equation for $M_t \theta_t^{-1}$ becomes

$$M_t \theta_t^{-1} = M_{t_0} + \sum_{j=1}^m \int_{t_0}^t \theta_s^{-1} (\xi_s^j - d_s^j M_s) d\tilde{W}_s^j. \quad (16.2.21)$$

Consequently, $M \theta^{-1}$ is an $(\underline{\mathcal{A}}, \tilde{P})$ -martingale. We can also verify the result, previously noted, that the process $M \theta^{-1}$ is an unbiased estimator under \tilde{P} for $E(M_T)$ at time $t \in [t_0, T]$. If we denote by $\tilde{\text{Var}}(M_t \theta_t^{-1})$ the variance of the estimator $M_t \theta_t^{-1}$ under \tilde{P} , then

$$\begin{aligned} \tilde{\text{Var}}(M_t \theta_t^{-1}) &= \tilde{E} \left(\left(\sum_{j=1}^m \int_{t_0}^t \theta_s^{-2} (\xi_s^j - d_s^j M_s)^2 d\tilde{W}_s^j \right)^2 \right) \\ &= \int_{t_0}^t \tilde{E} \left(\theta_s^{-2} \sum_{j=1}^m (\xi_s^j - d_s^j M_s)^2 \right) ds. \end{aligned} \quad (16.2.22)$$

This should be compared to the variance of $\tilde{M}_t \theta_t$ under P given by (16.2.13). If the inequality $M_t > 0$ holds \tilde{P} a.s. for all $t \in [t_0, T]$ and we choose

$$d_t^j = \frac{\xi_t^j}{M_t}$$

for each $j \in \{1, 2, \dots, m\}$, then the variance of $M_t \theta_t^{-1}$ under \tilde{P} is reduced to zero. As previously noted, the estimator $M \theta^{-1}$ is a martingale under \tilde{P} . Consequently, if the integrands in (16.2.21) satisfy a mean square integrability condition of the form (16.2.7), then we can repeat the above procedure to find a new measure $P^{(1)}$ and a new exponential process $\theta^{(1)}$ such that $M \theta^{-1}(\theta^{(1)})^{-1}$ is an unbiased variance reduced estimator under $P^{(1)}$ for $\tilde{E}(M_T \theta_T^{-1}) = E(M_T)$. As indicated previously, this methodology can be applied iteratively to build more refined estimators.

Note that compared to the estimators $\tilde{M} \theta, \tilde{M}^1 \theta^1, \dots$ obtained under P , we do not need to deal with successive iterates of the martingale process $\tilde{M}, \tilde{M}^1, \dots$, as the original martingale M remains the same. However, we may need to construct a new measure at each iteration step.

A Class of Estimators

For many derivative security pricing problems we are given an \mathcal{A}_T -measurable random variable $H : \Omega \rightarrow \mathbb{R}$ and a valuation martingale process of the form

$$M_t = E(H | \mathcal{A}_t)$$

for $t \in [t_0, T]$. This provides a general framework for modeling various types of asset dynamics, contingent claims and sources of uncertainty. Because our variance reduction procedures, obtained under \tilde{P} , require only M to be square integrable, they can therefore be applied to a wide class of evaluation problems covering risk neutral pricing and real world pricing.

We will now consider a more explicit form for the martingale process M given by (16.2.1). Our task will be to verify condition (16.2.8) and find another expression for $\text{Var}(M_t \theta_t)$ than given by (16.2.13). Let $\mathbf{X}^{t_0, \underline{x}} = \{\mathbf{X}_t^{t_0, \underline{x}}, t \in [t_0, T]\}$ be a d -dimensional diffusion process which satisfies the system of SDEs (15.1.1) starting at time t_0 with initial value $\underline{x} \in \mathbb{R}^d$. As usual, we assume that appropriate growth and smoothness conditions apply for the drift and diffusion coefficients so that the solution to (15.1.1) is unique and Markovian.

We also take the valuation function $u : \Gamma_0 \cup \Gamma_1 \rightarrow \mathbb{R}$, as given by (15.1.4), with $u = \bar{u}$ for some suitable choice of payoff functional $h : \Gamma_1 \rightarrow \mathbb{R}$ with $h = \bar{h}$. The regions Γ_0 and Γ_1 and stopping time τ are as given in Sect. 15.1. Our task is, therefore, to construct variance reduced estimators for the martingale M given by

$$M_t = E\left(h(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) \mid \mathcal{A}_t\right) = u\left(t \wedge \tau, \mathbf{X}_{t \wedge \tau}^{t_0, \underline{x}}\right) \quad (16.2.23)$$

for $t \in [t_0, T]$, see (15.1.11). We assume that appropriate growth bounds apply for h so that M is square integrable under P .

Suppose that the process $\mathbf{d} = (d^1, \dots, d^m)^\top$ has the form

$$d_t^j = d^j(t, \mathbf{X}_t^{t_0, \underline{x}}) \quad (16.2.24)$$

for some suitable choice of real-valued functions $d^j : [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ for $t \in [t_0, T]$, $j \in \{1, 2, \dots, m\}$.

We let $\tilde{\mathbf{X}}^{t_0, \underline{x}} = \{\tilde{\mathbf{X}}_t^{t_0, \underline{x}}, t \in [t_0, T]\}$ be a d -dimensional diffusion process, starting at time t_0 with initial value $\underline{x} \in \mathbb{R}^d$ and integral representation

$$\tilde{\mathbf{X}}_t^{t_0, \underline{x}} = \underline{x} + \int_{t_0}^t \mathbf{a}\left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}\right) ds + \sum_{j=1}^m \int_{t_0}^t \mathbf{b}^j\left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}\right) d\tilde{W}_s^j. \quad (16.2.25)$$

We can now express $\tilde{\mathbf{X}}^{t_0, \underline{x}}$ as a diffusion process using \mathbf{W} rather than $\tilde{\mathbf{W}}$. Thus, from (16.2.4) and (16.2.25) we have

$$\begin{aligned}\tilde{\mathbf{X}}_t^{t_0, \underline{x}} &= \underline{x} + \int_{t_0}^t \left[\mathbf{a}\left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}\right) - \sum_{j=1}^m \mathbf{b}^j\left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}\right) d^j\left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}\right) \right] ds \\ &\quad + \sum_{j=1}^m \int_{t_0}^t \mathbf{b}^j\left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}\right) dW_s^j.\end{aligned}\quad (16.2.26)$$

Define the stopping time $\tilde{\tau} : \Omega \rightarrow \mathbb{R}$ by

$$\tilde{\tau} = \inf \left\{ t > 0 : \left(t, \tilde{\mathbf{X}}_t^{t_0, \underline{x}}\right) \notin \Gamma_0 \right\},$$

the martingale process \tilde{M} by

$$\tilde{M}_t = \tilde{E} \left(h \left(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}} \right) \mid \mathcal{A}_t \right) \quad (16.2.27)$$

and the region $\tilde{\Gamma}_1$ by

$$\tilde{\Gamma}_1 = \left\{ \left(\tau(\omega), \tilde{\mathbf{X}}_{\tilde{\tau}(\omega)}^{t_0, \underline{x}}(\omega) \right) \in [t_0, T] \times \mathbb{R}^d : \omega \in \Omega \right\}$$

for $t \in [t_0, T]$. As is the case for the martingale M , we assume that \tilde{M} is square integrable with respect to \tilde{P} and that $\tilde{\Gamma}_1 \subseteq \Gamma_1$, so that the random variable $h(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}})$ is well defined by continuity of the sample paths of $\tilde{\mathbf{X}}^{t_0, \underline{x}}$.

Using equations (15.1.1) and (16.2.25) we see that $\mathbf{X}^{t_0, \underline{x}}$ and $\tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}}$ are solutions of formally the same system of SDEs. Only the probability spaces are different. Consequently, from (16.2.23) and (16.2.27) we see that

$$\tilde{E}(\tilde{M}_T) = \tilde{E} \left(h \left(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}} \right) \right) = E \left(h \left(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}} \right) \right) = E(M_T), \quad (16.2.28)$$

which verifies condition (16.2.8).

Approximating Function

Let $\bar{u} : \Gamma_0 \cup \Gamma_1 \rightarrow \mathbb{R}$ be some approximation \bar{u} to u with

$$\bar{u} \left(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}} \right) = u \left(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}} \right). \quad (16.2.29)$$

We assume that the functions u and \bar{u} are of class $C^{1,2}$, that is, these functions have continuous first order time and second order spatial partial derivatives.

To simplify the notation in what follows, we will use the operators L^0 , \tilde{L}^0 and L^j , $j \in \{1, 2, \dots, m\}$ defined by

$$\begin{aligned}
L^0 &= \frac{\partial}{\partial t} + \sum_{i=1}^d a^i \frac{\partial}{\partial x^i} + \frac{1}{2} \sum_{i,k=1}^d \sum_{j=1}^m b^{i,j} b^{k,j} \frac{\partial^2}{\partial x^i \partial x^k}, \\
\tilde{L}^0 &= \frac{\partial}{\partial t} + \sum_{i=1}^d \left(a^i - \sum_{j=1}^m b^{i,j} d^j \right) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,k=1}^d \sum_{j=1}^m b^{i,j} b^{k,j} \frac{\partial^2}{\partial x^i \partial x^k}, \\
L^j &= \sum_{i=1}^d b^{i,j} \frac{\partial}{\partial x^i}.
\end{aligned}$$

The Kolmogorov backward equation applied to u in the continuation region Γ_0 can be expressed in the form

$$L^0 u(s, \mathbf{X}_s^{t_0, \underline{x}}) = 0 \quad (16.2.30)$$

and

$$L^0 u(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) = 0$$

for $(s, \mathbf{X}_s^{t_0, \underline{x}}) \in \Gamma_0$ and $(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) \in \Gamma_0$, respectively.

If we expand $u(t \wedge \tilde{\tau}, \tilde{\mathbf{X}}_{t \wedge \tilde{\tau}}^{t_0, \underline{x}})$ using Itô's formula for semimartingales and (16.2.25), then we have from (16.2.30) the relation

$$u(t \wedge \tilde{\tau}, \tilde{\mathbf{X}}_{t \wedge \tilde{\tau}}^{t_0, \underline{x}}) = u(t_0, \underline{x}) + \sum_{j=1}^m \int_{t_0}^{t \wedge \tilde{\tau}} L^j u(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) d\tilde{W}_s^j. \quad (16.2.31)$$

Writing this as an integral equation using the Wiener process \mathbf{W} rather than $\tilde{\mathbf{W}}$, we obtain from (16.2.4) the expression

$$\begin{aligned}
u(t \wedge \tilde{\tau}, \tilde{\mathbf{X}}_{t \wedge \tilde{\tau}}^{t_0, \underline{x}}) &= u(t_0, \underline{x}) - \sum_{j=1}^m \int_{t_0}^{t \wedge \tilde{\tau}} d^j \left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}} \right) L^j u(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) ds \\
&\quad + \sum_{j=1}^m \int_{t_0}^{t \wedge \tilde{\tau}} L^j u(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) dW_s^j.
\end{aligned} \quad (16.2.32)$$

In a similar manner we can determine $\bar{u}(t \wedge \tilde{\tau}, \tilde{\mathbf{X}}_{t \wedge \tilde{\tau}}^{t_0, \underline{x}})$ using Itô's formula and (16.2.25) to obtain

$$\begin{aligned}
\bar{u}(t \wedge \tilde{\tau}, \tilde{\mathbf{X}}_{t \wedge \tilde{\tau}}^{t_0, \underline{x}}) &= u(t_0, \underline{x}) + \int_{t_0}^{t \wedge \tilde{\tau}} L^0 \bar{u}(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) ds \\
&\quad + \sum_{j=1}^m \int_{t_0}^{t \wedge \tilde{\tau}} L^j \bar{u}(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) d\tilde{W}_s^j,
\end{aligned} \quad (16.2.33)$$

which, from (16.2.4), can be written in the form

$$\begin{aligned}
\bar{u}(t \wedge \tilde{\tau}, \tilde{\mathbf{X}}_{t \wedge \tilde{\tau}}^{t_0, \underline{x}}) &= u(t_0, \underline{x}) + \int_{t_0}^{t \wedge \tilde{\tau}} \left(L^0 \bar{u}(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) \right. \\
&\quad \left. - \sum_{j=1}^m \int_{t_0}^{t \wedge \tilde{\tau}} d^j(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) L^j \bar{u}(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) \right) ds \\
&\quad + \sum_{j=1}^m \int_{t_0}^{t \wedge \tilde{\tau}} L^j \bar{u}(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) dW_s^j. \tag{16.2.34}
\end{aligned}$$

Applying, once again, Itô's formula using (16.2.32), (16.2.34) and (16.2.5) we have the relations

$$u(t \wedge \tilde{\tau}, \tilde{\mathbf{X}}_{t \wedge \tilde{\tau}}^{t_0, \underline{x}}) \theta_{t \wedge \tilde{\tau}} = u(t_0, \underline{x}) \tag{16.2.35}$$

$$+ \sum_{j=1}^m \int_{t_0}^{t \wedge \tilde{\tau}} \theta_s \left[L^j u(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) + u(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) d^j(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) \right] dW_s^j,$$

$$\bar{u}(t \wedge \tilde{\tau}, \tilde{\mathbf{X}}_{t \wedge \tilde{\tau}}^{t_0, \underline{x}}) \theta_{t \wedge \tilde{\tau}} = \bar{u}(t_0, \underline{x}) + \int_{t_0}^{t \wedge \tilde{\tau}} L^0 \bar{u}(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) \theta_s ds \tag{16.2.36}$$

$$+ \sum_{j=1}^m \int_{t_0}^{t \wedge \tilde{\tau}} \theta_s \left[L^j \bar{u}(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) + \bar{u}(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) d^j(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) \right] dW_s^j$$

for $(t, \mathbf{X}_t) \in \Gamma_0$ and $(t, \tilde{\mathbf{X}}_t) \in \tilde{\Gamma}_0$, respectively. These equations can also be obtained from Wagner-Platen expansions using (16.2.26) and (16.2.5).

Combining (16.2.27), (16.2.31) and the equality $h(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}}) = u(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}})$ we see that

$$\begin{aligned}
\tilde{M}_t &= \tilde{E} \left(h(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}}) \mid \mathcal{A}_t \right) \\
&= \tilde{E} \left(u(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}}) \mid \mathcal{A}_t \right) \\
&= u(t_0, \underline{x}) + \sum_{j=1}^m \int_{t_0}^{t \wedge \tilde{\tau}} L^j u(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}) d\tilde{W}_s^j \\
&= u(t \wedge \tilde{\tau}, \tilde{\mathbf{X}}_{t \wedge \tilde{\tau}}^{t_0, \underline{x}}). \tag{16.2.37}
\end{aligned}$$

This relation can also be derived using more general arguments as given in Sect. 15.1, see in particular (15.1.11).

Unbiased Estimator

If we expand $u(\tau, \mathbf{X}_{\tau}^{t_0, \underline{x}})$ using Itô's formula (16.2.30) and (15.1.1), then we have

$$u(\tau, \mathbf{X}_\tau^{t_0, \underline{x}}) = u(t_0, \underline{x}) + \sum_{j=1}^m \int_{t_0}^\tau L^j u(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s^j.$$

Consequently, from (16.2.37) and taking the expectation under P on both sides of this relation we see that

$$E(M_T) = E(u(\tau, \mathbf{X}_\tau^{t_0, \underline{x}})) = u(t_0, \underline{x}). \quad (16.2.38)$$

Applying this result and (16.2.35) we can infer that

$$E\left(u\left(t \wedge \tilde{\tau}, \tilde{\mathbf{X}}_{t \wedge \tilde{\tau}}^{t_0, \underline{x}}\right) \theta_{t \wedge \tilde{\tau}}\right) = u(t_0, \underline{x}) = E(M_T). \quad (16.2.39)$$

This expression shows that the process $\{u(t \wedge \tilde{\tau}, \tilde{\mathbf{X}}_{t \wedge \tilde{\tau}}^{t_0, \underline{x}}) \theta_{t \wedge \tilde{\tau}}, t \in [t_0, T]\}$ is an unbiased estimator under P for $E(M_T)$ at time $t \wedge \tilde{\tau}$.

Furthermore, using (16.2.29) and (16.2.39) with $t = T$ we have

$$E\left(\bar{u}\left(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}}\right) \theta_{\tilde{\tau}}\right) = E\left(u\left(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}}\right) \theta_{\tilde{\tau}}\right) = u(t_0, \underline{x}) = E(M_T),$$

so that the random variable $\bar{u}(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}}) \theta_{\tilde{\tau}}$ is an unbiased estimator for $E(M_T)$ at time $\tilde{\tau}$.

Variance of the Estimator

Assume that $\bar{u}(t, \mathbf{x}) > 0$ for all $(t, \mathbf{x}) \in \Gamma_0$ and choose the function $\mathbf{d} = (d^1, \dots, d^m)^\top$ by

$$d^j(t, \mathbf{x}) = -\frac{L^j \bar{u}(t, \mathbf{x})}{\bar{u}(t, \mathbf{x})} \quad (16.2.40)$$

for $(t, \mathbf{x}) \in \Gamma_0$ and $j \in \{1, 2, \dots, m\}$. Using this choice for \mathbf{d} , together with equation (16.2.36), we can compute the variance of $\bar{u}(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}}) \theta_{\tilde{\tau}}$ under P by the relation

$$\text{Var}\left(\bar{u}\left(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}}\right) \theta_{\tilde{\tau}}\right) = \text{Var}\left(\int_{t_0}^{\tilde{\tau}} L^0 \bar{u}\left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}\right) \theta_s ds\right).$$

By application of (16.2.30) this formula can also be written in the form

$$\text{Var}\left(\bar{u}\left(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}}\right) \theta_{\tilde{\tau}}\right) = \text{Var}\left(\int_{t_0}^{\tilde{\tau}} \left(L^0 \bar{u}\left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}\right) - L^0 u\left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}}\right)\right) \theta_s ds\right), \quad (16.2.41)$$

which is a more convenient representation of the variance for some applications. This should be compared with the variance obtained from (16.2.35), which is given by the equation

$$\begin{aligned}
& \text{Var} \left(\bar{u} \left(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}} \right) \theta_{\tilde{\tau}} \right) \\
&= E \left(\left(\sum_{j=1}^m \int_{t_0}^{\tilde{\tau}} \theta_s \left[L^j u \left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}} \right) - L^j \bar{u} \left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}} \right) \frac{u \left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}} \right)}{\bar{u} \left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}} \right)} \right] dW_s^j \right)^2 \right) \\
&= E \left(\int_{t_0}^{\tilde{\tau}} \theta_s^2 \sum_{j=1}^m \left[L^j u \left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}} \right) - L^j \bar{u} \left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}} \right) \frac{u \left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}} \right)}{\bar{u} \left(s, \tilde{\mathbf{X}}_s^{t_0, \underline{x}} \right)} \right]^2 ds \right). \quad (16.2.42)
\end{aligned}$$

This relation can also be obtained by substituting (16.2.40) into (16.2.13). Thus, for a suitable choice of the function \mathbf{d} we can make the random variable $\bar{u}(\tilde{\tau}, \tilde{\mathbf{X}}_{\tilde{\tau}}^{t_0, \underline{x}}) \theta_{\tilde{\tau}}$ an unbiased variance reduced estimator for $E(M_T)$ at time τ .

The methods described above provide the basis for powerful variance reduction and error minimization procedures. The explicit formulas for the variance of estimators given by (16.2.13), (16.2.22), (16.2.41) and (16.2.42) mean that exact controls of the variance can be built and theoretical estimates can be given.

We emphasize that the smoothness conditions used in the derivation of (16.2.41) and (16.2.42) are not required in the more general formulations leading to (16.2.13) and (16.2.22). Also, condition (16.2.8), required for the estimator $\tilde{M}\theta$, can often be easily verified in practice for reasonable choices of \tilde{M} . The approximating function $\bar{u} : \Gamma_0 \cup \Gamma_1 \rightarrow \mathbb{R}$ can be any choice which satisfies the terminal payoff condition (16.2.29) and is $C^{1,2}$ smooth. This approximating function may be constructed iteratively using some combination of analytic solutions, numerical methods and appropriate interpolation routines. For example, it may correspond to the pricing function for a similar model, where analytic or more accurate pricing can be obtained.

16.3 Discrete-Time Variance Reduced Estimators

In this section we consider the application of discrete-time numerical methods that can be used to either approximate a given continuous time estimator or to build new variance reduced estimators. We also find discrete-time approximations for the integrands appearing in the Itô integral representations of the discounted price processes of derivative securities. These methods are of interest theoretically because they use only basic properties of stochastic processes and are of practical value because of their close association with the actual implementation.

Discrete-Time Estimators

To illustrate the principles involved, we will first consider the problem of discretizing the product estimator $\tilde{M}\theta$ for $E(M_T)$ considered in Sect. 16.2, where M is a given square integrable (\mathcal{A}, P) -martingale with representation

(16.2.1) and satisfying the integrability condition (16.2.2). We take \tilde{M} to be an (\mathcal{A}, \tilde{P}) -martingale expressible in the form (16.2.6) and satisfying (16.2.7), where θ is the Radon-Nikodym derivative given by (16.2.5). The process $d = (d^1, \dots, d^m)^\top$ is assumed to be a free parameter process, satisfying (16.2.3), which is used to construct θ , to reduce the variance. Note that the definitions of unbiased and variance reduced unbiased estimators can be formulated for discrete-time processes in a similar fashion to that described in Sect. 16.2.

Let us consider an equidistant time discretization of the interval $[t_0, T]$ of the form

$$t_0 = \tau_0 < \tau_1 < \dots < \tau_N = T \quad (16.3.1)$$

with step size

$$\Delta = \frac{(T - t_0)}{N}$$

for $N \in \mathcal{N}$. Furthermore, let $M^\Delta = \{M_k^\Delta, k \in \{0, 1, \dots, N-1\}\}$, \tilde{M}_k^Δ and θ_k^Δ , represent weak Euler approximations of the processes M , \tilde{M} and θ given by (16.2.1), (16.2.11) and (16.2.5), respectively, of the form

$$M_{k+1}^\Delta = M_k^\Delta + \sum_{j=1}^m \tilde{\xi}_k^j \Delta W_k^j, \quad (16.3.2)$$

$$\tilde{M}_{k+1}^\Delta = \tilde{M}_k^\Delta - \sum_{j=1}^m d_k^j \tilde{\xi}_k^j \Delta + \sum_{j=1}^m \tilde{\xi}_k^j \Delta W_k^j, \quad (16.3.3)$$

$$\theta_{k+1}^\Delta = \theta_k^\Delta + \sum_{j=1}^m d_k^j \theta_k^\Delta \Delta W_k^j, \quad (16.3.4)$$

for $k \in \{0, 1, \dots, N-1\}$ with initial values $M_0^\Delta = M_0$, $\tilde{M}_0^\Delta = M_0$ and $\theta_0^\Delta = 1$. We will take the increments $(\Delta W_k^1, \dots, \Delta W_k^m)^\top$, $k \in \{0, 1, \dots, N-1\}$, to be either a collection of independent Gaussian random variables with expectation zero and variance Δ under P , or a collection of multi-point random variables, again with expectation zero and variance Δ under P . We require values for only the first and second moments of these increments in what follows. For example, we can take a set of two-point distributed random variables with

$$P(\Delta W_k^j = \pm \sqrt{\Delta}) = \frac{1}{2} \quad (16.3.5)$$

for $k \in \{0, 1, \dots, N-1\}$ and $j \in \{1, 2, \dots, m\}$.

Multiplying (16.3.3) by (16.3.4) the product $\tilde{M}_{k+1}^\Delta \theta_{k+1}^\Delta$ can be expressed in the form

$$\begin{aligned}
\tilde{M}_{k+1}^{\Delta} \theta_{k+1}^{\Delta} &= \tilde{M}_k^{\Delta} \theta_k^{\Delta} - \sum_{j=1}^m d_k^j \tilde{\xi}_k^j \theta_k^{\Delta} \Delta + \sum_{j=1}^m \tilde{\xi}_k^j \theta_k^{\Delta} \Delta W_k^j \\
&+ \sum_{j=1}^m \tilde{M}_k^{\Delta} d_k^j \theta_k^{\Delta} \Delta W_k^j - \left(\sum_{j=1}^m d_k^j \tilde{\xi}_k^j \Delta \right) \left(\sum_{j=1}^m d_k^j \theta_k^{\Delta} \Delta W_k^j \right) \\
&+ \left(\sum_{j=1}^m \tilde{\xi}_k^j \Delta W_k^j \right) \left(\sum_{j=1}^m d_k^j \theta_k^{\Delta} \Delta W_k^j \right). \tag{16.3.6}
\end{aligned}$$

We can write

$$\begin{aligned}
\left(\sum_{j=1}^m \tilde{\xi}_k^j \Delta W_k^j \right) \left(\sum_{j=1}^m d_k^j \theta_k^{\Delta} \Delta W_k^j \right) &= \sum_{j=1}^m \tilde{\xi}_k^j d_k^j \theta_k^{\Delta} (\Delta W_k^j)^2 \\
&+ \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^m \tilde{\xi}_k^{j_2} d_k^{j_1} \theta_k^{\Delta} \Delta W_k^{j_1} \Delta W_k^{j_2} \tag{16.3.7}
\end{aligned}$$

and consequently, using the fact that $E((\Delta W_k^j)^2) = \Delta$ and $E(\Delta W_k^j \Delta W_k^{\ell}) = 0$ for $j, \ell \in \{1, 2, \dots, m\}$, $k \in \{0, 1, \dots, N-1\}$ and $j \neq \ell$, the latter following from the independence property of the increments ΔW_k^j , we have the relation

$$E \left(\left(\sum_{j=1}^m \tilde{\xi}_k^j \Delta W_k^j \right) \left(\sum_{j=1}^m d_k^j \theta_k^{\Delta} \Delta W_k^j \right) \right) = E \left(\theta_k^{\Delta} \sum_{j=1}^m \tilde{\xi}_k^j d_k^j \right) \Delta.$$

From this result, again using the independence property of the increments ΔW_k^j and taking expectations on both sides of (16.3.6) and the initial value $\theta_0^{\Delta} = 1$, we obtain

$$E \left(\tilde{M}_{k+1}^{\Delta} \theta_{k+1}^{\Delta} \right) = E \left(\tilde{M}_k^{\Delta} \theta_k^{\Delta} \right) = E \left(\tilde{M}_0^{\Delta} \theta_0^{\Delta} \right) = M_0 \tag{16.3.8}$$

for $k \in \{0, 1, \dots, N-1\}$. This means that $\tilde{M}^{\Delta} \theta^{\Delta}$ is an unbiased estimator for $E(M_N) = M_0$ at time τ_k , $k \in \{0, 1, \dots, N-1\}$.

Variance of Discrete-Time Estimators

Applying once again the independence property of the increments ΔW_k^j , we see that for all $k, \ell \in \{0, 1, \dots, N-1\}$, $j_1, j_2 \in \{1, 2, \dots, m\}$ with $\ell < k$ the random variables $\Delta W_k^{j_1}$ and $\Delta W_k^{j_1} \Delta W_k^{j_2}$ are both independent of any Borel measurable functions of the variates $d_k^{j_1}, \tilde{\xi}_k^{j_1}, \theta_k^{\Delta}, M_k^{\Delta}, d_{\ell}^{j_2}, \tilde{\xi}_{\ell}^{j_2}, \theta_{\ell}^{\Delta}, M_{\ell}^{\Delta}, \Delta W_{\ell}^{j_1}, \Delta_{\ell}^{j_2}$.

In particular, $(\Delta W_k^{j_1})^2 - \Delta$ is independent of $d_k^{j_1}, \tilde{\xi}_k^{j_1}, \theta_k^{\Delta}, d_{\ell}^{j_2}, \tilde{\xi}_{\ell}^{j_2}, \theta_{\ell}^{\Delta}[(\Delta W_{\ell}^{j_2})^2 - \Delta]$.

Therefore, if we denote by $C_{k,\ell}$ the covariance

$$C_{k,\ell} = \text{Cov} \left(\tilde{M}_{k+1} \theta_{k+1}^\Delta - \tilde{M}_k \theta_k^\Delta, \tilde{M}_{\ell+1} \theta_{\ell+1}^\Delta - \tilde{M}_\ell \theta_\ell^\Delta \right),$$

then from (16.3.6), (16.3.7), (16.3.8) and the independence properties stated above and for integers $\ell < k$, we have after simplification the equations

$$\begin{aligned} C_{k,\ell} &= E \left(\left(\sum_{j=1}^m d_k^j \tilde{\xi}_k^j \theta_k^\Delta [(\Delta W_k^j)^2 - \Delta] \right) \left(\sum_{j=1}^m d_\ell^j \tilde{\xi}_\ell^j \theta_\ell^\Delta [(\Delta W_\ell^j)^2 - \Delta] \right) \right) \\ &= \sum_{j_1, j_2=1}^m E \left(d_k^{j_1} \tilde{\xi}_k^{j_1} \theta_k^\Delta d_\ell^{j_2} \tilde{\xi}_\ell^{j_2} \theta_\ell^\Delta [(\Delta W_k^{j_1})^2 - \Delta] [(\Delta W_\ell^{j_2})^2 - \Delta] \right) \\ &= \sum_{j_1, j_2=1}^m E \left(d_k^{j_1} \tilde{\xi}_k^{j_1} \theta_k^\Delta d_\ell^{j_2} \tilde{\xi}_\ell^{j_2} \theta_\ell^\Delta [(\Delta W_\ell^{j_2})^2 - \Delta] \right) E \left([(\Delta W_k^{j_1})^2 - \Delta] \right) \\ &= 0. \end{aligned} \tag{16.3.9}$$

Also, for any $k \in \{0, 1, \dots, N-1\}$ and using the initial value $\theta_0 = 1$ we can write

$$\tilde{M}_k^\Delta \theta_k^\Delta = \sum_{\ell=0}^{k-1} \left(\tilde{M}_{\ell+1}^\Delta \theta_{\ell+1}^\Delta - \tilde{M}_\ell^\Delta \theta_\ell^\Delta \right) + M_0$$

so that by applying (16.3.9), the variance of $\tilde{M}_k^\Delta \theta_k^\Delta$, $k \in \{1, 2, \dots, N\}$ can be computed from the formula

$$\text{Var} \left(\tilde{M}_k^\Delta \theta_k^\Delta \right) = \sum_{\ell=0}^{k-1} E \left(\left(\tilde{M}_{\ell+1}^\Delta \theta_{\ell+1}^\Delta - \tilde{M}_\ell^\Delta \theta_\ell^\Delta \right)^2 \right). \tag{16.3.10}$$

For small values of Δ we can ignore all terms in the product $(\tilde{M}_{\ell+1}^\Delta \theta_{\ell+1}^\Delta - \tilde{M}_\ell^\Delta \theta_\ell^\Delta)^2$, $\ell \in \{1, \dots, k-1\}$ obtained from (16.3.6) whose expectation under P is zero or of order Δ^q for $q \geq \frac{3}{2}$. This means that all terms on the right hand side of (16.3.6), except

$$\sum_{j=1}^m \tilde{\xi}_k^j \theta_k^\Delta \Delta W_k^j \quad \text{and} \quad \sum_{j=1}^m \tilde{M}_k^\Delta d_k^j \theta_k^\Delta \Delta W_k^j,$$

can be ignored, and using the property $E((\Delta W_k^j)^2) = \Delta$ we can approximate the variance of $\tilde{M}_k^\Delta \theta_k^\Delta$, $k \in \{1, \dots, N\}$ by the expression

$$\text{Var}(\tilde{M}_k^\Delta \theta_k^\Delta) \approx \sum_{\ell=0}^{k-1} E \left((\theta_\ell^\Delta)^2 \left(\sum_{j=1}^m \left(\tilde{\xi}_\ell^j + d_\ell^j \tilde{M}_\ell^\Delta \right)^2 \right) \right) \Delta. \tag{16.3.11}$$

As in Sect. 16.2 we let $\hat{\xi}_\ell = (\hat{\xi}_\ell^1, \dots, \hat{\xi}_\ell^m)^\top$ and \hat{M}_ℓ^Δ be approximations for $\tilde{\xi}_\ell = (\tilde{\xi}_\ell^1, \dots, \tilde{\xi}_\ell^m)^\top$ and \tilde{M}_ℓ^Δ , $\ell \in \{0, 1, \dots, N-1\}$, respectively. If the inequality $\tilde{M}_\ell^\Delta > 0$ holds P -a.s. for all $\ell \in \{0, 1, \dots, N-1\}$, then we can choose $d_\ell^j = -\frac{\hat{\xi}_\ell^j}{\hat{M}_\ell^\Delta}$. Substituting this value into (16.3.11) one obtains

$$\text{Var}(\tilde{M}_k^\Delta \theta_k^\Delta) \approx \sum_{\ell=0}^{k-1} E \left[(\theta_\ell^\Delta)^2 \sum_{j=1}^m \left(\tilde{\xi}_k^j - \hat{\xi}_\ell^j \frac{\tilde{M}_\ell^\Delta}{\hat{M}_\ell^\Delta} \right)^2 \right] \Delta \quad (16.3.12)$$

for $k \in \{1, 2, \dots, N\}$.

The variance formulas (16.3.11) and (16.3.12) for the estimator $\tilde{M}^\Delta \theta^\Delta$ should be compared to the variance of the continuous time version of the estimator given by (16.2.13) and (16.2.15), respectively. Note that with the above formulation we have used only very basic properties of discrete-time approximation processes. In the case where the increments ΔW_k^j are two-point random variables, with probabilities given by (16.3.5), these calculations can be further simplified since in this case $(\Delta W_k^j)^2 = \Delta$ for $j \in \{1, 2, \dots, m\}$, $k \in \{0, 1, \dots, N-1\}$. Similar expressions for the variance given by (16.3.11) and (16.3.12) are obtained if we replace the Euler approximations (16.3.2) to (16.3.4) by other higher order weak approximations, although for these approximations the above computations become more involved.

The variance reduction procedures discussed in the previous sections can also be formulated in a discrete-time framework. However, in this case we need to be more careful regarding how the problem should be formulated as some choices lead to biased estimates of $E(M_T)$.

Construction of an Unbiased Discrete-Time Estimator

To see how an unbiased estimator can be constructed, let M^Δ and θ^Δ be weak Euler approximations of the processes M and θ as given by (16.3.2) and (16.3.4), respectively. By observing the form of (16.3.4) and the initial condition $\theta_0^\Delta = 1$, an induction argument shows that

$$\theta_k^\Delta = \prod_{\ell=0}^{k-1} \left(1 + \sum_{j=1}^m d_\ell^j \Delta W_\ell^j \right) \quad (16.3.13)$$

for $k \in \{1, 2, \dots, N-1\}$.

Define the measure $\tilde{P}^\Delta : \mathcal{A}_T \rightarrow [0, 1]$ by

$$\tilde{P}^\Delta(A) = \int_A \theta_N^\Delta dP \quad (16.3.14)$$

for $A \in \mathcal{A}_T$.

We will now assume that the increments $(\Delta W_k^1, \dots, \Delta W_k^m)^\top$, $k \in \{0, 1, \dots, N-1\}$, form a vector of two-point distributed, independent random variables

with mean zero and variance Δ under P with probabilities given by (16.3.5). In addition, we assume that

$$\left| \sum_{j=1}^m d_k^j \Delta W_k^j \right| \leq K_1 \sqrt{\Delta} \quad (16.3.15)$$

for some constant $K_1 \in \mathbb{R}^+$ for all $k \in \{0, 1, \dots, N-1\}$ and a time discretization with maximum step size Δ . This assumption clearly depends on the properties of the random variables d_k^j , $j \in \{1, 2, \dots, m\}$, $k \in \{0, 1, \dots, N-1\}$. For a discussion on how the growth of these variates can be constrained, see the commentary following (16.2.16) in Sect. 16.2. We restrict our attention to two-point random increments because it simplifies the calculations in what follows. However, this analysis can, in fact, be extended to a wider class of multi-point random variables.

Applying (16.3.13) and the independence property of the increments ΔW_k^j we see that $E(\theta_N^\Delta) = 1$. Also for sufficiently fine time discretizations with maximum step size Δ relation (16.3.15) shows that $\theta_k^\Delta > 0$ for all $k \in \{0, 1, \dots, N-1\}$. This means that both the probability measure \tilde{P}^Δ and the quotient $\frac{M_k^\Delta}{\theta_k^\Delta}$, $k \in \{0, 1, \dots, N-1\}$, are well defined.

Using the definition of \tilde{P}^Δ given by (16.3.14), and again the independence property of the increments ΔW_ℓ^j , and (16.3.13) we can infer that

$$\begin{aligned} \tilde{E}^\Delta \left(\frac{M_k^\Delta}{\theta_k^\Delta} \right) &= E \left(\frac{M_k^\Delta}{\theta_k^\Delta} \theta_N^\Delta \right) \\ &= E(M_k^\Delta) E \left(\prod_{\ell=k}^{N-1} \left(1 + \sum_{j=1}^m d_\ell^j \Delta W_\ell^j \right) \right) \\ &= E(M_k^\Delta) \\ &= M_0 \end{aligned} \quad (16.3.16)$$

for $k \in \{0, 1, \dots, N\}$. Here \tilde{E}^Δ denotes expectation with respect to the measure \tilde{P}^Δ . This shows that $\frac{M_k^\Delta}{\theta_k^\Delta}$ is an unbiased estimator under \tilde{P}^Δ for $E(M_N) = M_0$ at time τ_k , $k \in \{0, 1, \dots, N\}$.

It is important to know the probabilities of the outcomes of $(\Delta W_k^1, \dots, \Delta W_k^m)^\top$, $k \in \{0, 1, \dots, N-1\}$, if we undertake a Monte Carlo simulation to approximate the value $E(M_N) = M_0$ using the estimator $\frac{M_N^\Delta}{\theta_N^\Delta}$ under the measure \tilde{P}^Δ . To determine these probabilities we let $A_k^j = \{\omega : \Delta W_k^j = +\sqrt{\Delta}\}$ for $j \in \{1, 2, \dots, m\}$, $k \in \{0, 1, \dots, N-1\}$. From the definition of \tilde{P}^Δ given by (16.3.14), the independence property of the increments ΔW_ℓ^p , and (16.3.13) we have

$$\begin{aligned}
\tilde{P}^\Delta(A_k^j) &= \tilde{E}^\Delta(\mathbf{1}_{A_k^j}) \\
&= E(\mathbf{1}_{A_k^j} \theta_N^\Delta) \\
&= E\left(\mathbf{1}_{A_k^j} \prod_{\ell=1}^{N-1} \left(1 + \sum_{p=1}^m d_\ell^p \Delta W_\ell^p\right)\right) \\
&= E\left(\mathbf{1}_{A_k^j} \left(1 + \sum_{p=1}^m d_k^p \Delta W_\ell^p\right)\right) \\
&= \frac{1 + d_k^j \sqrt{\Delta}}{2}.
\end{aligned}$$

A similar expression holds for $\tilde{P}^\Delta(\{\omega : \Delta W_k^j = -\sqrt{\Delta}\})$. These results can be summarized in the form

$$\tilde{P}^\Delta(\Delta W_k^j = \pm\sqrt{\Delta}) = \left(\frac{1 \pm \sqrt{\Delta} d_k^j}{2}\right). \quad (16.3.17)$$

These probabilities can also be used to show that

$$\begin{aligned}
\tilde{E}^\Delta(\Delta W_k^j) &= d_k^j \Delta, \\
\tilde{E}^\Delta((\Delta W_k^j)^2) &= \Delta,
\end{aligned} \quad (16.3.18)$$

for $j \in \{1, 2, \dots, m\}$, $k \in \{0, 1, \dots, N\}$.

We now introduce the random variables η_k^Δ , ϕ_k^Δ , ψ_k^Δ defined by

$$\begin{aligned}
\eta_k^\Delta &= \sum_{j=1}^m \xi_k^j \Delta W_k^j, \\
\psi_k^\Delta &= \sum_{j=1}^m d_k^j \Delta W_k^j, \\
\phi_k^\Delta &= \eta_k^\Delta - M_k^\Delta \psi_k^\Delta,
\end{aligned}$$

for $k \in \{1, 2, \dots, N-1\}$. From the definition of ψ_k^Δ , equation (16.3.13) can be expressed in the form

$$\theta_k^\Delta = \prod_{\ell=0}^{k-1} (1 + \psi_\ell^\Delta) \quad (16.3.19)$$

so that $\theta_{k+1}^\Delta = \theta_k^\Delta(1 + \psi_k^\Delta)$. Combining this result and the Euler approximation (16.3.2) we can write for $k \in \{0, 1, \dots, N-1\}$

$$\begin{aligned} \frac{M_{k+1}^\Delta}{\theta_{k+1}^\Delta} - \frac{M_k^\Delta}{\theta_k^\Delta} &= \frac{M_{k+1}^\Delta - M_k^\Delta(1 + \psi_\ell^\Delta)}{\theta_{k+1}^\Delta} \\ &= \frac{\phi_k^\Delta}{\theta_{k+1}^\Delta}. \end{aligned} \quad (16.3.20)$$

Variance of Discrete-Time Estimators

For integers $k, \ell \in \{0, 1, \dots, N-1\}$, let

$$\tilde{\text{Cov}}^\Delta \left(\frac{\phi_k^\Delta}{\theta_{k+1}^\Delta}, \frac{\phi_\ell^\Delta}{\theta_{\ell+1}^\Delta} \right) = \tilde{E}^\Delta \left(\left[\frac{\phi_k^\Delta}{\theta_{k+1}^\Delta} - \tilde{E} \left(\frac{\phi_k^\Delta}{\theta_{k+1}^\Delta} \right) \right] \left[\frac{\phi_\ell^\Delta}{\theta_{\ell+1}^\Delta} - \tilde{E} \left(\frac{\phi_\ell^\Delta}{\theta_{\ell+1}^\Delta} \right) \right] \right)$$

denote the covariance of the random variables $\phi_k^\Delta/\theta_{k+1}^\Delta$ and $\phi_\ell^\Delta/\theta_{\ell+1}^\Delta$ under the measure \tilde{P}^Δ . From (16.3.16) and (16.3.20) we see that

$$\tilde{E}^\Delta \left(\frac{\phi_k^\Delta}{\theta_{k+1}^\Delta} \right) = \tilde{E}^\Delta \left(\frac{M_{k+1}^\Delta}{\theta_{k+1}^\Delta} \right) - \tilde{E}^\Delta \left(\frac{M_k^\Delta}{\theta_k^\Delta} \right) = 0. \quad (16.3.21)$$

Consequently, applying the representation (16.3.19) we have for integers $k, \ell \in \{0, 1, \dots, N-1\}$ with $\ell < k$

$$\begin{aligned} \tilde{\text{Cov}}^\Delta \left(\frac{\phi_k^\Delta}{\theta_{k+1}^\Delta}, \frac{\phi_\ell^\Delta}{\theta_{\ell+1}^\Delta} \right) &= \tilde{E}^\Delta \left(\frac{\phi_k^\Delta}{\theta_{k+1}^\Delta} \frac{\phi_\ell^\Delta}{\theta_{\ell+1}^\Delta} \right) \\ &= E \left(\frac{\phi_k^\Delta \phi_\ell^\Delta}{\theta_{\ell+1}^\Delta} \left(\frac{\theta_N^\Delta}{\theta_{k+1}^\Delta} \right) \right) \\ &= E \left(\frac{(\eta_k^\Delta - M_k^\Delta \psi_k^\Delta) \phi_\ell^\Delta}{\theta_{\ell+1}^\Delta} \left(\frac{\theta_N^\Delta}{\theta_{k+1}^\Delta} \right) \right). \end{aligned} \quad (16.3.22)$$

Noting that the increments ΔW_k^j , $j \in \{1, 2, \dots, m\}$, appear only in the terms η_k^Δ and ψ_k^Δ and not in the terms M_k^Δ , $\theta_{\ell+1}^\Delta$, ϕ_ℓ^Δ and

$$\frac{\theta_N^\Delta}{\theta_{k+1}^\Delta} = \prod_{p=k+1}^{N-1} (1 + \psi_p^\Delta),$$

and using the independence property of the increments ΔW_k^j , we can infer from (16.3.22) that for $k, \ell \in \{0, 1, \dots, N-1\}$ with $\ell < k$

$$\begin{aligned} \tilde{\text{Cov}}^\Delta \left(\frac{\phi_k^\Delta}{\theta_{k+1}^\Delta}, \frac{\phi_\ell^\Delta}{\theta_{\ell+1}^\Delta} \right) &= E(\eta_k^\Delta) E \left(\frac{\psi_\ell^\Delta}{\theta_{\ell+1}^\Delta} \left(\frac{\theta_N^\Delta}{\theta_{k+1}^\Delta} \right) \right) \\ &\quad - E(\psi_k^\Delta) E \left(\frac{M_k^\Delta \phi_\ell^\Delta}{\theta_{\ell+1}^\Delta} \left(\frac{\theta_N^\Delta}{\theta_{k+1}^\Delta} \right) \right) \\ &= 0. \end{aligned} \quad (16.3.23)$$

Also for $k \in \{0, 1, \dots, N-1\}$ we can write

$$\frac{M_k^\Delta}{\theta_k^\Delta} = \sum_{\ell=0}^{k-1} \left(\frac{M_{\ell+1}^\Delta}{\theta_{\ell+1}^\Delta} - \frac{M_\ell^\Delta}{\theta_\ell^\Delta} \right) + M_0$$

since $\theta_0^\Delta = 1$. Combining (16.3.20), (16.3.21) and (16.3.23) we obtain

$$\begin{aligned} \tilde{\text{Var}}^\Delta \left(\frac{M_k^\Delta}{\theta_k^\Delta} \right) &= \tilde{\text{Var}}^\Delta \left(\sum_{\ell=0}^{k-1} \left(\frac{M_{\ell+1}^\Delta}{\theta_{\ell+1}^\Delta} - \frac{M_\ell^\Delta}{\theta_\ell^\Delta} \right) + M_0 \right) \\ &= \sum_{\ell=0}^{k-1} \tilde{E}^\Delta \left(\left(\frac{\phi_\ell^\Delta}{\theta_{\ell+1}^\Delta} \right)^2 \right). \end{aligned} \quad (16.3.24)$$

If we assume that we can ignore all terms in the expansions of $(\frac{\phi_\ell^\Delta}{\theta_{\ell+1}^\Delta})^2$ given by

$$\left(\frac{\phi_\ell^\Delta}{\theta_{\ell+1}^\Delta} \right)^2 = \left(\frac{\phi_\ell^\Delta}{\theta_\ell^\Delta} \right)^2 \left(\frac{1}{1 + \psi_\ell} \right)^2 = \left(\frac{\phi_\ell^\Delta}{\theta_\ell^\Delta} \right)^2 (1 - \psi_\ell + (\psi_\ell)^2 - \dots)^2,$$

whose expectation under \tilde{P}^Δ is zero or of order Δ^q for $q \geq \frac{3}{2}$, then the variance of $\frac{M_k^\Delta}{\theta_k^\Delta}$ under \tilde{P}^Δ , as given by (16.3.24), can be approximated by

$$\begin{aligned} \tilde{\text{Var}}^\Delta \left(\frac{M_k^\Delta}{\theta_k^\Delta} \right) &\approx \sum_{\ell=0}^{k-1} \tilde{E}^\Delta \left(\left(\frac{\phi_\ell^\Delta}{\theta_\ell^\Delta} \right)^2 \right) \\ &= \sum_{\ell=0}^{k-1} \tilde{E}^\Delta \left(\left(\frac{1}{\theta_\ell^\Delta} \right)^2 \left(\sum_{j=1}^m \left(\xi_\ell^j - M_\ell^\Delta d_\ell^j \right) \Delta W_\ell^j \right)^2 \right). \end{aligned} \quad (16.3.25)$$

We also know from the independence property of the increments ΔW_k^j and (16.3.19) that

$$\begin{aligned} \tilde{E}^\Delta \left(\Delta W_k^{j_1} \Delta W_k^{j_2} \right) &= E \left(\Delta W_k^{j_1} \Delta W_k^{j_2} \theta_N^\Delta \right) \\ &= E \left(\Delta W_k^{j_1} \Delta W_k^{j_2} \prod_{\ell=0}^{N-1} (1 + \psi_\ell^\Delta) \right) \\ &= E \left(\Delta W_k^{j_1} \Delta W_k^{j_2} (1 + \psi_k^\Delta) \right) \\ &= E \left(\Delta W_k^{j_1} \Delta W_k^{j_2} \right) + E \left(\Delta W_k^{j_1} \Delta W_k^{j_2} \Delta W_k^{j_2} \right) \\ &\quad + E \left(\Delta W_k^{j_1} \Delta W_k^{j_2} \Delta W_k^{j_2} \right) \\ &= 0 \end{aligned}$$

for $k \in \{0, 1, \dots, N-1\}$, $j_1, j_2 \in \{1, 2, \dots, m\}$ with $j_1 \neq j_2$. Consequently, since for any $\ell \in \{0, 1, \dots, N-1\}$ and $j \in \{1, 2, \dots, m\}$ the random variable ΔW_ℓ^j will be independent of the variates $(\frac{1}{\theta_\ell^\Delta})^2$, ξ_ℓ^j , M_ℓ^Δ and d_ℓ^j , as these variates include the increments ΔW_p^j , $j \in \{1, 2, \dots, m\}$, with index values p up to, but excluding the value ℓ , then from (16.3.25) and (16.3.18) we have

$$\tilde{\text{Var}}^\Delta \left(\frac{M_k^\Delta}{\theta_k^\Delta} \right) \approx \sum_{\ell=0}^{k-1} \tilde{E}^\Delta \left(\left(\frac{1}{\theta_\ell^\Delta} \right)^2 \sum_{j=1}^m (\xi_\ell^j - M_\ell^\Delta d_\ell^j)^2 \right) \Delta. \quad (16.3.26)$$

This should be compared to the variance obtained from (16.2.22) for the estimator $M\theta^{-1}$.

Discrete Estimation of Integrands

We will now consider the problem of directly estimating the integrands in the martingale representation (16.2.1) using discrete-time methods. Let $\mathbf{Y}_k^\Delta = (Y_k^{\Delta,1}, \dots, Y_k^{\Delta,d})^\top$, $k \in \{0, 1, \dots, N-1\}$ be an Euler approximation of the diffusion process \mathbf{X} given by (15.1.1) of the form

$$Y_{k+1}^{\Delta,i} = Y_k^{\Delta,i} + a^i(\tau_k, \mathbf{Y}_k^\Delta) \Delta + \sum_{j=1}^m b^{i,j}(\tau_k, \mathbf{Y}_k^\Delta) \Delta \hat{W}_k^j \quad (16.3.27)$$

for $i \in \{1, 2, \dots, d\}$, $k \in \{0, 1, \dots, N-1\}$ with initial value $\mathbf{Y}_0^\Delta = \mathbf{X}_0 = \underline{x} \in \Re^m$, where $\Delta \hat{W}_k^j$ are two-point random variables with probabilities given by (16.3.5), and we use the equidistant time discretization of the interval $[t_0, T]$ given by (16.3.1). For simplicity, we consider a valuation functional of the form $h(\mathbf{Y}_N^\Delta)$.

That is, we consider only functionals of the terminal value of the approximation \mathbf{Y}^Δ corresponding to European style derivative securities. We assume that \mathbf{Y}^Δ converges weakly to \mathbf{X} for the functional h , see Chap. 11. For $k \in \{0, 1, \dots, N-1\}$ define

$$M_k^\Delta = E \left(h \left(\mathbf{Y}_N^\Delta \right) \mid \mathcal{A}_{\tau_k} \right). \quad (16.3.28)$$

As in Chap. 15, see (15.1.11), we let $u : \{\tau_0, \dots, \tau_N\} \times \Re^d \rightarrow \Re$, be a valuation function of the form

$$u(\tau_k, \mathbf{Y}_k^\Delta) = M_k^\Delta. \quad (16.3.29)$$

This expression and the law of iterated conditional expectations, see (1.1.10), applied to the discrete-time martingale M^Δ means that

$$u(\tau_k, \mathbf{Y}_k^\Delta) = E \left(u \left(\tau_{k+1}, \mathbf{Y}_{k+1}^\Delta \right) \mid \mathcal{A}_{\tau_k} \right). \quad (16.3.30)$$

We will now estimate the function u at time τ_k , $k \in \{0, 1, \dots, N\}$, using a backward numerical technique. The basic idea is as follows: From (16.3.30) we

can evaluate $u(\tau_k, \mathbf{Y}_k^\Delta)$ if we know or have previously estimated $u(\tau_{k+1}, \mathbf{Y}_{k+1}^\Delta)$. The functional h determines the values of u at time $\tau_N = T$. Consequently, we can evaluate u at the earlier times $\tau_{N-1}, \tau_{N-2}, \dots, \tau_0$.

In practice, we cannot evaluate u at all nodes of the tree or lattice formed from the two-point random variables $\Delta\hat{W}_k^j$, $j \in \{1, 2, \dots, m\}$, $k \in \{0, 1, \dots, N-1\}$. For example, if $N = 256$ and $m = 2$, then the total number of outcomes or paths for the increments $\Delta\hat{W}_k^j$ equals $2^{512} > 10^{153}$. However, we can estimate u at certain points and use interpolation techniques to compute values between the points. For $n \geq 2$ multi-dimensional interpolation methods are needed. We may also require these interpolation procedures to produce $C^{1,2}$ smooth functions, if these estimates of u are to be used together with some other variance reduction techniques.

This method for calculating u using neighboring points obtained from the two-point random variables $\Delta\hat{W}_k^j$ at the next time step is related to a more general Markov chain approximation technique proposed by [Platen \(1992\)](#), which will be presented in the next chapter.

Suppose k and \mathbf{Y}_k^Δ are fixed with $k \in \{0, 1, \dots, N-1\}$. Let $(\Delta\hat{W}_k^1, \dots, \Delta\hat{W}_k^m)^\top$ and $Y_{k+1}^{\Delta,i}(\Delta\hat{W}_k)$ be the value of $Y_{k+1}^{\Delta,i}$ corresponding to the outcomes for $\Delta\hat{W}_k$ and the i th component of \mathbf{Y}^Δ . By (16.3.30) and the property, obtained from (16.3.5), that the vector of increments $\Delta\hat{W}_k$ takes values in the set $Q = \{-\sqrt{\Delta}, \sqrt{\Delta}\}^{\{1, 2, \dots, m\}}$ we have

$$u(\tau_k, \mathbf{Y}_k^\Delta) = \frac{1}{2^m} \sum_{\Delta\hat{W}_k \in Q} u\left(\tau_{k+1}, \mathbf{Y}_{k+1}^\Delta(\Delta\hat{W}_k)\right). \quad (16.3.31)$$

We can now attempt to compute the integrands in an approximate representation of $M_{k+1}^\Delta = u(\tau_{k+1}, \mathbf{Y}_{k+1}^\Delta)$ of the form

$$u(\tau_{k+1}, \mathbf{Y}_{k+1}^\Delta) \approx u(\tau_k, \mathbf{Y}_k^\Delta) + \sum_{j=1}^m \xi_k^j \Delta\hat{W}_k^j. \quad (16.3.32)$$

This representation can be obtained from the Wagner-Platen expansion, see Chap. 4, or the Euler approximation (16.3.2) with $u(\tau_k, \mathbf{Y}_k^\Delta)$ replacing M_k^Δ . In the case $m \geq 2$ we may not be able to find processes ξ_k^j , which solve (16.3.32) exactly. This is because the vector of increments $\Delta\hat{W}_k$ has 2^m outcomes, hence equation (16.3.32), involves solving a system of 2^m linear equations with m unknown variables ξ_k^1, \dots, ξ_k^m . In fact, the discrete framework here leads to a form of incompleteness and several possible choices for ξ_k^j . For a discussion on these issues the interested reader is referred to [Hofmann, Platen & Schweizer \(1992\)](#). Using real world pricing a natural choice for ξ_k^1, \dots, ξ_k^m results, where the hedgable part of the claim is valued by its minimal possible price and the benchmarked unhedgeable part has minimal variance.

If we use the criterion of minimization of squares

$$\left(u\left(\tau_{k+1}, \mathbf{Y}_{k+1}^{\Delta}\right) - u\left(\tau_k, \mathbf{Y}_k^{\Delta}\right) - \sum_{j=1}^m \xi_k^j \Delta \hat{W}_k^j \right)^2,$$

then we can find the corresponding optimal vector of coefficients $(\xi_k^1, \dots, \xi_k^m)^\top$ using standard least-squares analysis. We remark that, as in the case for the valuation function u , we can estimate the integrands ξ_k^j , $j \in \{1, 2, \dots, m\}$, $k \in \{0, 1, \dots, N-1\}$, at certain points and use multi-dimensional interpolation methods to determine the values at intermediate points. This provides an effective mechanism for approximating the integrand $\boldsymbol{\xi} = (\xi^1, \dots, \xi^m)^\top$ for a general valuation martingale M with representation (16.2.1). It is then of the form

$$M_t = E\left(h\left(\mathbf{X}_T^{t_0, \underline{x}}\right) \mid \mathcal{A}_t\right),$$

for $t \in [t_0, T]$, where $\mathbf{X}^{t_0, \underline{x}}$ is a d -dimensional diffusion process which satisfies (15.1.1). Obviously, the above discrete-time methodology can be generalized to the case with jumps.

16.4 Control Variates

In this section we consider the application of control variate methods and introduce a class of Itô integral control variates which extend and provide a continuous time formulation of popular martingale variates proposed, for instance, by Clewlow & Carverhill (1992, 1994). We show how the variance of linear combinations of these control variates can be minimized using least-squares analysis and orthogonalization techniques. Of course, this methodology can be easily applied to the case with jumps.

Basic Control Variate Method

We consider again a general d -dimensional diffusion process $\mathbf{X}^{t_0, \underline{x}} = \{\mathbf{X}_t^{t_0, \underline{x}} = (X_t^{1, t_0, \underline{x}}, \dots, X_t^{d, t_0, \underline{x}})^\top, t \in [t_0, T]\}$ with initial value $\underline{x} = (\underline{x}_1, \dots, \underline{x}_d)^\top \in \mathbb{R}^d$ satisfying (15.1.1). We assume, as given by (15.1.11) with $h = \bar{h}$ and $\tau = T$, that for a payoff functional $h : \mathbb{R}^d \rightarrow \mathbb{R}$ there is a corresponding valuation martingale M and function $u : ([t_0, T], \mathbb{R}^d) \rightarrow \mathbb{R}$ with

$$M_t = u\left(t, \mathbf{X}_t^{t_0, \underline{x}}\right) = E\left(h\left(\mathbf{X}_T^{t_0, \underline{x}}\right) \mid \mathcal{A}_t\right) \quad (16.4.1)$$

for $t \in [t_0, T]$. Our aim will be to find control variate formulations that will enable us to construct an accurate and fast estimate of $E(h(\mathbf{X}_T^{t_0, \underline{x}})) = u(t_0, \underline{x})$.

The *classical control variate method*, see for example Ross (1990) and Law & Kelton (1991), is based on finding a random variable Y with known mean $E(Y)$ and estimating $E(h(\mathbf{X}_T^{t_0, \underline{x}}))$ by using the unbiased estimator

$$Z = h(\mathbf{X}_T^{t_0, \underline{x}}) - \alpha(Y - E(Y))$$

for some suitable choice of $\alpha \in \Re$, rather than $h(\mathbf{X}_T^{t_0, \underline{x}})$ directly. The parameter α is chosen to minimize the variance of Z . Because

$$E(Z) = E(h(\mathbf{X}_T^{t_0, \underline{x}})),$$

Z will be an unbiased estimator for $E(h(\mathbf{X}_T^{t_0, \underline{x}}))$. We assume, of course, that both $h(\mathbf{X}_T^{t_0, \underline{x}})$ and Y can be evaluated for any realization $\omega \in \Omega$. With this type of formulation the random variable Y is called a *control variate* for the estimation of $E(\mathbf{X}_T^{t_0, \underline{x}})$.

The *basic control variate method* is simple but powerful. Suppose $\hat{\mathbf{X}}_T^{t_0, \underline{x}} = \{\mathbf{X}_t^{t_0, \underline{x}} = (\hat{X}_t^{1, t_0, \underline{x}}, \dots, \hat{X}_t^{d, t_0, \underline{x}})^\top, t \in [t_0, T]\}$ is some d -dimensional diffusion process which also satisfies an equation of the form (15.1.1) and which approximates $\mathbf{X}_T^{t_0, \underline{x}}$. The valuation function $\hat{u} : [t_0, T] \times \Re^d \rightarrow \Re$ for $h(\hat{\mathbf{X}}_T^{t_0, \underline{x}})$ is assumed to be known and satisfies

$$\hat{u}(t, \hat{\mathbf{X}}_t^{t_0, \underline{x}}) = E\left(h\left(\hat{\mathbf{X}}_T^{t_0, \underline{x}}\right) \mid \mathcal{A}_t\right) \quad (16.4.2)$$

for $t \in [t_0, T]$. Here h is the same payoff functional as used in (16.4.1). Then the random variable

$$\begin{aligned} \hat{Z}_T &= h\left(\mathbf{X}_T^{t_0, \underline{x}}\right) - \alpha\left(h\left(\hat{\mathbf{X}}_T^{t_0, \underline{x}}\right) - E\left(h\left(\hat{\mathbf{X}}_T^{t_0, \underline{x}}\right)\right)\right) \\ &= u\left(T, \mathbf{X}_T^{t_0, \underline{x}}\right) - \alpha\left(\hat{u}\left(T, \hat{\mathbf{X}}_T^{t_0, \underline{x}}\right) - \hat{u}(t_0, \underline{x})\right) \end{aligned} \quad (16.4.3)$$

will be an unbiased estimator for $E(h(\mathbf{X}_T^{t_0, \underline{x}}))$ and will usually be a variance reduced estimator if $\hat{\mathbf{X}}_T^{t_0, \underline{x}}$ is close to $\mathbf{X}_T^{t_0, \underline{x}}$. Since

$$\begin{aligned} \text{Var}(\hat{Z}_T) &= \text{Var}\left(h\left(\mathbf{X}_T^{t_0, \underline{x}}\right)\right) + \alpha^2 \text{Var}\left(h\left(\hat{\mathbf{X}}_T^{t_0, \underline{x}}\right)\right) \\ &\quad - 2\alpha \text{Cov}\left(h\left(\mathbf{X}_T^{t_0, \underline{x}}\right), h\left(\hat{\mathbf{X}}_T^{t_0, \underline{x}}\right)\right) \end{aligned}$$

the value of α , which will minimize the variance of \hat{Z}_T , is

$$\alpha_{\min} = \frac{\text{Cov}\left(h\left(\mathbf{X}_T^{t_0, \underline{x}}\right), h\left(\hat{\mathbf{X}}_T^{t_0, \underline{x}}\right)\right)}{\text{Var}\left(h\left(\hat{\mathbf{X}}_T^{t_0, \underline{x}}\right)\right)}. \quad (16.4.4)$$

Note that if a Monte Carlo simulation using the variate \hat{Z}_T is being performed, then we can estimate simultaneously the best value of α that will minimize the variance of \hat{Z}_T , as given by (16.4.4). This type of calculation can be performed as the simulation proceeds, see Clewlow & Carverhill (1992, 1994). We do not need to store the values of the variates $h(\mathbf{X}_T^{t_0, \underline{x}})$ and $h(\hat{\mathbf{X}}_T^{t_0, \underline{x}})$ for each path or realization of the simulation. This is a simple but useful form of least-squares analysis which can be extended in a straightforward manner to include linear combinations of control variates of the form $\alpha_1 Y_1 + \dots + \alpha_n Y_n$.

Stochastic Volatility Example

As an example of this method we consider a stochastic volatility model with a vector process $\mathbf{X} = \{\mathbf{X}_t = (S_t, \sigma_t)^\top, t \in [t_0, T]\}$ that satisfies the two-dimensional SDE

$$\begin{aligned} dS_t &= \sigma_t S_t dW_t^1 \\ d\sigma_t &= (\kappa - \sigma_t) dt + \xi \sigma_t dW_t^2 \end{aligned} \quad (16.4.5)$$

for $t \in [t_0, T]$ with initial values $S_{t_0} = s$, $\sigma_{t_0} = \sigma$. We take $(W^1, W^2)^\top$ to be a two-dimensional Wiener process defined on the probability space $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$, where we consider the European call payoff

$$h(S_T, \sigma_T) = (S_T - K)^+$$

for some constant strike $K > 0$ and maturity T . These type of models have been considered, for instance, by [Hofmann et al. \(1992\)](#) and [Heston \(1993\)](#).

Let $\hat{\mathbf{X}} = \{\hat{\mathbf{X}}_t = (\hat{S}_t, \hat{\sigma}_t), t \in [t_0, T]\}$ be an adjusted price process, which evolves according to the equation

$$\begin{aligned} d\hat{S}_t &= \hat{\sigma}_t \hat{S}_t dW_t^1 \\ d\hat{\sigma}_t &= (\kappa - \hat{\sigma}_t) dt \end{aligned} \quad (16.4.6)$$

for $t \in [t_0, T]$ with the same initial conditions. That is, we have $\hat{S}_{t_0} = s$ and $\hat{\sigma}_{t_0} = \sigma$. The adjusted valuation function $\hat{u}(\cdot, \cdot)$, corresponding to the payoff structure

$$h(\hat{S}_T, \hat{\sigma}_T) = (\hat{S}_T - K)^+,$$

is assumed to be given by

$$\hat{u}(t, \hat{S}_t, \hat{\sigma}_t) = E \left((\hat{S}_T - K)^+ \mid \mathcal{A}_t \right) \quad (16.4.7)$$

for $t \in [t_0, T]$. The above conditional expectation (16.4.7) can be computed explicitly for the vector process $\hat{\mathbf{X}}$. This result shows that the variate

$$\begin{aligned} \hat{Z}_T &= (S_T - K)^+ - \alpha((\hat{S}_T - K)^+ - E(\hat{S}_T - K)^+) \\ &= (S_T - K)^+ - \alpha((\hat{S}_T - K)^+ - \hat{u}(t_0, s, \sigma)) \end{aligned}$$

has outcomes that are easily computed. For this particular valuation problem this control variate estimator constitutes a powerful unbiased variance reduced estimator for $E((S_T - K)^+)$.

General Control Variates

The above control variate technique can be extended to include approximations $\tilde{h} : \mathbb{R}^d \rightarrow \mathbb{R}$ of the payoff function h . Thus, if the valuation function $\tilde{u} : [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ given by

$$\tilde{u}\left(t, \hat{\mathbf{X}}_t^{t_0, \underline{x}}\right) = E\left(\tilde{h}\left(\hat{\mathbf{X}}_T^{t_0, \underline{x}}\right) \mid \mathcal{A}_t\right) \quad (16.4.8)$$

for $t \in [t_0, T]$ is known, then a control variate can be constructed from $\tilde{h}(\hat{\mathbf{X}}_T^{t_0, \underline{x}})$ in a similar manner to that outlined for $h(\hat{\mathbf{X}}_T^{t_0, \underline{x}})$. That is, we use the unbiased estimator $h(\mathbf{X}_T^{t_0, \underline{x}}) - \alpha(\tilde{h}(\hat{\mathbf{X}}_T^{t_0, \underline{x}}) - E(\tilde{h}(\hat{\mathbf{X}}_T^{t_0, \underline{x}})))$, as is suggested by (16.4.3). We assume that the expectation $E(\tilde{h}(\hat{\mathbf{X}}_T^{t_0, \underline{x}}))$ is known explicitly or can be accurately approximated. Consequently, this control variate method extends to include changes both in the underlying diffusion $\mathbf{X}^{t_0, \underline{x}}$ and the payoff function h .

Using the martingale property of the process $\tilde{M} = \{\tilde{u}(t, \hat{\mathbf{X}}_t^{t_0, \underline{x}}), t \in [t_0, T]\}$, as given by (15.4.5), we can write, subject to certain integrability conditions applying for \tilde{M} , the martingale

$$\tilde{M}_t = \tilde{u}\left(t, \hat{\mathbf{X}}_t^{t_0, \underline{x}}\right) = \tilde{u}(t_0, \underline{x}) + \sum_{j=1}^m \int_{t_0}^t \tilde{\xi}_s^j dW_s^j, \quad (16.4.9)$$

where $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}^1, \dots, \tilde{\xi}^m)^\top$ is a vector of \mathcal{A} -predictable processes satisfying (16.2.2).

The following formulation is useful for certain types of path dependent options, for example, American options or barrier options with stochastic volatility. Let $\tau : \Omega \rightarrow \mathfrak{R}$ be a stopping time as given by (15.1.2) with continuation region Γ_0 and exercise boundary Γ_1 . We take $h : \Gamma_1 \rightarrow \mathfrak{R}$ to be some payoff function and we aim to approximate the expectation $E(h(\mathbf{X}_\tau^{t_0, \underline{x}}))$ assuming \tilde{u} to be known. Using the Itô formula we have from (16.4.9) the relation

$$\tilde{M}_\tau = \tilde{u}\left(\tau, \tilde{\mathbf{X}}_\tau^{t_0, \underline{x}}\right) = \tilde{u}(t_0, \underline{x}) + \sum_{j=1}^m \int_{t_0}^\tau \tilde{\xi}_s^j dW_s^j. \quad (16.4.10)$$

This means that

$$E\left(\tilde{u}\left(\tau, \tilde{\mathbf{X}}_\tau^{t_0, \underline{x}}\right)\right) = \tilde{u}(t_0, \underline{x})$$

therefore, the random variable

$$\tilde{Z}_\tau = h\left(\mathbf{X}_\tau^{t_0, \underline{x}}\right) - \alpha\left(\tilde{u}\left(\tau, \tilde{\mathbf{X}}_\tau^{t_0, \underline{x}}\right) - \tilde{u}(t_0, \underline{x})\right) \quad (16.4.11)$$

is an unbiased estimator for $E(h(\mathbf{X}_\tau^{t_0, \underline{x}}))$. Note that \tilde{u} can be any valuation function of the form (16.4.9). For example, it may correspond to a European style security with

$$M_t = E\left(h'\left(\mathbf{X}_T^{t_0, \underline{x}}\right) \mid \mathcal{A}_t\right)$$

for $t \in [t_0, T]$, for some payoff function h' and may have no direct relationship with the stopping time τ , which is obtained for a different problem.

Discrete-Time Control Variates

The above control variate formulations can be conveniently applied in a discrete-time setting. For example, if we intend to use the control variate given by (16.4.3), we may replace \mathbf{X} and $\hat{\mathbf{X}}$ by discrete-time approximations \mathbf{Y}^Δ and $\hat{\mathbf{Y}}^\Delta$, respectively, say of the form (16.3.27). In fact, there is often more flexibility with discrete-time formulations. If, for instance, there is no natural choice for the process $\hat{\mathbf{X}}$ in (16.4.3), then a control variate can be obtained from \mathbf{Y}^Δ itself, since $E(\mathbf{Y}^\Delta)$ can usually be easily calculated for discrete-time Monte Carlo schemes.

For one-dimensional diffusions Clewlow & Carverhill (1992, 1994) used linear combinations of discrete-time martingale control variates. If we use a time discretization of the interval $[t_0, T]$, as given by (16.3.1), then these control variates in several dimensions take the form

$$C_N^{\Delta,i} = \sum_{k=0}^{N-1} \phi_k^i (\Delta Y_k^{\Delta,i} - E(\Delta Y_k^{\Delta,i}))$$

for $i \in \{1, 2, \dots, d\}$, $k \in \{0, 1, \dots, N-1\}$, where $\Delta Y_k^{\Delta} = \mathbf{Y}_{k+1}^\Delta - \mathbf{Y}_k^\Delta$. Here ϕ_k^i is \mathcal{A}_{t_k} -measurable and is chosen as an approximation to a hedging parameter such as the delta or gamma sensitivities for the i th component of the diffusion process \mathbf{X} given by (15.1.1). Since

$$E(C_k^\Delta | \mathcal{A}_{t_\ell}) = C_\ell^\Delta$$

for $\ell, k \in \{0, 1, \dots, N-1\}$, with $\ell \leq k$, C^Δ is a discrete-time martingale. The expectation $E(\Delta Y_k^{\Delta,i})$, $i \in \{1, 2, \dots, d\}$, $k \in \{0, 1, \dots, N-1\}$, can usually be calculated in a straightforward manner. For example, with the Euler scheme (16.3.27)

$$E(\Delta Y_k^{\Delta,i}) = a^i(t_k, \mathbf{Y}_k^\Delta) \Delta.$$

If we consider the SDE (15.1.1), which defines our underlying diffusion process, a continuous time version of the above martingale control variate would be

$$\begin{aligned} C_T^i &= \int_{t_0}^T \psi_s^i (dX_s^{i,t_0,\underline{x}} - a^i(s, \mathbf{X}_s^{t_0,\underline{x}}) ds) \\ &= \int_{t_0}^T \psi_s^i \sum_{j=1}^m b^{ij}(s, \mathbf{X}_s^{t_0,\underline{x}}) dW_s^j \end{aligned} \tag{16.4.12}$$

for $i \in \{1, 2, \dots, d\}$, where $\psi = (\psi^1, \dots, \psi^d)^\top$ is some vector of \mathcal{A} -predictable processes, which satisfy appropriate integrability conditions. This representation shows that the control variate C_T^i is expressible as an Itô integral such that its expectation is zero. It also follows that another natural choice for a control variate is simply an Itô integral of the form

$$\bar{C}_T = \sum_{j=1}^m \int_{t_0}^T \bar{\xi}_s^j dW_s^j, \quad (16.4.13)$$

where $\bar{\xi} = (\bar{\xi}^1, \dots, \bar{\xi}^d)^\top$ is a vector of \mathcal{A} -predictable processes. We emphasize that there are, in principle, no limits in designing control variates.

Variance of Control Variates

From the representation (16.2.1) we can now compute the variance under P of the unbiased estimator \bar{Z}_T of $E(h(\mathbf{X}_T^{t_0, \underline{x}}))$ given by

$$\bar{Z}_T = h\left(\mathbf{X}_T^{t_0, \underline{x}}\right) - \alpha \bar{C}_T, \quad (16.4.14)$$

as

$$\begin{aligned} \text{Var}(\bar{Z}_T) &= \text{Var}\left(\sum_{j=1}^m \int_{t_0}^T (\xi_s^j - \alpha \bar{\xi}_s^j) dW_s^j\right) \\ &= E\left[\left(\sum_{j=1}^m \int_{t_0}^T (\xi_s^j - \alpha \bar{\xi}_s^j) dW_s^j\right)^2\right] \\ &= \sum_{j=1}^m \int_{t_0}^T E(\xi_s^j - \alpha \bar{\xi}_s^j)^2 ds. \end{aligned} \quad (16.4.15)$$

This shows that if we can find good approximations $\bar{\xi}^j$, $j \in \{1, 2, \dots, m\}$, for the integrands ξ^j , then with $\alpha = 1$ the variance of the unbiased estimator \bar{Z}_T will be small.

Note that to produce an arbitrarily small variance we require only an approximation to the integrand $\xi = (\xi^1, \dots, \xi^m)^\top$, which is related to the deltas of the underlying security. The use of control variates based on gammas or other greeks is, therefore, in a theoretical sense not required.

Suppose the valuation function u and its approximation \hat{u} , as given by (16.4.1) and (16.4.2), respectively, are of class $C^{1,2}$. Then expanding both $u(T, \mathbf{X}_T^{t_0, \underline{x}})$ and $\hat{u}(T, \hat{\mathbf{X}}_T^{t_0, \underline{x}})$ using Itô's formula and applying the Kolmogorov backward equation we obtain

$$u\left(T, \mathbf{X}_T^{t_0, \underline{x}}\right) = u(t_0, \underline{x}) + \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^T b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) \frac{\partial}{\partial x_i} u(s, \mathbf{X}_s^{t_0, \underline{x}}) dW_s^j$$

and

$$\hat{u}\left(T, \hat{\mathbf{X}}_T^{t_0, \underline{x}}\right) = \hat{u}(t_0, \underline{x}) + \sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^T b^{i,j}(s, \hat{\mathbf{X}}_s^{t_0, \underline{x}}) \frac{\partial}{\partial x_i} \hat{u}(s, \hat{\mathbf{X}}_s^{t_0, \underline{x}}) dW_s^j. \quad (16.4.16)$$

The variance of the estimator \hat{Z}_T given by (16.4.3) is, therefore,

$$\begin{aligned} \text{Var}(\hat{Z}_T) &= \text{Var} \left(\sum_{i=1}^d \sum_{j=1}^m \int_{t_0}^T \left(b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) \frac{\partial}{\partial x_i} u(s, \mathbf{X}_s^{t_0, \underline{x}}) \right. \right. \\ &\quad \left. \left. - \alpha b^{i,j}(s, \hat{\mathbf{X}}_s^{t_0, \underline{x}}) \frac{\partial}{\partial x_i} \hat{u}(s, \hat{\mathbf{X}}_s^{t_0, \underline{x}}) \right) dW_s^j \right) \\ &= \sum_{j=1}^m \int_{t_0}^T E \left(\sum_{i=1}^d \left(b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) \frac{\partial}{\partial x_i} u(s, \mathbf{X}_s^{t_0, \underline{x}}) \right. \right. \\ &\quad \left. \left. - \alpha b^{i,j}(s, \hat{\mathbf{X}}_s^{t_0, \underline{x}}) \frac{\partial}{\partial x_i} \hat{u}(s, \hat{\mathbf{X}}_s^{t_0, \underline{x}}) \right) \right)^2 ds. \quad (16.4.17) \end{aligned}$$

If we set

$$\bar{\xi}_s^j = \sum_{i=1}^d b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) \frac{\partial}{\partial x_i} \hat{u}(s, \mathbf{X}_s^{t_0, \underline{x}}), \quad (16.4.18)$$

as used in (16.4.13) for $j \in \{1, 2, \dots, m\}$, then the variance of the estimator \bar{Z}_T given by (16.4.14) can be calculated from (16.4.15). Thus, one obtains

$$\text{Var}(\bar{Z}_T) = \sum_{j=1}^m \int_{t_0}^T E \left(\sum_{i=1}^d b^{i,j}(s, \mathbf{X}_s^{t_0, \underline{x}}) \frac{\partial}{\partial x_i} (u - \alpha \hat{u})(s, \mathbf{X}_s^{t_0, \underline{x}}) \right)^2 ds. \quad (16.4.19)$$

We observe that the variance obtained for \hat{Z}_T is similar to, but different from that obtained for \bar{Z}_T with the above choice for $\bar{\xi} = (\bar{\xi}^1, \dots, \bar{\xi}^m)^\top$. In general, we can expect a lower variance for \bar{Z}_T as can be seen by comparing (16.4.17) with (16.4.19). The computational loads corresponding to the two estimators can also vary. For example, Monte Carlo estimations using the variate \hat{Z}_T involve only the calculation of payoff structures but require the simulation of two separate diffusion processes $\mathbf{X}^{t_0, \underline{x}}$ and $\hat{\mathbf{X}}^{t_0, \underline{x}}$. This can be compared with estimates of \bar{Z}_T , which require simulation of only one explicit diffusion process $\mathbf{X}^{t_0, \underline{x}}$. However, evaluation of the Itô integrals (16.4.13) effectively involves the simulation of another diffusion process, namely the Itô integral itself. Choosing the best estimator for a given valuation problem often involves preliminary simulation studies and clearly depends on the specific structure of the underlying model and payoff functional.

Minimizing the Variance

We can extend the above analysis, for the estimator \bar{Z}_T to include linear combinations of control variates $\sum_{\ell=1}^L \alpha_\ell \bar{C}_T^\ell$, where each \bar{C}_T^ℓ , $\ell \in \{1, 2, \dots, L\}$, is an Itô integral of the form (16.4.13). Using these Itô integrals the random variable

$$\bar{Z}_T = h\left(\mathbf{X}_T^{t_0, \underline{x}}\right) - \sum_{\ell=1}^L \alpha_\ell \bar{C}_T^\ell \quad (16.4.20)$$

will be an unbiased estimator for $E(h(\mathbf{X}_T^{t_0, \underline{x}}))$ with variance

$$\begin{aligned} \text{Var}(\bar{Z}_T) &= \text{Var}\left(h\left(\mathbf{X}_T^{t_0, \underline{x}}\right)\right) + \sum_{\ell, k=1}^L \alpha_\ell \alpha_k \text{Cov}(\bar{C}_T^\ell, \bar{C}_T^k) \\ &\quad - 2 \sum_{\ell=1}^L \alpha_\ell \text{Cov}\left(h\left(\mathbf{X}_T^{t_0, \underline{x}}\right), \bar{C}_T^\ell\right). \end{aligned} \quad (16.4.21)$$

This variance will be minimized if

$$\sum_{k=1}^L \alpha_k \text{Cov}(\bar{C}_T^\ell, \bar{C}_T^k) = \text{Cov}\left(h\left(\mathbf{X}_T^{t_0, \underline{x}}\right), \bar{C}_T^\ell\right) \quad (16.4.22)$$

for each $\ell \in \{1, 2, \dots, L\}$. This system of linear equations will admit a unique solution if the set of control variates $\{\bar{C}_T^\ell, \ell \in \{1, 2, \dots, L\}\}$ is linearly independent. As is the case for a single control variate, we can estimate the quantities $\text{Cov}(\bar{C}_T^\ell, \bar{C}_T^k)$ and $\text{Cov}(h(\mathbf{X}_T^{t_0, \underline{x}}), \bar{C}_T^\ell)$, $\ell, k \in \{1, 2, \dots, L\}$, for a given simulation and simultaneously the optimal vector of coefficients $\underline{\alpha} = (\alpha_1, \dots, \alpha_L)^\top$ to minimize the variance as given by (16.4.22). Also, as has been noted for a single control variate, we can calculate the optimal vector of coefficients $\underline{\alpha}$ progressively during the simulation. We do not need to store output data for the variates $\mathbf{X}_T^{t_0, \underline{x}}, \bar{C}_T^1, \dots, \bar{C}_T^L$ for each path of the simulation.

This formulation is simplified if we assume that $\text{Cov}(\bar{C}_T^\ell, \bar{C}_T^k) = 0$ for $\ell \neq k$, $\ell, k \in \{1, 2, \dots, L\}$. That is, the random variables \bar{C}_T^ℓ , $\ell \in \{1, 2, \dots, L\}$, are mutually orthogonal when considered as elements of the Hilbert space $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$ with inner product $(X, Y) = E(XY)$ for $X, Y \in L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$. In this case (16.4.22) reduces to

$$\begin{aligned} \alpha_\ell &= \frac{\text{Cov}\left(h\left(\mathbf{X}_T^{t_0, \underline{x}}\right), \bar{C}_T^\ell\right)}{\text{Var}(\bar{C}_T^\ell)} \\ &= \frac{E\left(\left(h\left(\mathbf{X}_T^{t_0, \underline{x}}\right) - E\left(h\left(\mathbf{X}_T^{t_0, \underline{x}}\right)\right)\right) \bar{C}_T^\ell\right)}{\|\bar{C}_T^\ell\|_2^2} \end{aligned} \quad (16.4.23)$$

for $\ell \in \{1, 2, \dots, L\}$, where $\|\cdot\|_2$ denotes the norm in $L^2(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$.

A mutually orthogonal set of control variates \bar{C}_T^ℓ , $\ell \in \{1, 2, \dots, L\}$, can be constructed using Gram-Schmidt orthogonalization. For example, if \bar{C}_T^1 and \bar{C}_T^2 are not orthogonal we can replace \bar{C}_T^2 with

$$\bar{C}_T^2 - \frac{\text{Cov}(\bar{C}_T^1, \bar{C}_T^2) \bar{C}_T^1}{\text{Var}(\bar{C}_T^1)}.$$

There are many ways of constructing the control variate process $\bar{\mathbf{C}} = (\bar{C}^1, \dots, \bar{C}^L)^\top$. If we have an equidistant time discretization of the interval $[t_0, T]$ of the form (16.3.1), then a good choice for $\bar{\mathbf{C}}$ using control variates of the form (16.4.13), for certain types of valuation problems, is given by

$$\bar{C}_T^\ell = \sum_{j=1}^m \int_{t_0}^T \mathbf{1}_{\{s \in [t_{\ell-1}, t_\ell]\}} \bar{\xi}_s^i dW_s^i \quad (16.4.24)$$

for $\ell \in \{1, 2, \dots, L\}$. For $\ell_1 \neq \ell_2$ it follows

$$\begin{aligned} (\bar{C}_T^{\ell_1}, \bar{C}_T^{\ell_2}) &= \sum_{j=1}^m \int_{t_0}^T E \left(\mathbf{1}_{\{s \in [t_{\ell_1-1}, t_{\ell_1}]\}} \mathbf{1}_{\{s \in [t_{\ell_2-1}, t_{\ell_2}]\}} (\bar{\xi}_s^i)^2 \right) ds \\ &= 0. \end{aligned}$$

Therefore, any two control variates $\bar{C}_T^{\ell_1}$ and $\bar{C}_T^{\ell_2}$, $\ell_1 \neq \ell_2$, are orthogonal.

These control variate methods can be applied iteratively or in combination with other techniques. For example, we may find an initial approximation $u_1 : [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ for u from finite difference methods. The integrands given by (16.4.18) together with appropriate interpolation routines could then be used to construct an Itô integral control variate of the form (16.4.13) and from this a new approximation u_2 for u would be obtained. Clearly, this procedure can be repeated until sufficient accuracy is achieved.

The control variate method is an efficient method which can deliver significant variance reductions. However, its potential is limited since it does not exploit fully the stochastic structure of the underlying problem. The generalization of the above results to the case with jumps is rather obvious.

16.5 Heath-Platen Variance Reduction Method

As previously pointed out, raw Monte Carlo simulation can often be significantly improved with the addition of appropriate variance reduction techniques. In this section a powerful variance reduction technique is presented. The method is based directly on the stochastic analytic structure of the underlying dynamics and can be used to find systematically highly efficient unbiased variance reduced estimators. The method presented has wide applicability. We illustrate how the method can be applied by considering the pricing of European style derivative securities for a class of stochastic volatility models. The case with jumps can be handled analogously.

The technique described in this section appeared in [Heath & Platen \(2002d\)](#) and is here called *Heath-Platen (HP) method*. It is related to the measure transformation and Itô integral representation methods that we presented in previous sections. All mentioned methods, in their basic form, provide a reasonable level of variance reduction. However, the approach described here

is superior in that it can be applied to a wider range of valuation problems and can benefit flexibly from any extra knowledge about the underlying dynamics. It can also be used to develop iterative approximation methods and is easily adapted to, or employed in conjunction with, finite difference methods, which we will discuss towards the end of the next chapter.

HP Estimator

First, we formulate the main theoretical result of the section. Let $T \in (0, \infty)$ be fixed and Γ be an open connected subset of \Re^d . Consider a general d -dimensional diffusion process $\mathbf{X}^{s,x} = \{\mathbf{X}_t^{s,x}, t \in [s, T]\}$, which satisfies the SDE

$$d\mathbf{X}_t^{s,x} = \mathbf{a}(t, \mathbf{X}_t^{s,x}) dt + \sum_{j=1}^m \mathbf{b}^j(t, \mathbf{X}_t^{s,x}) dW_t^j \quad (16.5.1)$$

for $t \in [s, T]$ with initial value $\mathbf{X}_s^{s,x} = \mathbf{x} \in \Gamma$ for $s \in [0, T]$.

Here the drift coefficient $\mathbf{a} : [0, T] \times \Gamma \rightarrow \Re^d$ and diffusion coefficient $\mathbf{b}^j : [0, T] \times \Gamma \rightarrow \Re^d$, $j \in \{1, 2, \dots, m\}$, satisfy appropriate conditions so that (16.5.1) admits a unique strong solution and is Markovian. The vector process $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \in [0, T]\}$ is an m -dimensional Wiener process defined on the filtered probability space $(\Omega, \mathcal{A}_T, \underline{\mathcal{A}}, P)$, where $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0, T]}$ fulfills the usual conditions and $\mathcal{A}_0 = \{\phi, \Omega\}$ is the trivial sigma-algebra.

Denote by $\tau : \Omega \rightarrow [s, T]$ the first exit time of $(t, \mathbf{X}_t^{s,x})$ from $[s, T] \times \Gamma$, that is

$$\tau = \inf\{t \geq s : (t, \mathbf{X}_t^{s,x}) \notin [s, T] \times \Gamma\}. \quad (16.5.2)$$

We consider here a stopping time formulation, similar to that in Chap. 15, to ensure that our results can be applied to a range of path dependent options.

Define the operators L^0 and L^j , $j \in \{1, 2, \dots, m\}$, on a sufficiently smooth function $f : [0, T] \times \Gamma \rightarrow \Re$ by

$$\begin{aligned} L^0 f(t, \mathbf{x}) &= \frac{\partial f(t, \mathbf{x})}{\partial t} + \sum_{i=1}^d a^i(t, \mathbf{x}) \frac{\partial f(t, \mathbf{x})}{\partial x^i} \\ &+ \frac{1}{2} \sum_{i,k=1}^d \sum_{j=1}^m b^{i,j}(t, \mathbf{x}) b^{k,j}(t, \mathbf{x}) \frac{\partial^2 f(t, \mathbf{x})}{\partial x^i \partial x^k} \end{aligned} \quad (16.5.3)$$

and

$$L^j f(t, \mathbf{x}) = \sum_{i=1}^d b^{i,j}(t, \mathbf{x}) \frac{\partial f(t, \mathbf{x})}{\partial x^i} \quad (16.5.4)$$

for $(t, \mathbf{x}) \in (0, T) \times \Gamma$.

Let $\partial\Gamma$ denote the boundary of Γ . That is $\partial\Gamma = \{\mathbf{x} \notin \Gamma : \exists(\mathbf{x}_n)_{n \in \{1, 2, \dots\}}$ with $\mathbf{x}_n \in \Gamma$, $n \in \{1, 2, \dots\}$ and $\lim_{n \rightarrow \infty} |\mathbf{x} - \mathbf{x}_n| = 0\}$. Consider a *payoff function* $h : B \rightarrow \Re$, where

$$B = [0, T] \times \partial\Gamma \cup \{T\} \times (\Gamma \cup \partial\Gamma) \quad (16.5.5)$$

and a valuation function $u : [0, T] \times \Gamma \rightarrow \mathbb{R}$ given by

$$u(t, \mathbf{x}) = E(h(\tau, \mathbf{X}_\tau^{t,x})) \quad (16.5.6)$$

for $(t, \mathbf{x}) \in [0, T] \times \Gamma$. We assume that h satisfies appropriate integrability conditions so that the process $M = \{M_t, t \in [0, T]\}$ with

$$M_t = E(h(\tau, \mathbf{X}_\tau^{0,x}) \mid \mathcal{A}_t) \quad (16.5.7)$$

for $t \in [0, T]$ is a square integrable (\mathcal{A}, P) -martingale. Under conditions, as described in Chap. 15, by using the martingale representation theorem together with the Markov property for \mathbf{X} , it can be inferred that there exists an m -dimensional \mathcal{A} -predictable integrand $\boldsymbol{\xi} = \{\boldsymbol{\xi}_t = (\xi_t^1, \dots, \xi_t^m)^\top, t \in [0, T]\}$ with

$$\begin{aligned} M_t &= u(t, \mathbf{X}_{t \wedge \tau}^{0,x}) \\ &= u(0, \mathbf{x}) + \sum_{j=1}^m \int_0^{t \wedge \tau} \xi_s^j dW_s^j \end{aligned} \quad (16.5.8)$$

for $t \in [0, T]$.

Our aim will be to find an unbiased variance reduced estimator for $u(0, \mathbf{x})$ given an approximation $\bar{u} : [0, T] \times \Gamma \rightarrow \mathbb{R}$ to u . Examples will be provided on how to construct such approximate functions. We assume that $\bar{u} \in C^{1,2}([0, T] \times \Gamma)$ is from the class of functions $f : [0, T] \times \Gamma \rightarrow \mathbb{R}$ with $\frac{\partial f}{\partial t}(\cdot, \mathbf{x})$ continuous on $(0, T)$ for $\mathbf{x} \in \Gamma$ and $\frac{\partial^2 f}{\partial x^i \partial x^k}(t, \cdot)$ continuous on Γ for $t \in (0, T)$ and $i, k \in \{1, 2, \dots, d\}$. We also assume that $L^j \bar{u}$ satisfies appropriate integrability conditions so that the process $\bar{M}^j = \{\bar{M}_t^j, t \in [0, T]\}$ with

$$\bar{M}_t^j = \int_0^{t \wedge \tau} L^j \bar{u}(s, \mathbf{X}_s^{0,x}) dW_s^j \quad (16.5.9)$$

for $t \in [0, T]$, $j \in \{1, 2, \dots, m\}$, is a square integrable (\mathcal{A}, P) -martingale. In addition, it is assumed that at maturity

$$\bar{u}(\tau, \mathbf{X}_\tau^{0,x}) = u(\tau, \mathbf{X}_\tau^{0,x}) = h(\tau, \mathbf{X}_\tau^{0,x}). \quad (16.5.10)$$

That is, \bar{u} is an approximation to u that matches the payoff for u on B . Applying the Itô formula to \bar{u} and using (16.5.1) yields

$$\bar{u}(\tau, \mathbf{X}_\tau^{0,x}) = \bar{u}(0, \mathbf{x}) + \int_0^\tau L^0 \bar{u}(t, \mathbf{X}_t^{0,x}) dt + \sum_{j=1}^m \int_0^\tau L^j \bar{u}(t, \mathbf{X}_t^{0,x}) dW_t^j. \quad (16.5.11)$$

Consequently, by (16.5.6), (16.5.10), (16.5.11), Fubini's theorem and the martingale property of \bar{M}^j one obtains

$$\begin{aligned}
u(0, \mathbf{x}) &= E(h(\tau, \mathbf{X}_\tau^{0,x})) \\
&= E(\bar{u}(\tau, \mathbf{X}_\tau^{0,x})) \\
&= \bar{u}(0, \mathbf{x}) + E\left(\int_0^\tau L^0 \bar{u}(t, \mathbf{X}_t^{0,x}) dt\right) \\
&= \bar{u}(0, \mathbf{x}) + \int_0^T E\left(\mathbf{1}_{\{t<\tau\}} L^0 \bar{u}(t, \mathbf{X}_t^{0,x})\right) dt \quad (16.5.12)
\end{aligned}$$

for $\mathbf{x} \in \Gamma$. Here $\mathbf{1}_{\{t<\tau\}}$ denotes the indicator function applied to the event $\{t < \tau\}$. Due to (16.5.12) the random variable

$$\bar{Z}_\tau = \bar{u}(0, \mathbf{x}) + \int_0^\tau L^0 \bar{u}(t, \mathbf{X}_t^{0,x}) dt \quad (16.5.13)$$

is an unbiased estimator for $u(0, \mathbf{x})$. We will refer to this estimator as the *HP estimator*.

Variance of the HP Estimator

To compute the variance of \bar{Z}_τ note that from (16.5.13), (16.5.11), (16.5.10) and (16.5.8) we have

$$\begin{aligned}
\bar{Z}_\tau &= \bar{u}(\tau, \mathbf{X}_\tau^{0,x}) - \sum_{j=1}^m \int_0^\tau L^j \bar{u}(t, \mathbf{X}_t^{0,x}) dW_t^j \\
&= u(\tau, \mathbf{X}_\tau^{0,x}) - \sum_{j=1}^m \int_0^\tau L^j \bar{u}(t, \mathbf{X}_t^{0,x}) dW_t^j \\
&= u(0, \mathbf{x}) + \sum_{j=1}^m \int_0^\tau \left(\xi_t^j - L^j \bar{u}(t, \mathbf{X}_t^{0,x})\right) dW_t^j. \quad (16.5.14)
\end{aligned}$$

Since \bar{M}^j is square integrable for each $j \in \{1, 2, \dots, m\}$ the variance of \bar{Z}_τ is, therefore, given by

$$\begin{aligned}
\text{Var}(\bar{Z}_\tau) &= E\left[\left(\sum_{j=1}^m \int_0^\tau \left(\xi_t^j - L^j \bar{u}(t, \mathbf{X}_t^{0,x})\right) dW_t^j\right)^2\right] \\
&= \sum_{j=1}^m \int_0^T E\left(\mathbf{1}_{\{t<\tau\}} \left(\xi_t^j - L^j \bar{u}(t, \mathbf{X}_t^{0,x})\right)^2\right) dt. \quad (16.5.15)
\end{aligned}$$

If the valuation function u is sufficiently smooth to permit an application of the Itô formula, then the integrands ξ^j in (16.5.8) take the form

$$\xi_t^j = L^j u(t, \mathbf{X}_t^{0,x}) \quad (16.5.16)$$

for $j \in \{1, 2, \dots, m\}$ and $t \in [0, T \wedge \tau]$. In this case the expression in (16.5.15) can be replaced by the formula

$$\text{Var}(\bar{Z}_\tau) = \sum_{j=1}^m \int_0^T E \left(\mathbf{1}_{\{t < \tau\}} \left((L^j u - L^j \bar{u})(t, \mathbf{X}_t^{0,x}) \right)^2 \right) dt. \quad (16.5.17)$$

This shows that if a good approximation \bar{u} to u can be found, so that $L^j u$ is close to $L^j \bar{u}$ for each $j \in \{1, 2, \dots, m\}$, then the variance of \bar{Z}_τ will be small. The form of (16.5.17) indicates that obtaining a small variance for \bar{Z}_τ is related to finding suitable approximations to the hedge ratio for the underlying security.

Let us now construct an additional unbiased estimator. By (16.5.6), (16.5.11), (16.5.13) and since \bar{M}^j is a square integrable martingale, it can be seen that the random variable

$$\begin{aligned} \bar{Z}_{\tau,\alpha} &= \bar{u}(\tau, \mathbf{X}_\tau^{0,x}) - \alpha \sum_{j=1}^m \int_0^\tau L^j \bar{u}(t, \mathbf{X}_t^{0,x}) dW_t^j \\ &= \bar{Z}_\tau + (1 - \alpha) \sum_{j=1}^m \int_0^\tau L^j \bar{u}(t, \mathbf{X}_t^{0,x}) dW_t^j \end{aligned} \quad (16.5.18)$$

is an unbiased estimator for $u(0, \mathbf{x})$. This variance reduced estimator can be interpreted as a control variate that is based on an Itô integral representation, see Sect. 16.4. Inspection of (16.5.18) shows that for $\alpha = 1$ the HP estimator and the above unbiased estimator are the same. This explains the relationship between the two variance reduction methods. In (16.5.18) the parameter α can be chosen in an optimal manner similar as in Sect. 16.4.

Note that we have provided precise estimates of the level of variance reduction that can be obtained, not in terms of improvements to, or specification of the order of weak convergence, but by explicitly computing the variance of the corresponding estimators, see (16.5.15) and (16.5.17). Clearly, the degree of variance reduction obtained depends on the choice of the approximation function \bar{u} . As in standard Monte Carlo methods, the reduced variance decreases with the inverse of the square root of the number of simulated trajectories. For a given sample size the absolute variance can be made drastically smaller by the above method when compared with a raw Monte Carlo estimate.

Extensions of the HP Method

We now consider applications of the HP variance reduction technique to the approximation of solutions of partial differential equations (PDEs) that are pricing functions in finance but could also originate from other applications. In addition, some extensions of the method are presented.

In this subsection the domain Γ is also assumed to be a bounded subset of \Re^d with $B \subseteq \Re^d$ as given by (16.5.5). Our task will be to find an approximation to the solution at an initial point $(0, \mathbf{x})$ with $\mathbf{x} \in \Gamma$ to the PDE

$$L^0 u(t, \mathbf{x}) = 0 \quad (16.5.19)$$

for $(t, \mathbf{x}) \in (0, T) \times \Gamma$ with boundary condition

$$u(t, \mathbf{x}) = h(t, \mathbf{x}) \quad (16.5.20)$$

for $(t, \mathbf{x}) \in B$ and some measurable function $h : B \rightarrow \Re$, where L^0 is the diffusion operator in (16.5.3). The link between this PDE formulation and the stochastic representation (16.5.6) is provided by the Feynman-Kac formula, see Sect. 2.7. Sufficient conditions ensuring that (16.5.6) is a solution to (16.5.19)–(16.5.20), are for example, provided in Sect. 2.7, see also Friedman (1975), Krylov (1980) and Heath & Schweizer (2000). In this case, for any solution u of (16.5.19)–(16.5.20), an application of the Itô formula shows that

$$u(t, \mathbf{x}) = u(\tau, \mathbf{X}_\tau^{t,x}) - \sum_{j=1}^m \int_t^\tau L^j u(s, \mathbf{X}_s^{t,x}) dW_s^j \quad (16.5.21)$$

for $(t, \mathbf{x}) \in [0, T] \times \Gamma$. The assumption that Γ is a bounded subset of \Re^d implies that

$$E \left(\sup_{x \in \Gamma} |L^j u(t, \mathbf{x})|^2 \right) < \infty$$

a.s. for $t \in [0, T]$ and $j \in \{1, 2, \dots, m\}$, therefore,

$$E \left(\sum_{j=1}^m \int_t^\tau L^j u(s, \mathbf{X}_s^{t,x}) dW_s^j \right) = 0.$$

Combining this result with (16.5.21) ensures that the representation (16.5.6) applies for u . Consequently, the solution to (16.5.19)–(16.5.20) must be unique.

Suppose there is some approximation $\bar{u} \in \mathcal{C}^{1,2}([0, T] \times \Gamma)$ with $\bar{u}(t, \mathbf{x}) = h(t, \mathbf{x})$ for $(t, \mathbf{x}) \in B$. Then the difference $u^* = u - \bar{u}$ satisfies the PDE

$$L^0 u^*(t, \mathbf{x}) + L^0 \bar{u}(t, \mathbf{x}) = 0 \quad (16.5.22)$$

with boundary condition

$$u^*(t, \mathbf{x}) = 0 \quad (16.5.23)$$

for $(t, \mathbf{x}) \in B$. If \bar{u} is close to u , then it may be computationally more efficient to solve (16.5.22)–(16.5.23) rather than (16.5.19)–(16.5.20). This provides a simple but powerful illustration of how the ideas used to produce variance reduced estimators can be directly applied in a PDE setting.

In some cases it may not be possible or convenient to use only PDE methods to solve (16.5.19)–(16.5.20). For example, the dimension d may be simply too high. A hybrid method, which combines features of a PDE approach

together with the HP variance reduction technique, would first apply a numerical solver to find an approximation \bar{u} to u . Then by simulation, using the formula (16.5.12) with estimator (16.5.13), a better approximation of the solution $u(0, \mathbf{x})$, $\mathbf{x} \in \Gamma$, at the point $(0, \mathbf{x})$, could be obtained. Typically, the approximate solution \bar{u} , possibly the output from an implicit finite difference solver, would only be available at discrete points on a grid. This may require some additional smoothing or multi-dimensional interpolation technique to ensure that values for $L^0 \bar{u}(t, \mathbf{X}_t^{0,x})$, $t \in [0, \bar{T}]$, can be computed for simulated trajectories, see (16.5.13).

Iterative HP Methods

For many practical applications that arise in a financial modeling context, a systematic way of obtaining an approximate solution \bar{u} satisfying (16.5.6) is to first find an approximation to the diffusion process \mathbf{X} . Let $\bar{\mathbf{X}} = \{\bar{\mathbf{X}}_t, t \in [s, T]\}$ be a d -dimensional diffusion process that is governed by the SDE

$$d\bar{\mathbf{X}}_t^{s,x} = \bar{\mathbf{a}}(t, \bar{\mathbf{X}}_t^{s,x}) dt + \sum_{j=1}^m \bar{\mathbf{b}}^j(t, \bar{\mathbf{X}}_t^{s,x}) dW_t^j \quad (16.5.24)$$

for $t \in [s, T]$, $s \in [0, T]$ with coefficient functions $\bar{\mathbf{a}} : [0, T] \times \Gamma \rightarrow \Re^d$ and $\bar{\mathbf{b}}^j : [0, T] \times \Gamma \rightarrow \Re^d$, $j \in \{1, 2, \dots, m\}$. As is the case for (16.5.1) it is assumed that (16.5.24) admits a unique strong solution which is Markovian. Denote by \bar{L}^0 the corresponding PDE operator, as given in (16.5.3), with $\bar{\mathbf{a}}^i$ and $\bar{\mathbf{b}}^{i,j}$, $i \in \{1, 2, \dots, d\}$, $j \in \{1, 2, \dots, m\}$ replacing \mathbf{a}^i and $\mathbf{b}^{i,j}$, respectively. Define the function $\bar{u} : [0, T] \times \Gamma \rightarrow \Re$ by replacing the diffusion process \mathbf{X} appearing in (16.5.6) with $\bar{\mathbf{X}}$, that is

$$\bar{u}(t, \mathbf{x}) = E \left(h(\bar{\mathbf{X}}_{\bar{\tau}}^{t,x}) \right) \quad (16.5.25)$$

for $(t, \mathbf{x}) \in [0, T] \times \Gamma$. Here $\bar{\tau} : \Omega \rightarrow [s, T]$ is the first exit time of $(t, \bar{\mathbf{X}}_t^{s,x})$ from $[s, T] \times \Gamma$, that is

$$\bar{\tau} = \inf \{t \geq s : (t, \bar{\mathbf{X}}_t^{s,x}) \notin [s, T] \times \Gamma\}. \quad (16.5.26)$$

If \bar{u} is sufficiently smooth, then an application of the Kolmogorov backward equation yields the PDE

$$\bar{L}^0 \bar{u}(t, \mathbf{x}) = 0 \quad (16.5.27)$$

for $(t, \mathbf{x}) \in [0, T] \times \Gamma$ with boundary condition

$$\bar{u}(t, \mathbf{x}) = h(t, \mathbf{x}) \quad (16.5.28)$$

for $(t, \mathbf{x}) \in B$. Therefore, with this choice for \bar{u} and using (16.5.27), the second to last equation in (16.5.12) takes the form

$$\begin{aligned} u(0, \mathbf{x}) &= \bar{u}(0, \mathbf{x}) + E \left(\int_0^\tau L^0 \bar{u}(t, \mathbf{X}_t^{0,x}) dt \right) \\ &= \bar{u}(0, \mathbf{x}) + E \left(\int_0^\tau (L^0 - \bar{L}^0) \bar{u}(t, \mathbf{X}_t^{0,x}) dt \right). \end{aligned} \quad (16.5.29)$$

This shows that if \bar{L}^0 is close to L^0 , then the variance of the unbiased estimator \bar{Z}_τ given by (16.5.13) will be small. The estimator \bar{Z}_τ is given by

$$\bar{Z}_\tau = \bar{u}(0, \mathbf{x}) + \int_0^\tau (L^0 - \bar{L}^0) \bar{u}(t, \mathbf{X}_t^{0,x}) dt. \quad (16.5.30)$$

Consider an additional approximation function $\bar{\bar{u}} : [0, T] \times \Gamma \rightarrow \mathfrak{R}$ defined by

$$\bar{\bar{u}}(t, \mathbf{x}) = \bar{u}(t, \mathbf{x}) + z(t, \mathbf{x}), \quad (16.5.31)$$

where

$$z(t, \mathbf{x}) = E \left(\int_t^\tau L^0 \bar{u}(s, \bar{\mathbf{X}}_s^{t,x}) ds \right) \quad (16.5.32)$$

for $(t, \mathbf{x}) \in [0, T] \times \Gamma$. Let $\hat{z} : [0, T] \times \Gamma \times (0, \infty) \rightarrow \mathfrak{R}$ be given by

$$\hat{z}(t, \mathbf{x}, y) = y + z(t, \mathbf{x}) \quad (16.5.33)$$

for $(t, \mathbf{x}, y) \in [0, T] \times \Gamma \times (0, \infty)$ with the process $Y^{s,y} = \{Y_t^{s,y}, t \in [s, T]\}$ defined by the relation

$$Y_t^{s,y*} = y + \int_s^\tau \mathbf{1}_{\{v < \tau\}} L^0 \bar{u}(v, \bar{\mathbf{X}}_v^{s,x}) dv \quad (16.5.34)$$

for $t \in [s, T]$, $y \in (0, \infty)$, $\mathbf{x} = (x_1, \dots, x_d)^\top \in \Gamma$, $\mathbf{y}^* = (x_1, \dots, x_d, y)^\top$ and $s \in [0, T]$. Combining (16.5.32), (16.5.33) and (16.5.34) implies that

$$\hat{z}(t, \mathbf{x}, y) = E \left(Y_T^{t,y*} \right). \quad (16.5.35)$$

Let us define the extended operator \hat{L}^0 on sufficiently smooth functions $f : [0, T] \times \Gamma \times (0, \infty) \rightarrow \mathfrak{R}$ by

$$\begin{aligned} \hat{L}^0 f(t, \mathbf{x}, y) &= \frac{\partial f(t, \mathbf{x}, y)}{\partial t} + \sum_{i=1}^d a^i(t, \mathbf{x}) \frac{\partial f(t, \mathbf{x}, y)}{\partial x^i} \\ &\quad + \frac{1}{2} \sum_{i,k=1}^d \sum_{j=1}^m b^{i,j}(t, \mathbf{x}) b^{k,j}(t, \mathbf{x}) \frac{\partial^2 f(t, \mathbf{x}, y)}{\partial x^i \partial x^k} \\ &\quad + L^0 \bar{u}(t, \mathbf{x}) \frac{\partial f(t, \mathbf{x}, y)}{\partial y} \end{aligned} \quad (16.5.36)$$

for $(t, \mathbf{x}, y) \in (0, T) \times \Gamma \times (0, \infty)$. Application of the Kolmogorov backward equation to \hat{z} using (16.5.35) and the $(d+1)$ -dimensional diffusion process, consisting of \mathbf{X} together with the component Y , shows that

$$\hat{L}^0 \hat{z}(t, \mathbf{x}, y) = 0$$

for $(t, \mathbf{x}, y) \in (0, T) \times \Gamma \times (0, \infty)$. Using (16.5.33) and (16.5.36) this equation can be rewritten in the form

$$\bar{L}^0 z(t, \mathbf{x}) + L^0 \bar{u}(t, \mathbf{x}) = 0 \quad (16.5.37)$$

for $(t, \mathbf{x}) \in (0, T) \times \Gamma$. Combining (16.5.31), (16.5.37), (16.5.32) and (16.5.27) yields

$$\begin{aligned} L^0 \bar{u}(t, \mathbf{x}) &= L^0 \bar{u}(t, \mathbf{x}) + L^0 z(t, \mathbf{x}) \\ &= (L^0 - \bar{L}^0) z(t, \mathbf{x}) \\ &= (L^0 - \bar{L}^0) \left(E \left(\int_t^\tau L^0 \bar{u}(s, \bar{\mathbf{X}}_s^{t,x}) ds \right) \right) \\ &= E \left(\int_t^\tau (L^0 - \bar{L}^0) L^0 \bar{u}(s, \bar{\mathbf{X}}_s^{t,x}) ds \right) \\ &= E \left(\int_t^\tau (L^0 - \bar{L}^0) \left[(L^0 - \bar{L}^0) \bar{u}(s, \bar{\mathbf{X}}_s^{t,x}) \right] ds \right). \end{aligned} \quad (16.5.38)$$

Consequently, these calculations provide the basis for the construction of an iterative series of approximations for u .

Let \bar{u} be a smooth approximation function to u but which does not match the payoff or boundary values for u on the set B . In this case by (16.5.6) and (16.5.11) the HP variance reduction equation (16.5.12) takes the form

$$\begin{aligned} u(0, x) &= E(h(\tau, \mathbf{X}_\tau^{0,x})) \\ &= \bar{u}(0, \mathbf{x}) + E(h(\tau, \mathbf{X}_\tau^{0,x}) - \bar{u}(\tau, \mathbf{X}_\tau^{0,x})) + E \left(\int_0^\tau L^0 \bar{u}(t, \mathbf{X}_t^{0,x}) dt \right) \\ &= \bar{u}(0, \mathbf{x}) + E(h(\tau, \mathbf{X}_\tau^{0,x}) - \bar{u}(\tau, \mathbf{X}_\tau^{0,x})) + \int_0^\tau E(\mathbf{1}_{\{t < \tau\}} L^0 \bar{u}(t, \mathbf{X}_t^{0,x})) dt. \end{aligned} \quad (16.5.39)$$

The corresponding unbiased estimator \bar{Z}_τ^h for $u(0, \mathbf{x})$ is given by

$$\bar{Z}_\tau^h = \bar{u}(0, \mathbf{x}) + h(\tau, \mathbf{X}_\tau^{0,x}) - \bar{u}(\tau, \mathbf{X}_\tau^{0,x}) + \int_0^\tau L^0 \bar{u}(t, \mathbf{X}_t^{0,x}) dt. \quad (16.5.40)$$

Note that (16.5.39) reduces to (16.5.12) if $h(\tau, \mathbf{X}_\tau^{0,x}) = \bar{u}(\tau, \mathbf{X}_\tau^{0,x})$.

An Example for the Heston Model

In this section the HP estimator is utilized to price European call options for the Heston stochastic volatility model, see Sect. 2.5 and [Heston \(1993\)](#). In the original treatment by Heston the prices of European style derivatives are computed via an integral representation using characteristic functions. The available almost exact solution made this stochastic volatility model very attractive and popular. The HP method, described in this section, is more general and can be applied to a wide class of stochastic volatility models. The popular Heston model is simply a well-known example, which we have chosen to illustrate the HP method. Consider the vector diffusion process $\mathbf{X} = \{\mathbf{X}_t = (S_t^{s,x^1}, v_t^{s,x^2})^\top, t \in [s, T]\}$, whose real world dynamics are governed by the system of SDEs

$$\begin{aligned} dS_t^{s,x^1} &= \mu S_t^{s,x^1} dt + \sqrt{v_t^{s,x^2}} S_t^{s,x^1} dW_t^1 \\ dv_t^{s,x^2} &= \kappa (\psi - v_t^{s,x^2}) dt + \xi \sqrt{v_t^{s,x^2}} (\varrho dW_t^1 + \sqrt{1-\varrho^2} dW_t^2) \end{aligned} \quad (16.5.41)$$

for $t \in [s, T]$ and $s \in [0, T]$ with initial values $S_s^{s,x^1} = x^1 > 0$ and $v_s^{s,x^2} = x^2 > 0$ and nonnegative constants μ, κ, ψ, ξ and $\varrho \in [-1, 1]$ with $\frac{\kappa\psi}{\xi^2} > \frac{1}{2}$. The vector $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, W_t^2)^\top, t \in [0, T]\}$ is a two-dimensional Wiener process under P . Here the symbols S^{s,x^1} and v^{s,x^2} denote the stock and squared volatility processes, respectively. The savings account process $\bar{B} = \{\bar{B}_t, t \in [0, T]\}$ is given by $\bar{B}_t = e^{rt}$, where $r > 0$ is, for simplicity, a constant short rate.

The GOP $S_t^{\delta_*}$ in this market satisfies the SDE

$$dS_t^{\delta_*} = S_t^{\delta_*} (r dt + \theta_t^1 (\theta_t^1 dt + dW_t^1)) \quad (16.5.42)$$

with market prices of risk

$$\theta_t^1 = \frac{\mu - r}{\sqrt{v_t^{s,x^2}}} \quad (16.5.43)$$

and $\theta_t^2 = 0$. The benchmarked stock $\hat{S}_t^{s,x^1} = \frac{S_t^{s,x^1}}{S_t^{\delta_*}}$ satisfies then the SDE

$$d\hat{S}_t^{s,x^1} = \hat{S}_t^{s,x^1} \left(\sqrt{v_t^{s,x^2}} - \theta_t^1 \right) dW_t^1. \quad (16.5.44)$$

Under the benchmark approach no further assumptions need to be made and one prices a contingent claim H_T with maturity T via the real world pricing formula (3.3.5) such that

$$c(t, \mathbf{x}) = E \left(\frac{S_t^{\delta_*}}{S_T^{\delta_*}} H_T \mid \mathcal{A}_t \right) \quad (16.5.45)$$

yields the fair price of H_T at time t .

For this type of market model, which is incomplete if $|\varrho| \neq 1$, real world pricing turns out to be equivalent to choosing the minimal equivalent martingale measure \tilde{P} as pricing measure, see for example, Föllmer & Schweizer (1991), Hofmann et al. (1992) and Heath, Platen & Schweizer (2001). This measure has the property that it is both equivalent to P and such that any local $(\underline{\mathcal{A}}, \tilde{P})$ -martingale, which is orthogonal to $\frac{S}{B}$, remains a local $(\underline{\mathcal{A}}, P)$ -martingale.

The Radon-Nikodym derivative Λ_t of \tilde{P} is given as the stochastic exponential

$$\Lambda_t = \frac{\bar{B}_t S_0^{\delta_*}}{S_t^{\delta_*}} = \exp \left\{ -\frac{1}{2} \int_0^t (\theta_t^1)^2 dt + \int_0^t \theta_t^1 dW_t^1 \right\}, \quad (16.5.46)$$

which is required to be an $(\underline{\mathcal{A}}, P)$ -martingale.

The real world pricing formula (3.3.5), (16.5.45) and the Bayes rule yield then also the traditional formula for the risk neutral price

$$\begin{aligned} c(t, \mathbf{x}) &= E \left(\frac{\bar{B}_T S_0^{\delta_*}}{S_T^{\delta_*}} \frac{S_t^{\delta_*}}{\bar{B}_t S_0^{\delta_*}} \frac{\bar{B}_t H_T}{\bar{B}_T} \mid \mathcal{A}_t \right) \\ &= E \left(\frac{\Lambda_T}{\Lambda_t} \frac{\bar{B}_t H_T}{\bar{B}_T} \mid \mathcal{A}_t \right) = \tilde{E} \left(\frac{\bar{B}_t H_T}{\bar{B}_T} \mid \mathcal{A}_t \right), \end{aligned} \quad (16.5.47)$$

where \tilde{E} denotes expectation under \tilde{P} .

Under the measure \tilde{P} the dynamics for S^{s,x^1} and v^{s,x^2} becomes

$$\begin{aligned} dS_t^{s,x^1} &= r S_t^{s,x^1} dt + \sqrt{v_t^{s,x^2}} S_t^{s,x^1} d\tilde{W}_t^1 \\ dv_t^{s,x^2} &= \kappa \left(\tilde{\psi} - v_t^{s,x^2} \right) dt + \xi \sqrt{v_t^{s,x^2}} \left(\varrho d\tilde{W}_t^1 + \sqrt{1-\varrho^2} d\tilde{W}_t^2 \right) \end{aligned} \quad (16.5.48)$$

for $t \in [s, T]$ and $s \in [0, T]$, where $\tilde{\psi} = \psi - \frac{\xi \varrho (\mu - r)}{\kappa}$ and

$$\begin{aligned} d\tilde{W}_t^1 &= \theta_t^1 dt + dW_t^1 \\ d\tilde{W}_t^2 &= dW_t^2. \end{aligned}$$

Here $\tilde{\mathbf{W}} = \{\tilde{\mathbf{W}}_t = (\tilde{W}_t^1, \tilde{W}_t^2)^\top, t \in [0, T]\}$ is a two-dimensional Wiener process under \tilde{P} . The minimal equivalent martingale measure transforms the stock price such that its appreciation rate becomes r . However, it leaves the untraded noise W_t^2 unchanged under \tilde{P} . We assume that $\frac{\kappa \tilde{\psi}}{\xi^2} > \frac{1}{2}$ so that the square root process v^{s,x^2} remains strictly positive under \tilde{P} . Furthermore, one needs to assume the parameters are chosen such that \tilde{P} is an equivalent risk neutral probability measure. Let us also ensure that the stock price process S^{s,x^1} remains strictly positive under \tilde{P} and P .

Even that it requires stronger assumptions, let us work under the risk neutral approach since this is common to most readers. However, the method below works equally well when applied directly under the benchmark approach with less restrictive modeling assumptions and without measure change. We now describe a version of the HP variance reduction method, which can be utilized to approximate the option price

$$c(0, \mathbf{x}) = e^{-rT} u(0, \mathbf{x}), \quad (16.5.49)$$

where

$$u(0, \mathbf{x}) = \tilde{E} \left((S_T^{0,x^1} - K)^+ \right).$$

This means, we consider a European call with strike K under the Heston model with risk neutral dynamics given by (16.5.48) and $\mathbf{x} = (x^1, x^2)^\top = (S_0^{s,x^1}, v_0^{s,x^2})^\top$. The main idea will be to use the estimate (16.5.29) with the HP estimator (16.5.30). As previously mentioned, one way of utilizing the HP method is to find an approximation $\bar{\mathbf{X}}$ for the diffusion process \mathbf{X} . A convenient choice, based on the Black-Scholes model, is as follows: Let $\bar{\mathbf{X}} = \{\bar{\mathbf{X}}_t = (\bar{S}_t^{s,x^1}, \bar{v}_t^{s,x^2})^\top, t \in [s, T]\}$ be the two-dimensional diffusion process which satisfies the SDE

$$\begin{aligned} d\bar{S}_t^{s,x^1} &= r \bar{S}_t^{s,x^1} dt + \sqrt{\bar{v}_t^{s,x^2}} \bar{S}_t^{s,x^1} d\tilde{W}_t^1 \\ d\bar{v}_t^{s,x^2} &= \kappa \left(\tilde{\psi} - \bar{v}_t^{s,x^2} \right) dt \end{aligned} \quad (16.5.50)$$

for $t \in [s, T]$ and $s \in [0, T]$ with initial values $\bar{S}_s^{s,x^1} = x^1 > 0$ and $\bar{v}_s^{s,x^2} = x^2 > 0$. For this system of SDEs the solution for \bar{v} can be explicitly computed and is given by

$$\bar{v}_t^{s,x^2} = \tilde{\psi} + (x^2 - \tilde{\psi}) e^{-\kappa(t-s)} \quad (16.5.51)$$

for $t \in [s, T]$ and $s \in [0, T]$. Denote by $BS(x^1, K, r, \sigma, T)$ the Black-Scholes price for a European call option with spot price x^1 , strike K , short rate r , constant volatility σ and maturity T . Using (16.5.51) and setting $\tau = T$, the approximate function $\bar{u} : [0, T] \times (0, \infty)^2 \rightarrow \mathbb{R}$ given in (16.5.25) takes the form

$$\begin{aligned} \bar{u}(t, \mathbf{x}) &= \tilde{E} \left((\bar{S}_T^{t,x^1} - K)^+ \right) \\ &= e^{r(T-t)} BS(x^1, K, r, \bar{\sigma}_t, T-t) \end{aligned} \quad (16.5.52)$$

for $(t, \mathbf{x}) \in [0, T] \times (0, \infty)^2$, where $\mathbf{x} = (x^1, x^2)^\top = (\bar{S}_t^{s,x^1}, \bar{v}_t^{s,x^2})^\top$ and

$$\begin{aligned} \bar{\sigma}_t &= \sqrt{\frac{1}{T-t} \int_t^T \bar{v}_z^{t,x^2} dz} \\ &= \sqrt{\tilde{\psi} - (x^2 - \tilde{\psi}) \frac{e^{-\kappa(T-t)} - 1}{\kappa(T-t)}}. \end{aligned} \quad (16.5.53)$$

Here $BS(\cdot, \cdot, \cdot, \cdot, \cdot)$ denotes the European call price according to the Black-Scholes formula.

Evaluation of the expectation appearing in (16.5.29) requires the calculation of the values $(L^0 - \bar{L}^0)\bar{u}(t, \mathbf{X}_t^{0,x})$ for $t \in [0, T]$. Using (16.5.48) and (16.5.50) and noting that the coordinates for $\mathbf{X} = (X^1, X^2)^\top$ and $\bar{\mathbf{X}} = (\bar{X}^1, \bar{X}^2)^\top$ correspond to the vectors $(S^{s,x^1}, v^{s,x^2})^\top$ and $(\bar{S}^{s,x^1}, \bar{v}^{s,x^2})^\top$, respectively, the operator $(L^0 - \bar{L}^0)$ can be determined. Thus, for sufficiently smooth functions $f : [0, T] \times (0, \infty)^2 \rightarrow \mathbb{R}$ we have

$$(L^0 - \bar{L}^0)f(t, \mathbf{x}) = \xi x^2 \left(\varrho \frac{\partial^2 f(t, \mathbf{x})}{\partial x^1 \partial x^2} + \frac{1}{2} \xi \frac{\partial^2 f(t, \mathbf{x})}{\partial (x^2)^2} \right) \quad (16.5.54)$$

for $(t, \mathbf{x}) \in (0, T) \times (0, \infty)^2$, where $\mathbf{x} = (x^1, x^2)^\top$.

The corresponding values for $(L^0 - \bar{L}^0)\bar{u}(t, \mathbf{x})$ can now be computed in an explicit form because $\bar{u}(t, \mathbf{x})$ can be expressed as a scaled Black-Scholes price with volatility $\bar{\sigma}_t$, see (16.5.52) and (16.5.53). Substitution of the appropriate partial derivatives for \bar{u} into (16.5.54) yields

$$\begin{aligned} (L^0 - \bar{L}^0)\bar{u}(t, \mathbf{x}) &= \xi x^2 e^{r(T-t)} \left[\varrho \frac{\partial^2 BS(x^1, K, r, \bar{\sigma}_t, T-t)}{\partial x^1 \partial \bar{\sigma}_t} \frac{\partial \bar{\sigma}_t}{\partial x^2} \right. \\ &\quad + \frac{1}{2} \xi \left\{ \frac{\partial^2 BS(x^1, K, r, \bar{\sigma}_t, T-t)}{\partial \bar{\sigma}_t^2} \left(\frac{\partial \bar{\sigma}_t}{\partial x^2} \right)^2 \right. \\ &\quad \left. \left. + \frac{\partial BS(x^1, K, r, \bar{\sigma}_t, T-t)}{\partial \bar{\sigma}_t} \frac{\partial^2 \bar{\sigma}_t}{\partial (x^2)^2} \right\} \right]. \end{aligned} \quad (16.5.55)$$

The partial derivatives $\frac{\partial BS}{\partial \bar{\sigma}_t}$, $\frac{\partial^2 BS}{\partial x^1 \partial \bar{\sigma}_t}$, $\frac{\partial^2 BS}{\partial \bar{\sigma}_t^2}$, $\frac{\partial \bar{\sigma}_t}{\partial x^2}$ and $\frac{\partial^2 \bar{\sigma}_t}{\partial (x^2)^2}$ can be computed in a straightforward manner using the Black-Scholes formula together with the expression for $\bar{\sigma}_t$ given in (16.5.53). For example, $\frac{\partial BS}{\partial \bar{\sigma}_t}$ is often referred to as the Black-Scholes vega.

HP Method for other Models

To see how the above HP method can be adapted for other stochastic volatility models, it is important to note that the approximate diffusion process $\bar{\mathbf{X}}$ does not depend on the diffusion term for the variance component appearing in (16.5.48). Therefore, our choice of \bar{u} can be directly employed for other diffusion coefficients for v , for example, a linear or constant diffusion term. If the drift term for the variance component in (16.5.48) is changed, then (16.5.50) can be modified to equal this changed drift term. Alternatively, a constant or average variance can be used to construct \bar{u} .

In addition, for some types of stochastic volatility models, including the Heston model and other affine models, the moments for $\ln(S_t)$ can be explicitly computed. Consequently, another HP estimator based method would be

to approximate $(S_t, v_t)^\top$, $t \in [0, T]$, with a geometric Brownian motion $\bar{\mathbf{X}}$, where the first two moments coincide for $t \in [0, T]$. The approximate pricing function \bar{u} is then constructed using the corresponding option price for geometric Brownian motion.

Let $0 = \tau_0 < \tau_1 < \dots < \tau_N = T$ be an equidistant time discretization of the interval $[0, T]$ with step size $\Delta = \frac{T}{N}$. We could use for the Heston model the almost exact simulation method proposed in Sect. 2.6. However, this is a strong approximation method and also not generally applicable for other stochastic volatility methods. For the simulation results described in this section, a predictor-corrector method of weak order 1.0, see (11.5.17), was employed. Although in theoretical terms this scheme has the same weak order of convergence as the Euler method, it exhibits much better numerical stability properties and usually produces more accurate results. Furthermore, it allows one to play with different degrees of implicitness θ and η , which gives a sense for the numerical stability of the implemented method. For a general d -dimensional diffusion process, which satisfies (16.5.1), this scheme takes the form

$$\begin{aligned} Y_{n+1} = Y_n + & \left\{ \theta \hat{a}(\tau_{n+1}, \hat{Y}_{n+1}) + (1 - \theta) \hat{a}(\tau_n, Y_n) \right\} \Delta \\ & + \sum_{j=1}^m \left\{ \eta b^j(\tau_{n+1}, \hat{Y}_{n+1}) + (1 - \eta) b^j(\tau_n, Y_n) \right\} \Delta W_n^j \end{aligned} \quad (16.5.56)$$

for $n \in \{0, 1, \dots, N-1\}$ with predictor

$$\hat{Y}_{n+1} = Y_n + a(\tau_n, Y_n) \Delta + \sum_{j=1}^m b^j \Delta W_n^j$$

and modified drift coefficient values

$$\hat{a}(\tau_n, Y_n) = a(\tau_n, Y_n) - \eta \sum_{i=1}^d \sum_{j=1}^m b^{i,j}(\tau_n, Y_n) \frac{\partial b^j(\tau_n, Y_n)}{\partial x^i}.$$

Here $\theta, \eta \in [0, 1]$ and ΔW_n^j , $j \in \{1, 2, \dots, m\}$, $n \in \{0, 1, \dots, N-1\}$ are independent $N(0, \Delta)$ Gaussian random variables. Note that one could have used two-point distributed random variables $\Delta \hat{W}_n^j$, which in this example give almost the same outcomes.

The corresponding discrete-time approximation for the estimator \bar{Z}_T , given in (16.5.30) with $\tau = T$, can be obtained by adding this component to the system of equations (16.5.48). With this choice of components the numerical scheme (16.5.56) can be applied to construct an approximation $\mathbf{Y}_N = (Y_N^1, Y_N^2, Y_N^3)^\top$ to the random vector $(S_T^{0,x^1}, v_T^{0,x^2}, \bar{Z}_T)^\top$ with initial values $Y_0^1 = x^1$, $Y_0^2 = x^2$ and $Y_0^3 = \bar{u}(0, x)$, $\mathbf{x} = (x^1, x^2)^\top$. The expectation $\tilde{E}(Y_N^3)$, therefore, provides an approximation to the value $u(0, \mathbf{x})$, see (16.5.29), (16.5.48) and (16.5.49).

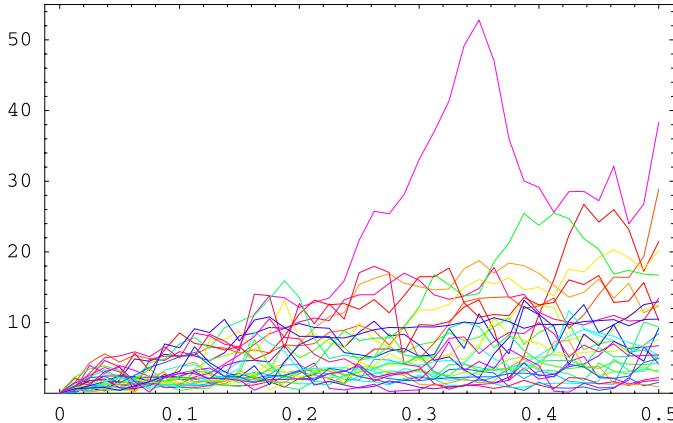


Fig. 16.5.1. Simulated outcomes for the intrinsic value $(S_t^{0,x^1} - K)^+$, $t \in [0, T]$

For all simulation experiments the results have been enhanced by the use of antithetic variates, see Sect. 16.2. This consisted of full reflection of both Wiener components. That is, sample paths computed via the $2N$ outcomes $(\Delta W_n^1, \Delta W_n^2)$, $n \in \{0, 1, \dots, N-1\}$ were combined with additional sample paths generated from the pairs $(\Delta W_n^1, -\Delta W_n^2)$, $(-\Delta W_n^1, \Delta W_n^2)$ and $(-\Delta W_n^1, -\Delta W_n^2)$ for $n \in \{0, 1, \dots, N-1\}$.

Simulation Results

To provide a graphical illustration of the level of variance reduction possible with the HP method, Fig. 16.5.1 displays a group of $10 \times 4 = 40$ simulated sample paths for the intrinsic value $(S_t^{0,x^1} - K)^+$, $t \in [0, T]$, with $N = 40$ time discretization points. The results include sample paths generated from the antithetically produced outcomes as explained above. For this and subsequent plots the following default parameter values were used: $\kappa = 0.6$, $\tilde{\psi} = 0.04$, $\xi = 0.2$, $r = 0.04$, $T = 0.5$, $K = 100$ with initial values $S_0^{0,x^1} = 100$ and $v_0^{0,x^2} = 0.04$. Note that for these parameter values $\frac{\kappa \tilde{\psi}}{\xi^2} = 0.6 > 0.5$. These choices produce a strong stochastic volatility effect, sufficiently constrained to ensure that S^{0,x^1} and v^{0,x^2} remain strictly positive under \tilde{P} .

Fig. 16.5.2 shows the same set of sample paths for the estimator \bar{Z}_t , $t \in [0, T]$. Note that simulated outcomes for $(S_t^{0,x^1} - K)^+$, $t \in [0, T]$, are spread over the wide interval $[0.00, 50.00]$, whereas those for \bar{Z}_t are contained in the extremely small interval $[6.60, 6.77]$. For a larger sample of $100 \times 4 = 400$ trajectories the standard error for the means $\tilde{E}((S_T^{0,x^1} - K)^+)$ and $\tilde{E}(\bar{Z}_T)$ were 0.608 and 0.004, respectively. A comparison of the two figures, therefore, demonstrates the dramatic reduction in variance that can be obtained with an implementation of the HP method. For this example the method produces

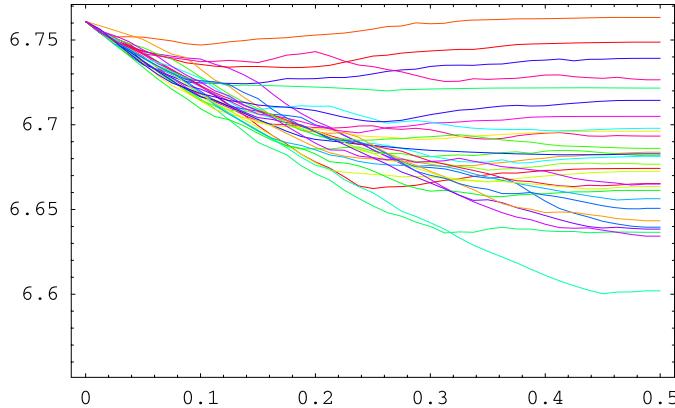


Fig. 16.5.2. Simulated outcomes for the estimator \bar{Z}_t , $t \in [0, T]$

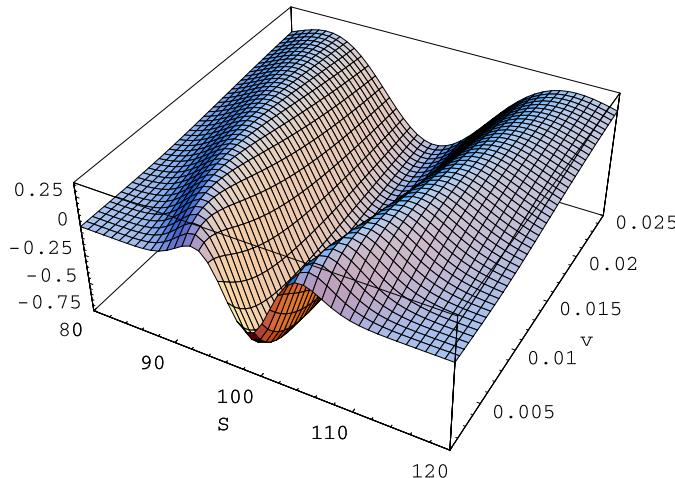


Fig. 16.5.3. Diffusion operator values $(L - \bar{L}^0) \bar{u}$ as a function of asset price S and squared volatility v

a variance reduction of approximately 23000 times compared to the standard Monte Carlo method with antithetic variates and even a more significant variance reduction when compared with a raw Monte Carlo method.

The degree of variance reduction achieved with the HP method is closely related to the variance of the integrand $(L^0 - \bar{L}^0) \bar{u}(t, \mathbf{x}_t^{0,x})$ appearing in (16.5.30). Fig. 16.5.3 shows the function values $(L^0 - \bar{L}^0) \bar{u}(t, \mathbf{x})$, for a fixed value of $t = 0.2$, using different values of the coordinates x^1 and x^2 corresponding to the components S^{0,x^1} and v^{0,x^2} , respectively.

The quantity $e^{-r(T-t)} (u(t, \mathbf{x}) - \bar{u}(t, \mathbf{x}))$ measures the difference in call option prices obtained from the Heston model and the corresponding Black-

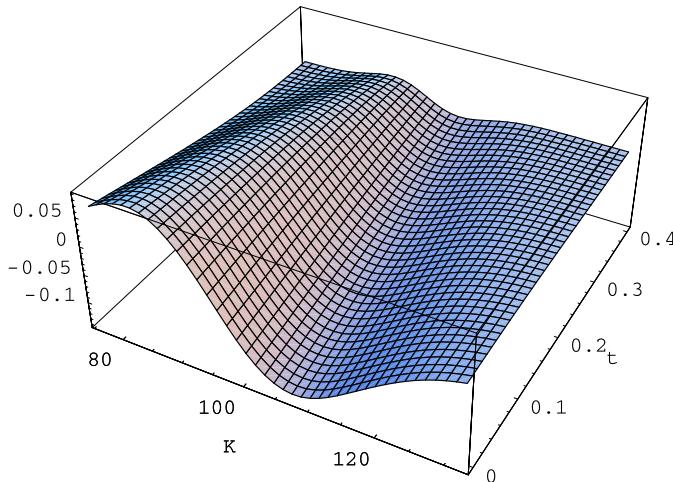


Fig. 16.5.4. Price differences between the Heston and corresponding Black-Scholes model as a function of strike K and time t

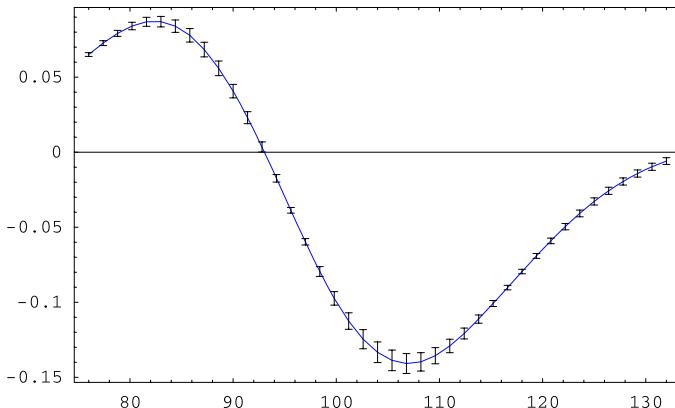


Fig. 16.5.5. Prices and corresponding error bounds as a function of strike K

Scholes model with a time dependent deterministic volatility, see (16.5.49) and (16.5.52). Fig. 16.5.4 displays these price differences as a function of time $t \in [0, T]$ and strike K . A correlation coefficient of $\rho = -0.15$ was used so that risky asset returns are negatively correlated with volatility. These results were obtained by using 256×4 sample paths and $N = 20$ discretization points.

For $t = 0$, corresponding error bounds at a 99% confidence level are shown in Fig. 16.5.5. These results indicate that for a typical at-the-money call option the HP method, with approximately 1000 sample paths and 20 discretization points, can be utilized to compute the corresponding price with a relative error of 0.04% at a 99% confidence level. For the same parameter values, as used in

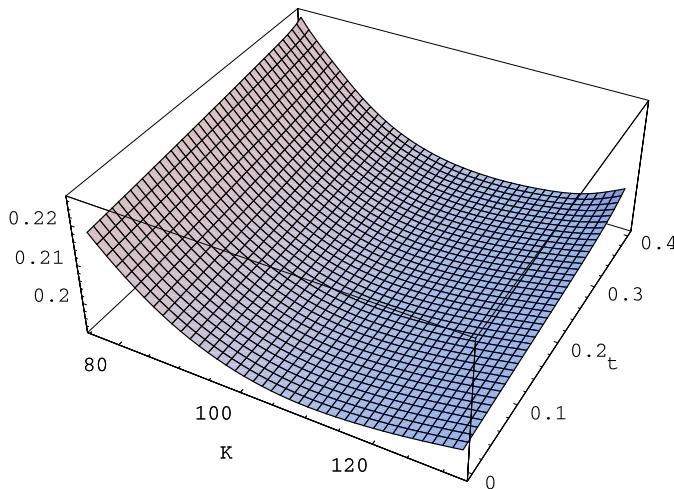


Fig. 16.5.6. Implied volatility term structure for the Heston model

Fig. 16.5.4, the corresponding implied volatilities are illustrated in Fig. 16.5.6. This implied volatility surface shows a negative skew with some smile effect, a feature which is observed in many markets. For the Heston model choosing $\varrho > 0$ leads to a positive skew. Note that the choice $\varrho = 0$ produces a smile.

The above described HP method is very powerful since it derives the variance reduction for each path using deep stochastic analytic relationships. It is straightforward to generalize the HP method to the case with jumps and to develop particular versions of it.

16.6 Exercises

16.1. Construct an antithetic variance reduction method, where you simulate the expectation $E(1 + Z + \frac{1}{2}(Z)^2)$ for a standard Gaussian random variable $Z \sim N(0, 1)$. For this purpose combine the raw Monte Carlo estimate

$$V_N^+ = \frac{1}{N} \sum_{k=1}^N \left(1 + Z(\omega_k) + \frac{1}{2} (Z(\omega_k))^2 \right)$$

using outcomes of Z with the antithetic one

$$V_N^- = \frac{1}{N} \sum_{k=1}^N \left(1 - Z(\omega_k) + \frac{1}{2} (-Z(\omega_k))^2 \right)$$

that also uses the same outcomes but with a negative sign. Demonstrate the variance reduction that can be achieved for the estimator

$$\hat{V}_N = \frac{1}{2} (V_N^+ + V_N^-)$$

instead of using only V_N^+ . Is \hat{V}_N an unbiased estimator?

16.2. For calculating the expectation $E(1+Z+\frac{1}{2}(Z)^2)$ use the control variate

$$V_N^* = \frac{1}{N} \sum_{k=1}^N (1 + Z(\omega_k)).$$

Analyze the variance reduction that can be achieved by the estimate

$$\tilde{V}_N = V_N^* + \alpha(\gamma - V_N^*)$$

for $\alpha \in \Re$. For which choice of $\gamma \in \Re$ does one obtain an unbiased estimator? For which $\alpha \in \Re$ does one achieve the minimum variance?

16.3. Check for the measure transformation variance reduction method and the payoff $g(z) = z^2$ that $g(\tilde{X}_T^{0,x}) \frac{\theta_T}{\theta_0}$ is non-random and equals

$$u(t, x) = E \left(g \left(X_T^{0,x} \right) \mid \mathcal{A}_t \right)$$

for $t = 0$, assuming for $X_T^{0,x}$ and $\tilde{X}_T^{0,x}$ the linear SDEs

$$dX_t^{0,x} = \alpha X_t^{0,x} dt + \beta X_t^{0,x} dW_t$$

and

$$d\tilde{X}_t^{0,x} = \alpha \tilde{X}_t^{0,x} dt + \beta \tilde{X}_t^{0,x} d\tilde{W}_t$$

for $t \in [0, T]$ with $X_T^{0,x} = \tilde{X}_T^{0,x} = x_0$, respectively, where

$$d\tilde{W}_t = dW_t - d(t, \tilde{X}_t^{0,x}) dt$$

and

$$d\theta_t = \theta_t d(t, \tilde{X}_t^{0,x}) dW_t$$

with

$$d(t, \tilde{X}_t^{0,x}) = -\frac{\beta \tilde{X}_t^{0,x}}{u(t, \tilde{X}_t^{0,x})} \frac{\partial u(t, \tilde{X}_t^{0,x})}{\partial x}.$$

16.4. Formulate a drift implicit simplified weak Euler scheme for the two-factor Heston model

$$\begin{aligned} dX_t &= (r_d - r_f) X_t + k \sqrt{v_t} X_t dW_t^1 \\ dv_t &= \kappa(v_t - \bar{v}) dt + p \sqrt{v_t} \left(\varrho dW_t^1 + \sqrt{1 - \varrho^2} dW_t^2 \right), \end{aligned}$$

for $t \in [0, T]$ with $t \in [0, T^*]$ $X_0 > 0$ and $v_0 > 0$, where W^1 and W^2 are independent standard Wiener processes.

16.5. Consider the Heston model of Exercise 16.4 for $\varrho = 0$. Make also the diffusion term in the X component of the simplified weak Euler scheme implicit and correct appropriately the drift term in the resulting scheme.

Trees and Markov Chain Approximations

This final chapter discusses numerical effects on tree methods. Furthermore, binomial, trinomial and multinomial trees will be interpreted as Markov chain approximations of solutions of SDEs. General higher order Markov chain approximations with given weak order of convergence will be described. In addition, the relationship with finite difference methods will be highlighted towards the end of the chapter. It is worth pointing out that jump diffusions can be rather easily approximated in a similar manner as will be described.

17.1 Numerical Effects of Tree Methods

This section discusses important numerical effects that can typically be observed when implementing trees for the approximation of functionals of solutions of SDEs.

Option Prices from Binomial Trees

We now consider the implementation of binomial trees with Cox-Ross-Rubinstein (CRR) steps, similar to that described in (3.7.34)–(3.7.35). At first we construct the tree with n time steps. Here one observes that at odd and even time steps we have different levels that can be reached. Let us number at each time $t = n\Delta$ the nodes starting at the bottom with $j = 1$ and ending with $j = n$. With each new time step there is one additional possible state in the binomial tree. The probability of the underlying state variable to be at time $t = n\Delta$, at the j th node is given by the well-known *binomial probability*

$$p_j(n) = \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \quad (17.1.1)$$

for $n \in \{1, 2, \dots\}$ and $j \in \{1, 2, \dots, n\}$ with $p \in (0, 1)$.

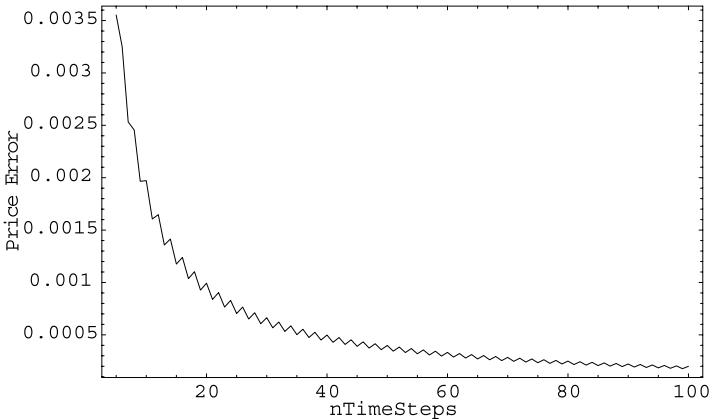


Fig. 17.1.1. Error of at-the-money European call in dependence on number of steps

A corresponding European option is now priced on such a binomial tree by setting at maturity the option value equal to the European payoff. By stepping backwards, one step in time, one obtains at each node the option value corresponding to the one step expectation of the terminal payoff by using the one step transition probabilities (3.7.42) of the tree, equalling p for moving up and $(1-p)$ for moving down. Next one interprets the obtained option price as a new payoff, then takes one further step back in time to calculate again the one time step expectations. In this manner one steps backward through the tree to finally obtain the price of the option, which can be found at the root of the tree, at $t = t_0$. This is a straightforward procedure that can be easily implemented and has been widely used in option pricing.

Let us now analyze the difference that arises between option prices obtained from the Black-Scholes formula (3.7.46) and those generated by using a CRR binomial tree. We have via the Black-Scholes pricing formula the limiting value for comparison. If one interprets the BS model as the exact model, then the difference in price can be interpreted as pricing error of the binomial model.

Let us consider a CRR binomial tree with initial value $S_0 = 1$ of the stock, strike $\hat{K} = 1$, maturity $T = 1$ and volatility $\sigma = 0.2$. We show in Fig. 17.1.1 the error of the corresponding at-the-money European call option price dependent on the number of time steps, as it arises for the binomial tree. It is clearly visible that the odd and even numbers of steps have an oscillating impact on the accuracy of the at-the-money binomial call option price. One also notes that for increased number of steps the error reduces on average. Since the number of time steps determines the position of the nodes with respect to the strike price, it is of interest to also see the dependence of the error for an out-of-the money call option, which we show in Fig. 17.1.2. This graph displays an increased oscillatory behavior of the error. The presence of this effect is important to know when valuing options via binomial trees. We will observe a similar oscillatory behavior for other trees later.

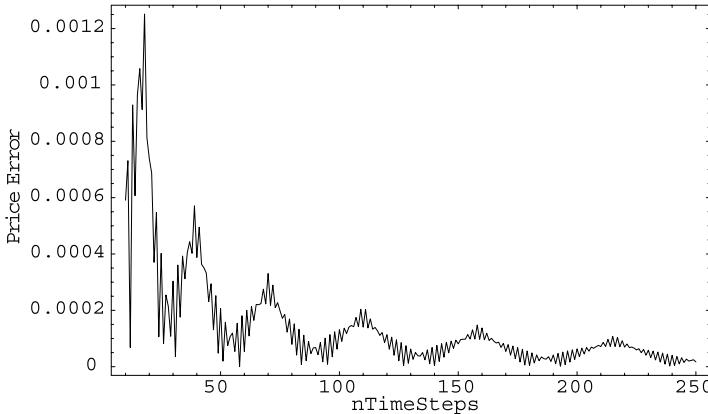


Fig. 17.1.2. Error for out-of-the money European call in dependence on number of steps

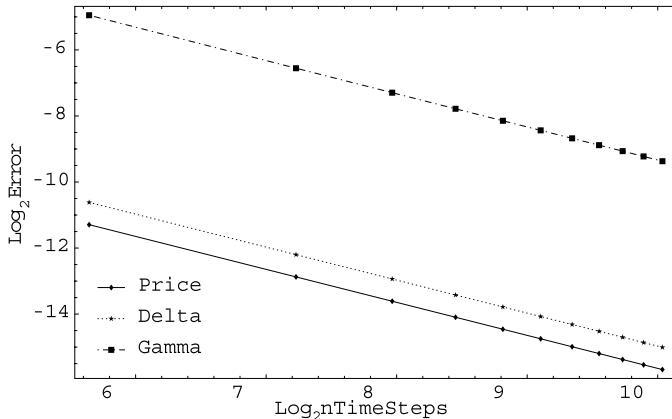


Fig. 17.1.3. Log-log plot of weak error for at-the-money price, delta and gamma

Another important characteristic is the weak order of convergence that a binomial tree achieves. The graphs in Fig. 17.1.1 and Fig. 17.1.2 indicate a systematic decline of the weak error

$$\mu(\Delta) = |E(g(X_T)) - E(g(Y_T^\Delta))|,$$

see (11.1.2), where $E(g(Y_T^\Delta))$ denotes here the binomial option price. In Fig. 17.1.3 we provide a log-log plot for the weak error of the binomial at-the-money European call price against the log of the number of time steps. One notes the almost linear decline of $\ln(\mu(\Delta))$ against $\ln(\frac{T}{\Delta})$. A linear regression of the graph yields the function

$$\ln(\mu(\Delta)) \approx -3.91695 - 0.999291 \ln\left(\frac{T}{\Delta}\right),$$

which shows experimentally that the binomial tree achieves for a European call a weak order of about $\beta \approx 1.0$.

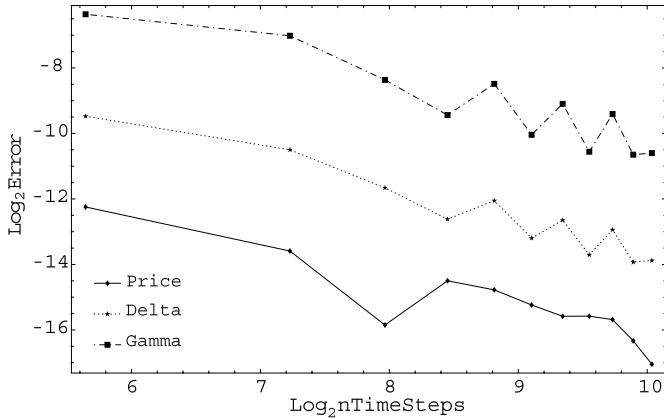


Fig. 17.1.4. Log-log plot of absolute weak error for out-of-the money price, delta and gamma of CRR binomial European call

It is interesting to study the numerical behavior of some sensitivities in binomial option pricing. For instance, the delta, which is the hedge ratio of the binomial European call price, is simply obtained by taking the numerical spatial derivative between two neighboring initial call prices. In Fig. 17.1.3 we show the log of the absolute error, of the hedge ratio dependent on the logarithm of the number of time steps. One notes here that the slope is again close to one, which suggests experimentally the weak order $\beta \approx 1.0$.

A similar log-log plot is provided in Fig. 17.1.3 for the error of the gamma, the sensitivity of the hedge ratio with respect to the underlying. One observes again a slope of about $\beta \approx 1.0$. In Fig. 17.1.4 we display in a log-log plot the weak error, for an out-of-the money binomial option of the price, delta and gamma. One notes that the decline with the log number of steps is not as linear as in the previous figure. The same applies to the corresponding log-log plot for an in-the-money binomial call option shown in Fig. 17.1.5. This figure suggests that some caution should be taken when applying binomial or other tree methods. In particular, for small step sizes and when modeling martingales we have seen in Chap. 14 that for decreasing time step size such an explicit discrete-time approximation can become numerically unstable.

Binomial American Pricing

A considerable advantage of tree methods is that one can easily handle the pricing of American style options. As mentioned previously, the only change in the binomial method compared to European option pricing is that one has to compare at each time step and node the values associated with continuing to hold the American option or exercising it early. In other words, one has to decide whether one is in the continuation region or in the stopping region, see Sect. 15.1. In Fig. 17.1.6 we show the logarithm of the binomial American

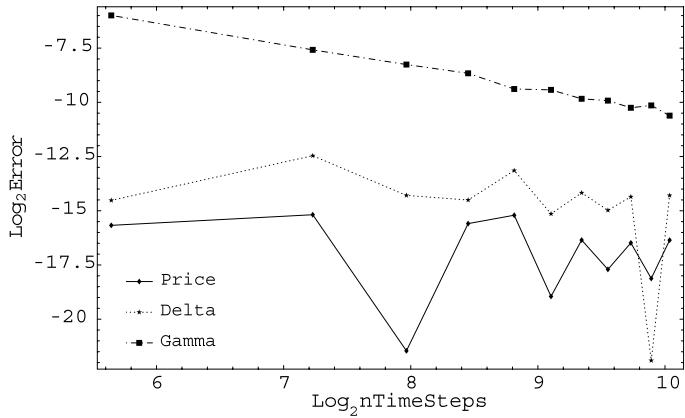


Fig. 17.1.5. Log-log plot of the absolute error for in-the-money price, delta and gamma of CRR binomial European call

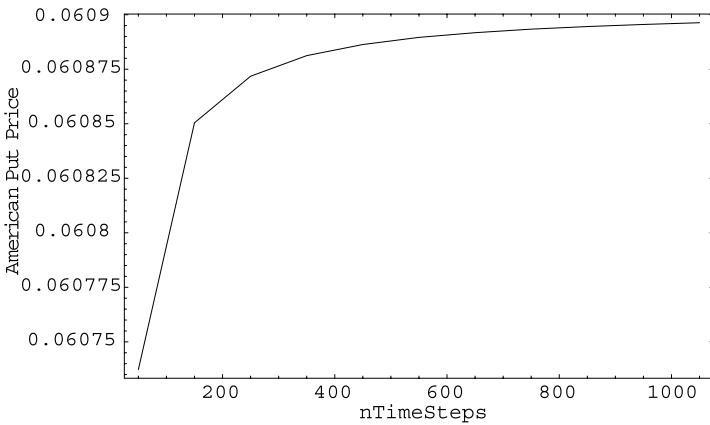


Fig. 17.1.6. CRR binomial American put prices

put price against the log of the number of time steps. One notes that the binomial American put price appears to converge to a limiting value. For American put options an early exercise boundary emerges, which describes the level at which the American option price equals the payoff function. As an illustration, in Fig. 17.1.7 we show for the above default parameters for an at-the-money American put option the resulting early exercise boundary when $n = 250$ time steps are used. One notes the oscillatory behavior of the early exercise boundary due to the nature of the binomial tree with odd and even time steps.

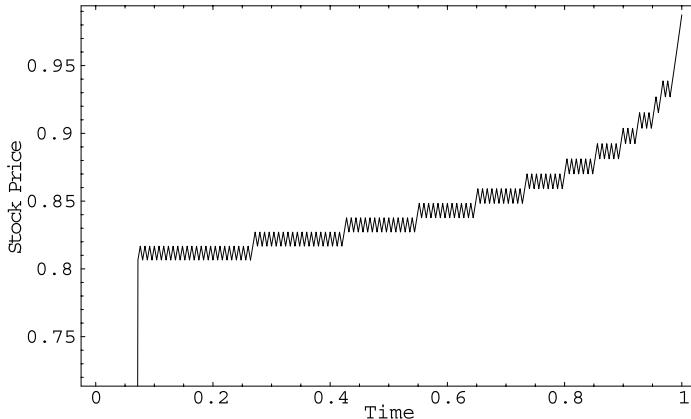


Fig. 17.1.7. Early exercise boundary for American put

Pricing on Other Trees

There are many versions of tree methods that have been suggested in the literature. This is understandable since there exist many possibilities to ensure weak convergence for the corresponding approximations.

In [Jarrow & Rudd \(1983\)](#) a binomial tree has been proposed with equal probabilities $p_d = p_u = \frac{1}{2}$ for an down and upward move

$$d = \exp \left\{ \left(r - \frac{\sigma^2}{2} \right) \Delta - \sigma \sqrt{\Delta} \right\} - 1,$$

$$u = \exp \left\{ \left(r - \frac{\sigma^2}{2} \right) \Delta + \sigma \sqrt{\Delta} \right\} - 1,$$

respectively. Here the behavior of the weak error is very similar to that of the CRR specification. Basically, only the center of the tree is slightly shifted. Experimentally, one also observes first weak order in this situation.

Let us now go further by allowing the tree to attain three possible values from each node at a new time step. This provides more freedom when constructing the tree. It also avoids the typical features that arise from odd and even time steps, as observed for a binomial tree. This raises the question whether the accuracy or order of weak convergence can be improved in such a trinomial tree.

To clarify this we form a trinomial tree, which is similar to the binomial tree. Its main property is that it is recombining, however, here we go to three nodes in each time step. This allows us to overcome the constraint of binomial trees where it is not possible to remain at the same level. Below we assign probabilities to the upward and the downward moves. With the remaining probability being that the tree stays at the present level. This means that the trinomial tree still remains very similar to the binomial tree, enriched by the

additional possibility of remaining at the current level, thus, a slightly finer Markov chain for approximating the underlying diffusion.

The underlying current asset price S will become in the next time step either $S(d + 1)$, S_m or $S(u + 1)$ where $d + 1 < m < u + 1$. One possible specification leads to the following steps

$$\begin{aligned} d &= e^{-\lambda \sigma \sqrt{\Delta}} - 1 \\ m &= 1 \\ u &= e^{\lambda \sigma \sqrt{\Delta}} - 1 \end{aligned} \tag{17.1.2}$$

with λ a free parameter. In a risk neutral setting, Boyle (1988) suggested for the above tree the following trinomial probabilities

$$\begin{aligned} p_d &= \frac{(W - R)(u + 1)^2 - (R - 1)(u + 1)^3}{u^2(u + 2)} \\ p_m &= 1 - p_d - p_u \\ p_u &= \frac{(W - R)(u + 1) - (R - 1)}{u^2(u + 2)} \end{aligned} \tag{17.1.3}$$

with

$$W = R^2 e^{\sigma^2 \Delta}$$

and

$$R = e^{r\Delta},$$

where r denotes the short rate and the underlying asset is denominated in the domestic currency. Note that for $\lambda = 1$ Boyle's trinomial tree model coincides with the original CRR binomial model.

The above probabilities and steps are chosen such that the first and second moment of the Black-Scholes model for the underlying security are matched, that is,

$$p_d(d + 1) + p_m m + p_u(u + 1) = R$$

and

$$p_d(d + 1)^2 + p_m m^2 + p_u(u + 1)^2 - R^2 = e^{2r\Delta}(e^{\sigma^2 \Delta} - 1).$$

Of course, other choices of trinomial probabilities are also possible. As we will see later on, one needs only to satisfy certain moment conditions to obtain some weak order of convergence, which allows for a large class of trinomial probabilities. The moment conditions are essentially those that we have already identified in Sect. 11.2 when studying simplified weak schemes.

For instance, Kamrad & Ritchken (1991) proposed the probabilities

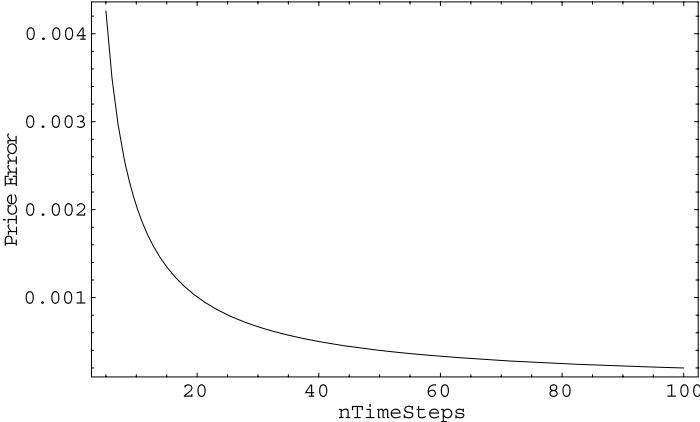


Fig. 17.1.8. Weak error for Boyle's trinomial at-the-money European call

$$\begin{aligned}
 p_d &= \frac{1}{2\lambda^2} - \frac{(r - \frac{\sigma^2}{2})\sqrt{\Delta}}{2\lambda\sigma} \\
 p_m &= 1 - \frac{1}{\lambda^2} \\
 p_u &= \frac{1}{2\lambda^2} + \frac{(r - \frac{\sigma^2}{2})\sqrt{\Delta}}{2\lambda\sigma}
 \end{aligned} \tag{17.1.4}$$

for $\lambda \geq 1$. Here one matches the first and second moment of the logarithm of the risk neutral asset price by the conditions

$$\begin{aligned}
 p_d \ln(d+1) + p_u \ln(u+1) &= \left(r - \frac{\sigma^2}{2}\right)\Delta \\
 p_d (\ln(d+1))^2 + p_u (\ln(u+1))^2 &\approx \sigma^2\Delta + \left(r - \frac{\sigma^2}{2}\right)^2\Delta^2.
 \end{aligned}$$

In the last condition the last term on its right hand side is of higher order and can still be neglected to obtain the previous probabilities. Numerical experiments show that if $p_m \approx \frac{1}{3}$, then errors in the approximation are somehow minimized.

For the same parameter settings, as used before for the binomial tree, and when applying Boyle's trinomial tree with $\lambda = \sqrt{3}$, we show in Fig. 17.1.8 for an at-the-money trinomial European call the weak error dependent on the number of time steps. It is remarkable that the odd-even oscillations, observed for the binomial tree, are now removed. In a corresponding log-log plot in Fig. 17.1.10 one realizes that Boyle's trinomial tree achieves experimentally the same order of weak convergence of about $\beta \approx 1.0$ as the binomial tree. We show in Fig. 17.1.9 for the same setting the weak error for an out-of-the money European call with strike $K = 1.1$. This figure reveals substantial oscillations of the absolute error, which generally decline as the number of steps increases.

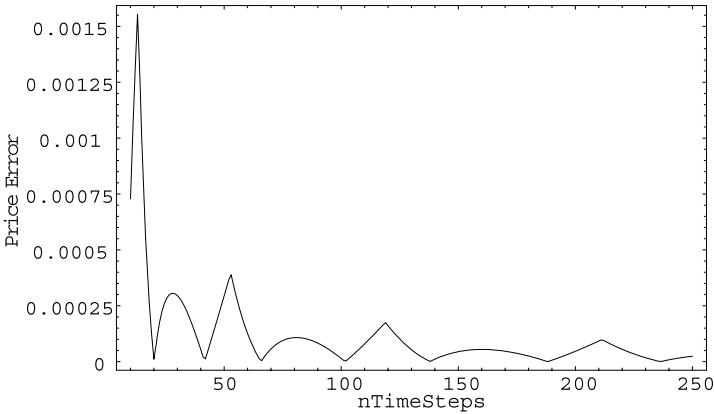


Fig. 17.1.9. Weak error for out-of-the money Boyle's trinomial European call with $K = 1.1$

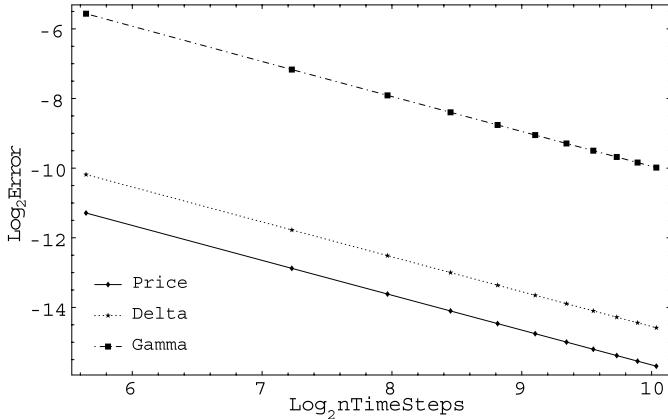


Fig. 17.1.10. Log-log plot of weak error for at-the-money price, delta and gamma of Boyle's trinomial European call

This behavior is a typical consequence of the discrete-time grid structure that a tree uses. We note that the odd-even oscillations, observable in Fig. 17.1.2, are not present here, removed by the use of a trinomial instead of a binomial tree structure.

For trinomial trees let us empirically study how the weak error for the hedge ratio, the delta, and its derivative, the gamma, decrease with increasing number of time steps. The weak error of price, delta and gamma for an at-the-money trinomial European call is shown in a log-log plot in Fig. 17.1.10 against the log of the number of time steps. This graph again suggests experimentally that the considered trinomial tree method is of a weak order of about $\beta \approx 1.0$, as was already observed for the binomial method. This means by introducing Boyle's trinomial tree, no higher order of weak convergence is obtained compared with the previously considered binomial trees. Having

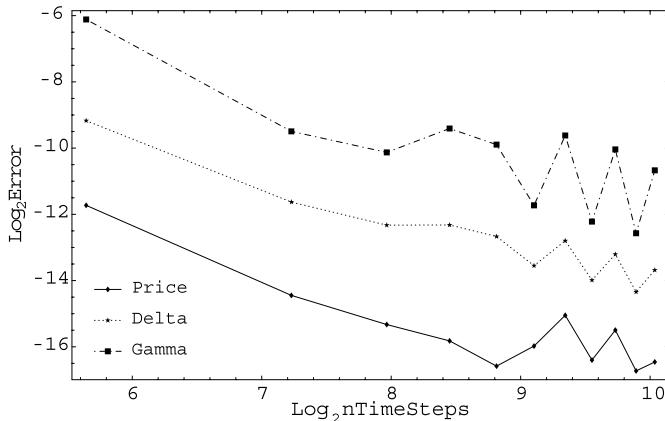


Fig. 17.1.11. Log-log plot of weak error for out-of-the money price, delta and gamma of Boyle's trinomial European call

three nodes in each time step smoothes the resulting pricing function when compared to that of a binomial tree. However, in the above case it does not improve the order of weak convergence. The reason for this will become clear when we formulate a corresponding weak convergence theorem.

As we have seen, for the at-the-money trinomial tree European option pricing there appear to be almost no oscillations in the prices, which is due to the matching of the strike with the corresponding node at maturity. In Fig. 17.1.11 we show a log-log plot for an out-of-the money trinomial European call with weak errors for price, delta and gamma. Here one notes oscillations when increasing the number of steps. If one adjusts the tree such that the in- or out-of-the money strike is located at a node, then one can reduce such in- and out-of-the money price oscillations.

Note, that this does not change the weak order of convergence of the method. These effects are of particular importance in the context of barrier options, see Boyle (1988). Obviously, American pricing can be performed on a trinomial tree similarly as on a binomial tree.

Black-Scholes Modification of a Tree

As we noticed above, the placing of the strike of an option at some node has an impact on the error behavior. Therefore it has been suggested in Broadie & Detemple (1997b) to smooth the call or put payoff function at-the-money and one step before maturity by using the corresponding price at that time obtained from the Black-Scholes formula. This *Black-Scholes modification* of a tree is then used to price options. To test this method we plot in Fig. 17.1.12 the resulting log-weak error for binomial European calls with and without the Black-Scholes modification against the log of the number of steps. It follows that there is no increase in the order of weak convergence.

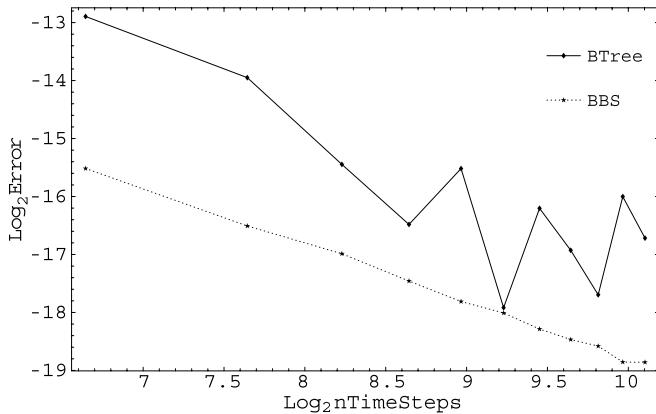


Fig. 17.1.12. Log-log plot of weak error with and without Black-Scholes modification for CRR binomial out-of-the money European call

However, the oscillations of the error, as seen in Fig. 17.1.4, are reduced and a better numerical performance of the method can be observed.

Smooth Payoff

Let us now check whether a smooth payoff, compared to the nondifferentiable European call or put payoff, provides some higher order of weak convergence. As a simple test case we consider the cubic payoff

$$g(S) = S^3$$

and compare the theoretical value

$$E(g(S_T)) = (S_0)^3 \exp\{3(r + \sigma^2)T\}$$

with that obtained from a trinomial tree. In Fig. 17.1.13 we show the log-log plot of the resulting absolute error dependent on the logarithm of the number of steps. One notes that there is no change in the order of weak convergence when using Boyle's trinomial method if compared to Fig. 17.1.10. The experimental weak order is about $\beta = 1.0$.

Now, let us consider the trinomial tree proposed in Heston & Zhou (2000). It uses

$$\begin{aligned} m &= \exp \left\{ \left(r - \frac{\sigma^2}{2} \right) \Delta \right\} \\ d &= m \exp \left\{ -\sigma \sqrt{3\Delta} \right\} - 1 \\ u &= m \exp \left\{ \sigma \sqrt{3\Delta} \right\} - 1 \end{aligned} \tag{17.1.5}$$

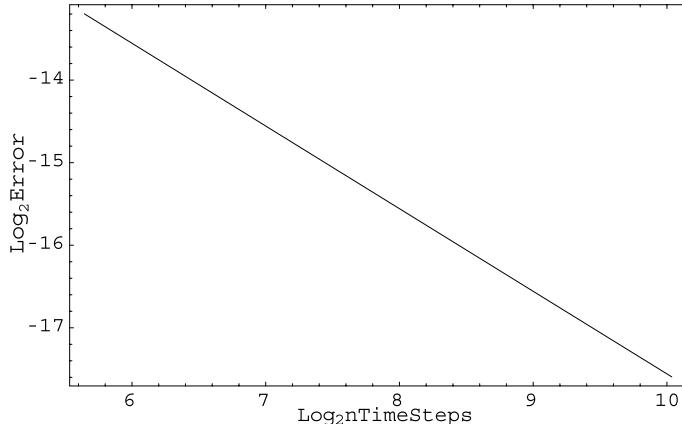


Fig. 17.1.13. Log-log plot for Boyle's trinomial tree with smooth payoff

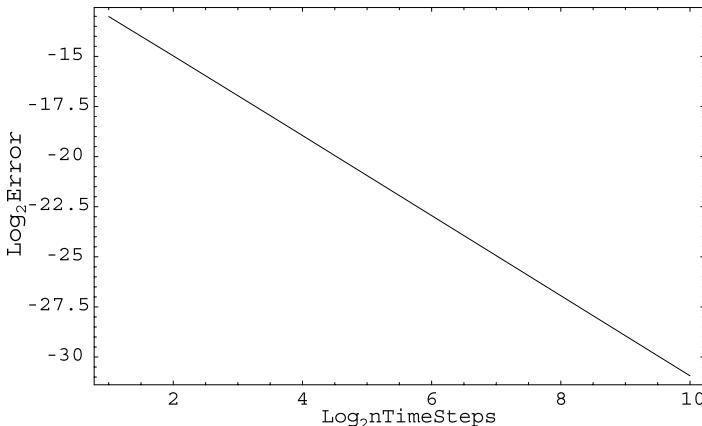


Fig. 17.1.14. Log-log plot for Heston-Zhou trinomial tree with smooth payoff

with probabilities

$$p_d = p_u = \frac{1}{6}. \quad (17.1.6)$$

In Fig. 17.1.14 we show the log-log plot of the weak error for the Heston-Zhou trinomial tree with the above smooth payoff. Here we note that we achieve experimentally the higher weak order of $\beta \approx 2.0$. We note that this particular tree matches exactly, for the logarithm of the given asset price model, those moments needed to achieve this weak order for smooth payoffs in accordance with what was shown in (11.2.8) for simplified weak schemes. Note that the probabilities and spatial steps in (17.1.5)–(17.1.6) resemble those in the three-point distributed random variables described in (11.2.9).

We also compute for this trinomial tree the weak error when calculating European calls. The corresponding log-log plot is shown in Fig. 17.1.15. This figure reveals that we obtain experimentally, due to the nonsmooth payoff,

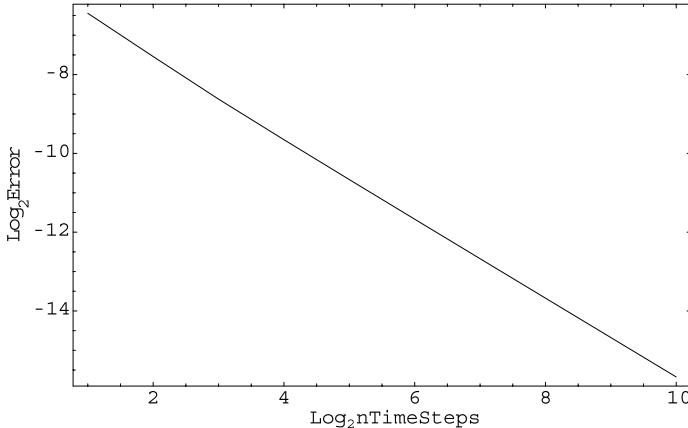


Fig. 17.1.15. Log-log plot for Heston-Zhou trinomial tree for European call

only a weak order of about $\beta \approx 1.0$. The reasoning for this effect will become clear when we discuss later a weak convergence theorem.

Pentanomial Trees

We have demonstrated empirically that the binomial and trinomial tree methods achieve, in general, a weak order of convergence of about $\beta \approx 1.0$. This raises the question of how one could systematically improve the weak order of convergence of tree methods. It turns out that the use of five nodes in each step in a tree or Markov chain approximation allows, quite generally, to approximate the transition probabilities to the extent which provides the weak order of convergence $\beta \approx 3.0$. One can construct a *pentanomial tree* similar to a trinomial tree by just adding at each time step two extra nodes that can be reached. As suggested in [Heston & Zhou \(2000\)](#), one may consider additionally to the steps d , m and u also d^* and u^* , where

$$\begin{aligned} m &= \exp \left\{ \left(r - \frac{\sigma^2}{2} \right) \Delta \right\} \\ d &= m \exp \left\{ -\sigma a \sqrt{\Delta} \right\} - 1 \\ u &= m \exp \left\{ \sigma a \sqrt{\Delta} \right\} - 1 \\ d^* &= m \exp \left\{ -3 \sigma a \sqrt{\Delta} \right\} - 1 \\ u^* &= m \exp \left\{ 3 \sigma a \sqrt{\Delta} \right\} - 1 \end{aligned} \tag{17.1.7}$$

with $a \approx 1.64947$. The probability to move to the direct lower and upper neighbors is each, set to $p_d = p_u \approx 0.181415$, and that to jump to the extreme

nodes is set to $p_{d^*} = p_{u^*} \approx 0.00261961$. The particular parameter choice is explained in [Heston & Zhou \(2000\)](#) and follows from solving the matching conditions for the moments.

With the same default parameters as before one can study the weak error of a pentanomial tree for the smooth payoff function $g(S_T) = S_T^3$. It turns out that experimentally one achieves a weak order of convergence of about $\beta \approx 3.0$. This raises the question. How many nodes in each step are necessary to achieve a certain weak order? This question will be answered by a convergence theorem which we present in the next section. Essentially, there must always be sufficiently many probabilities, which means many nodes, available to match the required moment conditions, which guarantee the desired weak order.

The results on Markov chain approximations that will be presented in the next section cover the above trees but also more general trees.

Extrapolation

Finally, let us explore in this section the possibilities of extrapolation when using trees. As was shown in Sect. 11.3, if the leading error term of a weak approximation is appropriate, then one can generate by extrapolation higher order weak schemes. To give an example, we consider the Richardson extrapolation, see (11.4.1). In this case we use approximations of weak order 1.0, with single and double time step size. Simply by combining the results using the formula

$$V_{g,2}^\Delta(T) = 2 E(g(Y_T^\Delta)) - E(g(Y_T^{2\Delta}))$$

one obtains a new valuation method, analogous to (11.4.1), that has theoretically weak order $\beta = 2.0$. For the above European call option pricing example we use the CRR binomial tree method and show in Fig. 17.1.16 the log-log plot of the resulting weak error (BTree) against the log of the time step size together with its Richardson extrapolation (Extrap). The regression of the curve for the error of the Richardson extrapolation reveals a slope of about $\beta \approx 1.0$, which indicates that only a first weak order is experimentally obtained. This is not what we theoretically expect. The reason for this is that the leading error coefficient for the considered nonsmooth payoff is not one. We also consider in Fig. 17.1.16, the Black-Scholes modification for a binomial tree (BBS) and its Richardson extrapolation (ExtrapBBS). Here we note a significant improvement due to extrapolation.

Now, let us study the case of a smooth payoff $g(S) = S^3$. Again with CRR binomial trees and Richardson extrapolation we calculate the weak error. In Fig. 17.1.17 we show the log-log plot of this error, which has now the slope of a second weak order method. This simple study demonstrates that extrapolation for a non-differentiable payoff may not be beneficial in the case of a European call option and may be questionable in other situations. However, for a smooth payoff the theoretically achievable order appears to be obtained.

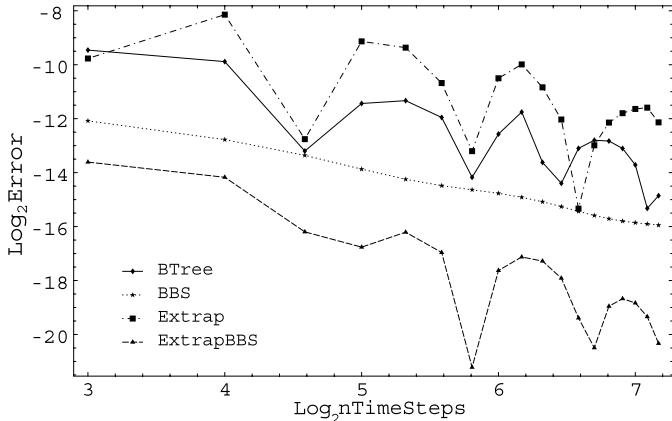


Fig. 17.1.16. Log-log plot for Richardson extrapolation with out-of-the money binomial call payoff and Black-Scholes modification

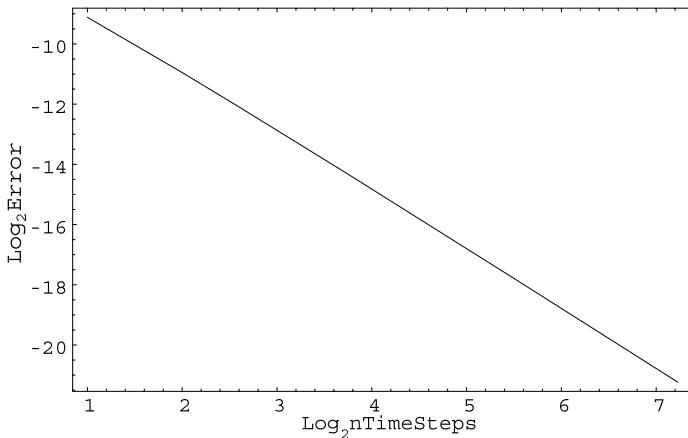


Fig. 17.1.17. Log-log plot for Richardson extrapolation with CRR binomial tree for smooth payoff

Also other weak first order methods can be extrapolated as long as their leading error coefficients are appropriate. To test this, we use the trinomial tree of [Heston & Zhou \(2000\)](#), which was introduced above and has second weak order for smooth payoffs. We apply it with step sizes Δ , 2Δ and 4Δ in the following extrapolation method

$$V_{g,4}^{\Delta}(T) = \frac{1}{21} \left[32 E(g(Y_T^{\Delta})) - 12 E(g(Y_T^{2\Delta})) + E(g(Y_T^{4\Delta})) \right],$$

see (11.4.2). In Fig. 17.1.18 we present a log-log plot for the absolute error of the resulting fourth order extrapolation to obtain the expectation for the smooth payoff $g(S) = S^3$. The figure shows that the slope of the log error curve is approximately four, which indicates that we experimentally have obtained a fourth weak order method. However, when one tries to extrapolate

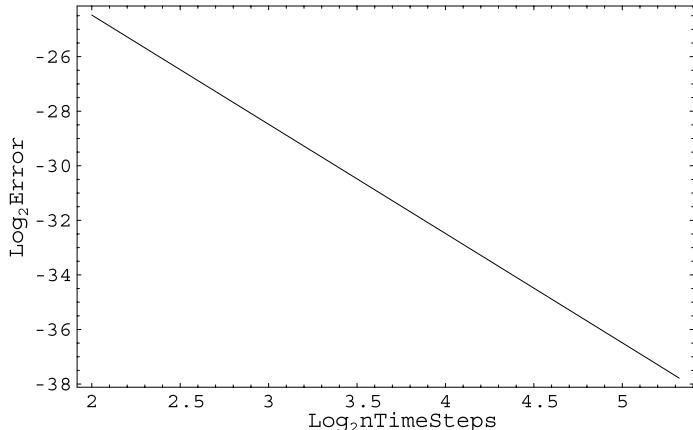


Fig. 17.1.18. Log-log plot for the weak error of a fourth order extrapolation using the second order Heston-Zhou trinomial tree for smooth payoff

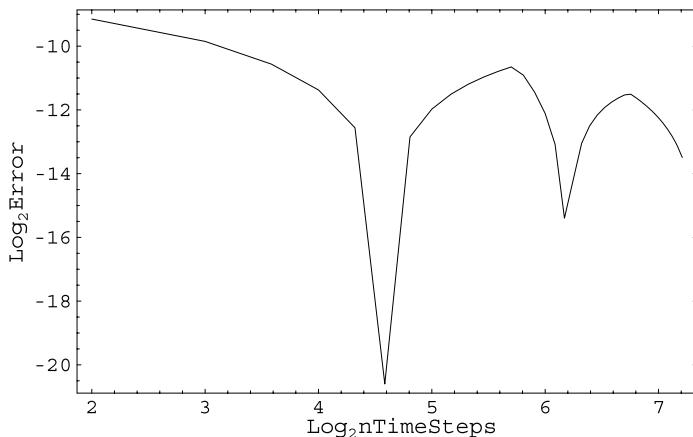


Fig. 17.1.19. Log-log plot for a fourth order extrapolation error for out-of-the money Heston-Zhou trinomial European calls

with this trinomial extrapolation method, European call option prices, this extrapolation method fails because of the nonsmooth payoffs, as can be seen from Fig. 17.1.19. This highlights the fact that extrapolation requires smooth payoffs to be efficient.

17.2 Efficiency of Simplified Schemes

The purpose of this section is to study the numerical efficiency of tree methods and to compare these with those of simplified weak schemes for the weak approximation of solutions of SDEs. In simplified schemes and trees, discrete random, instead of Gaussian increments, are used to approximate multiple

stochastic integrals from the Wagner-Platen expansion. This section presents results from [Bruti-Liberati & Platen \(2004\)](#) and shows that an implementation of simplified schemes based on random bit generators can significantly increase the computational speed. Simplified schemes also work in higher dimensions where trees have problems because of storage capacity.

Simplified Schemes and Trees

As described previously, to price a contingent claim, one does not require a pathwise approximation of the solution of the underlying SDE. Only an approximation of its probability distribution has to be constructed. Thus, an appropriate notion of convergence for such an approximation is the weak convergence, given in [\(11.1.2\)](#), instead of the strong convergence, described in [\(5.2.25\)](#). We have also shown above that in order to achieve a certain order of weak convergence one can approximate the random variables in a weak Taylor scheme by appropriate multi-point distributed random variables. For instance, in a Monte Carlo simulation one can employ, instead of a Gaussian increment in an Euler scheme, a much simpler two-point distributed random variable. In general, the simplified random variables have to match only some lower order moments of the random variables appearing in the corresponding Taylor schemes, as explained in [Chap. 11](#). For example, in the case of a weak Taylor scheme of first weak order, to construct a corresponding first order simplified method one can use a two-point distributed random variable. Similarly, it is sufficient to use a trinomial tree to achieve, in general, the weak order $\beta = 1.0$.

We emphasize that the simplified Euler method is equivalent to some kind of a random walk, which again is in some sense equivalent to a tree if the states are chosen such that they reside on a certain set of points. The possible states of the tree and that of a simplified Euler scheme can be made the same, if one chooses the random variables appropriately. However, while the tree can be used for a deterministic backward calculation of prices, the simplified method is a forward method which generates paths for a Monte Carlo estimator. As we will see, the numerical properties of simplified methods are similar to those of trees because of their similarity in approximating the underlying probabilistic structure. The widespread application of the tree methodology in finance motivated the following study of simplified schemes. The similarity between simplified schemes and tree methods helps in the understanding of the numerical properties for both types of methods. Later we will see that some finite difference methods for the corresponding partial differential equations (PDEs) are equivalent to trees, thus, can also be better understood by exploiting these similarities.

Simplified schemes, being forward algorithms, are not easily suited to the pricing of American options, even though corresponding algorithms have been developed involving such schemes, see for instance [Broadie & Glasserman \(1997b\)](#) and [Longstaff & Schwartz \(2001\)](#). Moreover, with simplified schemes one always has to face the typical statistical error from Monte Carlo simula-

tion. The major advantages of simplified schemes over tree methods are their flexibility and general applicability in higher dimensions.

As shown previously, implicit simplified methods can overcome certain numerical instabilities. Random bit generators, which will be presented below, can be efficiently applied to implicit schemes, while tree methods cannot easily be made implicit. On the other hand, simplified schemes can be interpreted as being, in some sense, equivalent to implicit finite difference methods. However, as with trees, these methods cannot be easily implemented for higher dimensions and one has an advantage when using simplified implicit methods with variance reduction in higher dimensions.

The order of convergence of simplified schemes is independent of the dimension of the problem. As shown in [Boyle, Broadie & Glasserman \(1997\)](#), simulation methods typically become more efficient than tree or finite difference methods around dimension three or four.

Weak Taylor Schemes and Simplified Methods

For the dynamics of the underlying security let us consider the scalar SDE

$$dX_t = a(t, X_t) dt + b(t, X_t) dW_t \quad (17.2.1)$$

for $t \in [0, T]$, with $X_0 \in \mathfrak{R}$. A derivative pricing problem consists of computing an expectation of a payoff function $g(X_T)$ of the solution of the SDE (17.2.1) at maturity T .

As before, let us assume an equidistant time discretization with n th discretization time $t_n = n\Delta$ for $n \in \{0, 1, \dots, N\}$, where $\Delta = \frac{T}{N}$ and $N \in \{1, 2, \dots\}$. As a set of test functions we use the space $\mathcal{C}_P^\ell(\mathfrak{R}, \mathfrak{R})$, of the ℓ times continuously differentiable functions g , which together with their partial derivatives of orders up to and including ℓ , have polynomial growth.

As shown previously, if one uses in the Euler scheme, instead of Gaussian random variables, simpler two-point distributed random variables, then one still obtains the same weak order of convergence $\beta = 1.0$. For the Euler method these simpler random variables have to coincide in their first three moments with those of the Gaussian Wiener process increments $\Delta W_n = W_{t_{n+1}} - W_{t_n}$. For instance, this permits us to replace the Gaussian increment ΔW_n , by a two-point distributed random variable $\Delta \hat{W}_n$, where

$$P(\Delta \hat{W}_n = \pm \sqrt{\Delta}) = \frac{1}{2}. \quad (17.2.2)$$

We then obtain the simplified weak Euler scheme

$$Y_{n+1} = Y_n + a(t_n, Y_n) \Delta + b(t_n, Y_n) \Delta \hat{W}_n, \quad (17.2.3)$$

see (11.2.2). Here the first three moments of the Wiener process increments ΔW_n match those of $\Delta \hat{W}_n$, that is

$$\begin{aligned} E(\Delta W_n) &= E(\Delta \hat{W}_n) = 0 & E((\Delta W_n)^2) &= E((\Delta \hat{W}_n)^2) = \Delta \\ E((\Delta W_n)^3) &= E((\Delta \hat{W}_n)^3) = 0. \end{aligned} \quad (17.2.4)$$

This property satisfies condition (11.2.3). A similar result applies to the weak order 2.0 Taylor scheme

$$\begin{aligned} Y_{n+1} &= Y_n + a \Delta + b \Delta W_n + \frac{1}{2} b' b \{ (\Delta W_n^2) - \Delta \} + \frac{1}{2} \left(a a' + \frac{1}{2} a'' b^2 \right) \Delta^2 \\ &\quad + a' b \Delta Z_n + \left(a b' + \frac{1}{2} b'' b^2 \right) \{ \Delta W_n \Delta - \Delta Z_n \}, \end{aligned} \quad (17.2.5)$$

where ΔZ_n represents the double Itô integral

$$\Delta Z_n = \int_{t_n}^{t_{n+1}} \int_{t_n}^z dW_z ds,$$

see (11.2.6), and we suppress again the dependence of the coefficient functions on (t_n, Y_n) . Here we can replace the Gaussian random variables ΔW_n and ΔZ_n by expressions that use a three-point distributed random variable $\Delta \hat{W}_n$ with

$$P(\Delta \hat{W}_n = \pm \sqrt{3\Delta}) = \frac{1}{6}, \quad P(\Delta \hat{W}_n = 0) = \frac{2}{3},$$

see (11.2.9). Then we obtain the simplified weak order 2.0 Taylor scheme

$$\begin{aligned} Y_{n+1} &= Y_n + a \Delta + b \Delta \hat{W}_n + \frac{1}{2} b b' \left\{ (\Delta \hat{W}_n)^2 - \Delta \right\} + \frac{1}{2} \left(a a' + \frac{1}{2} a'' b^2 \right) \Delta^2 \\ &\quad + \frac{1}{2} \left(a' b + a b' + \frac{1}{2} b'' b^2 \right) \Delta \hat{W}_n \Delta, \end{aligned} \quad (17.2.6)$$

see (11.2.7). Since the three-point distributed random variable $\Delta \hat{W}_n$ is such that the first five moments of the increments of the schemes (17.2.5) and (17.2.6) are matched, see condition (11.2.8), the scheme (17.2.6) can be shown to achieve the weak order $\beta = 2.0$.

By using four or even five point distributed random variables when approximating the multiple stochastic integrals needed, we can obtain simplified weak Taylor schemes of weak order $\beta = 3.0$ or 4.0 , respectively, as shown in Kloeden & Platen (1999) and Hofmann (1994).

An important issue with weak approximations for SDEs is their numerical stability. As noticed in Chap. 14, when considering a test equation with multiplicative noise, the weak schemes described above show narrow regions of asymptotic p -stability. In order to improve the numerical stability of such methods one needs to introduce implicitness in the diffusion terms. This leads, for instance, to the fully implicit weak Euler scheme

$$Y_{n+1} = Y_n + \left\{ a(t_{n+1}, Y_{n+1}) - b(t_{n+1}, Y_{n+1}) \frac{\partial}{\partial y} b(t_{n+1}, Y_{n+1}) \right\} \Delta + b(t_{n+1}, Y_{n+1}) \Delta W_n, \quad (17.2.7)$$

which makes no sense for certain diffusion coefficients, as shown in (7.3.3) and (7.3.4). However, under the weak convergence criterion one can employ the two-point distributed random variable $\Delta \hat{W}_n$, see (17.2.2), instead of ΔW_n to obtain the simplified fully implicit Euler scheme

$$Y_{n+1} = Y_n + \left\{ a(t_{n+1}, Y_{n+1}) - b(t_{n+1}, Y_{n+1}) \frac{\partial}{\partial y} b(t_{n+1}, Y_{n+1}) \right\} \Delta + b(t_{n+1}, Y_{n+1}) \Delta \hat{W}_n, \quad (17.2.8)$$

that achieves an order $\beta = 1.0$ of weak convergence. It does not have the problems that the scheme (17.2.7) may exhibit due to the Gaussianity of ΔW_n .

Random Bit Generators

Let us now demonstrate, for simplified schemes, how to implement highly efficient random bit generators that exploit the architecture of a digital computer or can be used on parallel working hardware devices. The crucial advantage of simplified schemes is the fact that only random bits have to be generated instead of continuous random variables. These substitute the computationally demanding Gaussian random numbers needed for several weak Taylor schemes.

Within this book we have not focussed on random or quasi-random number generation. For this topic we refer to a large literature which includes Devroye (1986), Bratley et al. (1987), Niederreiter (1988b, 1992), L'Ecuyer (1994, 1999), Fishman (1996), Pagés & Xiao (1997), Matsumoto & Nishimura (1998), Brent (2004) and Panneton, L'Ecuyer & Matsumoto (2006).

A well-known and efficient method to generate a pair of independent standard Gaussian random variables is the polar Marsaglia-Bray method coupled with a linear congruential random number generator, as described in Press, Teukolsky, Vetterling & Flannery (2002). The following comparative study uses, as Gaussian random number generator, the routine `gasdev`, see Press et al. (2002).

For the simplified Euler scheme (17.2.3) and simplified fully implicit Euler scheme (17.2.8) we use a two-point distributed random variable at each time step, which is obtained from a *random bit generator*. This generator is an algorithm that generates a single bit 0 or 1 with probability 0.5. The method proposed is based on the theory of primitive polynomials modulo 2. These are polynomials satisfying particular conditions whose coefficients are zero or one. The important property is that every primitive polynomial modulo 2 of order n defines a recurrence relation for obtaining a new bit from the n preceding ones with maximal length. This means that the period length of the recurrence

relation is equal to $2^n - 1$. For a study on random number generators based on primitive polynomials modulo 2 we refer to [Tausworthe \(1965\)](#).

Since the random number generator for the polar Marsaglia-Bray method has a period of 2^{31} we use a random bit generator based on the following primitive polynomial modulo 2 of order 31:

$$y(x) = x^{31} + x^3 + 1. \quad (17.2.9)$$

An implementation of this generator is reported in [Bruti-Liberati & Platen \(2004\)](#). This method is extremely fast and suitable for hardware implementation. On the test computer the two-point random bit generator, described above, was almost 30 times faster than for the polar Marsaglia-Bray method.

For simplified methods of higher weak order similar multi-point random bit generators can be constructed. For the second order simplified method (17.2.6) it is sufficient to use a three-point random bit generator. A corresponding code is described in [Bruti-Liberati & Platen \(2004\)](#). It produces three bits coupled with an acceptance-rejection method. On the test computer the central processor unit (CPU) time needed is still five times less than for the polar Marsaglia-Bray method.

Now, we present some numerical results for the Euler, fully implicit Euler and weak order 2.0 Taylor schemes as well as their simplified versions. As test dynamics we choose again an SDE with multiplicative noise of the Black-Scholes type, where

$$dX_t = \mu X_t dt + \sigma X_t dW_t \quad (17.2.10)$$

for $t \in [0, T]$, see (13.7.1). Recall that the SDE admits a closed form solution, see (13.7.2). The CPU times needed to compute approximate paths with 64 time steps with the Euler, fully implicit Euler and weak order 2.0 Taylor schemes are compared to: For the Euler method the simplified version was roughly 28 times faster and the simplified fully implicit method was roughly 18 times faster than their Gaussian counterparts. For the second order simplified method one finds that it is roughly four times more efficient than the weak order 2.0 Taylor scheme.

A Smooth Payoff Function

We now analyze the weak convergence of Monte Carlo simulations when using a smooth payoff function. Here we choose simply the first moment for illustration since for geometric Brownian motion the higher moments behave similarly. In addition, we consider below a non smooth payoff, which will be that of a European call option.

First, we study the weak error for a fixed number of simulations and time steps. We also compare the CPU time needed to reach a given accuracy. In order to analyze the weak error $\mu(\Delta)$ given in (11.1.2), we run sufficiently many simulations such that the statistical error can be neglected. We use in the SDE (17.2.10) the following parameters: $X_0 = 1$, $\mu = 1.5$, $\sigma = 0.01$, $T = 1$.

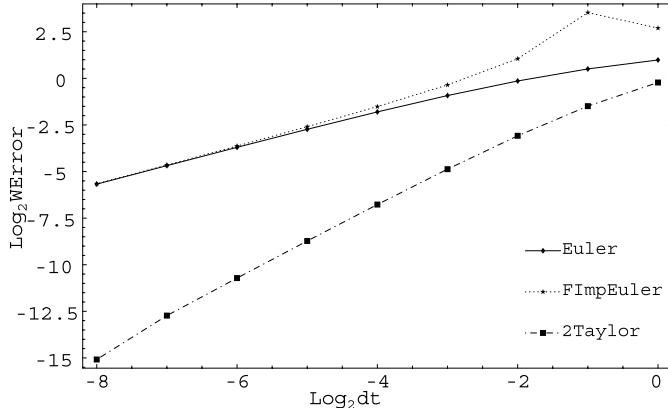


Fig. 17.2.1. Log-log plot of the weak error for the Euler, fully implicit Euler and weak order 2.0 Taylor schemes

An important application of Monte Carlo simulation is the calculation of Value at Risk (VaR) via the simulation of moments, as applied in Edgeworth expansions and saddle point methods. This is another reason why, as a simple test function, the first moment $E(X_T)$ of X_T is of interest. Other higher moments give similar numerical results due to the lognormal structure of the given example. We then estimate the weak error of the first moment by comparing the simulated Monte Carlo estimate, see (11.1.3), with the exact expectation $E(X_T) = X_0 e^{\mu T}$.

In Fig. 17.2.1 we show the logarithm $\log_2(\mu(\Delta))$ of the weak error for the Euler, fully implicit Euler, and weak order 2.0 Taylor method versus the logarithm $\log_2(\Delta)$ of the time step size. The number of simulated paths was again sufficiently high such that confidence intervals do not appear in Fig. 17.2.1.

We emphasize the important observation that the simplified methods achieve almost exactly the same accuracy as their Taylor counterparts. Note in Fig. 17.2.1 that the Euler and the fully implicit Euler scheme reproduce in the log-log plot the theoretically predicted weak order $\beta = 1.0$. Furthermore, the weak order 2.0 Taylor scheme achieves a weak order of about $\beta = 2.0$, as expected. Moreover, note in Fig. 17.2.1 that the fully implicit Euler scheme shows poor results for very large step sizes. However, as already explained in Chap. 14, the fully implicit method has better numerical stability properties than the explicit schemes, even for larger time step sizes.

What really matters, in practice, is the time needed to reach a given level of accuracy. In Fig. 13.7.7 we plotted the logarithm of the CPU time versus the negative of the logarithm of the weak error observed for the three methods described above and their simplified versions. Since the accuracy for a given time step size is almost identical for the schemes of the same order, the increase in efficiency simply reflects the fact that simplified schemes are computationally less intensive than their Gaussian counterparts. We recall that,

for instance, the simplified Euler scheme was 28 times faster than the Euler scheme.

By comparing all six methods, we conclude that the second order simplified scheme is significantly more efficient for the given example than any of the other schemes considered. This result reveals that efficient Monte Carlo simulation algorithms for smooth payoffs are obtained by using higher order simplified schemes.

An Option Payoff

In option pricing we are confronted with the computation of expectations of non smooth payoffs. Close to maturity and at-the-money, a European call option pricing function does not have the smoothness that is required in Theorem 11.2.1 for first order weak convergence. To give a simple example, let us compute the price of a European call option. Here we have a continuous but only piecewise differentiable payoff $(X_T - K)^+ = \max(X_T - K, 0)$ with strike price K . For the option price at time $t = 0$ we obtain, by the Black-Scholes formula,

$$E(e^{-rT}(X_T - K)^+) = X_0 N(d_1) - e^{-rT} K N(d_2),$$

where $d_1 = \frac{\ln(\frac{X_0}{K}) + (r + \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}$ and $d_2 = d_1 - \sigma\sqrt{T}$, see (13.7.16). For this non smooth payoff we study the absolute weak error for the Euler and the simplified Euler method, assuming the volatility $\sigma = 0.2$ and the short rate $r = 0.1$. Simulation experiments show that we observe no major gain by using higher order methods, which is due to the non smooth option payoff. Since the second order simplified method (17.2.6) is approximately equivalent to a trinomial tree, as discussed previously, this is consistent with an observation in Heston & Zhou (2000), where it was mentioned that in option pricing the order of convergence of trinomial trees is not superior to that of binomial trees.

In Fig. 17.2.2 we show the log-log absolute weak error plot for an at the money-forward option, with strike $K = X_0 e^{rT}$. The Euler method generates a weak order $\beta = 1.0$ with the log error forming a perfect line dependent on the log step size. As mentioned earlier, the simplified Euler method is approximately equivalent to a binomial tree. This method still achieves a weak order $\beta = 1.0$. However, its log-log error plot does not exhibit a perfect line, due to the discrete nature of the random variables used. This appears to be the same effect as we noticed for tree methods, see also Boyle & Lau (1994). We observed, for in the money and out of the money options, a similar order of convergence with similar log error patterns.

In Fig. 17.2.3 we show for the simplified Euler method the logarithm of the CPU time versus the negative logarithm of the weak error. For the considered non smooth payoff the increase in computational speed is still about 28 times. The simplified Euler method is significantly more efficient than the Euler scheme for every level of accuracy. We observed similar results also for in-the-money and out-of-the-money options. In summary, one can say that

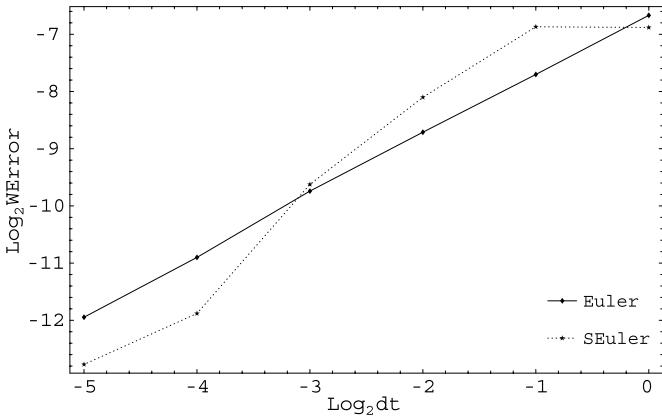


Fig. 17.2.2. Log-log plot of weak error for call option with Euler and simplified Euler scheme

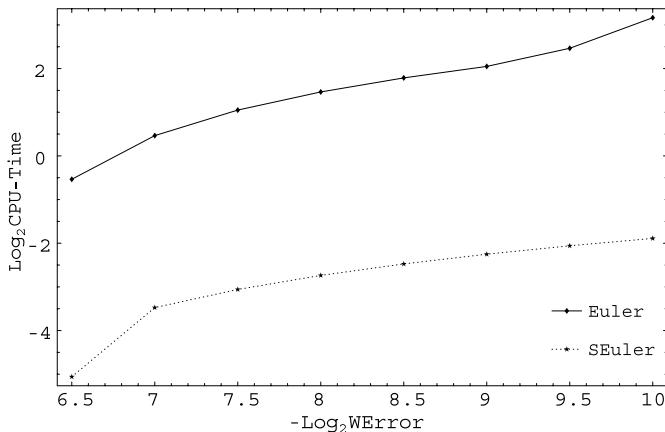


Fig. 17.2.3. Log-log plot of CPU time versus weak error for call option with Euler and simplified Euler scheme

the proposed, rather simple, random bit generators when combined with simplified schemes can significantly enhance the efficiency of typical Monte Carlo simulations in finance. It is worth emphasizing that these methods also work in higher dimensions and for implicit schemes. Since parallel hardware devices will become rather inexpensive in future, one can expect a significant increase of their use in Monte Carlo simulation, extending considerably the applicability of the kind of methods emphasized in this book.

17.3 Higher Order Markov Chain Approximations

This section presents a method that permits the systematic construction of discrete-time Markov chains for the weak approximation of solutions of SDEs.

The transition probabilities of Markov chains are chosen in such a way that functionals converge with a desired weak order with respect to vanishing step size.

Limits of Markov Chains

We have already seen in Chap. 3 how CRR binomial trees weakly approximate a given diffusion process, the geometric Brownian motion. Many problems in finance and economics can be on a basic level conveniently modeled by Markov chains. When small time steps are involved, then these Markov chains usually approximate limiting solutions of corresponding Itô SDEs. When an SDE is numerically approximated by trees or Markov chains, then one constructs an approximate characterization of the probability measure of the corresponding process. The same applies to simulation algorithms and even to finite difference methods for PDEs, as will be shown below. In all cases the validity and accuracy of such approximations are of importance. For Monte Carlo methods we have already discussed their weak order of convergence. A similar approach was indicated previously for trees, which are Markov chain approximations.

To illustrate the similarities in a simple example, let us examine the weak convergence of a sequence of random walks to a standard Wiener process on the time interval $[0, 1]$. This is one of the simplest cases, where a sequence of Markov chains approximates a diffusion process in a weak sense. For each $N \in \{1, 2, \dots\}$ we partition the time interval $[0, 1]$ at the instants $t_k = k\Delta$ with $\Delta = \frac{1}{N}$ for $k \in \{0, 1, \dots, N\}$. Then we generate a random walk Y^Δ , starting at $Y_0^\Delta = 0$ with equally probable spatial steps of height $\sqrt{\Delta} = N^{-\frac{1}{2}}$ or $-\sqrt{\Delta}$, for jumping up or down, respectively, from the current position $Y_{k\Delta}^\Delta$ to the next position $Y_{k\Delta+\Delta}^\Delta$ at the time instant t_{k+1} . That is, we evaluate

$$Y_{k\Delta+\Delta}^\Delta = Y_{k\Delta}^\Delta + \zeta_k \sqrt{\Delta} \quad (17.3.1)$$

for $k \in \{0, 1, \dots, N\}$. Here the ζ_k are independent two-point distributed random variables taking the values $+1$ and -1 with equal probabilities. We interpolate the random variables $Y_{k\Delta}^\Delta$ in a piecewise constant manner on the time intervals $[t_k, t_{k+1})$ for $k \in \{0, 1, \dots, N-1\}$ to define a process $Y^\Delta = \{Y_{n_t\Delta}^\Delta, t \in [0, 1]\}$. Here n_t is again the largest index n of the discretization times t_n before time t . We can then use the Central Limit Theorem to conclude that $Y_{n_t\Delta}^\Delta$ converges in distribution to the Gaussian distributed value W_t of a standard Wiener process W for $t \in [0, 1]$ as $N \rightarrow \infty$ or equivalently $\Delta \rightarrow 0$.

Clearly, the initial value W_0 of W_t for $t = 0$, is zero. The increments $W_{t+h} - W_t$ are Gaussian with mean zero and variance h . For disjoint intervals the resulting increments are also independent. Thus, by Definition 1.1.3 of a Wiener process it can be confirmed that the limiting continuous stochastic process is a standard Wiener process.

It is obvious that the $Y_{n\Delta}^{\Delta}$, generated by (17.3.1), for $n \in \{1, 2, \dots, N\}$ are the realizations of a Markov chain on the countably infinite state space

$$\mathcal{X}_{\sqrt{\Delta}}^1 = \left\{ x_k = k \sqrt{\Delta} : k \in \{0, \pm 1, \pm 2, \dots\} \right\},$$

which is called the $\sqrt{\Delta}$ -grid, with time homogeneous transition probabilities $p^{i,i-1} = p^{i,i+1} = \frac{1}{2}$ and $p^{i,j} = 0$ otherwise, for all $i, j \in \{0, \pm 1, \pm 2, \dots\}$, where $p^{i,j} = P(Y_{(k+1)\Delta}^{\Delta} = x_j | Y_{k\Delta}^{\Delta} = x_i)$. The above description also fits into that of a simplified Euler scheme for the trivial SDE of a Wiener process or a simple symmetric binomial tree, as described previously. In a weak sense these approximations of the standard Wiener process are somehow equivalent.

In the above random walk the spatial step size was the same at every point of the $\sqrt{\Delta}$ -grid $\mathcal{X}_{\sqrt{\Delta}}^1$. A natural generalization is to allow the Markov chain to depend on the current location $Y_{n\Delta}^{\Delta}$ with the next position being determined by

$$Y_{n\Delta+\Delta}^{\Delta} = Y_{n\Delta}^{\Delta} + a_{\Delta}(Y_{n\Delta}^{\Delta}) \Delta + b_{\Delta}(Y_{n\Delta}^{\Delta}) \sqrt{\Delta} \zeta_n \quad (17.3.2)$$

for given functions a_{Δ} and b_{Δ} defined on $\mathcal{X}_{\sqrt{\Delta}}^1$ such that

$$x + a_{\Delta}(x) \Delta \pm b_{\Delta}(x) \sqrt{\Delta} \in \mathcal{X}_{\sqrt{\Delta}}^1$$

falls again on the grid $\mathcal{X}_{\sqrt{\Delta}}^1$ for all $x \in \mathcal{X}_{\sqrt{\Delta}}^1$. This forms a recombining tree or lattice. The functions a_{Δ} and b_{Δ} can be derived rather freely and need to converge to the limiting functions a and b , respectively, on \mathfrak{N} in an appropriate manner where moments of increments are matched, as we will discuss below. In either case we can form an Itô SDE

$$dX_t = a(X_t) dt + b(X_t) dW_t \quad (17.3.3)$$

for $t \in [0, 1]$, the solution could become the limiting diffusion process if the coefficient functions a and b are sufficiently regular and the initial value X_0 is appropriate. In general, the increments $Y_{n\Delta+\Delta}^{\Delta} - Y_{n\Delta}^{\Delta}$ of the Markov chains defined by (17.3.2) will not be independent identically distributed random variables. So we cannot use the Central Limit Theorem to conclude that the approximations converge in distribution to a diffusion process satisfying (17.3.3). Some weak convergence can nevertheless still be established, but only if additional properties are satisfied by the Markov chain. Recall that we have already established weak convergence for simplified Euler schemes earlier in the book.

The concept of Markov chain approximations also applies to more complicated types of driving noises, including jump type noise, than the two-point random variables considered so far. Other relationships between the spatial and temporal step sizes may also be used. Furthermore, multi-dimensional Markov chains can be considered with great flexibility. Finally, a discrete-time approximating Markov chain does not even need to be recombining and

has not to live on a grid. What will be important under the convergence theorem, which we present at the end of the chapter, is that certain moments of the increments of the Markov chain match those of the limiting process. In more general cases it may, however, be difficult or cumbersome to write down explicitly in all detail the state space and the transition matrix for a particular Markov chain. Other descriptions may be more convenient. The trees of the previous section and the simplified Monte Carlo simulation schemes of Chap. 11 provide some examples. The following analysis will provide weak orders of convergence for discrete-time Markov chains that can be achieved when certain conditions are satisfied.

Markov Chain Approximations as Numerical Methods

As already indicated above, a usable and intuitively appealing numerical method for the approximation of functionals of diffusions and also jump diffusions are Markov chain approximations. These are, for instance, extensively discussed and applied in [Kushner \(1977\)](#) and [Kushner & Dupuis \(1992\)](#). The essential idea involves an approximation of the original process, which solves approximately a given SDE by a suitable Markov chain for which the corresponding approximate functional converges weakly to that of the diffusion process. Consistency conditions can ensure a type of weak convergence, as described in [Billingsley \(1968\)](#) and [Kushner \(1977\)](#), for the approximating Markov chain. Therefore, the convergence of a wide class of functionals is guaranteed. As early as in [Kushner \(1977\)](#), Markov chains that jump not only to the nearest neighbor but also to the next nearest neighbor nodes are discussed. Such approximations are expected to give somewhat better weak convergence. This refers, for instance, to the above described pentanomial trees when compared with binomial trees. The aim of this section is to present a flexible methodology that allows the systematic construction of approximate Markov chains for a range of numerical tasks. Following [Platen \(1992\)](#), a rich class of sufficiently smooth functionals will be shown to converge with given weak order to those of the underlying process if the maximum time step size converges to zero.

As in Monte Carlo simulation we will see that higher order Markov chain approximations require more smoothness of the drift and diffusion coefficients and also a more precise approximation of the transition probability of the increments of the diffusion by those of the Markov chain. The convergence of moments is again automatically covered by the weak convergence criterion being used.

It will turn out for Markov chains, as proposed in [Kushner \(1977\)](#) and [Kushner & Dupuis \(1992\)](#), and for binomial and trinomial trees that sufficient smoothness of the drift and diffusion coefficients, as well as of the payoff function, will give first order weak convergence. We will also explicitly describe examples of second- and third order Markov chain approximations. These can then be used to construct corresponding multinomial trees.

The main theorem will be formulated under rather general assumptions to provide the corresponding conditions for a desired order of weak convergence. This result will also be applicable to other discrete-time approximations of diffusions. Its proof is based on the Wagner-Platen expansion and uses substantially the smoothness of the functionals.

Higher Order Weak Approximation

Let us recall the typical setup under which we work for the remainder of this section. As usual let $(\Omega, \mathcal{A}_T, \mathcal{A}, P)$ denote a filtered probability space. We consider the d -dimensional process $\mathbf{X} = \{\mathbf{X}_t, t \in [0, T]\}$ defined by the Itô stochastic equation

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t \mathbf{a}(\mathbf{X}_s) ds + \int_0^t \mathbf{b}(\mathbf{X}_s) dW_s, \quad (17.3.4)$$

$t \in [0, T]$, where $\mathbf{X}_0 \in \Re^d$ denotes the initial value at time $t = 0$ and $\mathbf{W} = \{\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top, t \in [0, T]\}$ is an m -dimensional \mathcal{A} -adapted standard Wiener process. The vector $\mathbf{a}(\cdot) = [a^k(\cdot)]_{k=1}^d = (a^1(\cdot), \dots, a^d(\cdot))^\top$ denotes the d -dimensional drift vector function and $\mathbf{b}(\cdot) = [b^j(\cdot)]_{j=1}^m = [b^{k,j}(\cdot)]_{k,j=1}^{d,m}$ the $d \times m$ diffusion matrix function.

We assume throughout the remainder of this section that \mathbf{a} and \mathbf{b} are Lipschitz continuous and satisfy a linear growth condition such that the existence and uniqueness of the solution of equation (17.3.4) is guaranteed, see Chap. 1. Furthermore, we suppose that all initial moments

$$E(|\mathbf{X}_0|^q) < \infty$$

are finite for $q \in \{1, 2, \dots\}$ to ensure the existence of the moments of \mathbf{X}_t for $t \in [0, T]$.

As we have seen in option pricing, an important role is played by functionals of the form

$$u(s, \mathbf{x}) = E(g(\mathbf{X}_T^{s,x})), \quad (17.3.5)$$

where $\mathbf{X}_t^{s,x}$ is the value of the process at time $t \in [0, T]$ when starting at $s \in [0, t]$ in $\mathbf{x} \in \Re^d$ and g is a suitable real valued function. Many different attempts have been made to approximate such functionals by the use of discrete-time approximations of the process \mathbf{X} . The following approach will allow a unified treatment of several of these methods.

Within the following, we construct Markov chains and characterize more general discrete-time approximations that at time T converge with a given weak order $\beta \in \{1, 2, \dots\}$ with respect to vanishing step size $\Delta \rightarrow 0$.

Moments of Increments of Weak Taylor Schemes

To obtain systematically the appropriate conditions that will guarantee a desired weak order of convergence, we will construct Markov chains that

match the conditional moments of the corresponding weak Taylor scheme, see (11.2.2). For this purpose it is helpful to list, for simplicity, in the case $m = d = 1$, the conditional moment conditions for a given particular weak order.

Following the usual convention, we shall use the abbreviation $O(\Delta^k)$ to indicate that the remaining terms satisfy the limit

$$\lim_{\Delta \rightarrow 0} \frac{O(\Delta^k)}{\Delta^k} < \infty$$

almost surely for $k > 0$. For simplicity, we omit the dependence on the value Y_n of a Taylor scheme in the following conditions similarly to the description of weak schemes found earlier in the book.

In the case of weak order $\beta = 1.0$, it can be shown that the Euler scheme (11.2.1) satisfies the conditions:

$$\begin{aligned} E(Y_{n+1} - Y_n \mid \mathcal{A}_{t_n}) &= a \Delta + O(\Delta^2), \\ E(|Y_{n+1} - Y_n|^2 \mid \mathcal{A}_{t_n}) &= b^2 \Delta + O(\Delta^2), \\ E((Y_{n+1} - Y_n)^3 \mid \mathcal{A}_{t_n}) &= O(\Delta^2). \end{aligned} \quad (17.3.6)$$

For the second weak order $\beta = 2.0$, the weak order 2.0 Taylor scheme (11.2.6) satisfies the conditions:

$$\begin{aligned} E(Y_{n+1} - Y_n \mid \mathcal{A}_{t_n}) &= a \Delta + \frac{1}{2} \left(a a' + \frac{1}{2} b^2 a'' \right) \Delta^2 + O(\Delta^3), \\ E((Y_{n+1} - Y_n)^2 \mid \mathcal{A}_{t_n}) &= b^2 \Delta + \frac{1}{2} \left(2 a (a + b b') + b^2 (2 a' + b'^2 \right. \\ &\quad \left. + b b'') \right) \Delta^2 + O(\Delta^3), \\ E((Y_{n+1} - Y_n)^3 \mid \mathcal{A}_{t_n}) &= 3 b^2 (a + b b') \Delta^2 + O(\Delta^3), \\ E((Y_{n+1} - Y_n)^4 \mid \mathcal{A}_{t_n}) &= 3 b^4 \Delta^2 + O(\Delta^3), \\ E((Y_{n+1} - Y_n)^5 \mid \mathcal{A}_{t_n}) &= O(\Delta^3). \end{aligned} \quad (17.3.7)$$

Finally, in the case of third weak order $\beta = 3.0$, we find that the weak order 3.0 Taylor scheme (11.2.16) has for its increments the conditional moments:

$$E(Y_{n+1} - Y_n \mid \mathcal{A}_{t_n}) = a \Delta + \frac{1}{2} \left(a a' + \frac{1}{2} b^2 a'' \right) \Delta^2 + A \Delta^3 + O(\Delta^4) \quad (17.3.8)$$

with

$$\begin{aligned} A = \frac{1}{6} \left[a (a'^2 + a'' (a + b b')) + \frac{1}{2} b^2 (2 a a''' + 3 a' a'' + 2 a''' b b' \right. \\ \left. + a'' (b b'' + b'^2) + \frac{1}{2} a^{(iv)} b^2) \right], \end{aligned}$$

$$\begin{aligned} E((Y_{n+1} - Y_n)^2 | \mathcal{A}_{t_n}) &= b^2 \Delta + \frac{1}{2} \left(2a(a + bb') + b^2(2a' + b'^2 + bb'') \right) \Delta^2 \\ &\quad + B \Delta^3 + O(\Delta^4) \end{aligned}$$

with

$$\begin{aligned} B = \frac{1}{6} \left[&2a \left\{ a(3a' + b'^2 + bb'') + bb' (3a' + b'^2) \right\} + ab^2(7a'' + 8b'b'' + 2bb''') \right. \\ &+ b^2 \left\{ 4a'^2 + 7bb'a'' + b'^2(4a' + b'^2) + 8bb'^2b'' + 4bb''a' \right. \\ &\left. \left. + \frac{1}{2}b^2(4a''' + 5b''^2 + 8b'b'''' + bb^{(iv)}) \right\} \right], \end{aligned}$$

$$E((Y_{n+1} - Y_n)^3 | \mathcal{A}_{t_n}) = 3b^2(a + bb')\Delta^2 + C\Delta^3 + O(\Delta^4)$$

with

$$\begin{aligned} C = a^2(a + 3bb') + \frac{1}{2}ab^2(9a' + 11b'^2 + 7bb'') \\ + \frac{1}{2}b^2 \left\{ bb'(10a' + 8b'^2) + \frac{1}{2}b^2(7a'' + 28b'b'' + 4bb'') \right\}, \end{aligned}$$

$$E((Y_{n+1} - Y_n)^4 | \mathcal{A}_{t_n}) = 3b^4\Delta^2 + D\Delta^3 + O(\Delta^4)$$

with

$$D = 2b^2 \left[a(3a + 9bb') + \frac{1}{2}b^2(6a' + 19b'^2 + 7bb'') \right],$$

$$E((Y_{n+1} - Y_n)^5 | \mathcal{A}_{t_n}) = 15b^4(a + 2bb')\Delta^3 + O(\Delta^4),$$

$$E((Y_{n+1} - Y_n)^6 | \mathcal{A}_{t_n}) = 15b^6\Delta^3 + O(\Delta^4)$$

and

$$E((Y_{n+1} - Y_n)^7 | \mathcal{A}_{t_n}) = O(\Delta^4).$$

The above formulas can be very useful when designing approximate Markov chains of some given weak order.

A First Order Markov Chain Approximation

For simplicity, let us first approximate a one-dimensional time-homogeneous diffusion with $d = m = 1$ and uniformly bounded drift and diffusion coefficient functions $a(\cdot)$ and $b(\cdot)$. We assume that $a, b \in \mathcal{C}_P^4(\mathfrak{R}, \mathfrak{R})$ and $b(x) \geq b_0 > 0$ for all $x \in \mathfrak{R}$. Let \mathcal{X}_h^1 denote the h -grid on \mathfrak{R} , which is defined for $h > 0$ by

$$\mathcal{X}_h^1 = \{x = k h : k \in \{\dots, -1, 0, 1, \dots\}\}.$$

For an equidistant time discretization with time step size Δ , the Markov chain Y shall form a random walk on \mathcal{X}_h^1 where only jumps to the nearest neighbor points are possible. Let T be the last point of the time discretization and $Y_0 = X_0$. We define for all $x \in \mathcal{X}_h^1$ the functions

$$Q_h(x) = b^2(x)$$

$$\tilde{Q}_h = \sup_{x \in \mathcal{X}_h^1} Q_h(x),$$

$$\Delta = \frac{h^2}{\tilde{Q}_j}$$

and

$$a^+ = \begin{cases} a & \text{for } a \geq 0 \\ 0 & \text{otherwise,} \end{cases} \quad a^- = \begin{cases} -a & \text{for } a < 0 \\ 0 & \text{otherwise.} \end{cases}$$

In the following we concentrate on constructing appropriate transition probabilities. Let us choose the transition probabilities

$$p^h(x, y) = P(Y_{n+1} = y \mid Y_n = x)$$

in the form

$$p^h(x, x \pm h) = \left(\frac{b^2(x)}{2} \pm \frac{h}{2} a(x) \right)^+ \frac{1}{\tilde{Q}_h}$$

$$p^h(x, x) = 1 - p^h(x, x + h) - p^h(x, x - h)$$

for all $x \in \mathcal{X}_h^1$, $n \in \{0, 1, \dots\}$, and h sufficiently small.

From the convergence theorem that will be presented below it will follow that this Markov chain converges at time T with respect to $\Delta \rightarrow 0$ with weak order $\beta = 1$. Besides differentiability properties of a and b , the theorem will require the Markov chain to fulfill the following assumption on the first three conditional moments of its increments

$$\begin{aligned} & \left| E(Y_{n+1} - Y_n \mid Y_n = x) - a(x) \Delta \right| + \left| E((Y_{n+1} - Y_n)^2 \mid Y_n = x) - b^2(x) \Delta \right| \\ & + \left| E((Y_{n+1} - Y_n)^3 \mid Y_n = x) \right| \leq (1 + |x|^r) O(\Delta^2) \end{aligned} \tag{17.3.9}$$

for all $x \in \mathcal{X}_h^1$ and $n \in \{0, 1, \dots\}$ with some $r \in \mathcal{N}$.

The conditions in (17.3.9) can be easily verified for the preceding Markov chain. They basically represent consistency conditions, which ensure, together with some regularity conditions on a and b , that the Markov chain converges weakly towards a diffusion. They also reflect the conditions of the type (11.2.3)

that we imposed on the two-point random variables suitable for simplified Euler schemes. We obtain weak convergence of order $\beta = 1$ under these conditions when demanding enough smoothness of a , b and g . If we do not have enough smoothness for the drift and diffusion coefficients, then the weak order can only be less than one, as shown in [Mikulevicius & Platen \(1991\)](#) and [Kubilius & Platen \(2002\)](#).

When enough regularity for a and b is given, then the condition (17.3.9) provides the main restrictions for the construction of an approximating discrete-time Markov chain.

A Multi-dimensional Approximation

In the general *multi-dimensional* case let us consider a *Markov chain approximation* \mathbf{Y} , similar to those that we have seen for trees, see also [Kushner \(1977\)](#). Let \mathcal{X}_h^d denote the d -dimensional h -grid defined on \Re^d by

$$\mathcal{X}_h^d = \left\{ \mathbf{x} = \sum_{k=1}^d z_k \mathbf{e}_k h : z_k \in \{\dots, -1, 0, 1, \dots\} \quad \text{for all } k \in \{1, 2, \dots, d\} \right\},$$

where \mathbf{e}_k denotes the unit vector in the k th coordinate direction. Let T be the last point in the time discretization. We introduce for all $\mathbf{x} \in \Re^d$ the quantity

$$\delta^{k,\ell}(\mathbf{x}) = \sum_{j=1}^m b^{k,j}(\mathbf{x}) b^{\ell,j}(\mathbf{x}) \quad (17.3.10)$$

and assume that

$$\delta^{k,k}(\mathbf{x}) - \sum_{\substack{\ell=1 \\ \ell \neq k}}^d |\delta^{k,\ell}(\mathbf{x})| \geq 0$$

for all $k \in \{1, 2, \dots, d\}$ and $\mathbf{x} \in \Re^d$.

Furthermore, we assume that \mathbf{a} and \mathbf{b} are uniformly bounded, $a^k, b^{k,j} \in C_P^4(\Re^d, \Re)$, $b^{k,j}(\mathbf{x}) \geq b_0 > 0$ for all $\mathbf{x} \in \Re^d$, $k \in \{1, 2, \dots, d\}$ and $j \in \{1, 2, \dots, m\}$, and $\mathbf{Y}_0 = \mathbf{X}_0$. Now, we define for all $\mathbf{x} \in \Re^d$ the function

$$Q_h(\mathbf{x}) = \sum_{k=1}^d \delta^{k,k}(\mathbf{x}),$$

the time step size

$$\Delta t^h(\mathbf{x}) = \frac{h^2}{Q_h(\mathbf{x})} \leq \Delta \quad (17.3.11)$$

and the transition probabilities

$$\begin{aligned}
p^h(\mathbf{x}, \mathbf{x} \pm \mathbf{e}_k h) &= \left\{ \frac{\delta^{k,k}(\mathbf{x})}{2} - \sum_{\substack{i=1 \\ i \neq k}}^d \frac{|\delta^{k,\ell}(\mathbf{x})|}{2} \pm \frac{h}{2} a^k(\mathbf{x}) \right\} \frac{1}{Q_h(\mathbf{x})} \\
p^h(\mathbf{x}, \mathbf{x} + \mathbf{e}_k h + \mathbf{e}_\ell h) &= p^h(\mathbf{x}, \mathbf{x} - \mathbf{e}_k h - \mathbf{e}_\ell h) = \frac{1}{2Q_h(\mathbf{x})} (\delta^{k,\ell}(\mathbf{x}))^+ \\
p^h(\mathbf{x}, \mathbf{x} - \mathbf{e}_k h + \mathbf{e}_\ell h) &= p^h(\mathbf{x}, \mathbf{x} + \mathbf{e}_k h - \mathbf{e}_\ell h) = \frac{1}{2Q_h(\mathbf{x})} (\delta^{k,\ell}(\mathbf{x}))^- \quad (17.3.12)
\end{aligned}$$

for all $k \in \{1, 2, \dots, d\}$ and $\ell \in \{1, 2, \dots, d\} \setminus \{k\}$, where all other unlisted transition probabilities are zero.

This provides an example for a multi-dimensional Markov chain approximation. Note that in the convergence theorem the boundedness of \mathbf{a} and \mathbf{b} that we assume here will later not be required. Here it is used to allow a convenient construction of the Markov chain in the given example.

From the convergence theorem it will follow that this d -dimensional Markov chain converges with weak order $\beta = 1$ when $\Delta \rightarrow 0$. In this multi-dimensional case, the theorem will require the Markov chain to have conditional moments of its increments which satisfy the following properties:

$$\begin{aligned}
&\left| E(Y_{n+1}^k - Y_n^k \mid \mathbf{Y}_n = \mathbf{x}) - a^k(\mathbf{x}) \Delta t^h(\mathbf{x}) \right| \\
&+ \left| E((Y_{n+1}^k - Y_n^k)(Y_{n+1}^\ell - Y_n^\ell) \mid \mathbf{Y}_n = \mathbf{x}) - \delta^{k,\ell}(\mathbf{x}) \Delta t^h(\mathbf{x}) \right| \\
&+ \left| E((Y_{n+1}^k - Y_n^k)(Y_{n+1}^\ell - Y_n^\ell)(Y_{n+1}^p - Y_n^p) \mid \mathbf{Y}_n = \mathbf{x}) \right| \\
&\leq (1 + |\mathbf{x}|^r) O((\Delta t^h(\mathbf{x}))^2) \quad (17.3.13)
\end{aligned}$$

for all $\mathbf{x} \in \mathcal{X}_h^d$, $k, \ell, p \in \{1, 2, \dots, d\}$ and $n \in \{0, 1, \dots\}$ and some $r \in \mathcal{N}$. Here Y^k denotes the k th component of \mathbf{Y} . One notes that the condition (17.3.13) represents the multi-dimensional version of the condition (17.3.9) for first order weak convergence.

One obtains from the preceding Markov chain, with a slight modification, a Markov chain with constant time step size. One way to achieve this is to set

$$\tilde{Q}_h = \sup_{x \in \mathcal{X}_h^d} Q_h(\mathbf{x}) \quad (17.3.14)$$

and

$$\Delta = \frac{h^2}{\tilde{Q}_h} \quad (17.3.15)$$

to normalize all the transition probabilities in (17.3.12) by \tilde{Q}_h instead of $Q_h(\mathbf{x})$ and finally introduce the probability

$$p^h(\mathbf{x}, \mathbf{x}) = 1 - \sum_{k=1}^d \sum_{\ell=1}^d \sum_{u_k \in \{-1, 1\}} \sum_{v_\ell \in \{0, 1\}} p^h(\mathbf{x}, \mathbf{x} + u_k \mathbf{e}_k h + v_\ell \mathbf{e}_\ell h).$$

Again it will follow that this Markov chain converges for $\Delta \rightarrow 0$ with weak order $\beta = 1$, if enough smoothness for a and b is guaranteed.

A Second Order Markov Chain Approximation

Let us now describe a second order Markov chain approximation for the one-dimensional case, $d = m = 1$, for given $h > 0$ and $\Delta \in (0, 1)$ on the h -grid \mathcal{X}_h^1 . We assume that a and b are from $\mathcal{C}_P^6(\mathfrak{R}, \mathfrak{R})$ and together with their first and second derivatives are uniformly bounded and that $b(x) \geq b_0 > 0$ for all $x \in \mathfrak{R}$. Furthermore, we set $Y_0 = X_0$ and take T as the last point of the time discretization. For each pair $(x, \tilde{\Delta}) \in \mathfrak{R} \times (0, 1)$ let us introduce the following functions

$$\begin{aligned}\gamma_1(x, \tilde{\Delta}) &= a + \left(a a' + \frac{b^2}{2} a'' \right) \frac{\tilde{\Delta}}{2} \\ \gamma_2(x, \tilde{\Delta}) &= b^2 + (2 a(a + b b') + b^2(2 a' + b'^2 + b b'')) \frac{\tilde{\Delta}}{2} \\ \gamma_3(x, \tilde{\Delta}) &= 3 b^2 (a + b b') \tilde{\Delta} \\ \gamma_4(x, \tilde{\Delta}) &= 3 b^4 \tilde{\Delta} \\ \gamma_5(x, \tilde{\Delta}) &= 0,\end{aligned}$$

where we omit the dependence on x of the right hand sides of the equations.

We set for all $x \in \mathcal{X}_h^1$ and $\tilde{\Delta} \in (0, 1)$

$$Q_{h, \tilde{\Delta}}(x) = \frac{1}{3} \left(\frac{15}{4} \gamma_2 - \frac{3}{4 h^2} \gamma_4 \right),$$

where we again suppress the dependence on $\tilde{\Delta}$ and x of the right hand side and define

$$\tilde{Q}_{h, \tilde{\Delta}} = \sup_{x \in \mathcal{X}_h^1} Q_{h, \tilde{\Delta}}(x).$$

Now we choose $\Delta \in (0, 1)$ such that

$$\Delta = \frac{h^2}{\tilde{Q}_{h, \Delta}},$$

where the structure of $\tilde{Q}_{h, \tilde{\Delta}}$ allows this choice.

Finally, we define for all $x \in \mathcal{X}_h^1$ with $\tilde{\Delta} = \Delta$ the following transition probabilities

$$\begin{aligned}
p^h(x, x \pm h) &= \frac{1}{3} \left(2\gamma_2 - \frac{1}{2h^2} \gamma_4 \pm \frac{1}{2} \left(4h\gamma_1 - \frac{1}{h}\gamma_3 \right) \right) \frac{1}{\tilde{Q}_{h,\Delta}} \\
p^h(x, x \pm 2h) &= \frac{1}{3} \left(\frac{1}{8h^2} \gamma_4 - \frac{1}{8} \gamma_2 \pm \frac{1}{4} \left(\frac{1}{h} \gamma_3 - h\gamma_1 \right) \right) \frac{1}{\tilde{Q}_{h,\Delta}} \\
p^h(x, x) &= 1 - \sum_{k \in \{-2, -1, 1, 2\}} p^h(x, x + kh),
\end{aligned} \tag{17.3.16}$$

where all other transition probabilities are zero.

It will follow by the convergence theorem that this Markov chain converges with respect to $\Delta \rightarrow 0$ with weak order $\beta = 2$. To ensure this, we basically need the following properties of the conditional moments

$$\sum_{\ell=1}^5 \left| E((Y_{n+1} - Y_n)^\ell \mid Y_n = x) - \gamma_\ell(x, \Delta) \Delta \right| \leq (1 + |x|^r) O(\Delta^3) \tag{17.3.17}$$

for all $x \in \mathcal{X}_h^1$ and $n \in \{0, 1, \dots\}$, with some $r \in \mathcal{N}$. This condition can be verified to hold for the above Markov chain. We emphasize that the weak convergence theorem will not require the boundedness of a and b , and their derivatives, which in this example is assumed to ensure a finite $\tilde{Q}_{h,\Delta}$. The pentanomial tree, which we have constructed previously, can be shown to satisfy the above conditions.

One can systematically find second order Markov chain approximations in the general multi-dimensional case on the d -dimensional grid \mathcal{X}_h^d or for a more general discrete-time approximation. Here assumptions on the conditional moments, as will be stated in the convergence theorem, need to be satisfied. Note that one has to use, in general, $\sum_{k=1}^4 d^k$ “neighboring” points in a transition step of a Markov chain to achieve weak order $\beta = 2$. This means, in general, a second weak order tree needs to have in the one dimensional case five nodes to go to, whereas in the two dimensional case 31 nodes are needed. This indicates that, in general, higher order multi-dimensional trees may be difficult to implement unless the underlying structure of the diffusion process allows major simplifications.

A Third Order Markov Chain Approximation

As an example we will construct now a third order Markov chain approximation for the one-dimensional case, $d = m = 1$, for given $\Delta \in (0, 1)$ and $h > 0$, on the h -grid \mathcal{X}_h^1 . We assume that a and b are from $\mathcal{C}_P^8(\mathfrak{R}^1, \mathfrak{R}^1)$, $b(x) \geq b_0 > 0$ for all $x \in \mathfrak{R}$, and that they are uniformly bounded, together with their derivatives, up to the fourth order. Furthermore, we set $Y_0 = X_0$ and assume T as the last discretization point. For all $x \in \mathcal{X}_h^1$ and $\tilde{\Delta} > 0$ let us introduce the following functions

$$\beta_1(x, \tilde{\Delta}) = \beta_1 = \gamma_1 + a \left[(a'^2 + a''(a + b'b')) + \frac{b^2}{2} \left(2a a''' + 3a' a'' + 2a'''b b' \right. \right. \\ \left. \left. + a''(b b'' + b'^2) + a^{iv} \frac{b^2}{2} \right) \right] \frac{\tilde{\Delta}^2}{6}$$

$$\beta_2 = \gamma_2 + [2a(a(3a' + b'^2 + b b'') + b b'(3a' + b'^2)) \\ + b^2 \left(a(7a'' + 8b' b'' + 2b b''') + 4a'^2 + 7b b'a'' \right. \\ \left. + b'^2(4a' + b'^2) + 8b b'^2 b'' + b b'' 4a' \right. \\ \left. + \frac{b^2}{4}(4a''' + 5b''^2 + 8b' b''' + b b^{iv}) \right) \left. \right] \frac{\tilde{\Delta}^2}{6}$$

$$\beta_3 = \gamma_3 + \left[a^2(a + 3b b') + \frac{b^2}{2}(a(9a' + 11b'^2 + 7b b'') \right. \\ \left. + b b'(10a' + 8b'^2) + \frac{b^2}{2}(7a'' + 28b' b'' + 4b b'') \right] \tilde{\Delta}^2$$

$$\beta_4 = \gamma_4 + 2b^2 \left[a(3a + 9b b') + \frac{b^2}{2}(6a' + 19b'^2 + 7b b'') \right] \tilde{\Delta}^2$$

$$\beta_5 = \tilde{\Delta}^2 15b^4(a + 2b b')$$

$$\beta_6 = \tilde{\Delta}^2 15b^6$$

$$\beta_7 = 0,$$

where we omit the dependence on $\tilde{\Delta}$ and x , and the γ_k are as in the previous section. We define for all $x \in \mathcal{X}_h^1$ and $\tilde{\Delta} \in (0, 1)$ the function

$$Q_{h,\tilde{\Delta}}(x) = \frac{23}{18}\beta_2 - \frac{7}{18h^2}\beta_4 + \frac{1}{36h^4}\beta_6,$$

where we again omit on the right hand side the dependence on $\tilde{\Delta}$ and x . Furthermore, we introduce

$$\tilde{Q}_{h,\tilde{\Delta}} = \sup_{x \in \mathcal{X}_h^1} Q_{h,\tilde{\Delta}}(x)$$

and choose Δ such that

$$\Delta = \frac{h^2}{\tilde{Q}_{h,\Delta}},$$

which is again possible because of the structure of $Q_{h,\tilde{\Delta}}$.

Now, we define for all $x \in \mathcal{X}_h^1$ and $\tilde{\Delta} = \Delta$ the transition probabilities

$$\begin{aligned}
p^h(x, x \pm h) &= \left(\frac{1}{2} \beta_2 - \frac{39}{144 h^2} \beta_4 + \frac{1}{48 h^4} \beta_6 \right. \\
&\quad \left. \pm \frac{1}{2} \left(\frac{3}{2} h \beta_1 - \frac{157}{456 h} \beta_3 + \frac{1}{24 h^3} \beta_5 \right) \right) \frac{1}{\tilde{Q}_{h,\Delta}} \\
p^h(x, x \pm 2h) &= \left(\frac{1}{40} \beta_2 - \frac{1}{12 h^2} \beta_4 + \frac{1}{120 h^4} \beta_6 \right. \\
&\quad \left. \pm \frac{1}{2} \left(-\frac{3}{10} h \beta_1 + \frac{10}{57 h} \beta_3 - \frac{1}{30 h^3} \beta_5 \right) \right) \frac{1}{\tilde{Q}_{h,\Delta}} \\
p^h(x, x \pm 3h) &= \left(-\frac{1}{90} \beta_2 - \frac{1}{744 h^2} \beta_4 + \frac{1}{720 h^4} \beta_6 \right. \\
&\quad \left. \pm \frac{1}{2} \left(\frac{1}{30} h \beta_1 + \frac{1}{456 h} \beta_3 - \frac{1}{720 h^3} \beta_5 \right) \right) \frac{1}{\tilde{Q}_{h,\Delta}} \\
p^h(x, x) &= 1 - \sum_{\substack{k=-3 \\ k \neq 0}}^3 p^h(x, x + kh), \tag{17.3.18}
\end{aligned}$$

where all other transition probabilities are set to zero.

From the convergence theorem below it will follow that this Markov chain converges with weak order $\beta = 3$ as $\Delta \rightarrow 0$. The following properties of the conditional moments of the increments of the Markov chain are required to obtain third order weak convergence:

$$\sum_{\ell=1}^7 \left| E((Y_{n+1} - Y_n)^\ell \mid Y_n = x) - \beta_\ell(x, \Delta) \Delta \right| \leq (1 + |x|^r) O(\Delta^4) \tag{17.3.19}$$

for all $x \in \mathcal{X}_h^1$ and $n \in \{0, 1, \dots\}$ with some $r \in \mathcal{N}$.

In the multi-dimensional case it turns out that we need, in general, $\sum_{k=1}^6 d^k$ “neighboring” points for a d -dimensional Markov chain of weak order $\beta = 3$ to satisfy the corresponding moment matching conditions. Unfortunately, this leads to substantial complexity for higher dimensions. Thus, the curse of dimensionality is not avoided when using Markov chain approximations since the computational effort increases exponentially with the dimension. Therefore, it is once more that the efficiency of simplified weak schemes does not suffer greatly in higher dimensions. It is a core advantage of Monte Carlo methods that their numerical complexity increases polynomially with the dimension, whereas Markov chain approximations, lattice and grid methods suffer strongly under the curse of dimensionality, unless structural advantages of the underlying dynamics can be exploited.

17.4 Finite Difference Methods

Derivative Pricing via PDEs

In many practical situations in option pricing it is impossible to obtain an analytic solution for the transition density of the underlying diffusion process. This then typically applies also to the pricing function of the given derivative. However, in many cases of one or two factor dynamics it is possible to use numerical approximations for the solutions of the underlying PDE. One of the most common methods are *finite difference methods*. We will indicate that these methods are related to Markov chain approximations.

There is extensive literature on these methods that has evolved over several decades. The interested reader is referred to standard textbooks by Richtmeyer & Morton (1967) or Smith (1985). In the finance area the monographs by Tavella & Randall (2000), Shaw (1998) and Wilmott, Dewynne & Howison (1993) are popular references.

Let us consider a risk factor process $X = \{X_t, t \in [0, T]\}$ with SDE

$$dX_t = a(X_t) dt + b(X_t) dW_t \quad (17.4.1)$$

for $t \in [0, T]$ with $X_0 > 0$. Here W is denoting a standard Wiener process. A European option with discounted European payoff

$$u(t, X_T) = H(X_T) \quad (17.4.2)$$

at maturity T yields, under appropriate pricing, a pricing function $u(\cdot, \cdot)$ that satisfies the PDE

$$\frac{\partial u(t, x)}{\partial t} + a(x) \frac{\partial u(t, x)}{\partial x} + \frac{1}{2} (b(x))^2 \frac{\partial^2 u(t, x)}{\partial x^2} = 0 \quad (17.4.3)$$

for $t \in [0, T]$ and $x \in [0, \infty)$ with terminal condition

$$u(T, x) = H(x) \quad (17.4.4)$$

for $x \in [0, \infty)$. This is the Kolmogorov backward equation, which arises from the Feynman-Kac formula, see Sect. 2.7, when calculating the conditional expectation

$$u(t, X_t) = E(H(X_T) \mid \mathcal{A}_t). \quad (17.4.5)$$

Here E denotes expectation under the pricing measure P .

We recall, as explained in Chap. 3, that a benchmarked pricing function $u(t, X_t)$ is obtained as the fair benchmarked price of a benchmarked payoff $H(X_T)$. In this case the expectation (17.4.5) is taken with respect to the real world probability measure. When using risk neutral pricing an equivalent risk neutral probability measure needs to exist and the numéraire is the savings account.

As discussed in Sect. 2.7 it may be necessary to add to the characterization of the PDE the boundary condition for $u(t, x)$ when u goes to zero, where its benchmarked value should be absorbed. Typically, this arises at the boundary where the GOP approaches zero and one has a payoff that does not vanish at this boundary, see [Platen & Heath \(2006\)](#).

Spatial Differences

Let us now approach the problem of numerically solving the parabolic, second order PDE (17.4.3)–(17.4.4). The key idea of finite difference methods is to replace the partial derivatives in (17.4.3) by finite difference approximations. For example, by using the deterministic Taylor formula the first spatial partial derivative of $u(\cdot, \cdot)$ can be written as

$$\frac{\partial u(t, x)}{\partial x} = \frac{u(t, x + \Delta x) - u(t, x - \Delta x)}{2 \Delta x} + R_1(t, x) \quad (17.4.6)$$

and the second derivative as

$$\begin{aligned} \frac{\partial^2 u(t, x)}{\partial x^2} &= \frac{\frac{u(t, x + \Delta x) - u(t, x)}{\Delta x} - \frac{(u(t, x) - u(t, x - \Delta x))}{\Delta x}}{\Delta x} + R_2(t, x) \\ &= \frac{u(t, x + \Delta x) - 2u(t, x) + u(t, x - \Delta x)}{(\Delta x)^2} + R_2(t, x). \end{aligned} \quad (17.4.7)$$

Here $R_1(t, x)$ and $R_2(t, x)$ are corresponding remainder terms of the standard Taylor expansions, and $\Delta x > 0$ is a small increment in the spatial direction of x .

To place these differences let us introduce an equally spaced spatial grid

$$\mathcal{X}_{\Delta x}^1 = \{x_k = k \Delta x : k \in \{0, 1, \dots, N\}\} \quad (17.4.8)$$

for $N \in \{2, 3, \dots\}$. We refer to this grid as *spatial discretization*. Thus, we have the differences of neighboring grid points in the form

$$x_{k+1} - x_k = \Delta x.$$

When writing

$$u_k(t) = u(t, x_k)$$

we obtain from (17.4.6) and (17.4.7)

$$\frac{\partial u(t, x_k)}{\partial x} = \frac{u_{k+1}(t) - u_{k-1}(t)}{2 \Delta x} + R_1(t, x_k) \quad (17.4.9)$$

and

$$\frac{\partial^2 u(t, x_k)}{\partial x^2} = \frac{u_{k+1}(t) - 2u_k(t) + u_{k-1}(t)}{(\Delta x)^2} + R_2(t, x_k) \quad (17.4.10)$$

for $k \in \{1, 2, \dots, N-1\}$. The quantities $u_k(t)$ are for $k \in \{1, 2, \dots, N-1\}$ referred to as *interior values*, whereas $u_0(t)$ and $u_N(t)$ are called *boundary values* for $t \in [0, T]$. For the particular example let us linearly relate the boundary values with neighboring interior values such that

$$du_0(t) = \kappa(u_1(t) - u_0(t)) dt \quad (17.4.11)$$

and

$$du_N(t) = \varphi(u_{N-1}(t) - u_N(t)) dt \quad (17.4.12)$$

for $t \in [0, T]$, with $\kappa, \varphi \geq 0$. This creates a certain behavior of the PDE solution at the respective boundaries. As we indicated already in Sect. 2.7, when we discussed the Feynman-Kac formula, it can be rather important how the behavior at the boundaries is approximated.

For the vector $\mathbf{u}(t) = (u_0(t), u_1(t), \dots, u_N(t))^\top$ we can now transform the pricing equations that arise from the PDE (17.4.3) into an approximate system of ODEs that has the form

$$\frac{d\mathbf{u}(t)}{dt} + \frac{\mathbf{A}}{(\Delta x)^2} \mathbf{u}(t) + \mathbf{R}(t) = 0 \quad (17.4.13)$$

for $t \in [0, T]$. Note the appearance of the vector $\mathbf{R}(t)$ of resulting remainder terms and the corresponding matrix $\mathbf{A} = [A^{i,j}]_{i,j=0}^N$ of the above linear system. The matrix \mathbf{A} has the elements

$$\begin{aligned} A^{0,0} &= -\kappa(\Delta x)^2, & A^{N,N} &= -\varphi(\Delta x)^2 \\ A^{0,1} &= \kappa(\Delta x)^2, & A^{N,N-1} &= \varphi(\Delta x)^2 \\ A^{k,k} &= -(b(x_k))^2 \\ A^{k,k-1} &= -\frac{1}{2}(A^{k,k} + a(x_k)\Delta x) \\ A^{k,k+1} &= -\frac{1}{2}(A^{k,k} - a(x_k)\Delta x) \\ A^{k,j} &= 0 \end{aligned} \quad (17.4.14)$$

for $k \in \{1, 2, \dots, N-1\}$ and $|k-j| > 1$. The matrix \mathbf{A} thus has the form

$$A = \begin{pmatrix} -\kappa(\Delta x)^2 & \kappa(\Delta x)^2 & 0 & 0 & \dots & 0 & 0 & 0 \\ A^{1,0} & A^{1,1} & A^{1,2} & 0 & \dots & 0 & 0 & 0 \\ 0 & A^{2,1} & A^{2,2} & A^{2,3} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & A^{N-2,N-2} & A^{N-2,N-1} & 0 \\ 0 & 0 & 0 & 0 & \dots & A^{N-1,N-2} & A^{N-1,N-1} & A^{N-1,N} \\ 0 & 0 & 0 & 0 & \dots & 0 & \varphi(\Delta x)^2 & -\varphi(\Delta x)^2 \end{pmatrix}$$

By ignoring the vector of remainder terms in the system of ODEs (17.4.13) we obtain approximately the linear vector ODE

$$d\mathbf{u}(t) \approx \frac{\mathbf{A}}{(\Delta x)^2} \mathbf{u}(t) dt \quad (17.4.15)$$

for $t \in (0, T)$ with terminal condition

$$\mathbf{u}(T) = (H(x_0), \dots, H(x_N))^\top.$$

This ODE provides us with approximate values for $u(\cdot, \cdot)$ at the grid points over time.

In the above example we have used only one particular way of approximating the first and second order spatial derivatives. There are alternative ways of forming finite differences. These finite difference approximations are motivated by corresponding standard Taylor expansions. Along the boundaries and to achieve a certain order of convergence one can use also the following first order approximations instead of (17.4.9). For simplicity, we omit here, on the right hand sides, the dependence on t :

$$\begin{aligned} \frac{\partial u(t, x_k)}{\partial x} &= \frac{u_{k+1} - u_k}{\Delta x} + O(\Delta x) \\ \frac{\partial u(t, x_k)}{\partial x} &= \frac{u_k - u_{k-1}}{\Delta x} + O(\Delta x) \\ \frac{\partial u(t, x_k)}{\partial x} &= \frac{u_{k+1} - u_{k-1}}{2 \Delta x} + O((\Delta x)^2) \\ \frac{\partial u(t, x_k)}{\partial x} &= \frac{3 u_k - 4 u_{k+1} + u_{k-2}}{2 \Delta x} + O((\Delta x)^2) \\ \frac{\partial u(t, x_k)}{\partial x} &= \frac{-3 u_k + 4 u_{k+1} - u_{k-2}}{2 \Delta x} + O((\Delta x)^2). \end{aligned} \quad (17.4.16)$$

We also indicate in (17.4.16) the order of the leading error term, which is important for the order of overall convergence of the approximation. Some of these finite differences can be used at boundaries.

For the second order spatial derivatives the following finite differences can be used instead of (17.4.10) for the construction of a numerical method:

$$\begin{aligned} \frac{\partial^2 u(t, x_k)}{\partial x^2} &= \frac{u_k - 2 u_{k-1} + u_{k-2}}{(\Delta x)^2} + O(\Delta x) \\ \frac{\partial^2 u(t, x_k)}{\partial x^2} &= \frac{u_{k+2} - 2 u_{k+1} + u_k}{(\Delta x)^2} + O(\Delta x) \\ \frac{\partial^2 u(t, x_k)}{\partial x^2} &= \frac{u_{k+1} - 2 u_k + u_{k-1}}{(\Delta x)^2} + O((\Delta x)^2) \end{aligned}$$

$$\begin{aligned}\frac{\partial^2 u(t, x_k)}{\partial x^2} &= \frac{2u_k - 5u_{k-1} + 4u_{k-2} - u_{k-3}}{(\Delta x)^2} + O((\Delta x)^2) \\ \frac{\partial^2 u(t, x_k)}{\partial x^2} &= \frac{-u_{k+3} + 4u_{k+2} - 5u_{k+1} + 2u_k}{(\Delta x)^2} + O((\Delta x)^2).\end{aligned}\quad (17.4.17)$$

Finite Difference Method

So far we have only discretized the spatial coordinate. This is simply the first step. The finite difference approximation of the solution of the PDE (17.4.3) is then constructed in two steps. Applying any of the above spatial discretizations to approximate the spatial derivatives by finite differences, leads to the system (17.4.15) of coupled ODEs. One must then solve numerically the obtained ODEs. Here one uses discrete-time approximations for the ODEs, as described in Chap. 5.

We solve the ODE (17.4.15) numerically using *time discretization*. As in the case of SDEs for systems of ODEs, various numerical methods are available. Let us denote by Δ the time step size and use as n th time discretization point the time $t_n = n\Delta$, $n \in \{0, 1, \dots, n_T\}$, where n_T is the largest integer n for which t_n is less than or equal to t . Let us neglect any higher order terms and apply the Euler scheme to obtain the algorithm

$$\mathbf{u}(t_{n+1}) = \mathbf{u}(t_n) + \mathbf{A} \mathbf{u}(t_n) \frac{\Delta}{(\Delta x)^2} \quad (17.4.18)$$

for $n \in \{0, 1, \dots, n_T\}$. In our case we know the value of the vector $\mathbf{u}(T) = \mathbf{u}(t_{n_T})$, which is the payoff at maturity. This means, we have

$$u_k(T) = u(T, x_k) = H(x_k) \quad (17.4.19)$$

for all $k \in \{0, 1, \dots, N\}$. Now the discrete-time approximate system of difference equations (17.4.18) with terminal condition (17.4.19) can be solved in backward steps starting from the maturity date. Under appropriate conditions one obtains in this manner an approximate solution of the PDE (17.4.3). The method described above constitutes a *finite difference method*. This is a widely used method in derivative pricing and quite suitable for a range of models and exotic options.

Relation to Markov Chains

If one looks back into the previous section on Markov chain approximations, then one realizes that the above finite difference method resembles strongly a Markov chain approximation. The above finite difference method and the following Markov chain approximation can indeed be shown to be equivalent in a certain sense:

Note that (17.4.18) can be written as

$$\mathbf{u}(t_{n+1}) = \mathbf{p}^{\Delta x} \mathbf{u}(t_n) \quad (17.4.20)$$

where

$$\mathbf{p}^{\Delta x} = \mathbf{I} + \mathbf{A} \frac{\Delta}{(\Delta x)^2} = [p^{\Delta x}(x_i, x_j)]_{i,j=0}^N.$$

Here $p^{\Delta x}(x_i, x_j)$ denotes the transition probability for the Markov chain Y^Δ to move from the state $x_i \in \mathcal{X}_{\Delta x}^1$ at time t_n to the state $x_j \in \mathcal{X}_{\Delta x}^1$ at time $t_{n+1} = t_n + \Delta$. Due to (17.4.14) and (17.4.20) we have for $i \in \{1, 2, \dots, N-1\}$ the transition probabilities

$$p^\Delta(x_i, x_j) = \begin{cases} \frac{1}{2} (b(x_i))^2 + a(x_i) \Delta x \frac{\Delta}{(\Delta x)^2} & \text{for } j = i + 1 \\ 1 - b(x_i)^2 \frac{\Delta}{(\Delta x)^2} & \text{for } j = i \\ \frac{1}{2} (b(x_i))^2 - a(x_i) \Delta x \frac{\Delta}{(\Delta x)^2} & \text{for } j = i - 1 \\ 0 & \text{otherwise.} \end{cases} \quad (17.4.21)$$

Obviously, these are probabilities since they sum up to one. At the boundaries we obtain from (17.4.11) that

$$p^{\Delta x}(x_0, x_j) = \begin{cases} \kappa \Delta & \text{for } j = 1 \\ 1 - \kappa \Delta & \text{for } j = 0 \\ 0 & \text{otherwise} \end{cases}$$

and by (17.4.12) the relation

$$p^{\Delta x}(x_N, x_j) = \begin{cases} \varphi \Delta & \text{for } j = N - 1 \\ 1 - \varphi \Delta & \text{for } j = N \\ 0 & \text{otherwise.} \end{cases}$$

Also these quantities turn out to be probabilities for small enough Δ . However, it must be noted that a finite difference method can also be constructed such that there is no direct interpretation for an equivalent Markov chain. This theoretical possibility gives the method more freedom, as we will see later. Still, for an understanding of some of the numerical properties of these methods it is useful to draw parallels between the behavior of simplified Monte Carlo methods, Markov chain approximations and finite difference methods.

One notes that there are two sources of errors in a finite difference method. These are the truncation error caused by the spatial discretization and the truncation error resulting from the time discretization. Fortunately, there is no statistical error. We have seen previously that Markov chain approximations require particular relationships to be satisfied between the spatial and time discretization step sizes to obtain nonnegative probabilities. Due to the extra freedom in a finite difference method this only loosely applies for finite difference methods. In the literature it is suggested to use, for an explicit method,

a ratio of the spatial over the time step size such that $0 < \frac{\Delta}{(\Delta x)^2} \leq \frac{1}{2}$. This choice has a natural explanation because otherwise it is difficult to construct positive probabilities for an approximating Markov chain, see (17.4.21).

Similar to the moment and numerical stability conditions for Monte Carlo methods and Markov chain approximations, finite difference methods also require similar consistency and stability conditions. These have to ensure, as in the case of Markov chain approximations, that the approximate PDE solution evolves, locally, similar to that of the underlying PDE and does not propagate errors in an uncontrolled manner.

Additionally, when using a finite difference method one has to be aware of the fact that the underlying diffusion or jump diffusion process typically extends its dynamics beyond the boundaries that one uses for the approximation of the state space. Fortunately, one can show for most situations that this does not really harm significantly the computational results, see [Lamberton & Lapeyre \(2007\)](#), when the boundaries are chosen appropriately. Finally, we emphasize that as with numerical methods for ODEs and SDEs, the approximate solution that one generates by a finite difference method is clearly a different mathematical object, and only by weak convergence linked to the exact solution of the PDE that motivated its construction.

For the global and overall convergence we have already seen for ODE solvers and Monte Carlo schemes that the numerical stability is most important. In the literature on finite difference methods, for instance in Richtmeyer & Morton (1967) and [Raviart & Thomas \(1983\)](#), one can find corresponding conditions and convergence theorems that make this precise. For the case when the above approximate solution can be interpreted as a Markov chain approximation, Theorem 17.5.1 provides a corresponding weak order of convergence. Its numerical stability can be linked to that of the Markov chain approximation, using similar notions and methods as described in Chap. 14.

The Theta Method

We have seen in Chap. 11 that implicit methods are extremely important to achieve numerical stability. So far, when discretizing in time the system of ODEs (17.4.15), we have in (17.4.18) only used the simple explicit Euler scheme. For Markov chain approximations, including trees, it appears to be difficult to make these methods implicit. However, for the system of ODEs, determining the above finite difference PDE approximation, this is possible. We simply apply instead of the Euler scheme in (17.4.18) the more general family of implicit Euler schemes to the system of ODEs (17.4.15). This yields discrete-time approximations of the ODE that allow us to introduce implicitness. Let us use the ODE version of the family of implicit Euler schemes (11.5.6) to obtain the, so called, *theta method*

$$\mathbf{u}(t_{n+1}) = \mathbf{u}(t_n) + (\theta \mathbf{A} \mathbf{u}(t_{n+1}) + (1 - \theta) \mathbf{A} \mathbf{u}(t_n)) \frac{\Delta}{(\Delta x)^2} \quad (17.4.22)$$

with *degree of implicitness* θ . For instance, in the case $\theta = 0$ we recover the Euler method (17.4.18). We know from our discussions on the stability of ODE and SDE approximation methods in Chap. 14 that the region of stability for the Euler method with $\theta = 0$ is only the interior of an interval with center at -1 . When we choose $\theta = 1$, then we obtain the *fully implicit method*

$$\mathbf{u}(t_{n+1}) = \mathbf{u}(t_n) + \mathbf{A} \mathbf{u}(t_{n+1}) \frac{\Delta}{(\Delta x)^2}. \quad (17.4.23)$$

The fully implicit Euler method without stochasticity was shown to have a relatively large stability region, see Chap. 14.

Crank-Nicolson Method

Another important method is obtained by choosing in (17.4.22) the degree of implicitness $\theta = \frac{1}{2}$. This yields the popular *Crank-Nicolson method*

$$\mathbf{u}(t_{n+1}) = \mathbf{u}(t_n) + \frac{1}{2} \mathbf{A} (\mathbf{u}(t_{n+1}) + \mathbf{u}(t_n)) \frac{\Delta}{(\Delta x)^2}. \quad (17.4.24)$$

The Crank-Nicolson scheme is unconditionally stable. As we have discussed in Chap. 14, it is only for the above critical parameter choice $\theta = \frac{1}{2} \in [0, \frac{1}{2}]$ that the discrete-time approximation (17.4.22) becomes stable. If the solution of a PDE requires more numerical stability, then the parameter choice $\theta = 1$, yielding the fully implicit Euler method (17.4.23), is usually a better one. Commonly, the Crank-Nicolson scheme is recommended for calculating pricing functions because it achieves, due to its symmetric structure, a higher order of convergence than comparable methods. However, we have seen in previous chapters that numerical stability must have priority when attempting to reach any higher order of convergence. Despite the popularity of the Crank-Nicolson scheme, certain stability issues may suggest the use of a fully implicit method. For instance, this may be the case if the spatial or time step size needs to be large, as discussed in Shaw (1998), and will be discussed towards the end of this section.

With finite difference methods that have some degree of implicitness, one faces the problem that one has to solve at each time step a coupled system of equations because $\mathbf{u}(t_{n+1})$ appears also on the right hand side of the difference equation. What in practice happens is that in each time step one has to solve a large, linear system of equations. For instance, in the case of the Crank-Nicolson method (17.4.24) one obtains

$$\left(\mathbf{I} - \frac{1}{2} \mathbf{A} \frac{\Delta}{(\Delta x)^2} \right) \mathbf{u}(t_{n+1}) = \left(\mathbf{I} + \frac{1}{2} \mathbf{A} \frac{\Delta}{(\Delta x)^2} \right) \mathbf{u}(t_n). \quad (17.4.25)$$

With the large matrix

$$\mathbf{M} = \mathbf{I} - \frac{1}{2} \mathbf{A} \frac{\Delta}{(\Delta x)^2}$$

and the vector

$$\mathbf{B}_n = \left(\mathbf{I} + \frac{1}{2} \mathbf{A} \frac{\Delta}{(\Delta x)^2} \right) \mathbf{u}(t_n),$$

we can rewrite equation (17.4.25), and face at each time step the problem of solving the linear system of equations

$$\mathbf{M} \mathbf{u}(t_{n+1}) = \mathbf{B}_n. \quad (17.4.26)$$

The matrix \mathbf{M} and the vector \mathbf{B}_n are known. To obtain the vector $\mathbf{u}(t_{n+1})$ that satisfies (17.4.25) one needs to invert the matrix \mathbf{M} . However, this is, in general, a challenging numerical problem because \mathbf{M} is a large matrix. Fortunately, most of its elements consist of zeros. Additionally, the nonzero elements typically form diagonal bands. Such matrices are called *sparse matrices* and there is an area of scientific computing that has specialized on solving sparse linear systems.

For the solution of sparse, linear systems one uses either direct solvers or iterative solvers. A direct solver aims to achieve the solution to the given precision of the computer. A popular direct solver is the *tridiagonal solver*, which is also known as the *Gaussian elimination method*, see Barrett, Berry, Chan, Demmel, Dongarra, Eijkhout, Pozo, Romine & van der Vorst (1994). This can only be achieved in certain situations with reasonable effort. When the matrix \mathbf{A} is tridiagonal, reliable and efficient methods exist. On the other hand, an iterative solver satisfies only an accuracy criterion and provides more flexibility and efficiency. This is usually fully sufficient for a finite difference method because there is no need to obtain exact solutions for the approximate difference equation. Iterative solvers, see Barrett et al. (1994), start from an initial guess and iteratively improve the solution. Examples are the, so called, Jacobi method, the Gauss-Seidel method and the successive overrelaxation method.

Numerical Illustrations of Finite Difference Methods

We perform some numerical experiments to highlight a few features of finite difference methods. We concentrate on the theta method, as described in (17.4.22). Here this method is applied to the Black-Scholes dynamics for the underlying security S_t , satisfying the SDE

$$dS_t = S_t (r dt + \sigma dW_t) \quad (17.4.27)$$

for $t \in [0, T]$ with $S_0 > 0$. At first we need to decide on the spatial interval that we like to cover with our finite difference method. Here we choose zero to be the lowest grid point in S direction and a level S_{\max} as the upper level. This level is simply chosen such that a further increase does not change much the results in the targeted range of accuracy.

For the European call payoff that we will consider, we choose for simplicity, so called, Dirichlet boundary conditions. At maturity the approximate

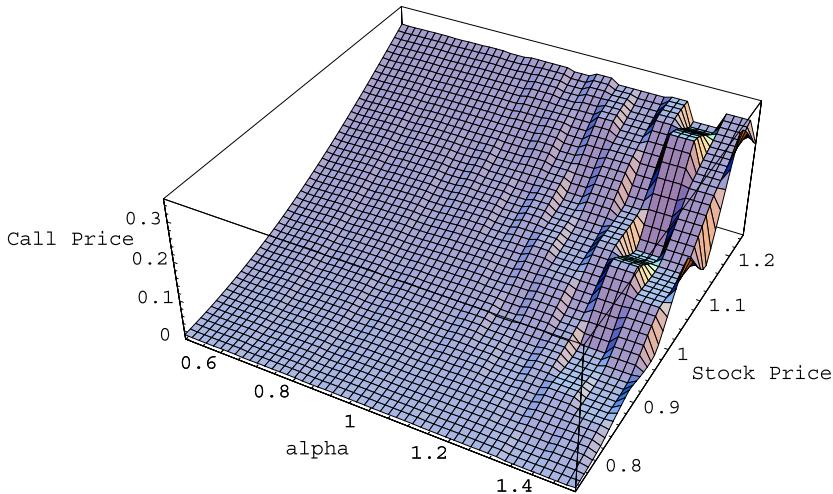


Fig. 17.4.1. Prices in dependence on α and S_0 for the explicit finite difference method

pricing function is set equal to the payoff function. To obtain the value of the approximate pricing function at one step back in time, we invert the resulting tridiagonal matrix, similar to that in (17.4.26), by using as direct solver the Gaussian elimination method. This then provides at each time step and each spatial grid point the approximate value of the European call option.

Since for European call options we have the explicit Black-Scholes formula, one can analyze for different parameter settings the differences between finite difference prices and theoretical prices from the Black-Scholes formula. The main requirement under the given Black-Scholes dynamics for the numerical stability of the explicit theta method, that is for $\theta = 0$, translates into the variable

$$\alpha = \left(\frac{\sigma^2 S^2}{(\Delta x)^2} + r \right) \Delta, \quad (17.4.28)$$

which needs to be less than one, see [Wilmott et al. \(1993\)](#). If this condition is satisfied, then errors will not be enhanced during the computation.

In Figure 17.4.1 we plot prices for the explicit finite difference method with $\theta = 0$ dependent on the parameter α when measured at the money. Here we set $\sigma = 0.2$, $r = 0.05$, $T = 1$ and $K = 1$ and consider a European call price in dependence on $S_0 \in (0, 3)$.

One notes that for α larger than one pricing errors become apparent that distort significantly the result. These errors are due to the fact that the time step size is too large compared to the spatial step size. From Figure 17.4.1 we notice some instability of the explicit finite difference method for too large time step sizes. If one compares this with Markov chain approximations, then one can see that for too large time step size one has, potentially, difficulties in obtaining positive probabilities when going to the next nodes, which, in

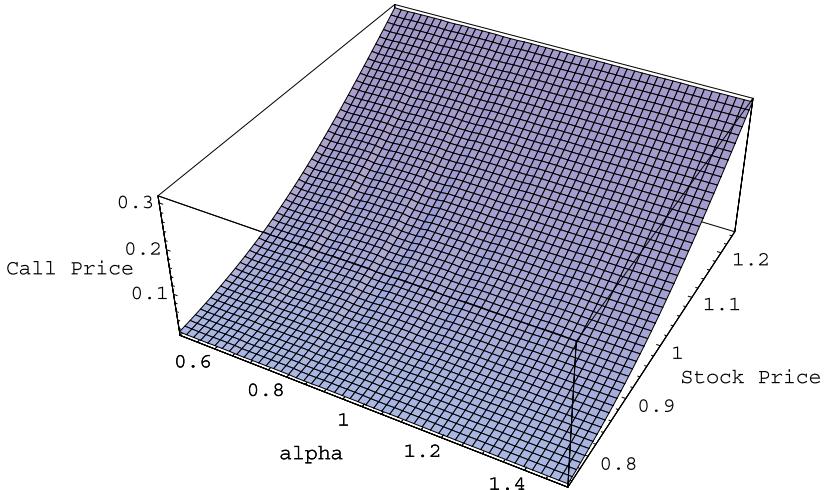


Fig. 17.4.2. Prices in dependence on α for implicit finite difference method

principle, also causes numerical problems for corresponding finite difference methods.

For comparison we plot in Figure 17.4.2 for the implicit finite difference method, that is $\theta = 1$, the prices obtained dependent on the parameter α as above. One notes that even for large time step size, when compared to the spatial step size, the implicit method controls well the propagation of errors when applying such a method. Also for the Crank-Nicolson method, that is with $\theta = \frac{1}{2}$, we obtain visually a similar graph as shown in Figure 17.4.2.

The Crank-Nicolson method arises for the choice $\theta = \frac{1}{2}$. This can be shown to be the smallest degree of implicitness θ that provides unconditional stability. However, since $\theta = \frac{1}{2}$ is just at the borderline of the stability region, in practice, it turns out that there may arise problems, see Shaw (1998) and Tavella & Randall (2000).

17.5 Convergence Theorem for Markov Chains

To formulate a weak convergence theorem for Markov chain approximations and a wider range of discrete-time approximations, we introduce the time discretization $(t)_\Delta$ of the interval $[0, T]$, with maximum step size $\Delta \in (0, 1)$ as a sequence of stopping times with

$$0 = t_0 < t_1 < \dots < t_{n_T} = T,$$

where

$$\sup_{n \in \{0, 1, \dots\}} (t_{n+1} - t_n) \leq \Delta$$

and $n_T < \infty$ P -a.s. Here we write again for all $t \in [0, T]$

$$n_t = \max\{n \in \{0, 1, \dots\} : t_n \leq t\}. \quad (17.5.1)$$

Furthermore, for all n the stopping time t_{n+1} is assumed to be \mathcal{A}_{t_n} -measurable.

For instance, $(t)_\Delta$ can be equidistant with step size Δ . But also a step size control is possible that is based on the information available until the previous discretization point. Such a step size control has been, for instance, indicated in (17.3.11).

We shall call a d -dimensional process $\mathbf{Y} = \{\mathbf{Y}_t, t \in [0, T]\}$, which is right-continuous with left hand limits, a discrete-time approximation if it is based on a time discretization $(t)_\Delta$ with maximum step size $\Delta \in (0, 1)$ such that $\mathbf{Y}_n = \mathbf{Y}_{t_n}$ is \mathcal{A}_{t_n} -measurable for each $n \in \{0, 1, \dots\}$ and \mathbf{Y}_{n+1} can be expressed as a function of $\mathbf{Y}_0, \dots, \mathbf{Y}_n, t_0, \dots, t_n, t_{n+1}$ and a finite number ℓ of $\mathcal{A}_{t_{n+1}}$ -measurable random variables $\mathbf{Z}_{n+1,j}$ for $j \in \{1, 2, \dots, \ell\}$.

This definition includes, for instance, the case of a Markov chain if we set $\ell = 1$ and $\mathbf{Z}_{n+1,1} = \mathbf{Y}_{n+1} - \mathbf{Y}_n$, where the random variable $\mathbf{Z}_{n+1,1}$ is characterized by the transition probabilities

$$P(\mathbf{Z}_{n+1,1} = \mathbf{y} - \mathbf{x} \mid \mathbf{Y}_n = \mathbf{x}) = p_n(\mathbf{x}, \mathbf{y}) = P(\mathbf{Y}_{n+1} = \mathbf{y} \mid \mathbf{Y}_n = \mathbf{x})$$

for $n \in \{0, 1, \dots\}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. It provides the possibility of a recursive simulation of \mathbf{Y} along the points of the time discretization. It also allows us to construct recombining trees or lattices. Obviously, it covers weak Monte Carlo simulation methods.

Various kinds of interpolations are possible between the time discretization points. For the case of a Markov chain, one can simply choose a piecewise constant interpolation such as

$$\mathbf{Y}_t = \mathbf{Y}_{t_n}$$

for all $t \in [0, T]$, see (17.5.1).

Furthermore, we introduce the operators

$$L^0 = \sum_{k=1}^d a^k \frac{\partial}{\partial x^k} + \frac{1}{2} \sum_{k,\ell=1}^d \delta^{k,\ell} \frac{\partial^2}{\partial x^k \partial x^\ell} \quad (17.5.2)$$

and

$$L^j = \sum_{k=1}^d b^{k,j} \frac{\partial}{\partial x^k} \quad (17.5.3)$$

for $j \in \{1, 2, \dots, m\}$, see (17.3.10), which are defined on the set of twice differentiable functions on \mathbb{R}^d .

We set for easier notation $W_t^0 = t$, $b^{k,0}(\mathbf{x}) = a^k(\mathbf{x})$ and write for given $\beta \in \mathcal{N}$ and $p_h \in \{1, 2, \dots, d\}$ the one step weak order β Taylor approximation, see (11.2.30), of its p_h th component

$$\eta_{n+1}^{p_h} = Y_n^{p_h} + \sum_{k=1}^{\beta} \sum_{j_1, \dots, j_k=0}^m L^{j_1} \dots L^{j_{k-1}} b^{p_h, j_k}(\mathbf{Y}_n) \int_{t_n}^{t_{n+1}} \dots \int_{t_n}^{s_2} dW_{s_1}^{j_1} \dots dW_{s_k}^{j_k}. \quad (17.5.4)$$

Throughout the rest of the section, $K < \infty$ and $r \in \mathcal{N}$ denote constants that are different at each place where they appear. The following weak convergence theorem is derived in [Platen \(1992\)](#).

Theorem 17.5.1. *Assume for a discrete-time approximation \mathbf{Y} defined with respect to a time discretization with maximum step size Δ that for $\beta \in \mathcal{N}$:*

- (i) \mathbf{Y}_0 converges weakly with respect to $\Delta \rightarrow 0$ to \mathbf{X}_0 with order β ,
- (ii) constants $K < \infty$ and $r \in \mathcal{N}$ exist that do not depend on Δ such that for all $\ell \in \{1, 2, \dots, 2\beta + 1\}$ and $(p_1, \dots, p_\ell) \in \{1, 2, \dots, d\}^\ell$ we have

$$\begin{aligned} & \left| E \left(\prod_{h=1}^{\ell} (Y_{n+1}^{p_h} - Y_n^{p_h}) - \prod_{h=1}^{\ell} (\eta_{n+1}^{p_h} - Y_n^{p_h}) \mid \mathcal{A}_{t_n} \right) \right| \\ & \leq K \left(1 + \max_{0 \leq k \leq n} |\mathbf{Y}_k|^{2r} \right) \Delta^\beta (t_{n+1} - t_n) \end{aligned} \quad (17.5.5)$$

for all $n \in \{0, 1, \dots\}$,

- (iii) $a^k, b^{k,j} \in \mathcal{C}^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ (17.5.6)

for all $k \in \{1, 2, \dots, d\}$ and $j \in \{1, 2, \dots, m\}$,

- (iv) for all $k \in \{1, 2, \dots, d\}$, $\ell \in \{1, 2, \dots, \beta\}$, $(j_1, \dots, j_\ell) \in \{0, 1, \dots, m\}^\ell$ and $\mathbf{x} \in \mathbb{R}^d$ the following linear growth condition holds

$$|L^{j_1} \dots L^{j_{\ell-1}} b^{k,j_\ell}(\mathbf{x})| \leq K(1 + |\mathbf{x}|), \quad (17.5.7)$$

- (v) for each $p \in \mathcal{N}$ there exist constants $K < \infty$ and $r \in \mathcal{N}$ that do not depend on Δ such that for all $q \in \{1, 2, \dots, p\}$ it holds that

$$E \left(\max_{0 \leq n \leq n_T} |\mathbf{Y}_n|^{2q} \mid \mathcal{A}_{t_0} \right) \leq K(1 + |\mathbf{Y}_0|^{2r}) \quad (17.5.8)$$

and

$$E(|\mathbf{Y}_{n+1} - \mathbf{Y}_n|^{2q} \mid \mathcal{A}_{t_n}) \leq K \left(1 + \max_{0 \leq k \leq n} |\mathbf{Y}_k|^{2r} \right) (t_{n+1} - t_n)^q \quad (17.5.9)$$

for all $n \in \{0, 1, \dots\}$,

then the discrete-time approximation \mathbf{Y} converges with weak order β to \mathbf{X} for $\Delta \rightarrow 0$.

The proof of this theorem will be given at the end of this section. The presented approach can be easily extended to the case of jump diffusion processes by using similar arguments as applied in previous chapters.

Comments on the Conditions of the Theorem

Let us discuss the conditions of the theorem. Assumption (ii) with the condition (17.5.5) provides the core rule on how to construct a Markov chain of weak order β . One has essentially to match the mixed moments up to order $2\beta + 1$ with those of the weak order β Taylor scheme. It is easy to check that the conditions of equations (17.3.9), (17.3.13), (17.3.17) and (17.3.19) are special cases of the inequality (17.5.5). Note that one needs at least, in general, $\sum_{k=1}^{2\beta} d^k$ “neighboring” points to construct a Markov chain of weak order β because one has to satisfy the same number of conditions as that contained in condition (ii). One has to admit that even for a weak scheme of order two and dimension ten this is a substantial number of the order of several hundreds.

Condition (iii) provides the necessary smoothness of the drift and diffusion coefficients of the underlying SDE, as previously required for weak Taylor approximations of corresponding order, see Theorem 11.2.1.

Condition (iv) ensures the existence and regularity of the solutions of the SDE together with those of the discrete-time approximations. In the preceding examples, we used much stronger boundedness conditions to fulfill condition (iv) in a convenient way. Such boundedness is not necessary, as can be seen.

Assumption (v) gives us the existence of higher moments of the approximations and thus the meaningfulness of the weak error criterion (11.1.2) and condition (ii).

It must be emphasized that the preceding weak convergence theorem covers not only the case of Markov chains approximating a diffusion but also Monte Carlo simulation schemes such as, for instance, the Euler scheme or the weak order 2.0 predictor-corrector method (11.5.20).

We have presented a general result implying higher order weak convergence of discrete-time approximations of diffusions. As we will see from the proof, the result relies heavily on the use of the Wagner-Platen expansion. Various examples of approximating Markov chains for orders up to $\beta = 3$ have been presented previously. There are more examples that can be found in the literature. Many other approximating Markov chains, satisfying the conditions of the theorem, can also be tailored for particular purposes. Basically, the moment conditions (17.5.5) provide the rule for the choice of the transition probabilities of the approximating Markov chain. In the case $\beta = 1$, they represent mainly the consistency conditions formulated in Kushner (1977), Platen & Rebolledo (1985) or Jacod & Shiryaev (2003) that establish some weak convergence. Here we also provide sufficient conditions to achieve some higher weak order.

Some Preparations for the Proof

Suppose for fixed weak order $\beta \in \mathcal{N}$ and payoff function $g \in \mathcal{C}_P^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ the assumptions of the theorem to be satisfied. In the following, we list some notations and results that we will need for the proof.

We introduce the diffusion process

$$\mathbf{X}^{s,y} = \mathbf{y} + \int_s^t \mathbf{a}(\mathbf{X}_z^{s,y}) dz + \int_s^t \mathbf{b}(\mathbf{X}_z^{s,y}) d\mathbf{W}_z \quad (17.5.10)$$

for $t \in [s, T]$, which starts at time $s \in [0, T]$ in $\mathbf{y} \in \Re^d$ and has the same drift and diffusion coefficients as that defined in (17.3.4).

For all $s \in [0, T]$ and $\mathbf{y} \in \Re^d$, we define the functional

$$u(s, \mathbf{y}) = E(g(\mathbf{X}_T^{s,y})) \quad (17.5.11)$$

and it is

$$u(0, \mathbf{X}_0) = E(g(\mathbf{X}_T^{0,X_0})) = E(g(\mathbf{X}_T)), \quad (17.5.12)$$

where \mathbf{X} was defined in (17.3.4). According to the Feynman-Kac formula, see Sect. 2.7, the function $u(\cdot, \cdot)$ is the unique solution of the Kolmogorov-backward equation

$$L^0 u(s, \mathbf{y}) = 0 \quad (17.5.13)$$

for all $(s, \mathbf{y}) \in (0, T) \times \Re^d$ with terminal condition

$$u(T, \mathbf{y}) = g(\mathbf{y}) \quad (17.5.14)$$

for $\mathbf{y} \in \Re^d$. Here the operator L^0 is given in (17.5.2). We have by a result in Mikulevicius (1983) that

$$u(s, \cdot) = \mathcal{C}_P^{2(\beta+1)}(\Re^d, \Re) \quad (17.5.15)$$

for all $s \in [0, T]$.

By use of the Itô formula, see Chap. 1, it follows with (17.5.13) for all $n \in \{1, 2, \dots, n_T\}$ and $\mathbf{y} \in \Re^d$ that

$$E \left(u \left(t_n, \mathbf{X}_{t_n}^{t_{n-2}, y} \right) - u(t_{n-1}, \mathbf{y}) \mid \mathcal{A}_{t_{n-1}} \right) = 0. \quad (17.5.16)$$

According to an estimate in Krylov (1980), there exists for each $p \in \mathcal{N}$ a constant $K < \infty$ such that for all $q \in \{1, 2, \dots, p\}$, all $\mathbf{y} \in \Re^d$ and all $n \in \{1, 2, \dots, n_T\}$ it holds that

$$E \left(|\mathbf{X}_{t_n}^{t_{n-1}, Y} - \mathbf{y}| \mid \mathcal{A}_{t_{n-1}} \right) \leq K (1 + |\mathbf{y}|^{2q}) (t_n - t_{n-1})^q. \quad (17.5.17)$$

To achieve a compact notation we introduce for all $\ell \in \mathcal{N}$ the set

$$P_\ell = \{1, 2, \dots, d\}^\ell,$$

and define for all $\mathbf{y} = (y^1, \dots, y^d)^\top \in \Re^d$ and $\mathbf{p} = (p_1, \dots, p_\ell) \in P_\ell$ the product

$$F_{\mathbf{p}}(\mathbf{y}) = \prod_{h=1}^{\ell} y_h^{p_h}, \quad (17.5.18)$$

where y^p denotes the p th component of \mathbf{y} .

In [Kloeden & Platen \(1999\)](#) it is shown that there exist constants $K < \infty$ and $r \in \mathcal{N}$ such that for all $n \in \{1, 2, \dots, n_T\}$, $\ell \in \{1, 2, \dots, 2(\beta + 1)\}$ and $\mathbf{p} \in P_\ell$ it holds that the estimate

$$\begin{aligned} E \left(\left| F_{\mathbf{p}}(\boldsymbol{\eta}_n - \mathbf{Y}_{n-1}) - F_{\mathbf{p}} \left(\mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}} - \mathbf{Y}_{n-1} \right) \right| \mid \mathcal{A}_{t_{n-1}} \right) \\ \leq K (1 + |\mathbf{Y}_{n-1}|^r) (t_n - t_{n-1})^{\beta+1}. \end{aligned} \quad (17.5.19)$$

Therefore, it is shown that for each $p \in \mathcal{N}$ there exist constants $K < \infty$ and $r \in \mathcal{N}$ such that for all $\ell \in \{1, 2, \dots, 2(\beta + 1)\}$, $q \in \{1, 2, \dots, p\}$, $\mathbf{p} \in P_\ell$ and $n \in \{1, 2, \dots, n_T\}$ it holds that

$$E \left(|F_{\mathbf{p}}(\boldsymbol{\eta}_n - \mathbf{Y}_{n-1})|^{2q} \mid \mathcal{A}_{t_{n-1}} \right) \leq K (1 + |\mathbf{Y}_{n-1}|^{2r}) (t_n - t_{n-1})^{q\ell}. \quad (17.5.20)$$

Using an estimate from [Krylov \(1980\)](#), p. 78, one has for $q \in \{1, 2, \dots\}$ and $i \in \{0, 1, \dots, n_T - 1\}$ a function f , with

$$R = \left(E \left(\sup_{t_n \leq s \leq t_{n+1}} |f(s)|^{2q} \mid \mathcal{A}_{t_{n-1}} \right) \right)^{\frac{1}{q}} < \infty$$

$\ell \in \{1, 2, \dots\}$ and $j_1, \dots, j_\ell \in \{0, 1, \dots, m\}$ the inequality

$$\left(E \left(\left| \int_{t_n}^{t_{n+1}} \cdots \int_{t_n}^{s_2} g(s_1) dW_{s_1}^{j_1} \cdots dW_{s_\ell}^{j_\ell} \right|^{2q} \right) \right)^{\frac{1}{q}} \leq K_{q,\ell} (t_{n+1} - t_n)^\ell \quad (17.5.21)$$

where $K_{q,\ell}$ is a finite constant depending on q and ℓ but not on Δ .

Proof of Theorem 17.5.1

We consider the weak error

$$\mu(\Delta) = |E(g(\mathbf{Y}_{n_T})) - E(g(\mathbf{X}_T))| \quad (17.5.22)$$

which is to be estimated. By the use of the terminal condition (17.5.14), the Kolmogorov-backward equation, and the relation in (17.5.14), we can write

$$\mu(\Delta) = |E(u(T, \mathbf{Y}_{n_T}) - u(0, \mathbf{X}_0))|.$$

Because \mathbf{Y}_0 converges weakly with order β to \mathbf{X}_0 , we have

$$\mu(\Delta) \leq \left| E \left(\sum_{n=1}^{n_T} (u(t_n, \mathbf{Y}_n) - u(t_{n-1}, \mathbf{Y}_{n-1})) \right) \right| + K \Delta^\beta.$$

The relation (17.5.16) yields

$$\begin{aligned}
\mu(\Delta) &\leq \left| E \left(\sum_{n=1}^{n_T} \left[(u(t_n, \mathbf{Y}_n) - u(t_{n-1}, \mathbf{Y}_{n-1})) \right. \right. \right. \\
&\quad \left. \left. \left. - (u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}}) - u(t_{n-1}, \mathbf{Y}_{n-1})) \right] \right) \right| + K \Delta^\beta \\
&\leq \left| E \left(\sum_{n=1}^{n_T} \left[(u(t_n, \mathbf{Y}_n) - u(t_{n-1}, \mathbf{Y}_{n-1})) \right. \right. \right. \\
&\quad \left. \left. \left. - (u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}}) - u(t_n, \mathbf{Y}_{n-1})) \right] \right) \right| + K \Delta^\beta \\
&= H_1 + H_2 + K \Delta^\beta
\end{aligned} \tag{17.5.23}$$

with

$$H_1 = \left| E \left(\sum_{n=1}^{n_T} \left[(u(t_n, \mathbf{Y}_n) - u(t_n, \mathbf{Y}_{n-1})) - (u(t_n, \boldsymbol{\eta}_n) - u(t_n, \mathbf{Y}_{n-1})) \right] \right) \right| \tag{17.5.24}$$

and

$$\begin{aligned}
H_2 = & \left| E \left(\sum_{n=1}^{n_T} \left[(u(t_n, \boldsymbol{\eta}_n) - u(t_n, \mathbf{Y}_{n-1})) \right. \right. \right. \\
& \quad \left. \left. \left. - (u(t_n, \mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}}) - u(t_n, \mathbf{Y}_{n-1})) \right] \right) \right|.
\end{aligned} \tag{17.5.25}$$

Using the smoothness of u , as expressed in (17.5.15), we can expand the increments of u , in (17.5.24) by the deterministic Taylor formula

$$\begin{aligned}
H_1 = & \left| E \left(\sum_{n=1}^{n_T} \left[\left(\sum_{\ell=2}^{2\beta+1} \frac{1}{\ell!} \sum_{\mathbf{p} \in P_\ell} (\partial_y^\mathbf{p} u(t_n, \mathbf{Y}_{n-1})) F_{\mathbf{p}}(\mathbf{Y}_n - \mathbf{Y}_{n-1}) + R_n(\mathbf{Y}_n) \right) \right. \right. \right. \\
& \quad \left. \left. \left. - \left(\sum_{\ell=1}^{2\beta+1} \frac{1}{\ell!} \sum_{\mathbf{p} \in P_\ell} (\partial_y^\mathbf{p} u(t_n, \mathbf{Y}_{n-1})) F_{\mathbf{p}}(\boldsymbol{\eta}_n - \mathbf{Y}_{n-1}) + R_n(\boldsymbol{\eta}_n) \right) \right] \right) \right|.
\end{aligned} \tag{17.5.26}$$

Here the remainder terms are of the form

$$\begin{aligned}
R_n(\mathbf{Z}) &= \frac{1}{(2(\beta+1))!} \sum_{\mathbf{p} \in P_{2(\beta+1)}} (\partial_y^\mathbf{p} u(t_n, \mathbf{Y}_{n-1} + \boldsymbol{\Theta}_{\mathbf{p}, n}(\mathbf{Z})(\mathbf{Z} - \mathbf{Y}_{n-1}))) \\
&\quad \times F_{\mathbf{p}}(\mathbf{Z} - \mathbf{Y}_{n-1})
\end{aligned} \tag{17.5.27}$$

for $\mathbf{Z} = \mathbf{Y}_n$ and $\boldsymbol{\eta}_n$, respectively, where we have the following elements of the $d \times d$ matrix $\boldsymbol{\Theta}_{\mathbf{p}, n}(Z)$

$$\hat{\Theta}_{\mathbf{p},n}^{k,j}(\mathbf{Z}) = \begin{cases} \hat{\Theta}_{\mathbf{p},n}^k(\mathbf{Z}) \in (0,1) & \text{for } k=j \\ 0 & \text{otherwise} \end{cases} \quad (17.5.28)$$

$k, j \in \{1, 2, \dots, d\}$. Furthermore, from (17.5.26) we can write

$$\begin{aligned} H_1 \leq & E \left(\sum_{n=1}^{n_T} \left[\sum_{\ell=1}^{2\beta+1} \frac{1}{\ell!} \sum_{\mathbf{p} \in P_\ell} |\partial_y^\mathbf{p} u(t_n, \mathbf{Y}_{n-1})| \right. \right. \\ & \times |E(F_{\mathbf{p}}(\mathbf{Y}_n - \mathbf{Y}_{n-1}) - F_{\mathbf{p}}(\boldsymbol{\eta}_n - \mathbf{Y}_{n-1}) | \mathcal{A}_{t_{n-1}})| \\ & \left. \left. + E(|R_n(\mathbf{Y}_n)| | \mathcal{A}_{t_{n-1}}) + E(|R_n(\boldsymbol{\eta}_n)| | \mathcal{A}_{t_{n-1}}) \right] \right). \quad (17.5.29) \end{aligned}$$

Now, we obtain, with the help of (17.5.27), (17.5.15), (17.5.25), (17.5.28) and (17.5.9), for all $n \in \{1, 2, \dots, n_T\}$

$$\begin{aligned} & E(|R_n(\mathbf{Y}_n)| | \mathcal{A}_{t_{n-1}}) \\ & \leq K \sum_{\mathbf{p} \in P_{2(\beta+1)}} \left(E \left(|\partial_y^\mathbf{p} u(t_n, \mathbf{Y}_{n-1}) + \boldsymbol{\Theta}_{\mathbf{p},n}(\mathbf{Y}_n)(\mathbf{Y}_n - \mathbf{Y}_{n-1})|^2 | \mathcal{A}_{t_{n-1}} \right) \right)^{\frac{1}{2}} \\ & \quad \times \left(E \left(|F_{\mathbf{p}}(\mathbf{Y}_n - \mathbf{Y}_{n-1})|^2 | \mathcal{A}_{t_{n-1}} \right) \right)^{\frac{1}{2}} \\ & \leq K \left(E \left(1 + |\mathbf{Y}_{n-1}|^{2r} + |\mathbf{Y}_n - \mathbf{Y}_{n-1}|^2 | \mathcal{A}_{t_{n-1}} \right) \right)^{\frac{1}{2}} \\ & \quad \times \left(E \left(|\mathbf{Y}_n - \mathbf{Y}_{n-1}|^{4(\beta+1)} | \mathcal{A}_{t_{n-1}} \right) \right)^{\frac{1}{2}} \\ & \leq K \left(1 + \max_{0 \leq k \leq n-1} |\mathbf{Y}_k|^{2r} \right) \Delta^\beta (t_n - t_{n-1}). \quad (17.5.30) \end{aligned}$$

In a similar way we obtain from (17.5.27), (17.5.4), (17.5.7) and the moment property (17.5.21) of multiple Itô integrals, for all $n \in \{1, 2, \dots, n_T\}$ the estimate

$$E(|R_n(\boldsymbol{\eta}_n)| | \mathcal{A}_{t_{n-1}}) \leq K \left(1 + |\mathbf{Y}_{n-1}|^{2r} \right) \Delta^\beta (t_n - t_{n-1}). \quad (17.5.31)$$

From (17.5.29), it follows now by application of equations (17.5.15), (17.5.19), (17.5.4), (17.5.5), (17.5.30), (17.5.31), (17.5.8), (17.5.9) the inequality

$$\begin{aligned}
H_1 &\leq E \left(\sum_{n=1}^{n_T} K \left(1 + \max_{0 \leq k \leq n_T} |\mathbf{Y}_k|^{2r} \right) \Delta^\beta (t_n - t_{n-1}) \right) \\
&\leq \Delta^\beta K \left(1 + E \left(\max_{0 \leq k \leq n_T} |\mathbf{Y}_k|^{2r} \right) \right) \\
&\leq \Delta^\beta K (1 + E(|\mathbf{Y}_0|^{2r})) \\
&\leq K \Delta^\beta.
\end{aligned} \tag{17.5.32}$$

By the use of (17.5.19), we can now estimate H_2 , as defined in (17.5.25), and in an analogous way as H_1 :

$$\begin{aligned}
H_2 &\leq E \left(\sum_{n=1}^{n_T} \left[\frac{1}{\ell!} \sum_{\mathbf{p} \in P_\ell} |\partial_y^{\mathbf{p}} u(t_n, \mathbf{Y}_{n-1})| \right. \right. \\
&\quad \times \left| E \left(F_{\mathbf{p}}(\boldsymbol{\eta}_n - \mathbf{Y}_{n-1}) - F_{\mathbf{p}}(\mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}} - \mathbf{Y}_{n-1}) \mid \mathcal{A}_{t_{n-1}} \right) \right| \\
&\quad \left. \left. + E(|R_n(\boldsymbol{\eta}_n)| \mid \mathcal{A}_{t_{n-1}}) + E(|R_n(\mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}})| \mid \mathcal{A}_{t_{n-1}}) \right] \right) \\
&\leq E \left(\sum_{n=1}^{n_T} \left[K (1 + |\mathbf{Y}_{n-1}|^{2r}) \Delta^\beta (t_n - t_{n-1}) + E(|R_n(\mathbf{X}_{t_n}^{t_{n-1}, Y_{n-1}})| \mid \mathcal{A}_{t_{n-1}}) \right] \right).
\end{aligned} \tag{17.5.33}$$

Because the remainder in (17.5.33) can be estimated similar to that in equation (17.5.30) by using (17.5.17) we obtain

$$\begin{aligned}
H_2 &\leq E \left(\sum_{n=1}^{n_T} K (1 + |\mathbf{Y}_{n-1}|^{2r}) \Delta^\beta (t_n - t_{n-1}) \right) \\
&\leq \Delta^\beta K \left(E \left(\max_{0 \leq k \leq n_T} |\mathbf{Y}_k|^{2r} \right) + 1 \right).
\end{aligned}$$

It follows with condition (17.5.8), and because \mathbf{Y}_0 has finite moments, that

$$H_2 = K \Delta^\beta. \tag{17.5.34}$$

Summarizing equations (17.5.22), (17.5.32) and (17.5.34), we finally obtain the desired result

$$\mu(\Delta) = |E(g(\mathbf{Y}_{n_T})) - E(g(\mathbf{X}_T))| \leq K_g \Delta^\beta,$$

which proves Theorem 17.5.1. \square

17.6 Exercises

17.1. For a European put option with benchmarked strike $\hat{K} > 0$ at maturity $T = i\Delta$, $i \in \{1, 2, \dots\}$, and time step size $\Delta \in (0, 1)$, write down the multi-period binomial tree option pricing formula at time $t = 0$ when a positive benchmarked return of the underlying security is of the form $u = \exp\{\sigma \sqrt{\Delta}\}$. We assume that the growth optimal portfolio has the value one at time $t = 0$.

17.2. What is the weak order of convergence that a recombining trinomial tree can, in general, achieve assuming sufficiently smooth drift and diffusion coefficients and payoff function?

17.3. How many nodes need to be reached from each node in one time step to attain, in general, under appropriate conditions second weak order for a one-dimensional recombining tree?

Solutions for Exercises

Solutions for Exercises of Chapter 1

1.1 By application of the Itô formula we obtain

$$dZ_t = Z_t \left(\mu + \frac{1}{2} \sigma^2 \right) dt + Z_t \sigma dW_t.$$

Applying again the Itô formula we obtain

$$\begin{aligned} d\ln(Z_t) &= \left(\mu + \frac{1}{2} \sigma^2 - \frac{1}{2} \sigma^2 \right) dt + \sigma dW_t \\ &= \mu dt + \sigma dW_t. \end{aligned}$$

1.2 We have by the Itô formula the stochastic differential

$$d(Y_t^1 Y_t^2) = (Y_t^2 a_1 + Y_t^1 a_2) dt + Y_t^2 b_1 dW_t^1 + Y_t^1 b_2 dW_t^2.$$

1.3 Applying the Itô formula we obtain

$$d(W_t)^2 = 2 W_t dW_t + dt.$$

Now by the covariation property (1.4.15) of Itô integrals we have

$$\begin{aligned} [W, (W)^2]_t &= \left[\int_0^t dW_s, 2 \int_0^t W_s dW_s + \int_0^t ds \right]_t \\ &= \int_0^t 2 W_s ds. \end{aligned}$$

1.4 The stochastic differential of X is

$$dX_t = \sigma dW_t + \xi dp(t)$$

for $t \in [0, T]$. By the Itô formula (1.5.15) it follows that

$$d \exp\{X_t\} = \exp\{X_t\} \left(\sigma dW_t + \frac{1}{2} \sigma^2 dt \right) + \exp\{X_{t-}\} (\exp\{\xi\} - 1) dp(t)$$

for $t \in [0, T]$.

1.5 The stochastic differential of X is

$$dX_t = a dp^1(t) + b dp^2(t)$$

for $t \in [0, T]$. Using the Itô formula (1.5.15) we obtain

$$\begin{aligned} d \exp\{X_t\} &= \exp\{X_{t-}\} (\exp\{a\} - 1) dp^1(t) \\ &\quad + \exp\{X_{t-}\} (\exp\{b\} - 1) dp^2(t) \end{aligned}$$

for $t \in [0, T]$.

1.6 We obtain with the choice $\gamma = 1$, $\bar{x} = 0$ and $\beta = \sqrt{2}$ from (1.7.5) or (1.7.6) the mean

$$\mu(t) = E(X_t) = \exp\{-(t - t_0)\}.$$

Furthermore we obtain then from (1.7.6) the variance

$$\begin{aligned} v(t) &= E((X_t - E(X_t))^2) \\ &= E\left(\left(\int_{t_0}^t \sqrt{2} \exp\{-(t-s)\} dW_s\right)^2\right) \\ &= 2 \int_{t_0}^t \exp\{-2(t-s)\} ds \\ &= 1 - \exp\{-2(t-t_0)\}. \end{aligned}$$

1.7 We apply formula (1.7.21), where

$$\begin{aligned} X_t &= X_0 \Psi_{t,0} \\ &= X_0 \exp \left\{ \left(-\frac{1}{2} - \frac{1}{2} - \frac{1}{2} \right) t + \sum_{l=1}^2 (W_t^l - W_0^l) \right\} \\ &= X_0 \exp \left\{ -\frac{3}{2} t + W_t^1 + W_t^2 \right\}. \end{aligned}$$

1.8 By the Itô formula one obtains

$$dX_t = X_{t-} \left[\left(k a + \frac{k^2 b^2}{2} \right) dt + k b dW_t + (\exp\{k c\} - 1) dN_t \right]$$

for $t \in [0, T]$ with $X_0 = 1$.

1.9 In the above linear SDE for X_t we take the expectation and obtain

$$d\mu(t) = \mu(t-) \left[k a + \frac{k^2 b^2}{2} + \lambda (\exp\{k c\} - 1) \right] dt$$

for $t \in [0, T]$ with $\mu(0) = 1$, which yields

$$\mu(t) = \exp \left\{ t \left[k a + \frac{k^2 b^2}{2} + \lambda (\exp\{k c\} - 1) \right] \right\}.$$

1.10 We obtain by the formula (1.1.20) for the Poisson probabilities that

$$\begin{aligned} E(N_t^2) &= \sum_{k=1}^{\infty} \frac{k^2}{k!} e^{-\lambda t} (\lambda t)^k \\ &= \lambda t \left(\sum_{k=1}^{\infty} (k-1) e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!} \right) \\ &= \lambda t (\lambda t + 1). \end{aligned}$$

1.11 By the independence of the marks from the Poisson process it follows that

$$E(Y_t) = E \left(\sum_{k=1}^{N_t} \xi_k \right) = E(N_t) E(\xi_k) = \frac{\lambda t}{2}.$$

1.12 The probability for a compound Poisson process with intensity $\lambda > 0$ of having no jumps until time $t > 0$ equals the probability of the Poisson process N of having no jumps until that time. Thus, by (1.1.20) we obtain

$$P(N_t = 0) = e^{-\lambda t}.$$

Solutions for Exercises of Chapter 2

2.1 The two-dimensional process with two correlated, drifted Wiener processes as components has a bivariate Gaussian transition density. If $\varrho \in (-1, 1)$ denotes the correlation and μ_1 the drift of the first and μ_2 the drift of the second component, then the transition density is of the form

$$p(s, x_1, x_2; t, y_1, y_2) = \frac{1}{2\pi(t-s)\sqrt{1-\varrho^2}} \exp\left\{-\frac{(y_1 - x_1 - \mu_1 t)^2 - 2(y_1 - x_1 - \mu_1 t)(y_2 - x_2 - \mu_2 t)\varrho + (y_2 - x_2 - \mu_2 t)^2}{2(t-s)(1-\varrho^2)}\right\}$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $x_1, x_2, y_1, y_2 \in \mathfrak{R}$.

2.2 We obtain the SDEs

$$dX_t^1 = \mu_1 dt + dW_t^1$$

and

$$dX_t^2 = \mu_2 dt + \varrho dW_t^1 + \sqrt{1-\varrho^2} dW_t^2,$$

for $t \geq 0$.

2.3 The logarithms of each component of the geometric Brownian motion are linearly transformed correlated Wiener processes. This leads to the transition density

$$\begin{aligned} p(s, x_1, x_2; t, y_1, y_2) &= \frac{1}{2\pi(t-s)\sqrt{1-\varrho^2} b^2 y_1 y_2} \\ &\times \exp\left\{-\frac{1}{2b^2(t-s)(1-\varrho^2)} \left[\left(\ln(y_1) - \ln(x_1) + \frac{1}{2}b^2(t-s) \right)^2 \right. \right. \\ &- 2\varrho \left(\ln(y_1) - \ln(x_1) + \frac{1}{2}b^2(t-s) \right) \left(\ln(y_2) - \ln(x_2) + \frac{1}{2}b^2(t-s) \right) \\ &\left. \left. + \left(\ln(y_2) - \ln(x_2) + \frac{1}{2}b^2(t-s) \right)^2 \right] \right\} \end{aligned}$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $x_1, x_2, y_1, y_2 \in \mathfrak{R}^+$.

2.4 The SDEs are the following

$$dX_t^1 = X_t^1 b dW_t^1$$

$$dX_t^2 = X_t^2 b \left(\varrho dW_t^1 + \sqrt{1-\varrho^2} dW_t^2 \right)$$

for $t \geq 0$ and $X_0^1, X_0^2 \in \mathfrak{R}^+$.

2.5 We apply the Itô formula to obtain the SDE for the function $u(t, X_t)$, where

$$dX_t = a X_t dt + b X_t dW_t + c X_{t-} dN_t.$$

This provides the equation

$$\begin{aligned} H(X_T) &= U(T, X_T) \\ &= u(t, X_t) + \int_t^T \left(\frac{\partial}{\partial t} u(s, X_s) + a X_s \frac{\partial}{\partial X} u(s, X_s) \right. \\ &\quad \left. + \frac{b^2}{2} X_s^2 \frac{\partial^2}{\partial X^2} u(s, X_s) + \lambda(u(s, X_s(1+c)) - u(s, X_s)) \right) ds \\ &\quad + \int_t^T b X_s \frac{\partial}{\partial X} u(s, X_s) dW_s \\ &\quad + \int_t^T (u(s, X_{s-}(1+c)) - u(s, X_{s-})) (dN_s - \lambda ds). \end{aligned}$$

Since $u(t, X_t)$ is a martingale the drift in the above SDE has to vanish and it needs to hold for all $(t, x) \in [0, T] \times \mathbb{R}^+$ that

$$\frac{\partial}{\partial t} u(t, x) + a x \frac{\partial}{\partial X} u(s, x) + \frac{b^2}{2} x^2 \frac{\partial^2}{\partial X^2} u(s, x) + \lambda(u(s, x(1+c)) - u(s, x)) = 0$$

and $H(x) = u(T, x)$, which confirms the Feynman-Kac formula.

2.6 We consider the square root process $Y = \{Y_t, t \in [0, \infty)\}$ of dimension $\delta > 2$ satisfying the SDE

$$dY_t = \left(\frac{\delta}{4} c^2 + b Y_t \right) dt + c \sqrt{Y_t} dW_t$$

for $t \in [0, \infty)$, $Y_0 > 0$, $c > 0$ and $b < 0$. The Itô integral

$$M_t = c \int_0^t \sqrt{Y_s} dW_s$$

forms a martingale due to Lemma 1.3.2 (iii), since the square root process, as a transformed time changed squared Bessel process, has moments of any positive order. Consequently,

$$dE(Y_t) = \left(\frac{\delta}{4} c^2 + b E(Y_t) \right) dt$$

for $t \in [0, \infty)$ with $E(Y_0) > 0$. Therefore, we obtain

$$E(Y_t) = E(Y_0) \exp\{bt\} + \frac{\delta c^2}{4b} (1 - \exp\{bt\}).$$

2.7 Using the time homogenous Fokker-Planck equation the stationary density of the ARCH diffusion model is of the form

$$\begin{aligned}\bar{p}(\theta^2) &= \frac{C}{\gamma^2 \theta^4} \exp \left\{ 2 \int_{\underline{\theta}^2}^{\theta^2} \frac{\kappa (\bar{\theta}^2 - u)}{\gamma^2 u^2} du \right\} \\ &= \frac{C}{\gamma^2 \theta^4} \exp \left\{ \frac{2 \kappa \bar{\theta}^2}{\gamma^2} \int_{\underline{\theta}^2}^{\theta^2} \frac{1}{u^2} du - \frac{2 \kappa}{\gamma^2} \int_{\underline{\theta}^2}^{\theta^2} \frac{1}{u} du \right\} \\ &= \frac{C}{\gamma^2 \theta^4} \exp \left\{ \frac{2 \kappa}{\gamma^2} \left(\bar{\theta}^2 \left(-\frac{1}{\theta^2} + \frac{1}{\underline{\theta}^2} \right) - (\ln(\theta^2) - \ln(\underline{\theta}^2)) \right) \right\} \\ &= C_1 \exp \left\{ -\frac{2 \kappa \bar{\theta}^2}{\gamma^2} \frac{1}{\theta^2} \right\} \left(\frac{1}{\theta^2} \right)^{\frac{2 \kappa}{\gamma^2} + 2},\end{aligned}$$

which is an inverse gamma density with an appropriate constant $C_1 > 0$.

2.8 The stationary density for the squared volatility needs to satisfy the expression

$$\begin{aligned}\bar{p}(\theta^2) &= \frac{C}{\gamma^2 \theta^6} \exp \left\{ 2 \int_{\underline{\theta}^2}^{\theta^2} \frac{\kappa (\bar{\theta}^2 - u) u}{\gamma^2 u^3} du \right\} \\ &= \frac{C}{\gamma^2 \theta^6} \exp \left\{ \frac{2 \kappa}{\gamma^2} \left(\bar{\theta}^2 \int_{\underline{\theta}^2}^{\theta^2} \frac{1}{u^2} du - \int_{\underline{\theta}^2}^{\theta^2} \frac{1}{u} du \right) \right\} \\ &= \frac{C}{\gamma^2 \theta^6} \exp \left\{ \frac{2 \kappa}{\gamma^2} \left(-\bar{\theta}^2 \left(\frac{1}{\theta^2} - \frac{1}{\underline{\theta}^2} \right) - (\ln(\theta^2) - \ln(\underline{\theta}^2)) \right) \right\} \\ &= \frac{C}{\gamma^2 \theta^6} \exp \left\{ \frac{2 \kappa}{\gamma^2} \left(-\bar{\theta}^2 \left(\frac{1}{\theta^2} - \frac{1}{\underline{\theta}^2} \right) - \ln(\theta^2) + \ln(\underline{\theta}^2) \right) \right\} \\ &= C_1 \exp \left\{ -\frac{2 \kappa}{\gamma^2} \bar{\theta}^2 \frac{1}{\theta^2} \right\} \left(\frac{1}{\theta^2} \right)^{\frac{2 \kappa}{\gamma^2} + 3},\end{aligned}$$

which is an inverse gamma density.

2.9 For the Heston model we obtain the stationary density for the squared volatility

$$\begin{aligned}\bar{p}(\theta^2) &= \frac{C}{\gamma^2 \theta^2} \exp \left\{ 2 \int_{\underline{\theta}^2}^{\theta^2} \frac{\kappa (\bar{\theta}^2 - u)}{\gamma^2 u} du \right\} \\ &= \frac{C}{\gamma^2 \theta^2} \exp \left\{ \frac{2 \kappa \bar{\theta}^2}{\gamma^2} \int_{\underline{\theta}^2}^{\theta^2} \frac{1}{u} du - \frac{2 \kappa}{\gamma^2} (\theta^2 - \underline{\theta}^2) \right\} = C_1 \exp \left\{ -\frac{2 \kappa}{\gamma^2} \bar{\theta}^2 \right\} (\theta^2)^{\frac{2 \kappa \bar{\theta}^2}{\gamma^2} - 1},\end{aligned}$$

which is a gamma density.

Solutions for Exercises of Chapter 3

3.1 The growth rate g_t^δ of a portfolio is defined as the drift of the SDE of the logarithm of the portfolio S^δ . By application of the Itô formula to $\ln(S_t^\delta)$ one obtains the SDE

$$\begin{aligned} d\ln(S_t^\delta) &= \frac{1}{S_t^\delta} dS_t^\delta - \frac{1}{2(S_t^\delta)^2} d[S_t^\delta] \\ &= \left(r_t + \sum_{k=1}^d \sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} \theta_t^k - \frac{1}{2} \sum_{k=1}^d \left(\sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} \theta_t^k \right)^2 \right) dt \\ &\quad + \sum_{k=1}^d \sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} dW_t^k. \end{aligned}$$

The drift of this SDE, which is the growth rate of S^δ , is then

$$g_t^\delta = r_t + \sum_{k=1}^d \left(\sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} \theta_t^k - \frac{1}{2} \left(\sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} \right)^2 \right).$$

3.2 We apply for the benchmarked value $\hat{S}_t^\delta = \frac{S_t^\delta}{S_{t^*}^{\delta^*}}$ the integration by parts formula (1.5.10) and obtain the SDE

$$\begin{aligned} d\hat{S}_t^\delta &= S_t^\delta d\left(\frac{1}{S_{t^*}^{\delta^*}}\right) + \frac{1}{S_{t^*}^{\delta^*}} dS_t^\delta + d\left[\frac{1}{S_{t^*}^{\delta^*}}, S^\delta\right]_t \\ &= \frac{S_t^\delta}{S_{t^*}^{\delta^*}} \left(-r_t dt - \sum_{k=1}^d \theta_t^k dW_t^k \right) + \frac{S_t^\delta}{S_{t^*}^{\delta^*}} \left(r_t dt + \sum_{k=1}^d \sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} (\theta_t^k dt + dW_t^k) \right) \\ &\quad + \frac{S_t^\delta}{S_{t^*}^{\delta^*}} \sum_{k=1}^d \left(-\theta_t^k \sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} \right) \\ &= \hat{S}_t^\delta \sum_{k=1}^d \left(\sum_{j=1}^d \pi_{\delta,t}^j b_t^{j,k} - \theta_t^k \right) dW_t^k. \end{aligned}$$

This SDE is driftless. Therefore, by (1.4.12) a square integrable \hat{S}^δ ensures that \hat{S}^δ is an $(\underline{\mathcal{A}}, P)$ -local martingale. However, this is not sufficient to guarantee that \hat{S}^δ is, in general, an $(\underline{\mathcal{A}}, P)$ -martingale, see Lemma 1.3.2 (iii). Since \hat{S}^δ is a nonnegative local martingale it is by Lemma 1.3.2 (i) an $(\underline{\mathcal{A}}, P)$ -supermartingale.

3.3 The SDE for the discounted GOP is of the form

$$d\bar{S}_t^{\delta_*} = \alpha_t^{\delta_*} dt + \sqrt{\bar{S}_t^{\delta_*} \alpha_t^{\delta_*}} dW_t.$$

By the Itô formula we obtain for $\ln(\bar{S}_t^{\delta_*})$ the SDE

$$d \ln \left(\bar{S}_t^{\delta_*} \right) = \frac{1}{2} \frac{\alpha_t^{\delta_*}}{\bar{S}_t^{\delta_*}} dt + \sqrt{\frac{\alpha_t^{\delta_*}}{\bar{S}_t^{\delta_*}}} dW_t,$$

which shows that the volatility of the discounted GOP equals $\sqrt{\frac{\alpha_t^{\delta_*}}{\bar{S}_t^{\delta_*}}}$.

3.4 By the Itô formula we obtain for $\sqrt{\bar{S}_t^{\delta_*}}$ the SDE

$$\begin{aligned} d\sqrt{\bar{S}_t^{\delta_*}} &= \left(\frac{\alpha_t^{\delta_*}}{2\sqrt{\bar{S}_t^{\delta_*}}} - \frac{1}{2} \frac{1}{4} \frac{1}{\left(\bar{S}_t^{\delta_*}\right)^{\frac{3}{2}}} \bar{S}_t^{\delta_*} \alpha_t^{\delta_*} \right) dt + \frac{\sqrt{\bar{S}_t^{\delta_*} \alpha_t^{\delta_*}}}{2\sqrt{\bar{S}_t^{\delta_*}}} dW_t \\ &= \frac{3}{8} \frac{\alpha_t^{\delta_*}}{\sqrt{\bar{S}_t^{\delta_*}}} dt + \frac{1}{2} \sqrt{\alpha_t^{\delta_*}} dW_t, \end{aligned}$$

3.5 The differential equation for $\alpha_t^{\delta_*}$ is of the form

$$d\alpha_t^{\delta_*} = \eta_t \alpha_t^{\delta_*} dt.$$

Together with the SDE of the discounted GOP and the Itô formula it follows

$$\begin{aligned} dY_t &= Y_t \left(\frac{\alpha_t}{\bar{S}_t^{\delta_*}} - \eta_t \right) dt + Y_t \sqrt{\frac{\alpha_t}{\bar{S}_t^{\delta_*}}} dW_t \\ &= (1 - \eta_t Y_t) dt + \sqrt{Y_t} dW_t, \end{aligned}$$

which is a square root process.

3.6 The squared volatility of the discounted GOP equals $|\theta_t|^2 = \frac{1}{Y_t}$, and, thus the inverse of the normalized GOP. This means, we obtain by the Itô formula the SDE

$$\begin{aligned} d|\theta_t|^2 &= d\left(\frac{1}{Y_t}\right) = \left(-\left(\frac{1}{Y_t}\right)^2 (1 - \eta_t Y_t) + \frac{Y_t}{(Y_t)^3}\right) dt - \left(\frac{1}{Y_t}\right)^2 \sqrt{Y_t} dW_t \\ &= \eta_t \frac{1}{Y_t} dt - \left(\frac{1}{Y_t}\right)^{\frac{3}{2}} dW_t = \eta_t |\theta_t|^2 dt - (|\theta_t|^2)^{\frac{3}{2}} dW_t. \end{aligned}$$

Solutions for Exercises of Chapter 4

4.1 By application of the Wagner-Platen expansion one directly obtains

$$X_{t_0+h} - X_{t_0} = a X_{t_0} dt + b X_{t_0} (W_{t_0+h} - W_{t_0}) + R.$$

4.2 Using the Wagner-Platen expansion it follows that

$$\begin{aligned} X_{t_0+h} - X_{t_0} &= X_{t_0} \left\{ a h + b (W_{t_0+h} - W_{t_0}) \right. \\ &\quad + a^2 \frac{h^2}{2} + a b \left(\int_{t_0}^{t_0+h} \int_{t_0}^s dz dW_s + \int_{t_0}^{t_0+h} \int_{t_0}^s dW_s dz \right) \\ &\quad \left. + \frac{b^2}{2} ((W_{t_0+h} - W_{t_0})^2 - h) \right\} + R. \end{aligned}$$

By application of the Itô formula to $(W_t - W_{t_0})(t - t_0)$ we can write

$$\begin{aligned} &= X_{t_0} \left\{ a h + b (W_{t_0+h} - W_{t_0}) + a^2 \frac{h^2}{2} + a b (W_{t_0+h} - W_{t_0}) h \right. \\ &\quad \left. + \frac{b^2}{2} ((W_{t_0+h} - W_{t_0})^2 - h) \right\} + R. \end{aligned}$$

4.3 According to the definitions of the multi-index notation we get for

$$\begin{aligned} \alpha &= (0, 0, 0, 0) : -\alpha = (0, 0, 0), \alpha- = (0, 0, 0), \ell(\alpha) = 4, n(\alpha) = 4; \\ \alpha &= (1, 0, 2) : -\alpha = (0, 2), \alpha- = (1, 0), \ell(\alpha) = 3, n(\alpha) = 1; \\ \alpha &= (0, 2, 0, 1) : -\alpha = (2, 0, 1), \alpha- = (0, 2, 0), \ell(\alpha) = 4, n(\alpha) = 2. \end{aligned}$$

4.4 By using the introduced notation one obtains

$$I_{(0,0),t} = \int_0^t \int_0^s dz ds = \int_0^t s ds = \frac{t^2}{2},$$

$$I_{(1,0),t} = \int_0^t \int_0^s dW_z ds$$

$$I_{(1,1),t} = \int_0^t \int_0^s dW_z dW_s = \int_0^t W_s dW_s$$

and it follows that $I_{(1,1),t} = \frac{1}{2} ((W_t)^2 - t)$. Finally, we have

$$I_{(1,2),t} = \int_0^t \int_0^s dW_z^1 dW_s^2 = \int_0^t W_s^1 dW_s^2.$$

4.5 It follows from Lemma 4.2.3 that

$$I_{(1)} I_{(0)} = I_{(1,0)} + I_{(0,1)}$$

and thus

$$I_{(0,1)} = I_{(1)} I_{(0)} - I_{(1,0)}.$$

4.6

$$E(I_{(1,0),\Delta}) = E \left(\int_0^\Delta \int_0^s dW_z ds \right) = \int_0^\Delta E(W_s) ds = 0,$$

$$\begin{aligned} E((I_{(0,1),\Delta})^2) &= E \left(\left(\int_0^\Delta \int_0^s dz dW_s \right)^2 \right) = \int_0^\Delta \left(\int_0^s dz \right)^2 ds \\ &= \int_0^\Delta s^2 ds = \frac{\Delta^3}{3}, \end{aligned}$$

$$E(I_{(1),\Delta} I_{(0,1),\Delta}) = E \left(\int_0^\Delta dW_s \int_0^\Delta s dW_s \right) = \int_0^\Delta s ds = \frac{\Delta^2}{2},$$

$$\begin{aligned} E((I_{(1,0),\Delta})^2) &= E \left((I_{(0),\Delta} I_{(1),\Delta} - I_{(0,1),\Delta})^2 \right) \\ &= \Delta^2 E((I_{(1),\Delta})^2) - 2\Delta E(I_{(1),\Delta} I_{(0,1),\Delta}) + E((I_{(0,1),\Delta})^2) \\ &= \Delta^3 - \Delta^3 + \frac{\Delta^3}{3} = \frac{\Delta^3}{3}. \end{aligned}$$

4.7 It follows from the given notation that

$$f_{(1,0)} = b a' \quad \text{and} \quad f_{(1,1,1)} = b((b')^2 + b b'').$$

4.8 The following sets are not hierarchical sets: \emptyset , $\{(1)\}$, $\{v, (0), (0, 1)\}$.

4.9 The remainder sets are:

$$\mathcal{B}(\{v, (1)\}) = \{(0), (0, 1), (1, 1)\}$$

$$\mathcal{B}(\{v, (0), (1), (0, 1)\}) = \{(1, 1), (1, 0), (0, 0), (1, 0, 1), (0, 0, 1)\}.$$

4.10 The expansion has the form

$$\begin{aligned} X_t &= X_0 + a(0, X_0) \int_0^t ds + b(0, X_0) \int_0^t dW_s \\ &\quad + b(0, X_0) \frac{\partial}{\partial x} b(0, X_0) \int_0^t \int_0^{s_2} dW_{s_1} dW_{s_2} \\ &= X_0 + a(0, X_0) t + b(0, X_0) W_t \\ &\quad + b(0, X_0) \frac{\partial}{\partial x} b(0, X_0) \frac{1}{2} ((W_t)^2 - t). \end{aligned}$$

4.11 It follows that

$$\ell((1, 0, -1)) = 3$$

$$n((1, 0, -1)) = 1$$

$$s((1, 0, -1)) = 1.$$

4.12 One has for $-\alpha$ with $\alpha = (1, 0, -1)$ simply to delete the first component, that is,

$$-\alpha = (0, -1).$$

Solutions for Exercises of Chapter 5

5.1 The Euler scheme is given by

$$Y_{n+1} = Y_n + (\mu Y_n + \eta) \Delta + \gamma Y_n \Delta W,$$

where $\Delta W = W_{\tau_{n+1}} - W_{\tau_n}$. The Milstein scheme has the form

$$Y_{n+1} = Y_n + (\mu Y_n + \eta) \Delta + \gamma Y_n \Delta W + \frac{\gamma^2}{2} Y_n ((\Delta W)^2 - \Delta).$$

5.2 Due to the additive noise of the Vasicek model the Euler and Milstein schemes are identical and of the form

$$Y_{n+1} = Y_n + \gamma(\bar{r} - Y_n) \Delta + \beta \Delta W,$$

where $\Delta W = W_{\tau_{n+1}} - W_{\tau_n}$.

5.3 The strong order 1.5 Taylor scheme has for the given SDE the form

$$\begin{aligned}
Y_{n+1} = & Y_n + \mu(Y_n + \eta) \Delta + \gamma Y_n \Delta W + \frac{Y_n}{2} \gamma^2 ((\Delta W)^2 - \Delta) \\
& + \mu \gamma Y_n \Delta Z + \frac{1}{2} \mu (\mu Y_n + \eta) \Delta^2 \\
& + (\mu Y_n + \eta) \gamma (\Delta W \Delta - \Delta Z) \\
& + \frac{1}{2} \gamma^3 Y_n \left(\frac{1}{3} (\Delta W)^2 - \Delta \right) \Delta W,
\end{aligned}$$

where $\Delta W = W_{\tau_{n+1}} - W_{\tau_n}$ and

$$\Delta Z = \int_{\tau_n}^{\tau_{n+1}} (W_s - W_{\tau_n}) ds.$$

5.4 The explicit strong order 1.0 scheme has the form

$$\begin{aligned}
Y_{n+1} = & Y_n + Y_n \mu \Delta + Y_n \sigma \Delta W \\
& + \frac{\sigma Y_n}{2 \sqrt{\Delta}} \left(\mu \Delta + \sigma \sqrt{\Delta} \right) ((\Delta W)^2 - \Delta),
\end{aligned}$$

where $\Delta W = W_{\tau_{n+1}} - W_{\tau_n}$.

Solutions for Exercises of Chapter 6

6.1 It follows that the diffusion coefficients for the first Wiener process W^1 are

$$b^{1,1} = 1 \quad \text{and} \quad b^{2,1} = 0$$

and that of the second Wiener process are

$$b^{1,2} = 0 \quad \text{and} \quad b^{2,2} = X_t^1.$$

Therefore we obtain that

$$L^1 b^{2,2} = b^{1,1} \frac{\partial}{\partial x^1} b^{2,2} + b^{2,1} \frac{\partial}{\partial x^2} b^{2,2} = 1$$

and

$$L^2 b^{2,1} = b^{1,2} \frac{\partial}{\partial x^1} b^{2,1} + b^{2,2} \frac{\partial}{\partial x^2} b^{2,1} = 0.$$

Since the above values are not equal, the SDE does not satisfy a diffusion commutativity condition.

6.2 The Milstein scheme applied to the given SDE is of the form

$$\begin{aligned}
Y_{n+1}^1 = & Y_n^1 + \Delta W^1 \\
Y_{n+1}^2 = & Y_n^2 + Y_n^1 \Delta W^2 + I_{(1,2)}
\end{aligned}$$

with

$$I_{(1,2)} = \int_{\tau_n}^{\tau_{n+1}} dW_{s_1}^1 dW_{s_2}^2,$$

$$\Delta W^1 = W_{\tau_{n+1}}^1 - W_{\tau_n}^1$$

$$\Delta W^2 = W_{\tau_{n+1}}^2 - W_{\tau_n}^2.$$

6.3 According to the jump commutativity condition (6.3.2) we need to consider the case of $b(t, x) = 2x$ and $c(t, x) = 2x$. Then it follows

$$\frac{\partial c(t, x)}{\partial x} = 2$$

and we have the relation

$$b(t, x) \frac{\partial c(t, x)}{\partial x} = 2(x + 2x) - 2x = b(t, x + c(t, x)) - b(t, x).$$

This shows that we have jump commutativity.

Solutions for Exercises of Chapter 7

7.1 The drift implicit Euler scheme has for the BS model the form

$$Y_{n+1} = Y_n + \mu Y_{n+1} \Delta + \sigma Y_n \Delta W$$

and we get

$$Y_{n+1} = \frac{Y_n (1 + \sigma \Delta W)}{1 - \mu \Delta}.$$

7.2 The drift implicit strong order 1.0 Runge-Kutta method for the Vasicek model yields the equation

$$Y_{n+1} = Y_n + \gamma (\bar{r} - Y_{n+1}) \Delta + \beta \Delta W$$

and thus the algorithm

$$Y_{n+1} = \frac{Y_n + \gamma \bar{r} \Delta + \beta \Delta W}{1 + \gamma \Delta},$$

where $\Delta W = W_{\tau_{n+1}} - W_{\tau_n}$.

7.3 A balanced implicit method for the BS model is obtained by setting

$$Y_{n+1} = Y_n + \mu Y_n \Delta + \sigma Y_n \Delta W + \sigma (Y_n - Y_{n+1}) |\Delta W|.$$

Thus one has

$$Y_{n+1} (1 + \sigma |\Delta W|) = Y_n (1 + \mu \Delta + \sigma (\Delta W + |\Delta W|))$$

and obtains the algorithm

$$Y_{n+1} = \frac{Y_n (1 + \mu \Delta + \sigma (\Delta W + |\Delta W|))}{1 + \sigma |\Delta W|}.$$

Obviously, the denominator and nominator in the algorithms remain always nonnegative and thus also the approximation.

7.4 We obtain from (7.1.9) the scheme

$$\begin{aligned} Y_{n+1} &= Y_n + a Y_n \Delta_n + b Y_n \Delta W_n + c Y_n \Delta p_n \\ &+ \frac{b(\bar{Y}_n - Y_n)}{2\sqrt{\Delta_n}} \{(\Delta W_n)^2 - \Delta_n\} + b c Y_n \Delta p_n \Delta W_n \\ &+ \frac{c^2 Y_n}{2} \{(\Delta p_n)^2 - \Delta p_n\} \end{aligned}$$

with supporting value

$$\bar{Y}_n = Y_n + b Y_n \sqrt{\Delta_n}.$$

7.5 By application of (7.2.6) one obtains

$$\begin{aligned} Y_{n+1} &= \frac{Y_n}{1 - a \Delta_n} \left\{ 1 + b \Delta W_n + c \Delta p_n + \frac{1}{2} b^2 \{(\Delta W_n)^2 - \Delta_n\} \right. \\ &\quad \left. + b c \Delta p_n \Delta W_n + \frac{1}{2} c^2 \{(\Delta p_n)^2 - \Delta p_n\} \right\}. \end{aligned}$$

Solutions for Exercises of Chapter 8

8.1 By (8.9.17) we have the strong order 1.0 Taylor approximation

$$Y_{n+1} = Y_n + (a + b Y_n) \Delta N_n + \frac{b}{2} (a + b Y_n) (\Delta N_n) (\Delta N_n - 1).$$

8.2 We obtain from (8.9.20) the strong order 2.0 Taylor approximation

$$\begin{aligned} Y_{n+1} &= Y_n + (a + b Y_n) \Delta N_n + b(a + b Y_n) \binom{\Delta N_n}{2} \\ &+ b^2(a + b Y_n) \binom{\Delta N_n}{3} \\ &+ b^3(a + b Y_n) \binom{\Delta N_n}{4}. \end{aligned}$$

Solutions for Exercises of Chapter 9

9.1 Let us introduce the centralized short rate $X_t = r_t - \bar{r}$. Then its SDE is of the form

$$dX_t = -\gamma X_t dt + \beta dW_t.$$

It then follows from (9.1.8) that we obtain for γ over the period $[0, T]$ the maximum likelihood estimator

$$\begin{aligned}\hat{\gamma}(T) &= \frac{-\int_0^T X_t dX_t}{\int_0^T (X_t)^2 dt} = \frac{-\int_0^T (r_t - \bar{r}) dr_t}{\int_0^T (r_t - \bar{r})^2 dt} \\ &= \frac{-\int_0^T r_t dr_t + \bar{r} (r_T - r_0)}{\int_0^T r_t^2 dt - 2 \bar{r} \int_0^T r_t dt + (\bar{r})^2 T}.\end{aligned}$$

9.2 One can calculate the quadratic variation of $\ln(X) = \{\ln(X_t), t \in [0, T]\}$, where

$$d\ln(X_t) = \left(a - \frac{1}{2} \sigma^2 \right) dt + \sigma W_t$$

for $t \in [0, T]$. Then the quadratic variation is of the form

$$[\ln(X.)]_t = \sigma^2 t.$$

Consequently, to obtain the volatility one can use the expression

$$\sigma = \sqrt{\frac{[\ln(X.)]_T}{T}}.$$

Solutions for Exercises of Chapter 10

10.1 For continuous Gaussian observed and hidden quantities the Kalman-Bucy filter with its generalizations is applicable.

10.2 For continuous time Markov chains with finite number of states, describing hidden variables, the Wonham filter is applicable when the signal process has a drift, determined by the state of the Markov chain, and additive noise is driven by a Wiener process.

10.3 In hidden Markov chain filtering the unnormalized conditional probabilities of the Wonham filter satisfy the Zakai SDE which has no explicit solution. It can be solved numerically by strong discrete-time numerical schemes for SDEs using the discrete-time observations.

10.4 The balanced implicit method, as an implicit scheme, is adequate for the solution of the Zakai equation and potentially can overcome numerical stability problems.

10.5 Consider the logarithm $W_t = \ln(S_t)$, which satisfies by the Itô formula the SDE

$$dW_t = \eta dt + dW_t^*$$

for $t \in [0, T]$ with $W_0 = \ln(S_0)$. This is a Kalman-Bucy filter problem which has an explicit solution of the form

$$\hat{\eta}_t = \frac{\sigma^2}{1 + \sigma^2 t} W_t$$

for $t \in [0, T]$, see (10.1.19).

10.6 An Expectation Maximization algorithm can be used to iteratively estimate in a filter problem parameters that characterize the dynamics of the hidden and observed quantities.

Solutions for Exercises of Chapter 11

11.1 Due to symmetry of the probabilities of the two-point distributed random variable $\Delta\hat{W}$ one obtains

$$E(\Delta\hat{W}) = E((\Delta\hat{W})^3) = 0.$$

Furthermore, it follows

$$E((\Delta\hat{W})^2) = \frac{1}{2}\Delta + \frac{1}{2}\Delta = \Delta$$

and thus

$$E((\Delta\hat{W})^2) - \Delta = 0.$$

This proves that

$$|E(\Delta\hat{W})| + |E((\Delta\hat{W})^3)| + |E((\Delta\hat{W})^2 - \Delta)| = 0 \leq K\Delta^2.$$

11.2 By symmetry one obtains

$$E(\Delta\tilde{W}) = E((\Delta\tilde{W})^3) = E((\Delta\tilde{W})^5) = 0.$$

Furthermore, calculation of the expected values provides

$$E((\Delta\tilde{W})^2) = \frac{1}{6}3\Delta + \frac{1}{6}3\Delta = \Delta,$$

$$E \left((\Delta \tilde{W})^4 \right) = \frac{1}{6} 9 \Delta^2 + \frac{1}{6} 9 \Delta^2 = 3 \Delta^2,$$

which yields the required estimate

$$\begin{aligned} & \left| E(\Delta \tilde{W}) \right| + \left| E((\Delta \tilde{W})^3) \right| + \left| E((\Delta \tilde{W})^5) \right| \\ & + \left| E((\Delta \tilde{W})^2) - \Delta \right| + \left| E((\Delta \tilde{W})^4) - 3 \Delta^2 \right| = 0 \leq K \Delta^3. \end{aligned}$$

11.3 The simplified Euler scheme for the BS model is of the form

$$Y_{n+1} = Y_n \left(1 + \mu \Delta + \sigma \Delta \hat{W}_n \right),$$

where $\Delta = \tau_{n+1} - \tau_n$ and

$$P(\Delta \hat{W}_n = \pm \sqrt{\Delta}) = \frac{1}{2}$$

for $n \in \{0, 1, \dots\}$ with $Y_0 = X_0$. This scheme attains the weak order $\beta = 1.0$.

11.4 The simplified weak order 2.0 Taylor scheme has for the BS model the form

$$Y_{n+1} = Y_n \left(1 + \mu \Delta + \sigma \Delta \tilde{W}_n + \frac{\sigma^2}{2} \left((\Delta \tilde{W}_n)^2 - \Delta \right) + \mu \sigma \Delta \tilde{W}_n \Delta + \frac{\mu^2}{2} \Delta^2 \right)$$

with $\Delta = \tau_{n+1} - \tau_n$ and

$$P(\Delta \tilde{W}_n = \pm \sqrt{3 \Delta}) = \frac{1}{6} \quad \text{and} \quad P(\Delta \tilde{W}_n = 0) = \frac{2}{3}$$

for $n \in \{0, 1, \dots\}$ with $Y_0 = X_0$.

11.5 For the BS model the drift implicit Euler scheme has the form

$$Y_{n+1} = Y_n + \mu Y_{n+1} \Delta + \sigma Y_n \Delta W_n.$$

Therefore, we obtain the algorithm

$$Y_{n+1} = Y_n \frac{1 + \sigma \Delta W_n}{1 - \mu \Delta}$$

with $\Delta W_n \sim \mathcal{N}(0, \Delta)$ for $n \in \{0, 1, \dots\}$ and $Y_0 = X_0$.

11.6 For the BS model one obtains the fully implicit Euler scheme

$$Y_{n+1} = Y_n + Y_{n+1} (\mu - \sigma^2) \Delta + Y_{n+1} \sigma \Delta \hat{W}_n$$

yielding the algorithm

$$Y_{n+1} = Y_n \frac{1}{1 + (\sigma^2 - \mu) \Delta - \sigma \Delta \hat{W}_n},$$

where $\Delta = \tau_{n+1} - \tau_n$ and

$$P(\Delta \hat{W}_n = \pm \sqrt{\Delta}) = \frac{1}{2}$$

for $n \in \{0, 1, \dots\}$ and $Y_0 = X_0$.

Solutions for Exercises of Chapter 12

12.1 The weak order 1.0 Taylor scheme is simply the Euler scheme which has here the form

$$Y_{n+1} = Y_n + a Y_n \Delta + \sigma Y_n \Delta W_n + c Y_n \Delta N_n$$

where $\Delta = t_{n+1} - t_n$, $\Delta W_n = W_{t_{n+1}} - W_{t_n}$ and $\Delta N_n = N_{t_{n+1}} - N_{t_n}$.

12.2 The SDE is mark-independent. Furthermore, it satisfies the first jump commutativity condition (12.2.3) and the diffusion commutativity condition (12.2.8). Its weak order 2.0 Taylor scheme is of the form

$$\begin{aligned} Y_{n+1} &= Y_n + a Y_n \Delta + \sigma Y_n \Delta W_n + c Y_n \Delta N_n \\ &\quad + \frac{1}{2} \sigma^2 Y_n (\Delta W_n^2 - \Delta) + \sigma c Y_n \Delta N_n \Delta W_n \\ &\quad + \frac{1}{2} c^2 Y_n (\Delta N_n - 1) \Delta N_n + a \sigma c Y_n \Delta \Delta W_n \\ &\quad + \frac{1}{2} a^2 Y_n \Delta^2 + a c Y_n \Delta N_n \Delta, \end{aligned}$$

where $\Delta = t_{n+1} - t_n$, $\Delta W_n = W_{t_{n+1}} - W_{t_n}$ and $\Delta N_n = N_{t_{n+1}} - N_{t_n}$.

Solutions for Exercises of Chapter 13

13.1 First, we generate the jump times of the driving Poisson process N with intensity λ . The waiting times between jumps are independent and exponentially distributed with mean $\frac{1}{\lambda}$. By superimposing these jump times on an equidistant time discretization with step size $\Delta \in (0, 1)$ one obtains the jump-adapted sequence discretization times $0 = t_0 < t_1 < \dots$. A jump-adapted simplified weak order 2.0 scheme is obtained by the algorithm

$$\begin{aligned} Y_{t_{n+1}-} &= Y_{t_n} + a Y_{t_n} (t_{n+1} - t_n) + \sigma Y_{t_n} \Delta \hat{W}_{t_n} + \frac{\sigma^2}{2} \left((\Delta \hat{W}_{t_n})^2 - (t_{n+1} - t_n) \right) \\ &\quad + \frac{1}{2} a^2 Y_{t_n} (t_{n+1} - t_n)^2 + a \sigma Y_{t_n} \Delta \hat{W}_{t_n} (t_{n+1} - t_n) \end{aligned}$$

with

$$Y_{t_{n+1}} = Y_{t_{n+1}-} + c Y_{t_{n+1}} - \Delta N_n$$

where

$$P(\Delta \hat{W}_{t_n} = \pm \sqrt{3(t_{n+1} - t_n)}) = \frac{1}{6}$$

$$P(\Delta \hat{W}_{t_n} = 0) = \frac{2}{3}$$

and

$$\Delta N_n = \begin{cases} 1 & \text{if } t_{n+1} \text{ is a jump time of } N \\ 0 & \text{otherwise.} \end{cases}$$

13.2 A jump-adapted fully-implicit weak order predictor-corrector scheme can use the same jump-adapted time discretization as mentioned in the solution of Exercise 13.1. The algorithm for approximating the diffusion part between discretization times can take the following form

$$Y_{t_{n+1}-} = Y_{t_n} + (a - \sigma^2) \bar{Y}_{t_{n+1}-}(t_{n+1} - t_n) + \sigma \bar{Y}_{t_{n+1}-} \Delta \hat{W}_{t_n}$$

with

$$\bar{Y}_{t_{n+1}-} = Y_{t_n} + a Y_{t_n} (t_{n+1} - t_n) + \sigma Y_{t_n} \Delta \hat{W}_{t_n}$$

where

$$Y_{t_{n+1}} = Y_{t_{n+1}-} + c Y_{t_{n+1}-} \Delta N_n.$$

Here $P(\Delta \hat{W}_n = \pm \sqrt{t_{n+1} - t_n}) = \frac{1}{2}$ and $\Delta N_n = 1$ if t_{n+1} was a jump time of N .

Solutions for Exercises of Chapter 14

14.1 For $p = 2$ we obtain from the transfer function of the Euler scheme its second moment

$$\begin{aligned} E((G_{n+1}(\lambda\Delta, \alpha))^2) &= E\left(\left(1 + \left(1 - \frac{3}{2}\alpha\right)\lambda\Delta + \sqrt{-\alpha\lambda}\Delta W_n\right)^2\right) \\ &= 1 + 2\left(1 - \frac{3}{2}\alpha\right)\lambda\Delta + \left(1 - \frac{3}{2}\alpha\right)^2\lambda^2\Delta^2 - \alpha\lambda\Delta \\ &= 1 + 2(1 - 2\alpha)\lambda\Delta + \left(1 - \frac{3}{2}\alpha\right)^2\lambda^2\Delta^2. \end{aligned}$$

Therefore, we have by (14.1.8) for $\lambda\Delta < 0$ that

$$E((G_{n+1}(\lambda\Delta, \alpha))^2) < 1$$

if and only if

$$2(1 - 2\alpha)\lambda\Delta + \left(1 - \frac{3}{2}\alpha\right)^2\lambda^2\Delta^2 < 0,$$

that is,

$$2(1 - 2\alpha) > -\left(1 - \frac{3}{2}\alpha\right)^2\lambda\Delta.$$

This means, the boundary $B(\alpha)$ for values $\lambda\Delta$ of the stability region as a function of $\alpha \in [0, 1)$ is given by the function

$$B(\alpha) = -\left(\frac{2(1 - 2\alpha)}{(1 - \frac{3}{2}\alpha)^2}\right)^+.$$

14.2 It follows from (14.4.3) with $\alpha = \eta = \theta = 1$ that

$$G_{n+1}(\lambda\Delta, 1) = \frac{1}{1 + \frac{|\lambda|\Delta}{2} - \sqrt{|\lambda|}\Delta\hat{W}_n}.$$

Consequently, we obtain

$$E(|G_{n+1}(\lambda\Delta, 1)|) = \frac{1 + \frac{|\lambda|\Delta}{2}}{(1 + \frac{|\lambda|\Delta}{2})^2 - |\lambda|\Delta},$$

which is less than one for $\Delta > \frac{2}{|\lambda|}$.

Therefore, the fully implicit simplified Euler scheme is for $\alpha = 1$ asymptotically p -stable with $p = 1$ as long as the step size is not smaller than $\frac{2}{|\lambda|}$.

Solutions for Exercises of Chapter 15

15.1 The pricing function $u : [0, T] \times (0, \infty) \rightarrow (0, \infty)$ is according to risk neutral pricing given by the conditional expectation

$$u(t, S_t) = \tilde{E}(H_T \mid \mathcal{A}_t) = \tilde{E}((S_T)^q \mid \mathcal{A}_t)$$

for $t \in [0, T]$, where \tilde{E} denotes expectation under the risk neutral martingale measure \tilde{P} . For the given asset price we have

$$(S_T)^q = (S_t)^q \exp \left\{ q \left(-\frac{1}{2}\sigma^2(T-t) + \sigma(W_T - W_t) \right) \right\}$$

and therefore

$$\begin{aligned} u(t, S_t) &= (S_t)^q \exp \left\{ -\frac{q}{2}\sigma^2(T-t) + \frac{q^2\sigma^2}{2}(T-t) \right\} \\ &\quad \times \tilde{E} \left(\exp \left\{ -\frac{q^2\sigma^2}{2}(T-t) + q\sigma(W_T - W_t) \right\} \mid \mathcal{A}_t \right). \end{aligned}$$

Since the term under the conditional expectation forms with respect to T an $(\underline{A}, \tilde{P})$ -martingale we get the pricing function in the form

$$u(t, S) = (S)^q \exp \left\{ \frac{q}{2} (q-1) \sigma^2 (T-t) \right\}$$

for $t \in [0, T]$ and $S \in (0, \infty)$. The corresponding martingale representation is obtained by application of the Itô formula to $u(t, S_t)$, which yields

$$\begin{aligned} (S_T)^q &= u(T, S_T) \\ &= u(t, S_t) + \int_t^T \left(\frac{\partial u(z, S_z)}{\partial t} + \frac{1}{2} \frac{\partial^2 u(z, S_z)}{\partial S^2} \sigma^2 S_z^2 \right) dz \\ &\quad + \int_t^T \frac{\partial u(z, S_z)}{\partial S} \sigma S_z dW_z \\ &= u(t, S_t) + \int_t^T \left[(S_z)^q \exp \left\{ \frac{q}{2} (q-1) \sigma^2 (T-z) \right\} \left(\frac{q}{2} (1-q) \sigma^2 \right) \right. \\ &\quad \left. + \frac{1}{2} q (q-1) S_z^{(q-2)} \exp \left\{ \frac{q}{2} (q-1) \sigma^2 (T-z) \right\} \sigma^2 S_z^2 \right] dz \\ &\quad + \int_t^T \frac{\partial u(z, S_z)}{\partial S} \sigma S_z dW_z \\ &= u(t, S_t) + \int_t^T \frac{\partial u(z, S_z)}{\partial S} \sigma S_z dW_z. \end{aligned}$$

15.2 By using Theorem 15.4.1 the hedge ratio is obtained as the partial derivative

$$\begin{aligned} \frac{\partial u(t, S_t)}{\partial S} &= \frac{\partial}{\partial S} \left((S_t)^q \exp \left\{ \frac{q(q-1)}{2} \sigma^2 (T-t) \right\} \right) \\ &= q (S_t)^{q-1} \exp \left\{ \frac{q(q-1)}{2} \sigma^2 (T-t) \right\}. \end{aligned}$$

15.3 By the Black-Scholes formula we have

$$c_{T,K}(t, S_t) = S_t N(d_1(t)) - K N(d_2(t))$$

with

$$d_1(t) = \frac{\ln \left(\frac{S_t}{K} \right) + \frac{\sigma^2}{2} (T-t)}{\sigma \sqrt{T-t}}$$

and

$$d_2(t) = d_1(t) - \sigma \sqrt{T-t}.$$

Then, by application of the Itô formula and the BS PDE

$$\left(\frac{\partial}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2}{\partial S^2} \right) c_{T,K}(t, S) = 0$$

for all $(t, S) \in (0, T) \times (0, \infty)$, we obtain the martingale representation

$$\begin{aligned} c_{T,K}(T, S_T) &= (S_T - K)^+ \\ &= c_{T,K}(t, S_t) + \int_t^T \frac{\partial c_{T,K}(z, S_z)}{\partial S} dS_z. \end{aligned}$$

Therefore, the hedge ratio equals the well-known Black-Scholes delta

$$\frac{\partial c_{T,K}(t, S_t)}{\partial S} = N(d_1(t))$$

for $t \in (0, T)$.

Solutions for Exercises of Chapter 16

16.1 We have the expectation

$$E \left(1 + Z + \frac{1}{2} (Z)^2 \right) = \frac{3}{2}$$

and it follows that

$$E(V_N^+) = E(V_N^-) = E(\hat{V}_N) = \frac{3}{2}.$$

Therefore, \hat{V}_N is unbiased.

We calculate the variance of V_N^+ as

$$\begin{aligned} \text{Var}(V_N^+) &= E \left(\left(\frac{1}{N} \sum_{k=1}^N \left(1 + Z(\omega_k) + \frac{1}{2} (Z(\omega_k))^2 \right) - \frac{3}{2} \right)^2 \right) \\ &= E \left(\left(\frac{1}{N} \sum_{k=1}^N \left(Z(\omega_k) + \frac{1}{2} ((Z(\omega_k))^2 - 1) \right) \right)^2 \right) \\ &= \frac{1}{N} E \left(\left(Z + \frac{1}{2} ((Z)^2 - 1) \right)^2 \right) \\ &= \frac{1}{N} \left(E((Z)^2) + \frac{1}{4} (E((Z)^4) - 2 E((Z)^2) + 1) \right) \\ &= \frac{1}{N} \left(1 + \frac{1}{4} (3 - 2 + 1) \right) = \frac{3}{2N}. \end{aligned}$$

Alternatively, we obtain

$$\begin{aligned}
 \text{Var}(\hat{V}_N) &= E\left(\left(\frac{1}{2}\left(\frac{1}{N} \sum_{k=1}^N \left(1 + Z(\omega_k) + \frac{1}{2}(Z(\omega_k))^2\right)\right.\right. \\
 &\quad \left.\left.+ \frac{1}{N} \sum_{k=1}^N \left(1 - Z(\omega_k) + \frac{1}{2}(Z(\omega_k))^2\right)\right) - \frac{3}{2}\right)^2\Bigg) \\
 &= E\left(\left(\frac{1}{N} \sum_{k=1}^N \frac{1}{2}((Z(\omega_k))^2 - 1)\right)^2\right) \\
 &= \frac{1}{N} \frac{1}{4} E\left(\left((Z)^2 - 1\right)^2\right) = \frac{1}{4N} (E((Z)^4) - 2E((Z)^2) + 1) \\
 &= \frac{1}{4N} (3 - 2 + 1) = \frac{1}{2N}.
 \end{aligned}$$

This shows that the antithetic method provides a Monte Carlo estimator or yielding only a third of the variance of a raw Monte Carlo estimator.

16.2 For $E(V_N^*) = \gamma = 1$ it is $E(\tilde{V}_N) = \frac{2}{3}$, and \tilde{V}_N is an unbiased estimator. The variance of \tilde{V}_N is then obtained as

$$\begin{aligned}
 \text{Var}(\tilde{V}_N) &= E\left(\left(\frac{1}{N} \sum_{k=1}^N \left(1 + Z(\omega_k) + \frac{1}{2}(Z(\omega_k))^2\right)\right.\right. \\
 &\quad \left.\left.+ \alpha \left(1 - \frac{1}{N} \sum_{k=1}^N (1 + Z(\omega_k))\right) - \frac{3}{2}\right)^2\right) \\
 &= E\left(\left(\frac{1}{N} \sum_{k=1}^N \left(Z(\omega_k)(1 - \alpha) + \frac{1}{2}((Z(\omega_k))^2 - 1)\right)\right)^2\right) \\
 &= \frac{1}{N} E\left(\left(Z(1 - \alpha) + \frac{1}{2}((Z)^2 - 1)\right)^2\right) \\
 &= \frac{1}{N} \left(E((Z)^2(1 - \alpha)^2) + \frac{1}{4}(E((Z)^4) - 2E((Z)^2) + 1)\right) \\
 &= \frac{1}{N} \left((1 - \alpha)^2 + \frac{1}{4}(3 - 2 + 1)\right) = \frac{1}{N} \left(\frac{1}{2} + (1 - \alpha)^2\right).
 \end{aligned}$$

It turns out that the minimum variance can be achieved for $\alpha_{\min} = 1$, which yields $\text{Var}(\tilde{V}_N) = \frac{1}{2N}$.

16.3 It follows that

$$u(t, x) = E \left((X_T^{t,x})^2 \mid \mathcal{A}_t \right) = x^2 \exp\{(2\alpha + \beta^2)(T-t)\}.$$

Therefore, we obtain

$$d \left(t, \tilde{X}_t^{0,x} \right) = -2\beta.$$

This then yields

$$\left(\tilde{X}_T^{0,x} \right)^2 = x^2 \exp\{(2\alpha + 3\beta^2)T + 2\beta W_T\}$$

and

$$\theta_T = \theta_0 \exp\{-2\beta^2 T - 2\beta W_T\}.$$

Consequently, one has

$$\begin{aligned} g \left(\tilde{X}_T^{0,x} \right) \frac{\theta_T}{\theta_0} &= \left(\tilde{X}_T^{0,x} \right)^2 \frac{\theta_T}{\theta_0} \\ &= x^2 \exp\{(2\alpha + \beta^2)T\} \\ &= u(0, x) \\ &= E \left(g \left(X_T^{0,x} \right) \mid \mathcal{A}_0 \right). \end{aligned}$$

16.4 The fully drift implicit simplified weak Euler scheme for the Heston model, when approximating X by Y and v by V , has for an equi-distant time discretization with time step size Δ the form

$$Y_{n+1} = Y_n + (r_d + r_f) Y_{n+1} \Delta + k \sqrt{V_n} Y_n \Delta \hat{W}_n^1$$

$$V_{n+1} = V_n + \kappa (V_{n+1} - \bar{v}) \Delta + p \sqrt{V_n} \left(\varrho \Delta \hat{W}_n^1 + \sqrt{1 - \varrho^2} \Delta \hat{W}_n^2 \right)$$

for $n \in \{0, 1, \dots\}$ with $Y_0 = X_0$ and $V_0 = v_0$.

Therefore, we obtain

$$Y_{n+1} = \frac{Y_n \left(1 + k \sqrt{V_n} \Delta \hat{W}_n^1 \right)}{1 - (r_d + r_f) \Delta}$$

and

$$V_{n+1} = \frac{V_n - \kappa \bar{v} \Delta + p \sqrt{V_n} \left(\varrho \Delta \hat{W}_n^1 + \sqrt{1 - \varrho^2} \Delta \hat{W}_n^2 \right)}{1 - \kappa \Delta}$$

with independent $\Delta \hat{W}_n^1$ and $\Delta \hat{W}_n^2$ such that

$$P \left(\Delta \hat{W}_n^1 = \pm \sqrt{\Delta} \right) = P \left(\Delta \hat{W}_n^2 = \pm \sqrt{\Delta} \right) = \frac{1}{2}.$$

16.5 The simplified weak Euler scheme with implicit drifts and implicit diffusion term for X has the form

$$Y_{n+1} = Y_n + (r_d + r_f) Y_{n+1} \Delta - k^2 V_{n+1} Y_{n+1} \Delta + k \sqrt{V_{n+1}} Y_{n+1} \Delta \hat{W}_n^1$$

$$V_{n+1} = V_n + \kappa (V_{n+1} - \bar{v}) \Delta + p \sqrt{V_n} \Delta \hat{W}_n^2$$

Here we obtain

$$Y_{n+1} = \frac{Y_n}{1 - (r_d + r_f) \Delta + k^2 V_{n+1} \Delta - k \sqrt{V_{n+1}} \Delta \hat{W}_n^1}$$

and

$$V_{n+1} = \frac{V_n - \kappa \bar{v} \Delta + p \sqrt{V_n} \Delta \hat{W}_n^2}{1 - \kappa \Delta}.$$

Solutions for Exercises of Chapter 17

17.1 For the multi-period binomial tree we have the benchmarked return

$$u = \exp\{\sigma \sqrt{\Delta}\} - 1$$

with probability $p = \frac{-d}{u-d}$ and the benchmarked return

$$d = \exp\{-\sigma \sqrt{\Delta}\} - 1$$

with the remaining probability $1 - p$. The binomial European put price at time $t = 0$ is then given by the expression

$$\begin{aligned} S_0^{(\delta_{H_i \Delta})} &= \hat{S}_0^{(\delta_{H_i \Delta})} = E \left(\left(\hat{K} - \hat{S}_{i \Delta} \right)^+ \mid \mathcal{A}_0 \right) \\ &= \sum_{k=0}^i \frac{i!}{k! (i-k)!} p^k (1-p)^{i-k} \left(\hat{K} - (1+u)^k (1+d)^{i-k} \hat{S}_0 \right)^+ \\ &= \hat{K} \sum_{k=0}^{\bar{k}} \frac{i!}{k! (i-k)!} p^k (1-p)^{i-k} \\ &\quad - S_0 \sum_{k=0}^{\bar{k}} \frac{i!}{k! (i-k)!} p^k (1-p)^{i-k} (1+u)^k (1+d)^{i-k}, \end{aligned}$$

where \bar{k} denotes the first integer k for which $S_0 (1+u)^k (1+d)^{i-k} < \hat{K}$.

17.2 A trinomial recombining tree can achieve, in general, weak order $\beta = 1.0$ because three conditions can, in general, be satisfied to construct the transition probabilities to match the first two conditional moments of its increments.

17.3 For a recombining tree to achieve weak order $\beta = 2.0$ it is, in general, necessary to have five possible nodes at each time step where the Markov chain could go to in one time step. Otherwise, the transition densities could, in general, not be chosen to match the necessary moments.

Acknowledgements

Continuing support is acknowledged by the Quantitative Finance Research Centre, the Department of Mathematical Sciences and the School of Finance & Economics at the University of Technology Sydney. Research on the book was supported by the ARC grants DP1092933, DP0878155, DP0771882, DP0559879 and DP0343913. Research on numerical methods benefited greatly from visits by the first author to the ETH Zurich, Humboldt University Berlin, Isaac Newton Institute Cambridge, Kyoto University, Hitotsubashi University Tokyo, Maximilian University Munich, The New School New York. Generous support by these institutions is acknowledged. The research by the second author was supported by scholarships from Boconi University and the Quantitative Finance Research Centre at UTS.

The financial data for most of the figures were provided by Datastream Thomson Financial. Figs. 3.6.1–3.6.6 are based on reconstructed long term data from Global Financial Data.

With the permission of Routledge Publishers, a part of Taylor & Francis, Figs. 16.5.1–16.5.6 are taken from [Heath & Platen \(2002d\)](#) and Figs. 14.2.1–14.4.3 from [Platen & Shi \(2008\)](#). Walter de Gruyter Publishers permitted the inclusion of Figs. 10.4.1–10.4.2 from [Fischer & Platen \(1999\)](#). Finally, “Communications on Stochastic Analysis” allowed the use of Figs. 2.3.4–2.3.14 from [Platen & Rendek \(2009a\)](#).

Bibliographical Notes

Chap. 1: SDEs with Jumps

A comprehensive introduction into stochastic processes with jumps is given in [Cont & Tankov \(2004\)](#). Other references that treat jump processes include [Brémaud \(1981\)](#), [Ikeda & Watanabe \(1989\)](#), [Bertoin \(1996\)](#), [Protter \(2005\)](#), [Asmussen & Glynn \(2007\)](#) and [Øksendal & Sulem \(2005\)](#).

Chap. 2: Exact Simulation of SDEs

The chapter follows closely [Platen & Rendek \(2009a\)](#). Some explicit solutions of SDEs were listed in Section 4.4 in [Kloeden & Platen \(1999\)](#). Papers that relate to some kind of explicit solutions of functionals include [Lewis & Shedler \(1979\)](#), [Cheng & Feast \(1980\)](#), [Deelstra & Delbaen \(1998\)](#), [Diop \(2003\)](#), [Bossy & Diop \(2004\)](#), [Craddock & Platen \(2004\)](#), [Alfonsi \(2005\)](#), [Beskos & Roberts \(2005\)](#), [Broadie & Kaya \(2006\)](#), [Kahl & Jäckel \(2006\)](#), [Lord et al. \(2006\)](#), [Andersen & Piterbarg \(2007\)](#), [Asmussen & Glynn \(2007\)](#), [Smith \(2007\)](#), [Andersen \(2008\)](#), [Berkaoui, Bossy & Diop \(2008\)](#), [Burq & Jones \(2008\)](#) and [Chen \(2008\)](#).

Copulas are presented, for instance, in [Sklar \(1959\)](#), [Nelsen \(1999\)](#) and [McNeil et al. \(2005\)](#).

The literature on Lévy processes and their application in finance includes [Madan & Seneta \(1990\)](#), [Eberlein & Keller \(1995\)](#), [Geman, El Karoui & Rochet \(1995\)](#), [Bertoin \(1996\)](#), [Protter & Talay \(1997\)](#), [Barndorff-Nielsen & Shephard \(2001\)](#), [Eberlein \(2002\)](#), [Kou \(2002\)](#), [Rubenthaler \(2003\)](#), [Cont & Tankov \(2004\)](#), [Jacod et al. \(2005\)](#), [Klüppelberg et al. \(2006\)](#).

Squared Bessel, Wishart and affine processes are researched, for instance in [Bru \(1991\)](#), [Duffie & Kan \(1996\)](#), [Revuz & Yor \(1999\)](#), [Platen \(2001\)](#), [Borodin & Salminen \(2002\)](#), [Duffie, Filipović & Schachermayer \(2003\)](#), [Gouriéroux &](#)

Sufana (2004), Beskos, Papaspiliopoulos & Roberts (2006, 2008), Casella & Roberts (2007), Da Fonseca, Grasselli & Tebaldi (2007, 2008), Grasselli & Tebaldi (2008) and Chen (2008).

Chap. 3: Benchmark Approach to Finance and Insurance

This chapter summarizes an alternative framework for pricing and hedging beyond the classical risk neutral paradigm, see Delbaen & Schachermayer (1998, 2006). It is extending the numéraire portfolio idea, see Long (1990), Bajeux-Besnainou & Portait (1997) and Becherer (2001), and has been developed in the book Platen & Heath (2006) and a sequence of papers, including Platen (2001, 2002, 2004a, 2004b, 2004c, 2005a, 2005b, 2006), Bühlmann & Platen (2003), Platen & Stahl (2003), Christensen & Platen (2005, 2007), Hulley et al. (2005), Miller & Platen (2005, 2008, 2009), Platen & Runggaldier (2005, 2007), Le & Platen (2006), Christensen & Larsen (2007), Hulley & Platen (2008a, 2008b, 2009), Kardaras & Platen (2008), Bruti-Liberati et al. (2009), Filipović & Platen (2009), Galessos & Runggaldier (2010), Hulley (2010) and Hulley & Schweizer (2010).

Other literature that steps out of the classical pricing paradigm includes Rogers (1997), Sin (1998), Lewis (2000), Loewenstein & Willard (2000a, 2000b), Fernholz (2002), Cheridito (2003), Cox & Hobson (2005), Delbaen & Schachermayer (2006), Bender, Sottinen & Valkeila (2007), Heston, Loewenstein & Willard (2007), Jarrow, Protter & Shimbo (2007a, 2007b), Karatzas & Kardaras (2007) and Ekström & Tysk (2009).

Chap. 4: Stochastic Expansions

The Wagner-Platen expansion was first formulated in Wagner & Platen (1978) and developed further in Platen (1981, 1982a, 1982b, 1984), Azencott (1982), Liske (1982), Liske, Platen & Wagner (1982), Platen & Wagner (1982), Milstein (1988a, 1995a), Yen (1988, 1992), BenArous (1989), Kloeden & Platen (1991a, 1991b, 1992), Kloeden, Platen & Wright (1992), Hu & Meyer (1993), Hofmann (1994), Gaines (1994, 1995), Gaines & Lyons (1994), Castell & Gaines (1995) Li & Liu, (1997, 2000), Lyons (1998), Burrage (1998), Takahashi (1999), Kuznetsov (1998, 1999, 2009), Misawa (2001), Rydén & Wiktorsson (2001), Studer (2001), Tocino & Ardanuy (2002), Schoutens & Studer (2003), Kusuoka (2004), Lyons & Victoir (2004), Milstein & Tretjakov (2004) and Bruti-Liberati & Platen (2007b).

Related other work on stochastic expansions includes Stroud (1971), Engel (1982), Kohatsu-Higa (1997) and Bayer & Teichmann (2008).

Chap. 5: Introduction to Scenario Simulation

This chapter summarizes results on Euler schemes and some other discrete-time strong approximations relating to work in Maruyama (1955), Franklin

(1965), Clements & Anderson (1973), Allain (1974), Wright (1974), Milstein, (1974, 1978), Fahrmeir (1976), Yamada (1976), Doss (1977), Kloeden & Pearson (1977), Wagner & Platen (1978), Sussmann (1978), Yamato (1979), Gikhman & Skorokhod (1979), Gorostiza (1980), Clark & Cameron (1980), Rümelin (1982), Talay (1982c, 1983), Clark (1984, 1985), Janssen (1984a, 1984b), Klauder & Petersen (1985a), Atalla (1986), Liske & Platen (1987), Kanagawa (1988, 1989, 1995, 1996, 1997), Kaneko & Nakao (1988), Manella & Palleschi (1989), Golec & Ladde (1989, 1993), Newton (1990, 1991), Mikulevicius & Platen (1991), Kurtz & Protter (1991a, 1991b), Kohatsu-Higa & Protter (1994), Mackevicius (1994), Chan & Stramer (1998), Grecksch & Anh (1998, 1999), Gyöngy (1998), Jacod & Protter (1998), Shoji (1998), Schurz (1999), Hausenblas (2002), Higham, Mao & Stuart (2002), Konakov & Mammen (2002), Milstein, Repin & Tretjakov (2002), Malliavin & Thalmaier (2003), Talay & Zheng (2003), Yuan & Mao (2004a), Jacod et al. (2005), Guyon (2006), Berkaoui et al. (2008).

Systematic treatments, surveys and books on topics related to the simulation of solutions of SDEs, include Pardoux & Talay (1985), Gard (1988), Milstein, (1988a, 1995a), Kloeden & Platen (1999), Bouleau & Lépingle (1993), Kloeden et al. (2003), Janicki & Weron (1994), Talay (1995), Rogers & Talay (1997), Kuznetsov (1998), Platen (1999a, 2004d, 2008), Cyganowski, Kloeden & Ombach (2001), Higham (2001), Burrage, Burrage & Tian (2004), Glasserman (2004) and Bruti-Liberati & Platen (2007a).

Pseudo-random and quasi-random number generation are discussed, for instance, in Brent (1974, 2004), Devroye (1986), Bratley et al. (1987), Niederreiter (1988b, 1992), L'Ecuyer (1994, 1999), Paskov & Traub (1995), Fishman (1996), Joy et al. (1996), Hofmann & Mathé (1997), Pagés & Xiao (1997), Matsumoto & Nishimura (1998), Sloan & Wozniakowski (1998), Pivitt (1999), Bruti-Liberati, Platen, Martini & Piccardi (2005), Panneton et al. (2006) and Bruti-Liberati et al. (2008).

Stochastic differential delay equations have been studied, for instance, in Tudor & Tudor (1987), Tudor (1989), Mao (1991), Baker & Buckwar (2000), Buckwar (2000), Küchler & Platen (2000, 2002), Hu, Mohammed & Yan (2004), Yuan & Mao (2004b), Buckwar & Shardlow (2005) and Kushner (2005).

Papers on numerical methods for backward stochastic differential equations include Ma, Protter & Yong (1994), Douglas, Ma & Protter (1996), Chevance (1997), Ma, Protter, San Martin & Torres (2002), Nakayama (2002), Makarov (2003), Bouchard & Touzi (2004), Gobet, Lemor & Warin (2005) and Ma, Shen & Zhao (2008).

Chap. 6: Regular Strong Taylor Approximations with Jumps

Further discrete-time strong approximations of solutions of SDEs, potentially with jumps, have been studied, for instance, in Franklin (1965), Shinozuka (1971), Kohler & Boyce (1974), Rao, Borwanker & Ramkrishna (1974),

Dsagnidse & Tschitashvili (1975), Harris (1976), Gorenne (1977), Kloeden & Pearson (1977), Clark (1978, 1982, 1985), Nikitin & Razevig (1978), Helfand (1979), Platen (1980a, 1982a, 1982b, 1984, 1987), Razevig (1980), Rootzén (1980), Wright (1980), Greenside & Helfand (1981), Casasus (1982, 1984), Guo (1982, 1984), Talay (1982a, 1982b, 1982c, 1983, 1984, 1999), Drummond, Duane & Horgan (1983), Janssen (1984a, 1984b), Shimizu & Kawachi (1984), Tetzlaff & Zschiesche (1984), Unny (1984), Harris & Maghsoudi (1985), Averina & Artemiev (1986), Drummond, Hoch & Horgan (1986), Kozlov & Petryakov (1986), Greiner, Strittmatter & Honerkamp (1987), Liske & Platen (1987), Mikulevicius & Platen (1988, 1991), Milstein (1988a, 1988b), Golec & Ladde (1989), Ikeda & Watanabe (1989), Feng, Lei & Qian (1992), Artemiev (1993b), Kloeden et al. (1993), Saito & Mitsui (1993a, 1996), Komori, Saito & Mitsui (1994), Milstein & Tretjakov (1994a, 1994b, 1997, 2000), Petersen (1994), Török (1994), Ogawa (1995), Rascanu & Tudor (1995), Gelbrich & Rachev (1996), Grecksch & Wadewitz (1996), Mackevicius (1996), Maghsoudi (1996, 1998), Newton (1996), Schurz (1996b), Yannios & Kloeden (1996), Artemiev & Averina (1997), Bokor (1997, 2003), Denk & Schäffler (1997), Kohatsu-Higa (1997), Kohatsu-Higa & Ogawa (1997), Komori, Mitsui & Sugiura (1997), Protter & Talay (1997), Tudor & Tudor (1997), Abukhaled & Allen (1998a, 1998b), Schein & Denk (1998), Kurtz & Protter (1991b), Jansons & Lythe (2000), Glasserman & Zhao (2000), Grüne & Kloeden (2001b), Hofmann, Müller-Gronbach & Ritter (2001), Hausenblas (2002), Kubilius & Platen (2002), Lehn, Rössler & Schein (2002), Metwally & Atiya (2002), Glasserman & Merener (2003b), Jimenez & Ozaki (2003), Gardoñ (2004, 2006), Carbonell, Jimenez, Biscay & de la Cruz (2005), Cortés, Sevilla-Peris & Jódar (2005), Doering, Sagsyan & Smereka (2005), El-Borai, El-Nadi, Mostafa & Ahmed (2005), Jacod et al. (2005), Jódar, Cortés, Sevilla-Peris & Villafuerte (2005), Mao, Yuan & Yin (2005), Mora (2005), Bruti-Liberati et al. (2006), Buckwar & Winkler (2006), Buckwar, Bokor & Winkler (2006), Burrage, Burrage, Higham, Kloeden & Platen (2006), Carbonell, Jimenez & Biscay (2006), Jimenez, Biscay & Ozaki (2006), Bruti-Liberati & Platen (2007a, 2007b, 2008), Kloeden & Jentzen (2007), Kloeden & Neuenkirch (2007), Komori (2007) and Wang, Mei & Xue (2007).

Some work on discrete-time approximations for semimartingales, including Lévy processes and α -stable processes can be found, for instance, in Marcus (1978, 1981), Platen & Rebolledo (1985), Protter (1985), Jacod & Shiryaev (2003), Mackevicius (1987), Bally (1989a, 1989b, 1990), Gyöngy (1991), Kurtz & Protter (1991a, 1991b), Janicki & Weron (1994), Kohatsu-Higa & Protter (1994), Janicki (1996), Janicki, Michna & Weron (1996), Protter & Talay (1997) and Tudor & Tudor (1997).

The approximation of solutions of SDEs with boundary conditions or first exit times were considered, for instance, in Platen (1983, 1985), Gerardi, Marchetti & Rosa (1984), Slominski (1994, 2001), Asmussen, Glynn & Pitman (1995), Lépingle (1995), Petterson (1995), Ferrante, Kohatsu-Higa & Sanz-Solé (1996), Abukhaled & Allen (1998b), Hausenblas (1999a, 1999b, 2000a,

2000b, 2000c), Kanagawa & Saisho (2000), Makarov (2001), Alabert & Ferrante (2003), Bossy, Gobet & Talay (2004), Alcock & Burrage (2006), Kim, Lee, Hänggi & Talkner (2007) and Sickenberger (2008).

Chap. 7: Regular Strong Itô Approximations

Derivative-free and Runge-Kutta type methods including predictor-corrector and implicit schemes were studied, for instance, in Clements & Anderson (1973), Wright (1974), Kloeden & Pearson (1977), Rümelin (1982), Talay (1982b, 1984), Platen (1984), Klauder & Petersen (1985a), Pardoux & Talay (1985), Chang (1987), Gard (1988), McNeil & Craig (1988), Milstein (1988a, 1995a), Smith & Gardiner (1988), Petersen (1990), Artemiev & Shkurko (1991), Drummond & Mortimer (1991), Lépingle & Ribémont (1991), Newton (1991), Hernandez & Spigler (1992, 1993), Kloeden & Platen (1999), Artemiev (1993a, 1993b, 1994), Denk (1993), Saito & Mitsui (1993a, 1993b, 1995, 1996), Burrage & Platen (1994), Hofmann & Platen (1994, 1996), Komori et al. (1994), Milstein & Platen (1994), Hofmann (1995), Mitsui (1995), Biscay, Jimenez, Riera & Valdes (1996), Burrage & Burrage (1996, 1998, 1999, 2000), Schurz (1996a, 1996b, 1996c), Bokor (1997), Burrage, Burrage & Belward (1997), Denk & Schäffler (1997), Komori et al. (1997), Milstein & Tretjakov (1997), Ryashko & Schurz (1997), Burrage (1998), Denk, Penski & Schäffler (1998), Milstein et al. (1998), Petersen (1998), Fischer & Platen (1999), Liu & Xia (1999), Burrage, Burrage & Mitsui (2000), Burrage & Tian (2000, 2002), Higham (2000), Hofmann et al. (2000a, 2000b, 2001), Asmussen & Rosiński (2001), Fleury & Bernard (2001), Grüne & Kloeden (2001a, 2001b, 2006), Higham et al. (2002), Müller-Gronbach (2002), Tocino & Ardanuy (2002), Vanden-Eijnden (2003), Cao, Petzold, Rathinam & Gillespie (2004), Gardoñ (2004, 2006), Glasserman (2004), Lépingle & Nguyen (2004), Wu, Han & Meng (2004), Higham & Kloeden (2005, 2006), Rodkina & Schurz (2005), Alcock & Burrage (2006), Kahl & Schurz (2006), Bruti-Liberati & Platen (2008), Li, Abdulle & E (2008), Pang, Deng & Mao (2008), Rathinasamy & Balachandran (2008c, 2008a, 2008b) and Wang (2008).

Chap. 8: Jump-Adapted Strong Approximations

The use of jump-adapted time discretizations goes back to Platen (1982a) and has been employed, for instance, in Maghsoodi (1996), Glasserman (2004), Turner et al. (2004), Bruti-Liberati et al. (2006) and Bruti-Liberati & Platen (2007b, 2008).

Chap. 9: Estimating Discretely Observed Diffusions

There exists an extensive literature on estimation for stochastic differential equations, which includes Novikov (1972b), Taraskin (1974), Brown & Hewitt

(1975), Le Breton (1976), Balakrishnan (1977), Liptser & Shiryaev (1977, 2001), Lanska (1979), Hansen (1982), Kozin (1983), Linkov (1984), Kutoyants (1984), Dacunha-Castelle & Florens-Zmirou (1986), Godambe & Heyde (1987), Lo (1988), Heyde (1989, 1997), Kazimierczuk (1989), Küchler & Sørensen (1989, 1997), Sørensen (1991), Gouriéroux et al. (1993), Duffie & Singleton (1993), Barndorff-Nielsen & Sørensen (1994), Bibby (1994), Pedersen (1995), Bibby & Sørensen (1995, 1996), Ait-Sahalia (1996, 2002, 2004), Gallant & Tauchen (1996), Kloeden et al. (1996), Sørensen (1997), Shoji & Ozaki (1998a, 1998b), Kessler & Sørensen (1999), Poulsen (1999), Prakasa Rao (1999), Sørensen (1997, 1999, 2000, 2001), Genon-Catalot et al. (2000), Christensen et al. (2001), Elerain et al. (2001), Eraker (2001), Jiang & Knight (2002), Brandt & Santa-Clara (2002), Bibby et al. (2003), Barndorff-Nielsen & Shephard (2004), Carrasco, Chernov, Florens & Ghysels (2007), Kelly et al. (2004), Iacus (2008) and Bladt & Sørensen (2009).

Chap. 10: Filtering

The literature on filtering and approximate discrete time filters includes Kalman & Bucy (1961), Wonham (1965), Zakai (1969), Fujisaki et al. (1972), Liptser & Shiryaev (1977), Clark (1978, 1982), Kallianpur (1980), Clark & Cameron (1980), Talay (1982c), Newton (1984, 1986a, 1986b, 1991), Dembo & Zeitouni (1986), Hernandez (1988), Elliott & Glowinski (1989), Baras & Fineso (1991), Le Gland (1992), Kloeden & Platen (1992), Kloeden et al. (1993), Sun & Glowinski (1994), DiMasi et al. (1995), Elliott et al. (1995), Castell & Gaines (1996), Elliott & van der Hoek (1997), Kannan & Zhang (1998), Crisan, Gaines & Lyons (1998), Elliott, Fischer & Platen (1999a, 1999b), Fischer & Platen (1999), Fischer et al. (1999), Setiawaty (1999), Landen (2000), Chiarella et al. (2001), Elliott & Platen (2001), Frey & Rungaldier (2001), Gombani & Rungaldier (2001), Bhar, Chiarella & Rungaldier (2002, 2004), Platen & Rungaldier (2005, 2007) and Platen & Rendek (2009b).

Chap. 11: Monte Carlo Simulation of SDEs

First let us mention the increasing collection of books containing parts on Monte Carlo simulation of SDEs that is available, including Kloeden & Platen (1999), Bouleau & Lépingle (1993), Milstein (1988a, 1995a), Kloeden, Platen & Schurz (2003), Artemiev & Averina (1997), Cyganowski et al. (2001), Jäckel (2002), Cont & Tankov (2004), Glasserman (2004), Milstein & Tretjakov (2004), Higham (2004), Achdou & Pironneau (2005), McLeish (2005), Chan & Wong (2006), Malliavin & Thalmaier (2006), Platen & Heath (2006), Seydel (2006), Asmussen & Glynn (2007) and Iacus (2008).

Weak approximations have been studied, for instance, in Fahrmeir (1974), Boyle (1977), Milstein (1978, 1985, 1988a, 1988b, 1995a, 1995b, 1995c, 1996, 1997), Helfand (1979), Platen (1980b, 1984, 1987), Römisch (1983), Artemiev

(1984, 1985), Gladyshev & Milstein (1984), Talay (1984, 1986, 1987, 1990, 1991, 1995), Kanagawa (1985, 1989), Klauder & Petersen (1985a, 1985b), Pardoux & Talay (1985), Ventzel, Gladyshev & Milstein (1985), Averina & Artemiev (1986), Haworth & Pope (1986), Chang (1987), Petersen (1987, 1990), Römisch & Wakolbinger (1987), Mikulevicius & Platen (1988, 1991), Gelbrich (1989), Greenside & Helfand (1981), Wagner (1989a, 1989b), Grorud & Talay (1990, 1996), Talay & Tubaro (1990), Kloeden & Platen (1991b, 1992), Kloeden, Platen & Hofmann (1992, 1995), Goodlett & Allen (1994), Hofmann (1994), Mackevicius (1994, 1996), Komori & Mitsui (1995), Platen (1995), Arnold & Kloeden (1996), Bally & Talay (1996a, 1996b), Boyle et al. (1997), Broadie & Detemple (1997a), Kohatsu-Higa & Ogawa (1997), Milstein & Tretjakov (1997), Protter & Talay (1997), Jacod & Protter (1998), Hausenblas (1999a, 2002), Kurtz & Protter (1991a), Andersen & Andreasen (2000b), Hofmann et al. (2000a, 2000b), Liu & Li (2000), Hunter et al. (2001), Kohatsu-Higa (2001), Kusuoka (2001), Jansons & Lythe (2000), Jäckel (2002), Kubilius & Platen (2002), Metwally & Atiya (2002), Tocino & Vigo-Aguiar (2002), Glasserman & Melenner (2003a, 2003b), Jimenez & Ozaki (2003, 2006), Joshi (2003), Schoutens & Symens (2003), Cruzeiro, Malliavin & Thalmaier (2004), Glasserman (2004), Kusuoka (2004), Lyons & Victoir (2004), Brandt, Goyal, Santa-Clara & Stroud (2005), Jacod et al. (2005), Kebaier (2005), Malliavin & Thalmaier (2006), Mao et al. (2005), Mora (2005), Schmitz-Abe & Shaw (2005), Tocino (2005), Carbonell et al. (2006), Guyon (2006), Jimenez et al. (2006), Giles (2007, 2008), Siopacha & Teichmann (2007), Wang et al. (2007), Bruti-Liberati et al. (2008), Joshi & Stacey (2008), Mordecki, Szepessy, Tempone & Zouraris (2008), Ninomiya & Victoir (2008) and Giles, Higham & Mao (2009).

Chap. 12: Regular Weak Taylor Approximations

American option pricing via simulation or related methods has been studied, for instance, by Carriére (1996), Broadie & Glasserman (1997a, 1997b), Broadie & Detemple (1997b), Acworth, Broadie & Glasserman (1998), Broadie, Glasserman & Ha (2000), Thitsiklis & Van Roy (2001), Rogers (2002), Longstaff & Schwartz (2001), Clément, Lamberton & Protter (2002), Andersen & Broadie (2004), Stentoft (2004) and Chen & Glasserman (2007).

LIBOR market models and their implementation have been considered, for instance, in Andersen & Andreasen (2000b), Glasserman & Zhao (2000), Brigo & Mercurio (2005), Hunter et al. (2001), Rebonato (2002), Schoutens & Studer (2003), Pietersz, Pelsser & van Regenmortel (2004), Andersen & Brotherton-Ratcliffe (2005), Schoenmakers (2005) and Joshi & Stacey (2008).

Step size control and related questions were studied, for instance, in Artemiev (1985), Ozaki (1992), Müller-Gronbach (1996), Gaines & Lyons (1997), Shoji & Ozaki (1997, 1998a, 1998b), Burrage (1998), Mauthner (1998), Hofmann et al. (2000a, 2000b, 2001), Szepessy, Tempone & Zouraris (2001), Burrage & Burrage (2002), Lehn et al. (2002), Guerra & Sorini (2003), Lamba,

Mattingly & Stuart (2003), Burrage, Herdiana & Burrage (2004), Bossy et al. (2004), Römisch & Winkler (2006) and Lamba, Mattingly & Stuart (2007).

Chap. 13: Jump-Adapted Weak Approximations

Most of the literature mentioned already for Chap. 8, Chap. 11 and Chap. 12 is also relevant for Chap. 13.

Chap. 14: Numerical Stability

Questions related to numerical stability of simulations appear in a wide range of papers, including Talay (1982b, 1984), Klauder & Petersen (1985a), Pardoux & Talay (1985), McNeil & Craig (1988), Milstein (1988a, 1995a), Smith & Gardiner (1988), Petersen (1990), Artemiev & Shkurko (1991), Drummond & Mortimer (1991), Kloeden & Platen (1992), Hernandez & Spigler (1992, 1993), Artemiev (1993a, 1993b, 1994), Saito & Mitsui (1993a, 1993b, 1996), Hofmann & Platen (1994, 1996), Komori et al. (1994), Milstein & Platen (1994), Hofmann (1995), Komori & Mitsui (1995), Mitsui (1995), Schurz (1996a, 1996c), Bokor (1997, 1998), Komori et al. (1997), Ryashko & Schurz (1997), Burrage (1998), Milstein et al. (1998), Petersen (1998), Fischer & Platen (1999), Liu & Xia (1999), Burrage et al. (2000), Burrage & Tian (2000, 2002), Higham (2000), Vanden-Eijnden (2003), Cao, Petzold, Rathinam & Gillespie (2004), Wu et al. (2004), Higham & Kloeden (2005, 2006), Rodkina & Schurz (2005), Tocino (2005), Alcock & Burrage (2006), Bruti-Liberati & Platen (2008), Platen & Shi (2008), Li et al. (2008), Rathinasamy & Balachandran (2008a, 2008b, 2008c) and Pang et al. (2008).

Stability in stochastic differential equations with regime switching and potential delay have been studied, for instance, in Cao, Liu & Fan (2004), Liu, Cao & Fan (2004), Yuan & Mao (2004b), Ding, Wu & Liu (2006), Mao, Yin & Yuan (2007), Mao & Yuan (2006), Ronghua, Hongbing & Qin (2006), Ronghua & Yingmin (2006), Wang & Zhang (2006), Yuan & Glover (2006), Mao (2007), Mao, Yuan & Yin (2007), Ronghua & Zhaoguang (2007), Mao, Shen & Yuan (2008) and Zhang & Gan (2008).

Chap. 15: Martingale Representations

Methods that involve martingale representations or the simulation of hedge ratios can be found, for instance, in Heath (1995), Fournie, Lasry, Lebuchoux, Liou & Touzi (1999, 2001), Potters, Bouchaud & Sestovic (2001), Avellaneda & Gamba (2002), Heath & Platen (2002a, 2002b, 2002c), Milstein & Schoenmakers (2002), Takeuchi (2002), Bernis et al. (2003), Gobet & Kohatsu-Higa (2003), Bouchard, Ekeland & Touzi (2004), El-Khatib & Privault (2004), Kusuoka (2004), Løkka (2004), Victoir (2004), Bally et al. (2005), Detemple, Garcia & Rindisbacher (2005), Schoenmakers (2005),

Takahashi & Yoshida (2005), Tebaldi (2005), Bavouzet & Messaoud (2006), Davis & Johansson (2006), Giles & Glasserman (2006), Malliavin & Thalmaier (2006), Teichmann (2006) and Kampen, Kolodko & Schoenmakers (2008).

Chap. 16: Variance Reduction Techniques

References on variance reduction techniques include Hammersley & Handscomb (1964), Fosdick & Jordan (1968), Chorin (1973), Ermakov (1975), Maltz & Hitzl (1979), Sabelfeld (1979), Rubinstein (1981), Ermakov & Mikhailov (1982), Ripley (1983), Kalos (1984), Binder (1985), De Raedt & Lagendijk (1985), Ventzel et al. (1985), Kalos & Whitlock (1986), Bratley et al. (1987), Chang (1987), Wagner (1987, 1988a, 1988b, 1989a, 1989b), Hull & White (1988), Milstein (1988a, 1995a), Ross (1990), Law & Kelton (1991), Hofmann et al. (1992), Newton (1994), Melchior & Öttinger (1995, 1996), Fishman (1996), Fournie, Lebuchoux & Touzi (1997), Kohatsu-Higa (2001), Fouque & Tullie (2002), Heath & Platen (2002d), Kohatsu-Higa & Petterson (2002), Zou & Skeel (2004), Leobacher (2006), Leon, Sole, Utzet & Vives (2006) and Kebaier & Kohatsu-Higa (2008).

Variance reduction for the simulation of rare events has been discussed, for instance, in Glasserman, Heidelberger & Shahabuddin (2000), Glasserman (2004), Asmussen & Glynn (2007), Dupuis, Leder & Wang (2007) and Blanchet & Glynn (2008).

Chap. 17: Trees and Markov Chain Approximations

Papers on Markov chains and trees that are approximating jump diffusions include Kushner (1974, 1977, 1984), Kushner & DiMasi (1978), Cox et al. (1979), DiMasi & Runggaldier (1981), Black et al. (1990), Kushner & Dupuis (1992), Platen (1992), Boyle & Lau (1994), Rubinstein (1994), Pelsser & Vorst (1994), Heston (1995), Leisen & Reimer (1996), Jackwerth (1997), Pliska (1997), Brown & Toft (1999), Jensen (2000), Heston & Zhou (2000), Xu, Qian & Jiang (2003), van der Hoek & Elliott (2006) and Tanaka & Kohatsu-Higa (2009).

Finite difference methods have a wide literature and we refer here only to Richtmeyer & Morton (1967), Smith (1985), Wilmott et al. (1993), Shaw (1998) and Tavella & Randall (2000).

References

- Abramowitz, M. & Stegun, I. A. (eds) (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York.
- Abukhaled, M. I. & Allen, E. J. (1998a). A class of second-order Runge-Kutta methods for numerical solution of stochastic differential equations, *Stochastic Anal. Appl.* **16**(6): 977–991.
- Abukhaled, M. I. & Allen, E. J. (1998b). A recursive integration method for approximate solution of stochastic differential equations, *Internat. J. Comput. Math.* **66**(1-2): 53–66.
- Achdou, Y. & Pironneau, O. (2005). *Computational Methods for Option Pricing*, SIAM, Philadelphia.
- Acwirth, P., Broadie, M. & Glasserman, P. (1998). A comparison of some Monte Carlo and quasi-Monte Carlo methods for option pricing, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, Springer, pp. 1–18.
- Ait-Sahalia, Y. (1996). Testing continuous-time models for the spot interest rate, *Rev. Financial Studies* **9**: 385–426.
- Ait-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach, *Econometrica* **70**: 223–262.
- Ait-Sahalia, Y. (2004). Disentangling diffusion from jumps, *J. Financial Economics* **74**: 487–528.
- Ait-Sahalia, Y., Fan, J. & Peng, H. (2009). Nonparametric transition-based tests for jump diffusion, *J. Amer. Statist. Assoc.* **104**(487): 1102–1116.
- Ait-Sahalia, Y. & Jacod, J. (2009). Estimating the degree of activity of jumps in high-frequency data, *Ann. Statist.* **37**(5A): 2202–2244.
- Alabert, A. & Ferrante, M. (2003). Linear stochastic differential equations with functional boundary conditions, *Ann. Probab.* **31**(4): 2082–2108.
- Alcock, J. T. & Burrage, K. (2006). A note on the balanced method, *BIT Numerical Mathematics* **46**: 689–710.
- Alfonsi, A. (2005). On the discretization schemes for the CIR (and Bessel squared) processes, *Monte Carlo Methods Appl.* **11**(4): 355–384.
- Allain, M. F. (1974). *Sur quelques types d'approximation des solutions d'équations différentielles stochastiques*, PhD thesis, Univ. Rennes.
- Andersen, L. (2008). Simple and efficient simulation of the Heston stochastic volatility model, *J. Comput. Finance* **11**(3): 1–42.
- E. Platen, N. Bruti-Liberati, *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, Stochastic Modelling and Applied Probability 64, DOI [10.1007/978-3-642-13694-8](https://doi.org/10.1007/978-3-642-13694-8),
© Springer-Verlag Berlin Heidelberg 2010

- Andersen, L. & Andreasen, J. (2000a). Jump-diffusion processes: Volatility smile fitting and numerical methods for option pricing, *Rev. Derivatives Res.* **4**: 231–262.
- Andersen, L. & Andreasen, J. (2000b). Volatility skews and extensions of the LIBOR market model, *Appl. Math. Finance* **7**: 1–32.
- Andersen, L. & Broadie, M. (2004). A primal-dual simulation algorithm for pricing multi-dimensional American options, *Management Science* **50**: 1222–1234.
- Andersen, L. & Brotherton-Ratcliffe, R. (2005). Extended Libor market models with stochastic volatility, *J. Comput. Finance* **9**(1): 1–40.
- Andersen, L. & Piterbarg, V. (2007). Moment explosions in stochastic volatility models, *Finance Stoch.* **11**(1): 29–50.
- Ansel, J. P. & Stricker, C. (1994). Couverture des actifs contingents, *Ann. Inst. H. Poincaré Probab. Statist.* **30**: 303–315.
- Arnold, L. & Kloeden, P. E. (1996). Discretization of a random dynamical system near a hyperbolic point, *Math. Nachr.* **181**: 43–72.
- Artemiev, S. S. (1984). Comparison of certain methods of numerical solution of stochastic differential equations. Preprint 84-474, Akad. Nauk SSSR Sibirsk. Otdel., Novosibirsk Computing Center, 25 pp., (in Russian).
- Artemiev, S. S. (1985). A variable step algorithm for numerical solution of stochastic differential equations, *Chisl. Metody Mekh. Sploshn. Sredy* **16**(2): 11–23. (in Russian).
- Artemiev, S. S. (1993a). Certain aspects of application of numerical methods of solving SDE systems, *Numer. Anal.*, Vol. 1 of *Bull. of the Novosibirsk Computing Center*, NCC Publisher, pp. 1–16.
- Artemiev, S. S. (1993b). The stability of numerical methods for solving stochastic differential equations, *Numer. Anal.*, Vol. 2 of *Bull. of the Novosibirsk Computing Center*, NCC Publisher, pp. 1–10.
- Artemiev, S. S. (1994). The mean square stability of numerical methods for solving stochastic differential equations, *Russian J. Numer. Anal. Math. Modelling* **9**(5): 405–416.
- Artemiev, S. S. & Averina, A. T. (1997). *Numerical Analysis of Systems of Ordinary and Stochastic Differential Equations*, VSP, Utrecht.
- Artemiev, S. S. & Shkurko, I. O. (1991). Numerical analysis of dynamics of oscillatory stochastic systems, *Soviet J. Numer. Anal. Math. Modelling* **6**: 277–298.
- Ash, R. B. (1972). *Real Analysis and Probability*, Academic Press, New York.
- Asmussen, S. & Glynn, P. W. (2007). *Stochastic Simulation, Algorithms and Analysis*, Springer.
- Asmussen, S., Glynn, P. W. & Pitman, J. (1995). Discretization error in simulation of one-dimensional reflecting Brownian motion, *Ann. Appl. Probab.* **5**(4): 875–896.
- Asmussen, S. & Rosiński, J. (2001). Approximations of small jumps of Lévy processes with a view towards simulation, *J. Appl. Probab.* **38**: 482–493.
- Atalla, M. A. (1986). Finite-difference approximations for stochastic differential equations, *Probabilistic Methods for the Investigation of Systems with an Infinite Number of Degrees of Freedom*, Collection of Scientific Works, Kiev, pp. 11–16. (in Russian).
- Athreya, K. B., Kliemann, W. H. & Koch, G. (1988). On sequential construction of solutions of stochastic differential equations with jump terms, *Systems Control Lett.* **10**: 141–146.

- Avellaneda, M. & Gamba, R. (2002). Conquering the Greeks in Monte Carlo: Efficient calculation of the market sensitivities and hedge-ratios of financial assets by direct numerical simulation, *Mathematical Finance - Bachelier Congress 2000*, pp. 93–109.
- Averina, A. T. & Artemiev, S. S. (1986). A new family of numerical methods for solving stochastic differential equations, *Soviet. Math. Dokl.* **33**(3): 736–738.
- Azencott, R. (1982). Stochastic Taylor formula and asymptotic expansion of Feynman integrals, *Séminaire de Probabilités XVI, Supplément*, Vol. 921 of *Lecture Notes in Math.*, Springer, pp. 237–285.
- Babbs, S. & Webber, N. (1995). When we say jump ..., *Risk* **8**(1): 49–53.
- Bachelier, L. (1900). Théorie de la spéculation, *Annales de l'Ecole Normale Supérieure, Series 3* **17**: 21–86.
- Bajeux-Besnainou, I. & Portait, R. (1997). The numeraire portfolio: A new perspective on financial theory, *The European Journal of Finance* **3**: 291–309.
- Baker, C. T. H. & Buckwar, E. (2000). Numerical analysis of explicit one-step methods for stochastic delay differential equations, *LMS J. Comp. Math.* **3**: 315–335.
- Balakrishnan, A. V. (1977). Likelihood ratios for time-continuous data models: the white noise approach, *New Trends in Systems Analysis*, Vol. 2 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 102–110.
- Bally, V. (1989a). Approximation for the solution of stochastic differential equations. I: L^p -convergence, *Stochastics Stochastics Rep.* **28**: 209–246.
- Bally, V. (1989b). Approximation for the solution of stochastic differential equations. II: Strong-convergence, *Stochastics Stochastics Rep.* **28**: 357–385.
- Bally, V. (1990). Approximation for the solutions of stochastic differential equations. III: Jointly weak convergence, *Stochastics Stochastics Rep.* **30**: 171–191.
- Bally, V., Caramellino, L. & Zanette, A. (2005). Pricing and hedging American options by Monte Carlo methods using a Malliavin calculus approach, *Monte Carlo Methods Appl.* **11**: 97–133.
- Bally, V. & Talay, D. (1996a). The law of the Euler scheme for stochastic differential equations I. Convergence rate of the distribution function, *Probab. Theory Related Fields* **104**(1): 43–60.
- Bally, V. & Talay, D. (1996b). The law of the Euler scheme for stochastic differential equations II. Convergence rate of the density, *Monte Carlo Methods Appl.* **2**(2): 93–128.
- Baras, J. B. & Fineso, L. (1991). Consistent estimation of the order of hidden Markov chains, *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 26–39.
- Barndorff-Nielsen, O. & Shephard, N. (2001). Modelling by Lévy processes for financial econometrics, in O. E. Barndorff-Nielsen, T. Mikosch & S. Resnick (eds), *Lévy Processes - Theory and Applications*, Birkhäuser, Boston, pp. 283–318.
- Barndorff-Nielsen, O. & Shephard, N. (2004). Economic analysis of realised covariation: High frequency covariance, regression, and correlation in financial economics, *Econometrica* **72**(3): 885–925.
- Barndorff-Nielsen, O. & Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation, *J. Financial Economics* **4**: 1–30.
- Barndorff-Nielsen, O. & Sørensen, M. (1994). A review of some aspects of asymptotic likelihood theory for stochastic processes, *Internat. Statist. Rev.* **62**: 133–165.
- Barraquand, J. (1995). Numerical valuation of high dimensional multivariate European securities, *Management Science* **41**(12): 1882.

- Barrett, R., Berry, M., Chan, T. F., Demmel, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C. & van der Vorst, H. (1994). *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, SIAM, Philadelphia.
- Basle (1995). *Planned Supplement to the Capital Accord to Incorporate Market Risk*, Basle Committee on Banking and Supervision, Basle, Switzerland.
- Bavouzet, M.-P. & Messaoud, M. (2006). Computation of Greeks using Malliavin's calculus in jump type market models, *Electronic J. of Prob.* **11**(10): 276–300.
- Bayer, C. & Teichmann, J. (2008). Cubature on Wiener space in infinite dimension, *Proceeding of the Royal Society London*, Vol. 464, pp. 2493–2516.
- Becherer, D. (2001). The numeraire portfolio for unbounded semimartingales, *Finance Stoch.* **5**: 327–341.
- Bell, E. T. (1934). Exponential polynomials, *Ann. Math.* **35**: 258–277.
- BenArous, G. (1989). Flots et séries de Taylor stochastiques, *Probab. Theory Related Fields* **81**: 29–77.
- Bender, C., Sottinen, T. & Valkeila, E. (2007). Arbitrage with fractional Brownian motion?, *Theory of Stochastic Processes* **13**(1-2): 23–34.
- Berkaoui, A., Bossy, M. & Diop, A. (2005). Euler schemes for SDEs with non-Lipschitz diffusion coefficient: Strong convergence. INRIA working paper no.5637.
- Berkaoui, A., Bossy, M. & Diop, A. (2008). Euler scheme for SDEs with non-Lipschitz diffusion coefficient: Strong convergence, *ESAIM: Prob. and Statist.* **12**: 1–11.
- Bernis, G., Gobet, E. & Kohatsu-Higa, A. (2003). Monte Carlo evaluation of Greeks for multidimensional barrier and look-back options, *Math. Finance* **13**: 99–113.
- Bertoin, J. (1996). *Lévy Processes*, Cambridge University Press.
- Beskos, A., Papaspiliopoulos, O. & Roberts, G. (2006). Retrospective exact simulation of diffusion sample paths with applications, *Bernoulli* **12**: 1077–1098.
- Beskos, A., Papaspiliopoulos, O. & Roberts, G. (2008). A factorisation of diffusion measure and finite sample path constructions, *Methodology and Comp. Appl. Prob.* **10**: 85–104.
- Beskos, A. & Roberts, G. (2005). Exact simulation of diffusions, *Ann. Appl. Probab.* **15**: 2422–2444.
- Bhar, R., Chiarella, C. & Runggaldier, W. J. (2002). Estimation in models of the instantaneous short term interest rate by use of a dynamic Bayesian algorithm, in K. Sandmann & P. Schönbucher (eds), *Advances in Finance and Stochastics - Essays in Honour of Dieter Sondermann*, Springer, pp. 177–195.
- Bhar, R., Chiarella, C. & Runggaldier, W. J. (2004). Inferring the forward looking equity risk premium from derivative prices, *Studies in Nonlinear Dynamics & Econometrics* **8**(1): 1–24.
- Bibby, B. M. (1994). Optimal combination of martingale estimating functions for discretely observed diffusions, *Technical report*, University of Aarhus, Department of Theoretical Statistics. Research Report No. 298.
- Bibby, B. M., Jacobsen, M. & Sørensen, M. (2003). Estimating functions for discretely sampled diffusion-type models, in Y. Ait-Sahalia & L. P. Hansen (eds), *Handbook of Financial Economics*, Amsterdam: North Holland, chapter 4, pp. 203–268.
- Bibby, B. M. & Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes, *Bernoulli* **1**: 17–39.
- Bibby, B. M. & Sørensen, M. (1996). On estimation for discretely observed diffusions: A review, *Theory of Stochastic Processes* **2**(18): 49–56.

- Bielecki, T. & Rutkowski, M. (2002). *Credit Risk: Modelling, Valuation and Hedging*, Springer Finance.
- Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.
- Binder, K. (1985). The Monte-Carlo method for the study of phase transitions: a review of recent progress, *J. Comput. Phys.* **59**: 1–55.
- Biscay, R., Jimenez, J. C., Riera, J. J. & Valdes, P. A. (1996). Local linearization method for the numerical solution of stochastic differential equations, *Ann. Inst. Statist. Math.* **48**(4): 631–644.
- Björk, T. (1998). *Arbitrage Theory in Continuous Time*, Oxford University Press.
- Björk, T., Kabanov, Y. & Runggaldier, W. J. (1997). Bond market structure in the presence of marked point processes, *Math. Finance* **7**: 211–239.
- Black, F. (1976). The pricing of commodity contracts, *J. Financial Economics* **3**: 167–179.
- Black, F. (1995). Interest rates as options, *J. Finance* **1**(7): 1371–1376.
- Black, F., Derman, E. & Toy, W. (1990). A one-factor model of interest rates and its application to treasury bond options, *Financial Analysts J.* **46**(1): 33–39.
- Black, F. & Karasinski, P. (1991). Bond and option pricing when short rates are lognormal, *Financial Analysts J.* **47**(4): 52–59.
- Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities, *J. Political Economy* **81**: 637–654.
- Bladt, M. & Sørensen, M. (2005). Statistical inference for discretely observed Markov jump processes, *J. Roy. Statist. Soc. Ser. B* **67**(3): 395–410.
- Bladt, M. & Sørensen, M. (2009). Efficient estimation of transition rates between credit ratings from observations at discrete time points, *Quant. Finance* **9**(2): 147–160.
- Blanchet, J. & Glynn, P. W. (2008). Efficient rare-event simulation for the maximum of heavy-tailed random walks, *Ann. Appl. Probab.* **18**: 1351–1378.
- Bokor, R. H. (1997). On two-step methods for stochastic differential equations, *Acta Cybernet.* **13**(2): 197–207.
- Bokor, R. H. (1998). On stability of two-step methods for stochastic differential equations, *Operational Research Proc. KOI'98*, pp. 267–279.
- Bokor, R. H. (2003). Stochastically stable one-step approximations of solutions of stochastic ordinary differential equations, *Appl. Numer. Math.* **44**: 299–312.
- Borodin, A. N. & Salminen, P. (2002). *Handbook of Brownian Motion - Facts and Formulae*, 2nd edn, Birkhäuser.
- Bossy, M. & Diop, A. (2004). An efficient discretization scheme for one dimensional SDEs with a diffusion coefficient function of the form $|x|^\alpha$, $\alpha \in [1/2, 1]$. INRIA working paper no. 5396.
- Bossy, M., Gobet, E. & Talay, D. (2004). A symmetrized Euler scheme for an efficient approximation of reflected diffusions, *J. Appl. Probab.* **41**: 877–889.
- Bouchard, B., Ekeland, I. & Touzi, N. (2004). On the Malliavin approach to Monte Carlo approximation of conditional expectations, *Finance Stoch.* **8**(1): 45–71.
- Bouchard, B. & Touzi, N. (2004). Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations, *Stochastic Process. Appl.* **111**: 175–206.
- Bouleau, N. & Lépingle, D. (1993). *Numerical Methods for Stochastic Processes*, Wiley, New York.
- Boyle, P. P. (1977). A Monte Carlo approach, *J. Financial Economics* **4**: 323–338.
- Boyle, P. P. (1988). A lattice framework for option pricing with two state variables, *J. Financial and Quantitative Analysis* **23**: 1–12.

- Boyle, P. P., Broadie, M. & Glasserman, P. (1997). Monte Carlo methods for security pricing. Computational financial modelling, *J. Econom. Dynam. Control* **21**(8-9): 1267–1321.
- Boyle, P. P. & Lau, S. H. (1994). Bumping up against the barrier with the binomial method, *J. Derivatives* pp. 6–14.
- Brandt, M., Goyal, A., Santa-Clara, P. & Stroud, J. R. (2005). A simulation approach to dynamic portfolio choice with an application to learning about return predictability, *Rev. Financial Studies* **18**: 831–873.
- Brandt, M. & Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with application to exchange rate dynamics in incomplete markets, *J. Financial Economics* **63**(2): 161–210.
- Bratley, P., Fox, B. L. & Schrage, L. (1987). *A Guide to Simulation*, 2nd edn, Springer.
- Brémaud, P. (1981). *Point Processes and Queues*, Springer, New York.
- Brent, R. P. (1974). A Gaussian pseudo number generator, *Commun. Assoc. Comput. Mach.* **17**: 704–706.
- Brent, R. P. (2004). Note on Marsaglia's xorshift random number generators, *J. Statist. Software* **11**(5): 1–4.
- Breymann, W., Kelly, L. & Platen, E. (2006). Intraday empirical analysis and modeling of diversified world stock indices, *Asia-Pacific Financial Markets* **12**(1): 1–28.
- Brigo, D. & Mercurio, F. (2005). *Interest Rate Models - Theory and Practice*, 2nd edn, Springer Finance.
- Broadie, M. & Detemple, J. (1997a). Recent advances in numerical methods for pricing derivative securities, *Numerical Methods in Finance*, Cambridge Univ. Press, pp. 43–66.
- Broadie, M. & Detemple, J. (1997b). The valuation of American options on multiple assets, *Math. Finance* **7**(3): 241–286.
- Broadie, M. & Glasserman, P. (1997a). Estimating security price derivatives using simulation, *Management Science* **42**(2): 269–285.
- Broadie, M. & Glasserman, P. (1997b). Pricing American - style securities using simulation. Computational financial modelling, *J. Econom. Dynam. Control* **21**(8-9): 1323–1352.
- Broadie, M., Glasserman, P. & Ha, Z. (2000). Pricing American options by simulation using a stochastic mesh with optimized weights, *Probabilistic Constrained Optimization: Methodology and Applications*, Kluwer, Norwell, pp. 32–50.
- Broadie, M. & Kaya, O. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes, *Oper. Res.* **54**: 217–231.
- Brown, B. M. & Hewitt, J. T. (1975). Asymptotic likelihood theorem for diffusion processes, *J. Appl. Probab.* **12**: 228–238.
- Brown, G. & Toft, K. J. (1999). Constructing binomial trees from multiple implied probability distributions, *J. Derivatives* **7**(2): 83–100.
- Bru, M.-F. (1991). Wishart processes, *J. Theoret. Probab.* **4**(4): 725–751.
- Bruti-Liberati, N., Martini, F., Piccardi, M. & Platen, E. (2008). A hardware generator for multi-point distributed random numbers for Monte Carlo simulation, *Math. Comput. Simulation* **77**(1): 45–56.
- Bruti-Liberati, N., Nikitopoulos-Sklibosios, C. & Platen, E. (2006). First order strong approximations of jump diffusions, *Monte Carlo Methods Appl.* **12**(3-4): 191–209.

- Bruti-Liberati, N., Nikitopoulos-Sklibosios, C. & Platen, E. (2009). Real-world jump-diffusion term structure models, *Quant. Finance*. electronic in iFirst.
- Bruti-Liberati, N. & Platen, E. (2004). On the efficiency of simplified weak Taylor schemes for Monte Carlo simulation in finance, *Computational Science - ICCS 2004*, Vol. 3039 of *Lecture Notes in Comput. Sci.*, Springer, pp. 771–778.
- Bruti-Liberati, N. & Platen, E. (2007a). Approximation of jump diffusions in finance and economics, *Computational Economics* **29**(3-4): 283–312.
- Bruti-Liberati, N. & Platen, E. (2007b). Strong approximations of stochastic differential equations with jumps, *J. Comput. Appl. Math.* **205**(2): 982–1001.
- Bruti-Liberati, N. & Platen, E. (2008). Strong predictor-corrector Euler methods for stochastic differential equations, *Stochastics and Dynamics* **8**(3): 561–581.
- Bruti-Liberati, N., Platen, E., Martini, F. & Piccardi, M. (2005). An FPGA generator for multi-point distributed random variables, *Proceedings of the 2005 ACM/SIGDA 13th International Symposium on Field-Programmable Gate Arrays*, p. 280.
- Buchen, P. W. (2004). The pricing of dual-expiry exotics, *Quant. Finance* **4**(1): 101–108.
- Buckwar, E. (2000). Introduction of the numerical analysis of ssde, *J. Comput. Appl. Math.* **125**: 297–307.
- Buckwar, E., Bokor, R. & Winkler, R. (2006). Asymptotic mean-square stability of two-step methods for stochastic ordinary differential equations, *BIT Numerical Mathematics* **46**: 261–282.
- Buckwar, E. & Shardlow, T. (2005). Weak approximation of stochastic differential delay equations, *IMA J. Numer. Anal.* **25**: 57–86.
- Buckwar, E. & Winkler, R. (2006). Multistep methods for sdes and their application to problems with small noise, *SIAM J. Numer. Anal.* **44**(2): 779–803.
- Bühlmann, H. & Platen, E. (2003). A discrete time benchmark approach for insurance and finance, *ASTIN Bulletin* **33**(2): 153–172.
- Burq, Z. A. & Jones, O. D. (2008). Simulation of Brownian motion at first-passage times, *Math. Comput. Simulation* **77**: 64–81.
- Burrage, K. & Burrage, P. M. (1996). High strong order explicit Runge-Kutta methods for stochastic ordinary differential equations, *Appl. Numer. Math.* **22**: 81–101.
- Burrage, K. & Burrage, P. M. (1998). General order conditions for stochastic Runge-Kutta methods for both commuting and non-commuting stochastic ordinary differential equation systems, *Appl. Numer. Math.* **28**: 161–177.
- Burrage, K. & Burrage, P. M. (1999). High strong order methods for non-commutative stochastic ordinary differential equation system and the Magnus formula, *Physica D* **133**: 34–48.
- Burrage, K. & Burrage, P. M. (2000). Order conditions of stochastic Runge-Kutta methods by B-series, *SIAM J. Numer. Anal.* **38**: 1626–1646.
- Burrage, K., Burrage, P. M. & Belward, J. A. (1997). A bound on the maximum strong order of stochastic Runge-Kutta methods for stochastic ordinary differential equations, *BIT* **37**(4): 771–780.
- Burrage, K., Burrage, P. M., Higham, D. J., Kloeden, P. E. & Platen, E. (2006). Comment on “numerical methods for stochastic differential equations”, *Phys. Rev. E* **74**: 068701, 1–2.
- Burrage, K., Burrage, P. M. & Mitsui, T. (2000). Numerical solutions of stochastic differential equations - implementation and stability issues, *J. Comput. Appl. Math.* **125**: 171–182.

- Burrage, K., Burrage, P. M. & Tian, T. (2004). Numerical methods for strong solutions of SDES, *Proceeding of the Royal Society London*, Vol. 460, pp. 373–402.
- Burrage, K. & Platen, E. (1994). Runge-Kutta methods for stochastic differential equations, *Ann. Numer. Math.* **1**(1-4): 63–78.
- Burrage, K. & Tian, T. (2000). A note on the stability properties of the Euler method for solving stochastic differential equations, *New Zealand J. of Maths.* **29**: 115–127.
- Burrage, K. & Tian, T. (2002). Predictor-corrector methods of Runge-Kutta type for stochastic differential equations, *SIAM J. Numer. Anal.* **40**: 1516–1537.
- Burrage, P. M. (1998). *Runge-Kutta methods for stochastic differential equations*, PhD thesis, University of Queensland, Brisbane, Australia.
- Burrage, P. M. & Burrage, K. (2002). A variable stepsize implementation for stochastic differential equations, *SIAM J. Sci. Comput.* **24**(3): 848–864.
- Burrage, P. M., Herdiana, R. & Burrage, K. (2004). Adaptive stepsize based on control theory for stochastic differential equations, *J. Comput. Appl. Math.* **170**: 317–336.
- Cao, W. R., Liu, M. Z. & Fan, Z. C. (2004). MS-stability of the Euler-Maruyama method for stochastic differential delay equations, *Appl. Math. Comput.* **159**: 127–135.
- Cao, Y., Petzold, L., Rathinam, M. & Gillespie, D. T. (2004). The numerical stability of leaping methods for stochastic simulation of chemically reacting systems, *J. Chemical Physics* **121**: 12169–12178.
- Carbonell, F., Jimenez, J. C. & Biscay, R. (2006). Weak local linear discretizations for stochastic differential equations: Convergence and numerical schemes, *J. Comput. Appl. Math.* **197**: 578–596.
- Carbonell, F., Jimenez, J. C., Biscay, R. & de la Cruz, H. (2005). The local linearization method for numerical integration of random differential equations, *BIT* **45**: 1–14.
- Carr, P., Geman, H., Madan, D. & Yor, M. (2003). Stochastic volatility for Lévy processes, *Math. Finance* **13**(3): 345–382.
- Carrasco, M., Chernov, M., Florens, J.-P. & Ghysels, E. (2007). Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions, *J. Econometrics* **140**: 529–573.
- Carriére, J. (1996). Valuation of early-exercise price of options using simulations and non-parametric regression, *Insurance: Mathematics and Economics* **19**: 19–30.
- Casasus, L. L. (1982). On the numerical solution of stochastic differential equations and applications, *Proceedings of the Ninth Spanish-Portuguese Conference on Mathematics*, Vol. 46 of *Acta Salmanticensia. Ciencias*, Univ. Salamanca, pp. 811–814. (in Spanish).
- Casasus, L. L. (1984). On the convergence of numerical methods for stochastic differential equations, *Proceedings of the Fifth Congress on Differential Equations and Applications*, Univ. La Laguna, pp. 493–501. Puerto de la Cruz (1982), (in Spanish), Informes 14.
- Casella, B. & Roberts, G. (2007). Exact Monte Carlo simulation of killed diffusions, *Adv. in Appl. Probab.* **40**: 273–291.
- Castell, F. & Gaines, J. (1995). An efficient approximation method for stochastic differential equations by means of the exponential Lie series, *Math. Comput. Simulation* **38**: 13–19.

- Castell, F. & Gaines, J. (1996). The ordinary differential equation approach to asymptotically efficient schemes for solution of stochastic differential equations, *Ann. Inst. H. Poincaré Probab. Statist.* **32**(2): 231–250.
- Chan, K. S. & Stramer, O. (1998). Weak consistency of the Euler method for numerically solving stochastic differential equations with discontinuous coefficients, *Stochastic Process. Appl.* **76**: 33–44.
- Chan, N. H. & Wong, H. Y. (2006). *Simulation Techniques in Financial Risk Management*, Wiley.
- Chang, C. C. (1987). Numerical solution of stochastic differential equations with constant diffusion coefficients, *Math. Comp.* **49**: 523–542.
- Chen, N. (2008). Exact simulation of stochastic differential equations. Chinese Univ. of Hong Kong (working paper).
- Chen, N. & Glasserman, P. (2007). Malliavin Greeks without Malliavin calculus, *Stochastic Process. Appl.* **117**: 1689–1723.
- Cheng, R. & Feast, G. (1980). Gamma variate generators with increased shape parameter range, *Communications of the ACM* **23**(7): 389–394.
- Cheridito, P. (2003). Arbitrage in fractional Brownian motion models, *Finance Stoch.* **7**(4): 533–553.
- Chernov, M., Gallant, A. R., Ghysels, E. & Tauchen, G. (2003). Alternative models of stock price dynamics, *J. Econometrics* **116**: 225–257.
- Cherny, A. S. (2000). On the strong and weak solutions of stochastic differential equations governing Bessel processes, *Stochastics Stochastics Rep.* **70**: 213–219.
- Chevance, D. (1997). Numerical methods for backward stochastic differential equations, in L. C. G. Rogers & D. Talay (eds), *Numerical Methods in Finance*, Cambridge University Press, pp. 232–244.
- Chiarella, C., Pasquali, S. & Runggaldier, W. J. (2001). On filtering in Markovian term structure models (an approximation approach), *Adv. in Appl. Probab.* **33**(4): 794–809.
- Chorin, A. J. (1973). Accurate evaluation of Wiener integrals, *Math. Comp.* **27**: 1–15.
- Christensen, B. J., Poulsen, R. & Sørensen, M. (2001). Optimal inference in diffusion models of the short rate of interest, *Technical report*, University of Aarhus, Centre for Analytical Finance. Working Paper No. 102.
- Christensen, M. M. & Larsen, K. (2007). No arbitrage and the growth optimal portfolio, *Stochastic Anal. Appl.* **25**(1): 255–280.
- Christensen, M. M. & Platen, E. (2005). A general benchmark model for stochastic jump sizes, *Stochastic Anal. Appl.* **23**(5): 1017–1044.
- Christensen, M. M. & Platen, E. (2007). Sharpe ratio maximization and expected utility when asset prices have jumps, *Int. J. Theor. Appl. Finance* **10**(8): 1339–1364.
- Clark, J. M. C. (1970). The representation of functionals of Brownian motion by stochastic integrals, *Ann. Math. Statist.* **41**: 1281–1295.
- Clark, J. M. C. (1978). The design of robust approximations to the stochastic differential equations of nonlinear filtering, in J. K. Skwirzynski (ed.), *Communication Systems and Random Processes Theory*, Vol. 25 of *NATO Advanced Study Inst. Series. Series E: Applied Sciences*, Sijthoff and Noordhoff, Alphen aan den Rijn, pp. 721–734.
- Clark, J. M. C. (1982). An efficient approximation scheme for a class of stochastic differential equations, *Advances in Filtering and Optimal Stochastic Control*, Vol. 42 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 69–78.

- Clark, J. M. C. (1984). Asymptotically optimal ratio formulae for stochastic differential equations, *Proc. 23rd Conf. Decisions and Control*, pp. 712–715. Las Vegas. IEEE.
- Clark, J. M. C. (1985). A nice discretization for stochastic line integrals, *Stochastic Differential Systems*, Vol. 69 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 131–142.
- Clark, J. M. C. & Cameron, R. J. (1980). The maximum rate of convergence of discrete approximations for stochastic differential equations, in B. Grigelionis (ed.), *Stochastic Differential Systems*, Vol. 25 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 162–171.
- Clément, E., Lamberton, D. & Protter, P. (2002). An analysis of a least squares regression algorithm for American option pricing, *Finance Stoch.* **6**: 449–471.
- Clements, D. J. & Anderson, B. D. O. (1973). Well behaved Ito equations with simulations that always misbehave, *IEEE Trans. Automat. Control* **18**: 676–677.
- Clewlow, L. & Carverhill, A. (1992). Efficient Monte Carlo valuation and hedging of contingent claims, *Technical report*, Financial Options Research Center, University of Warwick. 92/30.
- Clewlow, L. & Carverhill, A. (1994). Quicker on the curves, *Risk* **7**(5).
- Colwell, D. B., Elliott, R. J. & Kopp, P. E. (1991). Martingale representation and hedging policies, *Stochastic Process. Appl.* **38**: 335–345.
- Cont, R. & Tankov, P. (2004). *Financial Modelling with Jump Processes*, Financial Mathematics Series, Chapman & Hall/CRC.
- Cont, R. & Voltchkova, E. (2005). Finite difference methods for option pricing in jump-diffusion and exponential Lévy models, *SIAM J. Numer. Anal.* **43**(4): 1596–1626.
- Cortés, J. C., Sevilla-Peris, P. & Jódar, L. (2005). Analytic-numerical approximating processes of diffusion equation with data uncertainty, *Comput. Math. Appl.* **49**: 1255–1266.
- Courtadon, G. (1982). The pricing of options on default-free bonds, *J. Financial and Quantitative Analysis* **17**: 75–100.
- Cox, A. M. G. & Hobson, D. G. (2005). Local martingales, bubbles and option prices, *Finance Stoch.* **9**(4): 477–492.
- Cox, J. C. (1975). Notes on option pricing I: constant elasticity of variance diffusions. Stanford University, (working paper, unpublished).
- Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985). A theory of the term structure of interest rates, *Econometrica* **53**: 385–407.
- Cox, J. C., Ross, S. A. & Rubinstein, M. (1979). Option pricing: A simplified approach, *J. Financial Economics* **7**: 229–263.
- Craddock, M. & Platen, E. (2004). Symmetry group methods for fundamental solutions, *J. of Differential Equations* **207**(2): 285–302.
- Crisan, D., Gaines, J. & Lyons, T. G. (1998). Convergence of a branching particle method to the solution of the Zakai equation, *SIAM J. Appl. Math.* **58**(5): 1568–1590.
- Crisan, D. & Rozovski, B. (2010). *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press.
- Cruzeiro, A. B., Malliavin, P. & Thalmaier, A. (2004). Geometrization of Monte-Carlo numerical analysis of an elliptic operator: Strong approximation, *C. R. Acad. Sci. Paris Sér. I Math* **338**: 481–486.
- Curran, M. (1994). Rata gems, *Risk* **7**(3).

- Cyganowski, S., Kloeden, P. E. & Ombach, J. (2001). *From Elementary Probability to Stochastic Differential Equations with MAPLE*, Universitext, Springer.
- Da Fonseca, J., Grasselli, M. & Tebaldi, C. (2007). Option pricing when correlations are stochastic: An analytic framework, *Rev. Derivatives Res.* **10**(2): 151–180.
- Da Fonseca, J., Grasselli, M. & Tebaldi, C. (2008). A multifactor volatility Heston model, *Quant. Finance* **8**(6): 591–604.
- Dacunha-Castelle, D. & Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations, *Stochastics* **19**(4): 263–284.
- Davis, M. H. A. (1980). Functionals of diffusion processes as stochastic integrals, *Math. Proc. Cambridge Philos. Soc.* **87**: 157–166.
- Davis, M. H. A. & Etheridge, A. (2006). *Louis Bachelier's Theory of Speculation: The Origins of Modern Finance*, Princeton University Press.
- Davis, M. H. A. & Johansson, M. P. (2006). Malliavin Monte Carlo Greeks for jumps via the Malliavin calculus, *Stochastic Process. Appl.* **116**: 101–129.
- De Raedt, H. & Lagendijk, A. (1985). Monte-Carlo simulation of quantum statistical lattice models, *Phys. Rep.* **127**: 233–307.
- Deelstra, G. & Delbaen, F. (1998). Convergence of discretized stochastic (interest rate) processes with stochastic drift term, *Appl. Stoch. Models Data Anal.* **14**: 77–84.
- Delbaen, F. & Schachermayer, W. (1998). The fundamental theorem of asset pricing for unbounded stochastic processes, *Math. Ann.* **312**: 215–250.
- Delbaen, F. & Schachermayer, W. (2006). *The Mathematics of Arbitrage*, Springer Finance, Springer.
- Delbaen, F. & Shirakawa, H. (2002). A note on option pricing for the constant elasticity of variance model, *Asia-Pacific Financial Markets* **9**(2): 85–99.
- Dembo, A. & Zeitouni, O. (1986). Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm, *Stochastic Process. Appl.* **23**: 91–113. Erratum, **31**, pp. 167–169, 1989.
- DeMiguel, V., Garlappi, L. & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy?, *Rev. Financial Studies* **22**(5): 1915–1953.
- Denk, G. (1993). A new numerical method for the integration of highly oscillatory second-order ordinary differential equations, *Appl. Numer. Math.* **13**(1-3): 57–67.
- Denk, G., Penski, C. & Schäffler, S. (1998). Noise analysis in circuit simulation with stochastic differential equations, *Z. Angew. Math. Mech.* **67**(3): S887–S890.
- Denk, G. & Schäffler, S. (1997). Adam's methods for the efficient solution of stochastic differential equations with additive noise, *Computing* **59**(2): 153–161.
- Detemple, J., Garcia, R. & Rindisbacher, M. (2005). Representation formulas for Malliavin derivatives of diffusion processes, *Finance Stoch.* **9**: 349–367.
- Devroye, L. (1986). *Non-Uniform Random Variate Generation*, Vol. 19 of *Graduate Studies in Mathematics*, Springer-Verlag. Out of Printing. Electronic copy made available by the author at <http://cg.scs.carleton.ca/~luc/rnbookindex.html>.
- D'Halluin, Y., Forsyth, P. A. & Vetzal, K. R. (2005). Robust numerical methods for contingent claims under jump diffusion processes, *IMA Journal of Numerical Analysis* **25**: 87–112.
- DiMasi, G., Platen, E. & Rungaldier, W. J. (1995). Hedging of options under discrete observation on assets with stochastic volatility, *Seminar on Stochastic Analysis, Random Fields and Applications*, Vol. 36 of *Progr. Probab.*, Birkhäuser Verlag, pp. 359–364.

- DiMasi, G. & Runggaldier, W. J. (1981). An approximation to optimal nonlinear filtering with discontinuous observations, *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, Vol. 78 of *NATO Adv. Study Inst. Ser. C: Math. Phys. Sci.*, Reidel, Dordrecht-Boston, Mass., pp. 583–590.
- Ding, X., Wu, K. & Liu, M. (2006). Convergence and stability of the semi-implicit Euler method for linear stochastic delay integro-differential equations, *Internat. J. Comput. Math.* **83**: 753–763.
- Diop, A. (2003). *Sur la discréétisation et le comportement à petit bruit d'EDS multidimensionnelles dont les coefficients sont à dérivées singulières*, PhD thesis, INRIA.
- Doering, C. R., Sargsyan, K. V. & Smereka, P. (2005). Numerical method for some stochastic differential equations with multiplicative noise, *Physics Letters A* **344**: 149–155.
- Donati-Martin, C., Doumerc, Y., Matsumoto, H. & Yor, M. (2004). Some properties of the Wishart processes and a matrix extension of the Hartman-Watson laws, *Publ. RIMS, Kyoto Univ.* **40**: 1385–1412.
- Doob, J. L. (1953). *Stochastic Processes*, Wiley, New York.
- Doss, H. (1977). Liens entre équations différentielles stochastiques et ordinaires, *Ann. Inst. H. Poincaré Sect. B (N.S.)* **13**(2): 99–125.
- Dothan, L. U. (1978). On the term structure of interest rates, *J. Financial Economics* **6**: 59–69.
- Douglas, J., Ma, J. & Protter, P. (1996). Numerical methods for forward-backward stochastic differential equations, *Ann. Appl. Probab.* **6**(3): 940–968.
- Doukhan, P. (1994). *Mixing, Properties and Examples*, Springer.
- Drummond, I. T., Duane, S. & Horgan, R. R. (1983). The stochastic method for numerical simulations: Higher order corrections, *Nuc. Phys. B* **220**(1): 119–136. FS 8.
- Drummond, I. T., Hoch, A. & Horgan, R. R. (1986). Numerical integration of stochastic differential equations with variable diffusivity, *J. Phys. A: Math. Gen.* **19**: 3871–3881.
- Drummond, P. D. & Mortimer, I. K. (1991). Computer simulation of multiplicative stochastic differential equations, *J. Comput. Phys.* **93**(1): 144–170.
- Dsagnidse, A. A. & Tschitashvili, R. J. (1975). Approximate integration of stochastic differential equations. Tbilisi State University, Inst. Appl. Math. ‘Trudy IV’, Tbilisi, pp. 267–279, (in Russian).
- Duffie, D., Filipović, D. & Schachermayer, W. (2003). Affine processes and applications in finance, *Ann. Appl. Probab.* **3**(13): 984–1053.
- Duffie, D. & Glynn, P. W. (1995). Efficient Monte Carlo simulation of security prices, *Ann. Appl. Probab.* **5**(4): 897–905.
- Duffie, D. & Kan, R. (1994). Multi-factor term structure models, *Philos. Trans. Roy. Soc. London Ser. A* **347**: 577–580.
- Duffie, D. & Kan, R. (1996). A yield-factor model of interest rates, *Math. Finance* **6**(4): 379–406.
- Duffie, D., Pan, J. & Singleton, K. (2000). Transform analysis and option pricing for affine jump diffusions, *Econometrica* **68**: 1343–1376.
- Duffie, D. & Singleton, K. (1993). Simulated moments estimation of Markov models of asset prices, *Econometrica* **61**: 929–952.
- Dupuis, P. G., Leder, K. & Wang, H. (2007). Importance sampling for sums of random variables with heavy tails, *ACM TOMACS* **17**: 1–21.

- Eberlein, E. (2002). Applications of generalized Lévy motions in finance, in O. E. Barndorff-Nielsen, T. Mikosch & S. Resnick (eds), *Lévy Processes: Theory and Application*, Birkhäuser Boston, pp. 319–336.
- Eberlein, E. & Keller, U. (1995). Hyperbolic distributions in finance, *Bernoulli* **1**: 281–299.
- Einstein, A. (1905). On the movement of small particles suspended in stationary liquid demanded by molecular-kinetic theory of heat, *Ann. Phys.* **17**: 549–560.
- Ekström, E. & Tysk, J. (2009). Bubbles, convexity and the Black-Scholes equation, *Ann. Appl. Probab.* **19**: 1369–1384.
- El-Borai, M. M., El-Nadi, K. E., Mostafa, O. L. & Ahmed, H. M. (2005). Numerical method for some nonlinear stochastic differential equations, *J. KSIAM* **9**(1): 79–90.
- El-Khatib, Y. & Privault, N. (2004). Computations of Greeks in a market with jumps via the Malliavin calculus, *Finance Stoch.* **8**: 161–179.
- Elerain, O., Chib, S. & Shephard, N. (2001). Likelihood inference for discretely observed non-linear diffusions, *Econometrica* **69**: 959–993.
- Elliott, R. J. (1982). *Stochastic Calculus and Applications*, Springer.
- Elliott, R. J., Aggoun, L. & Moore, J. B. (1995). *Hidden Markov Models: Estimation and Control*, Vol. 29 of *Appl. Math.*, Springer.
- Elliott, R. J., Fischer, P. & Platen, E. (1999a). Filtering and parameter estimation for a mean reverting interest rate model, *Canad. Appl. Math. Quarterly* **7**(4): 381–400.
- Elliott, R. J., Fischer, P. & Platen, E. (1999b). Hidden Markov filtering for a mean reverting interest rate model, *38th IEEE CDC meeting in Phoenix*, IEEE Control Systems Society, pp. 2782–2787.
- Elliott, R. J. & Glowinski, R. (1989). Approximations to solutions of the Zakai filtering equation, *Stochastic Anal. Appl.* **7**: 145–168.
- Elliott, R. J. & Kohlmann, M. (1988). A short proof of a martingale representation result, *Statist. Probab. Lett.* **6**: 327–329.
- Elliott, R. J. & Kopp, P. E. (2005). *Mathematics of Financial Markets*, Springer Finance, 2nd edn, Springer, New York.
- Elliott, R. J. & Platen, E. (2001). Hidden Markov chain filtering for generalised Bessel processes, *Stochastics in Finite and Infinite Dimensions*, Birkhäuser, pp. 123–143.
- Elliott, R. J. & van der Hoek, J. (1997). An application of hidden Markov models to asset allocation problems, *Finance Stoch.* **1**: 229–238.
- Elliott, R. J. & van der Hoek, J. (2006). *Binomial Models in Finance*, Vol. 14 of *Springer Finance*, Springer.
- Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events. For Insurance and Finance*, Vol. 33 of *Appl. Math.*, Springer.
- Engel, D. (1982). The multiple stochastic integral, *Mem. Amer. Math. Soc.* **38**: 265.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* **50**(4): 987–1007.
- Eraker, B. (2001). MCMC analysis of diffusion models with application to finance, *J. Bus. Econom. Statist.* **19**(2): 177–191.
- Eraker, B., Johannes, M. & Polson, N. (2003). The impact of jumps in equity index volatility and returns, *J. Finance* **58**: 1269–1300.
- Ermakov, S. M. (1975). *Die Monte Carlo-Methode und verwandte Fragen*, Hochschulbücher für Mathematik, Band 72, VEB Deutscher Verlag der Wissenschaften, Berlin. (in German) Translated from Russian by E. Schincke and M. Schleiff.

- Ermakov, S. M. & Mikhailov, G. A. (1982). *Statistical Modeling*, 2nd edn, Nauka, Moscow.
- Evans, L. C. (1999). *Partial Differential Equations*, Vol. 19 of *Graduate Studies in Mathematics*, American Mathematical Society. Second printing.
- Fahrmeir, L. (1974). Schwache Konvergenz gegen Diffusionsprozesse, *Z. Angew. Math. Mech.* **54**: 245.
- Fahrmeir, L. (1976). Approximation von stochastischen Differenzialgleichungen auf Digital- und Hybridrechnern, *Computing* **16**(4): 359–371.
- Feng, J. F., Lei, G. Y. & Qian, M. P. (1992). Second order methods for solving stochastic differential equations, *J. Comput. Math.* **10**(4): 376–387.
- Fernholz, E. R. (2002). *Stochastic Portfolio Theory*, Vol. 48 of *Appl. Math.*, Springer.
- Ferrante, M., Kohatsu-Higa, A. & Sanz-Solé, M. (1996). Strong approximations for stochastic differential equations with boundary conditions, *Stochastic Process. Appl.* **61**: 323–337.
- Filipović, D. (2001). A general characterization of one factor affine term structure models, *Finance Stoch.* **5**: 389–412.
- Filipović, D. & Platen, E. (2009). Consistent market extensions under the benchmark approach, *Math. Finance* **19**(1): 41–52.
- Fischer, P. & Platen, E. (1999). Applications of the balanced method to stochastic differential equations in filtering, *Monte Carlo Methods Appl.* **5**(1): 19–38.
- Fischer, P., Platen, E. & Rungaldier, W. J. (1999). Risk minimizing hedging strategies under partial observation, *Seminar on Stochastic Analysis, Random Fields and Applications*, Vol. 45 of *Progr. Probab.*, Birkhäuser, pp. 175–188.
- Fishman, G. S. (1996). *Monte Carlo: Concepts, Algorithms and Applications*, Springer Ser. Oper. Res., Springer.
- Fleming, W. H. & Sheu, S.-J. (1999). Optimal long term growth rate of expected utility of wealth, *Ann. Appl. Probab.* **9**: 871–903.
- Fleury, G. & Bernard, P. (2001). Convergence of numerical schemes for stochastic differential equations, *Monte Carlo Methods Appl.* **7**: 35–44.
- Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes, *Stochastics* **20**(4): 547–557.
- Föllmer, H. (1991). Probabilistic aspects of options. Nevanlinna Institute Research Report, (working paper).
- Föllmer, H. & Schweizer, M. (1991). Hedging of contingent claims under incomplete information, in M. H. A. Davis & R. J. Elliott (eds), *Applied Stochastic Analysis*, Vol. 5 of *Stochastics Monogr.*, Gordon and Breach, London/New York, pp. 389–414.
- Föllmer, H. & Schweizer, M. (1993). A microeconomic approach to diffusion models for stock prices, *Math. Finance* **3**: 1–23.
- Fosdick, L. D. & Jordan, H. F. (1968). Approximation of a conditional Wiener integral, *J. Comput. Phys.* **3**: 1–16.
- Fouque, J. P. & Tullie, T. A. (2002). Variance reduction for Monte Carlo simulation in a stochastic volatility environment, *Quant. Finance* **2**: 24–30.
- Fournie, E., Lasry, J. M., Lebuchoux, J., Lions, P. L. & Touzi, N. (1999). Applications of Malliavin calculus to Monte Carlo methods in finance, *Finance Stoch.* **3**(4): 391–412.
- Fournie, E., Lasry, J. M., Lebuchoux, J., Lions, P. L. & Touzi, N. (2001). Applications of Malliavin calculus to Monte Carlo methods in finance II, *Finance Stoch.* **5**: 201–236.

- Fournie, E., Lasry, J. M. & Touzi, N. (1997). Monte Carlo methods for stochastic volatility models, *Numerical Methods in Finance*, Cambridge Univ. Press, Cambridge, pp. 146–164.
- Fournie, E., Lebuchoux, J. & Touzi, N. (1997). Small noise expansion and importance sampling, *Asymptot. Anal.* **14**(4): 331–376.
- Franklin, J. N. (1965). Difference methods for stochastic ordinary differential equations, *Math. Comp.* **19**: 552–561.
- Frey, R. & Runggaldier, W. J. (2001). A nonlinear filtering approach to volatility estimation with a view towards high frequency data, *Int. J. Theor. Appl. Finance* **4**(2): 199–210.
- Friedman, A. (1975). *Stochastic Differential Equations and Applications, Vol. I*, Vol. 28 of *Probability and Mathematical Statistics*, Academic Press, New York.
- Fu, M. C. (1995). Pricing of financial derivatives via simulation, in C. Alexopoulos, K. Kang, W. Lilegdon & D. Goldsman (eds), *Proceedings of the 1995 Winter Simulation Conference*, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Fujisaki, M., Kallianpur, G. & Kunita, H. (1972). Stochastic differential equations for the non-linear filtering problem, *Osaka J. Math.* **9**: 19–40.
- Gaines, J. G. (1994). The algebra of iterated stochastic integrals, *Stochastics Stochastics Rep.* **49**: 169–179.
- Gaines, J. G. (1995). A basis for iterated stochastic integrals, *Math. Comput. Simulation* **38**(1-3): 7–11.
- Gaines, J. G. & Lyons, T. J. (1994). Random generation of stochastic area integrals, *SIAM J. Appl. Math.* **54**(4): 1132–1146.
- Gaines, J. G. & Lyons, T. J. (1997). Variable step size control in the numerical solution of stochastic differential equations, *SIAM J. Appl. Math.* **57**(5): 1455–1484.
- Galesto, G. & Runggaldier, W. J. (2010). Pricing without equivalent martingale measures under complete and incomplete observation, in C. Chiarella & A. Novikov (eds), *Contemporary Quantitative Finance - Essays in Honour of Eckhard Platen*, Springer.
- Gallant, A. R. & Tauchen, G. (1996). Which moments to match?, *Econometric Theory* **12**: 657–681.
- Gard, T. C. (1988). *Introduction to Stochastic Differential Equations*, Marcel Dekker, New York.
- Gardoù, A. (2004). The order of approximations for solutions of Itô-type stochastic differential equations with jumps, *Stochastic Anal. Appl.* **22**(3): 679–699.
- Gardoù, A. (2006). The order 1.5 approximation for solutions of jump-diffusion equations, *Stochastic Anal. Appl.* **24**(3): 1147–1168.
- Gelbrich, M. (1989). *L^p -Wasserstein-Metriken und Approximation stochastischer Differentialgleichungen*, PhD thesis, Dissert. A, Humboldt-Universität Berlin.
- Gelbrich, M. & Rachev, S. T. (1996). Discretization for stochastic differential equations, L^p Wasserstein metrics, and econometrical models, *Distributions with fixed marginals and related topics*, Vol. 28 of *IMS Lecture Notes Monogr. Ser.*, Inst. Math. Statist., Hayward, CA, pp. 97–119.
- Geman, H., Madan, D. & Yor, M. (2001). Asset prices are Brownian motion: Only in business time, *Quantitative Analysis in Financial Markets*, World Sci. Publishing, pp. 103–146.
- Geman, H. & Roncoroni, A. (2006). Understanding the fine structure of electricity prices, *J. Business* **79**(3): 1225–1261.

- Geman, S., El Karoui, N. & Rochet, J. C. (1995). Changes of numeraire, changes of probability measures and pricing of options, *J. Appl. Probab.* **32**: 443–458.
- Genon-Catalot, V., Jeantheau, T. & Laredo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications, *Bernoulli* **6**: 1051–1079.
- Gerardi, A., Marchetti, F. & Rosa, A. M. (1984). Simulation of diffusions with boundary conditions, *Systems Control Lett.* **4**: 253.
- Gikhman, I. I. & Skorokhod, A. V. (1979). *The Theory of Stochastic Processes*, Vol. I-III, Springer.
- Giles, M. B. (2007). Improved multi-level Monte Carlo convergence using Milstein scheme, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pp. 343–358.
- Giles, M. B. (2008). Multi-level Monte Carlo path simulation, *Oper. Res.* **56**: 607–617.
- Giles, M. B. & Glasserman, P. (2006). Smoking adjoints: Fast Monte Carlo Greeks, *Risk* **19**: 88–92.
- Giles, M. B., Higham, D. J. & Mao, X. (2009). Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff, *Finance Stoch.* **13**(3): 403–413.
- Gladyshev, S. A. & Milstein, G. N. (1984). The Runge-Kutta method for calculation of Wiener integrals of functionals of exponential type, *Zh. Vychisl. Mat. i. Mat. Fiz.* **24**(8): 1136–1149. (in Russian).
- Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*, Vol. 53 of *Appl. Math.*, Springer.
- Glasserman, P., Heidelberger, P. & Shahabuddin, P. (2000). Variance reduction techniques for estimating value-at-risk, *Management Science* **46**: 1349–1364.
- Glasserman, P. & Kou, S. G. (2003). The term structure of simple forward rates with jump risk, *Math. Finance* **13**(3): 383–410.
- Glasserman, P. & Merener, N. (2003a). Convergence of a discretization scheme for jump-diffusion processes with state-dependent intensities, *Proceedings of the Royal Society*, Vol. 460, pp. 111–127.
- Glasserman, P. & Merener, N. (2003b). Numerical solution of jump-diffusion LIBOR market models, *Finance Stoch.* **7**(1): 1–27.
- Glasserman, P. & Zhao, X. (2000). Arbitrage-free discretization of lognormal forward LIBOR and swap rate models, *Finance Stoch.* **4**: 35–68.
- Gleser, L. J. (1976). A canonical representation for the noncentral Wishart distribution useful for simulation, *J. Amer. Statist. Assoc.* **71**(355): 690–695.
- Glorennec, P. Y. (1977). Estimation a priori des erreurs dans la résolution numérique d'équations différentielles stochastiques, *Séminaire de Probabilités, Univ. Rennes* **1**: 57–93.
- Gobet, E. & Kohatsu-Higa, A. (2003). Computation of Greeks for barrier and look-back options using Malliavin calculus, *Electronic Commun. in Prob.* **8**: 51–62.
- Gobet, E., Lemor, J. P. & Warin, X. (2005). A regression-based Monte-Carlo method to solve backward stochastic differential equations, *Ann. Appl. Probab.* **15**(3): 2172–2002.
- Godambe, V. P. & Heyde, C. C. (1987). Quasi-likelihood and optimal estimation, *Internat. Statist. Rev.* **55**: 231–244.
- Golec, J. & Ladde, G. S. (1989). Euler-type approximation for systems of stochastic differential equations, *J. Appl. Math. Simul.* **2**(4): 239–249.
- Golec, J. & Ladde, G. S. (1993). On an approximation method for a class of stochastic singularly perturbed systems, *Dynam. Systems Appl.* **2**(1): 11–20.
- Gombani, A. & Runggaldier, W. J. (2001). A filtering approach to pricing in multifactor term structure models, *Int. J. Theor. Appl. Finance* **4**(2): 303–320.

- Goodlett, S. T. & Allen, E. J. (1994). A variance reduction technique for use with the extrapolated Euler method for numerical solution of stochastic differential equations, *Stochastic Anal. Appl.* **12**(1): 131–140.
- Gorostiza, L. G. (1980). Rate of convergence of an approximate solution of stochastic differential equations, *Stochastics* **3**: 267–276. Erratum in *Stochastics* **4** (1981), 85.
- Gorovoi, V. & Linetsky, V. (2004). Black's model of interest rates as options, eigenfunction expansions and Japanese interest rates, *Math. Finance* **14**(1): 49–78.
- Gouriéroux, C., Monfort, A. & Renault, E. (1993). Indirect inference, *J. Appl. Econometrics* **8**: 85–118.
- Gouriéroux, C. & Sufana, R. (2004). Derivative pricing with multivariate stochastic volatility: Application to credit risk. Working paper CREST.
- Grant, D., Vora, D. & Weeks, D. (1997). Path-dependent options: Extending the Monte Carlo simulation approach, *Management Science* **43**(11): 1589–1602.
- Grasselli, M. & Tebaldi, C. (2008). Solvable affine term structure models, *Math. Finance* **18**: 135–153.
- Grecksch, W. & Anh, V. V. (1998). Approximation of stochastic differential equations with modified fractional Brownian motion, *Z. Anal. Anwendungen* **17**(3): 715–727.
- Grecksch, W. & Anh, V. V. (1999). A parabolic stochastic differential equation with fractional Brownian motion input, *Statist. Probab. Lett.* **41**(4): 337–346.
- Grecksch, W. & Wadewitz, A. (1996). Approximation of solutions of stochastic differential equations by discontinuous Galerkin methods, *Z. Anal. Anwendungen* **15**(4): 901–916.
- Greenside, H. S. & Helfand, E. (1981). Numerical integration of stochastic differential equations, *Bell Syst. Techn. J.* **60**(8): 1927–1940.
- Greiner, A., Strittmatter, W. & Honerkamp, J. (1987). Numerical integration of stochastic differential equations, *J. Statist. Phys.* **51**(1/2): 95–108.
- Grorud, A. & Talay, D. (1990). Approximation of Lyapunov exponents of stochastic differential systems on compact manifolds, *Analysis and Optimization of Systems*, Vol. 144 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 704–713.
- Grorud, A. & Talay, D. (1996). Approximation of Lyapunov exponents of nonlinear stochastic differential equations, *SIAM J. Appl. Math.* **56**(2): 627–650.
- Grüne, L. & Kloeden, P. E. (2001a). Higher order numerical schemes for affinely controlled nonlinear systems, *Numer. Math.* **89**: 669–690.
- Grüne, L. & Kloeden, P. E. (2001b). Pathwise approximation of random ordinary differential equations, *BIT* **41**: 711–721.
- Grüne, L. & Kloeden, P. E. (2006). Higher order numerical approximation of switching systems, *Systems Control Lett.* **55**: 746–754.
- Guerra, M. L. & Sorini, L. (2003). Step control in numerical solution of stochastic differential equations, *Internat. J. Comput. & Num. Analysis & Appl.* **4**(3): 279–295.
- Guo, S. J. (1982). On the mollifier approximation for solutions of stochastic differential equations, *J. Math. Kyoto Univ.* **22**: 243–254.
- Guo, S. J. (1984). Approximation theorems based on random partitions for stochastic differential equations and applications, *Chinese Ann. Math.* **5**: 169–183.
- Guyon, J. (2006). Euler scheme and tempered distributions, *Stochastic Process. Appl.* **116**(6): 877–904.

- Gyöngy, I. (1991). Approximation of Ito stochastic equations, *Math. USSR-Sb* **70**(1): 165–173.
- Gyöngy, I. (1998). A note on Euler's approximations, *Potential Anal.* **8**: 205–216.
- Hammersley, J. M. & Handscomb, D. C. (1964). *Monte Carlo Methods*, Methuen, London.
- Hansen, L. P. (1982). Large sample properties of the generalized method of moments, *Econometrica* **50**: 1029–1054.
- Harris, C. J. (1976). Simulation of nonlinear stochastic equations with applications in modelling water pollution, in C. A. Brebbi (ed.), *Mathematical Models for Environmental Problems*, Pentech Press, London, pp. 269–282.
- Harris, C. J. & Maghsoudi, Y. (1985). Approximate integration of a class of stochastic differential equations, *Control Theory, Proc. 4th IMA Conf.*, Cambridge/Engl., pp. 159–168.
- Harrison, J. M. & Kreps, D. M. (1979). Martingale and arbitrage in multiperiod securities markets, *J. Economic Theory* **20**: 381–408.
- Hausenblas, E. (1999a). A Monte-Carlo method with inherited parallelism for solving partial differential equations with boundary conditions, *Lecture Notes Comput. Sci.*, Vol. 1557, Springer, pp. 117–126.
- Hausenblas, E. (1999b). A numerical scheme using Itô excursion for simulating local time resp. stochastic differential equations with reflection, *Osaka J. Math.* **36**(1): 105–137.
- Hausenblas, E. (2000a). Monte-Carlo simulation of killed diffusion, *Monte Carlo Methods Appl.* **6**(4): 263–295.
- Hausenblas, E. (2000b). A numerical scheme using excursion theory for simulating stochastic differential equations with reflection and local time at a boundary, *Monte Carlo Methods Appl.* **6**(2): 81–103.
- Hausenblas, E. (2000c). A numerical scheme using excursion theory for simulating stochastic differential equations with reflection and local time at a boundary, *Monte Carlo Methods Appl.* **6**: 81–103.
- Hausenblas, E. (2002). Error analysis for approximation of stochastic differential equations driven by Poisson random measures, *SIAM J. Numer. Anal.* **40**(1): 87–113.
- Haussmann, U. (1978). Functionals of Ito processes as stochastic integrals, *SIAM J. Control Optim.* **16**: 252–269.
- Haussmann, U. (1979). On the integral representation of functionals of Ito processes, *Stochastics* **3**: 17–28.
- Haworth, D. C. & Pope, S. B. (1986). A second-order Monte-Carlo method for the solution of the Ito stochastic differential equation, *Stochastic Anal. Appl.* **4**: 151–186.
- Heath, D. (1995). *Valuation of derivative securities using stochastic analytic and numerical methods*, PhD thesis, ANU, Canberra.
- Heath, D. & Platen, E. (2002a). Consistent pricing and hedging for a modified constant elasticity of variance model, *Quant. Finance* **2**(6): 459–467.
- Heath, D. & Platen, E. (2002b). Perfect hedging of index derivatives under a minimal market model, *Int. J. Theor. Appl. Finance* **5**(7): 757–774.
- Heath, D. & Platen, E. (2002c). Pricing and hedging of index derivatives under an alternative asset price model with endogenous stochastic volatility, in J. Yong (ed.), *Recent Developments in Mathematical Finance*, World Scientific, pp. 117–126.

- Heath, D. & Platen, E. (2002d). A variance reduction technique based on integral representations, *Quant. Finance* **2**(5): 362–369.
- Heath, D. & Platen, E. (2005). Currency derivatives under a minimal market model with random scaling, *Int. J. Theor. Appl. Finance* **8**(8): 1157–1177.
- Heath, D., Platen, E. & Schweizer, M. (2001). Numerical comparison of local risk-minimisation and mean-variance hedging, in E. Jouini, J. Cvitanić & M. Musiela (eds), *Option Pricing, Interest Rates and Risk Management*, Handbooks in Mathematical Finance, Cambridge University Press, pp. 509–537.
- Heath, D. & Schweizer, M. (2000). Martingales versus PDEs in finance: An equivalence result with examples, *J. Appl. Probab.* **37**(4): 947–957.
- Helfand, E. (1979). Numerical integration of stochastic differential equations, *Bell Syst. Techn. J.* **58**(10): 2289–2299.
- Hernandez, D. B. (1988). Computational studies in continuous time nonlinear filtering, in C. I. Byrne, C. E. Martin & R. E. Sacks (eds), *Analysis and Control of Nonlinear Systems*, North-Holland, pp. 239–246.
- Hernandez, D. B. & Spigler, R. (1992). A-stability of implicit Runge-Kutta methods for systems with additive noise, *BIT* **32**: 620–633.
- Hernandez, D. B. & Spigler, R. (1993). Convergence and stability of implicit Runge-Kutta methods for systems with multiplicative noise, *BIT* **33**: 654–669.
- Herz, C. S. (1955). Bessel functions of matrix argument, *Ann. of Math.* **61**(3): 474–523.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Rev. Financial Studies* **6**(2): 327–343.
- Heston, S. L. (1995). Discrete versions of continuous-time interest rate models, *J. Fixed Income* pp. 86–88.
- Heston, S. L., Loewenstein, M. & Willard, G. A. (2007). Options and bubbles, *Rev. Financial Studies* **20**(2): 359–390.
- Heston, S. L. & Zhou, G. (2000). On the rate of convergence of discrete-time contingent claims, *Math. Finance* **10**(1): 53–75.
- Heyde, C. C. (1989). Quasi-likelihood and optimality for estimating functions: some current unifying themes, *Bull. Inst. Internat. Statist.* **53**(1): 19–29.
- Heyde, C. C. (1997). *Quasi-Likelihood and Its Application. A General Approach to Optimal Parameter Estimation*, Springer Ser. Statist., Springer.
- Higham, D. J. (2000). Mean-square and asymptotic stability of numerical methods for stochastic ordinary differential equations, *SIAM J. Numer. Anal.* **38**: 753–769.
- Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations, *SIAM Rev.* **43**(3): 525–546.
- Higham, D. J. (2004). *An Introduction to Financial Option Valuation: Mathematics, Stochastics and Computation*, Cambridge University Press.
- Higham, D. J. & Kloeden, P. E. (2005). Numerical methods for nonlinear stochastic differential equations with jumps, *Numer. Math.* **110**(1): 101–119.
- Higham, D. J. & Kloeden, P. E. (2006). Convergence and stability of implicit methods for jump-diffusion systems, *Internat. J. of Numer. Analysis & Modeling.* **3**(2): 125–140.
- Higham, D. J. & Kloeden, P. E. (2007). Strong convergence rates for backward Euler on a class of nonlinear jump-diffusion problems, *J. Comput. Appl. Math.* **205**: 949–956.

- Higham, D. J., Mao, X. & Stuart, A. M. (2002). Strong convergence of Euler-type methods for nonlinear stochastic differential equations, *SIAM J. Numer. Anal.* **40**: 1041–1063.
- Ho, T. S. Y. & Lee, S. B. (1986). Term structure movements and pricing of interest rate claims, *J. Finance* **41**: 1011–1029.
- Hofmann, N. (1994). *Beiträge zur schwachen Approximation stochastischer Differentialgleichungen*, PhD thesis, Dissertation A, Humboldt Universität Berlin.
- Hofmann, N. (1995). Stability of weak numerical schemes for stochastic differential equations, *Math. Comput. Simulation* **38**(1-3): 63–68.
- Hofmann, N. & Mathé, P. (1997). On quasi-Monte Carlo simulation of stochastic differential equations, *Math. Comp.* **66**(218): 573–589.
- Hofmann, N., Müller-Gronbach, T. & Ritter, K. (2000a). Optimal approximation of stochastic differential equations by adaptive step-size control, *Math. Comp.* **69**: 1017–1034.
- Hofmann, N., Müller-Gronbach, T. & Ritter, K. (2000b). Step-size control for the uniform approximation of systems of stochastic differential equations with additive noise, *Ann. Appl. Probab.* **10**: 616–633.
- Hofmann, N., Müller-Gronbach, T. & Ritter, K. (2001). The optimal discretization of stochastic differential equations, *J. Complexity* **17**: 117–153.
- Hofmann, N. & Platen, E. (1994). Stability of weak numerical schemes for stochastic differential equations, *Comput. Math. Appl.* **28**(10-12): 45–57.
- Hofmann, N. & Platen, E. (1996). Stability of superimplicit numerical methods for stochastic differential equations, *Fields Inst. Commun.* **9**: 93–104.
- Hofmann, N., Platen, E. & Schweizer, M. (1992). Option pricing under incompleteness and stochastic volatility, *Math. Finance* **2**(3): 153–187.
- Honoré, P. (1998). Pitfalls in estimating jump-diffusion models. Centre for Analytical Finance, Aarhus, (working paper).
- Hu, Y. Z. & Meyer, P. A. (1993). On the approximation of multiple Stratonovich integrals, *Stochastic Processes*, Springer, pp. 141–147.
- Hu, Y. Z., Mohammed, S.-E. A. & Yan, F. (2004). Discrete-time approximations of stochastic differential delay equations: The Milstein scheme, *Ann. Probab.* **32**: 265–314.
- Hull, J. & White, A. (1987). The pricing of options on assets with stochastic volatilities, *J. Finance* **42**: 281–300.
- Hull, J. & White, A. (1988). The use of control variate techniques in option pricing, *J. Financial and Quantitative Analysis* **23**(3): 237–251.
- Hull, J. & White, A. (1990). Pricing interest rate derivative securities, *Rev. Financial Studies* **3**: 573–592.
- Hulley, H. (2010). The economic plausibility of strict local martingales in financial modelling, in C. Chiarella & A. Novikov (eds), *Contemporary Quantitative Finance - Essays in Honour of Eckhard Platen*, Springer.
- Hulley, H., Miller, S. & Platen, E. (2005). Benchmarking and fair pricing applied to two market models, *Kyoto Economic Review* **74**(1): 85–118.
- Hulley, H. & Platen, E. (2008a). Laplace transform identities for diffusions, with applications to rebates and barrier options, in L. Stettner (ed.), *Advances in Mathematical Finance*, Vol. 83 of *Banach Center Publications*, pp. 139–157.
- Hulley, H. & Platen, E. (2008b). A visual classification of local martingales, *Technical report*, University of Technology, Sydney. QFRC Research Paper 238.

- Hulley, H. & Platen, E. (2009). A visual criterion for identifying Itô diffusions as martingales or strict local martingales, *Technical report*, University of Technology, Sydney. QFRC Research Paper 263, in *Trends in Mathematics*, Birkhäuser (to appear).
- Hulley, H. & Schweizer, M. (2010). M^6 - On minimal market models and minimal martingale measures, in C. Chiarella & A. Novikov (eds), *Contemporary Quantitative Finance - Essays in Honour of Eckhard Platen*, Springer.
- Hunter, C. J., Jäckel, P. & Joshi, M. S. (2001). Getting the drift, *Risk* **14**(7): 81–84.
- Iacus, S. M. (2008). *Simulation and Inference for Stochastic Differential Equations With R Examples*, Springer Series in Statistics, Springer.
- Ikeda, N. & Watanabe, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd edn, North-Holland. (first edition (1981)).
- Ingersoll, J. E. (2000). Digital contracts: Simple tools for pricing complex derivatives, *The Journal of Business* **73**(1): 67–88.
- Jäckel, P. (2002). *Monte Carlo Methods in Finance*, Wiley.
- Jackwerth, J. C. (1997). Generalized binomial trees, *J. Derivatives* **5**(2): 7–17.
- Jacod, J., Kurtz, T. G., Méléard, S. & Protter, P. (2005). The approximate Euler method for Lévy driven stochastic differential equations, *Ann. Inst. H. Poincaré Probab. Statist.* **41**(3): 523–558.
- Jacod, J. & Protter, P. (1998). Asymptotic error distribution for the Euler method for stochastic differential equations, *Ann. Probab.* **26**(1): 267–307.
- Jacod, J. & Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd edn, Springer.
- Janicki, A. (1996). *Numerical and Statistical Approximation of Stochastic Differential Equations with Non-Gaussian Measures*, H. Steinhaus Center for Stochastic Methods in Science and Technology, Wrocław, Poland.
- Janicki, A., Michna, Z. & Weron, A. (1996). Approximation of stochastic differential equations driven by α -stable Lévy motion, *Appl. Math. (Warsaw)* **24**(2): 149–168.
- Janicki, A. & Weron, A. (1994). *Simulation and Chaotic Behavior of α -stable Stochastic Processes*, Vol. 178 of *Monogr. Textbooks Pure Appl. Math.*, Marcel Dekker, New York.
- Jansons, K. M. & Lythe, G. D. (2000). Efficient numerical solution of stochastic differential equations using exponential timestepping, *J. Statist. Phys.* **100**(5/6): 1097–1109.
- Janssen, R. (1984a). Difference-methods for stochastic differential equations with discontinuous coefficients, *Stochastics* **13**: 199–212.
- Janssen, R. (1984b). Discretization of the Wiener process in difference methods for stochastic differential equations, *Stochastic Process. Appl.* **18**: 361–369.
- Jarrow, R., Lando, D. & Turnbull, S. (1997). A Markov model for the term structure of credit risk spreads, *Rev. Financial Studies* **10**(2): 481–523.
- Jarrow, R., Protter, P. & Shimbo, K. (2007a). Asset price bubbles in complete markets, *Advances in Mathematical Finance*, Birkhäuser, Boston, pp. 97–121.
- Jarrow, R., Protter, P. & Shimbo, K. (2007b). Asset price bubbles in incomplete markets. Johnson School Research Paper Series, No. 03-07, Cornell University, June.
- Jarrow, R. & Rudd, A. (1983). *Option Pricing*, Irwin.
- Jeanblanc, M., Yor, M. & Chesney, M. (2009). *Mathematical Methods for Financial Markets*, Springer Finance, Springer.

- Jensen, B. (2000). Binomial models for the term structure of interest. Department of Finance, Copenhagen Business School (working paper).
- Jiang, G. J. & Knight, J. L. (2002). Estimation of continuous-time processes via the empirical characteristic function, *J. Bus. Econom. Statist.* **20**(2): 198–212.
- Jimenez, J. C., Biscay, R. & Ozaki, T. (2006). Inference methods for discretely observed continuous-time stochastic volatility models: A commented overview, *Asia-Pacific Financial Markets* **12**: 109–141.
- Jimenez, J. C. & Ozaki, T. (2003). Local linearization filters for nonlinear continuous-discrete state space models with multiplicative noise, *Int. J. Control.* **76**: 1159–1170.
- Jimenez, J. C. & Ozaki, T. (2006). An approximate innovation method for the estimation of diffusion processes from discrete data, *J. Time Series Analysis* **27**: 77–97.
- Jódar, L., Cortés, J. C., Sevilla-Peris, P. & Villafuerte, L. (2005). Approximating processes of stochastic diffusion models under spatial uncertainty, *Proc. ASMDA, 16-21 May, Brest, France*, pp. 1303–1312.
- Johannes, M. (2004). The statistical and economic role of jumps in continuous-time interest rate models, *J. Finance* **59**(1): 227–260.
- Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2 of *Wiley Series in Probability and Mathematical Statistics*, 2nd edn, John Wiley & Sons.
- Jorion, P. (1988). On jump processes in the foreign exchange and stock markets, *Rev. Financial Studies* **1**: 427–445.
- Joshi, M. S. (2003). *The Concept and Practice of Mathematical Finance*, Cambridge University Press.
- Joshi, M. S. & Stacey, A. M. (2008). New and robust drift approximations for the LIBOR market model, *Quant. Finance* **8**: 427–434.
- Joy, C., Boyle, P. P. & Tan, K. S. (1996). Quasi Monte Carlo methods in numerical finance, *Management Science* **42**(6): 926–938.
- Kahl, C. & Jäckel, P. (2006). Fast strong approximation Monte-Carlo schemes for stochastic volatility models, *Quant. Finance* **6**(6): 513–536.
- Kahl, C. & Schurz, H. (2006). Balanced Milstein methods for ordinary SDEs, *Monte Carlo Methods Appl.* **12**(2): 143–170.
- Kallianpur, G. (1980). *Stochastic Filtering Theory*, Springer.
- Kalman, R. E. & Bucy, R. S. (1961). New results in linear filtering and prediction theory, *Trans. ASME Ser. D. J. Basic Engrg.* **83**: 95–108.
- Kalos, M. H. (1984). *Monte Carlo Methods in Quantum Problems*, Vol. 125 of *NATO ASI, Series C*, Reidel, Boston.
- Kalos, M. H. & Whitlock, P. A. (1986). *Monte Carlo Methods. Vol. I. Basics*, Wiley, New York.
- Kampen, J., Kolodko, A. & Schoenmakers, J. (2008). Monte Carlo Greeks for financial products via approximative transition densities, *SIAM J. Sci. Comput.* **31**(1): 1–22.
- Kamrad, B. & Ritchken, P. (1991). Multinomial approximating models for options with k state variables, *Management Science* **37**: 1640–1652.
- Kanagawa, S. (1985). The rate of convergence in Maruyama's invariance principle, *Proc. 4th Inter. Vilnius Confer. Prob. Theory Math. Statist.*, pp. 142–144. Vilnius 1985.
- Kanagawa, S. (1988). The rate of convergence for Maruyama's approximate solutions of stochastic differential equations, *Yokohoma Math. J.* **36**: 79–85.

- Kanagawa, S. (1989). The rate of convergence for approximate solutions of stochastic differential equations, *Tokyo J. Math.* **12**: 33–48.
- Kanagawa, S. (1995). Error estimation for the Euler-Maruyama approximate solutions of stochastic differential equations, *Monte Carlo Methods Appl.* **1**(3): 165–171.
- Kanagawa, S. (1996). Convergence rates for the Euler-Maruyama type approximate solutions of stochastic differential equations, *Probability Theory and Mathematical Statistics*, World Scientific, Singapore, pp. 183–192. Proc. Seventh Japan Russia Symposium.
- Kanagawa, S. (1997). Confidence intervals of discretized Euler-Maruyama approximate solutions of SDEs, *Nonlinear Anal.* **30**(7): 4101–4104.
- Kanagawa, S. & Saisho, Y. (2000). Strong approximation of reflecting Brownian motion using penalty method and its application to computer simulation, *Monte Carlo Methods Appl.* **6**(2): 105–114.
- Kaneko, T. & Nakao, S. (1988). A note on approximations for stochastic differential equations, *Séminaire de Probabilités XXII*, Vol. 1321 of *Lecture Notes in Math.*, Springer, pp. 155–162.
- Kannan, D. & Zhang, Q. (1998). Nonlinear filtering of an interactive multiple model with small observation noise: Numerical methods, *Stochastic Anal. Appl.* **16**(4): 631–659.
- Karatzas, I. & Kardaras, C. (2007). The numeraire portfolio in semimartingale financial models, *Finance Stoch.* **11**(4): 447–493.
- Karatzas, I. & Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd edn, Springer.
- Karatzas, I. & Shreve, S. E. (1998). *Methods of Mathematical Finance*, Vol. 39 of *Appl. Math.*, Springer.
- Kardaras, C. & Platen, E. (2008). Minimizing the expected market time to reach a certain wealth level, *Technical report*, University of Technology, Sydney. QFRC Research Paper 230, to appear in *SIAM J. Financial Mathematics*.
- Kaufmann, R., Gadmer, A. & Klett, R. (2001). Introduction to dynamic financial analysis, *ASTIN Bulletin* **31**(1): 213–249.
- Kazimierczyk, P. (1989). Consistent ML estimator for drift parameters of both ergodic and nonergodic diffusions, *Stochastic Systems and Optimization*, Vol. 136 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 318–327.
- Kebaier, A. (2005). Statistical Romberg extrapolation: A new variance reduction method and applications to option pricing, *Ann. Appl. Probab.* **15**(4): 2681–2705.
- Kebaier, A. & Kohatsu-Higa, A. (2008). An optimal control variance reduction method for density estimation, *Stochastic Process. Appl.* **118**(12): 2143–2180.
- Kelly, J. R. (1956). A new interpretation of information rate, *Bell Syst. Techn. J.* **35**: 917–926.
- Kelly, L., Platen, E. & Sørensen, M. (2004). Estimation for discretely observed diffusions using transform functions, *Stochastic Methods and Their Applications*, Vol. 41A of *J. Appl. Probab.*, Applied Prob. Trust, pp. 99–118.
- Kendall, W. S. (1989). The diffusion of Euclidean shape., *Technical report*, Research Report 161, University of Warwick, Dept. of Statistics.
- Kessler, M. & Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process, *Bernoulli* **5**: 299–314.
- Kim, C., Lee, E. K., Hänggi, P. & Talkner, P. (2007). Numerical methods for stochastic differential equations with Poisson white shot noise, *Phys. Rev. E* p. 011109.

- Klauder, J. R. & Petersen, W. P. (1985a). Numerical integration of multiplicative-noise stochastic differential equations, *SIAM J. Numer. Anal.* **6**: 1153–1166.
- Klauder, J. R. & Petersen, W. P. (1985b). Spectrum of certain non-self-adjoint operators and solutions of Langevin equations with complex drift, *J. Statist. Phys.* **39**: 53–72.
- Klebaner, F. (2005). *Introduction to Stochastic Calculus with Applications*, Imperial College Press. Second edition.
- Kloeden, P. E. (2002). The systematic derivation of higher order numerical schemes for stochastic differential equations, *Milan Journal of Mathematics* **70**(1): 187–207.
- Kloeden, P. E. & Jentzen, A. (2007). Pathwise convergent higher order numerical schemes for random ordinary differential equations, *Proc. Roy. Soc. London A* **463**: 2929–2944.
- Kloeden, P. E. & Neuenkirch, A. (2007). The pathwise convergence of approximation schemes for stochastic differential equations, *LMS J. Comp. Math.* **10**: 235–253.
- Kloeden, P. E. & Pearson, R. A. (1977). The numerical solution of stochastic differential equations, *J. Austral. Math. Soc. Ser. B* **20**: 8–12.
- Kloeden, P. E. & Platen, E. (1991a). Relations between multiple Ito and Stratonovich integrals, *Stochastic Anal. Appl.* **9**(3): 311–321.
- Kloeden, P. E. & Platen, E. (1991b). Stratonovich and Ito stochastic Taylor expansions, *Math. Nachr.* **151**: 33–50.
- Kloeden, P. E. & Platen, E. (1992). Higher order implicit strong numerical schemes for stochastic differential equations, *J. Statist. Phys.* **66**(1/2): 283–314.
- Kloeden, P. E. & Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations*, Vol. 23 of *Appl. Math.*, Springer. Third printing, (first edition (1992)).
- Kloeden, P. E., Platen, E. & Hofmann, N. (1992). Stochastic differential equations: Applications and numerical methods, *Proc. 6th IAHR International Symposium on Stochastic Hydraulics*, pp. 75–81. National Taiwan University, Taipeh.
- Kloeden, P. E., Platen, E. & Hofmann, N. (1995). Extrapolation methods for the weak approximation of Ito diffusions, *SIAM J. Numer. Anal.* **32**(5): 1519–1534.
- Kloeden, P. E., Platen, E. & Schurz, H. (1992). Effective simulation of optimal trajectories in stochastic control, *Optimization* **1**: 633–644.
- Kloeden, P. E., Platen, E. & Schurz, H. (1993). Higher order approximate Markov chain filters, in S. Cambanis (ed.), *Stochastic Processes - A Festschrift in Honour of Gopinath Kallianpur*, Springer, pp. 181–190.
- Kloeden, P. E., Platen, E. & Schurz, H. (2003). *Numerical Solution of SDEs Through Computer Experiments*, Universitext, Springer. Third corrected printing, (first edition (1994)).
- Kloeden, P. E., Platen, E., Schurz, H. & Sørensen, M. (1996). On effects of discretization on estimators of drift parameters for diffusion processes, *J. Appl. Probab.* **33**: 1061–1076.
- Kloeden, P. E., Platen, E. & Wright, I. (1992). The approximation of multiple stochastic integrals, *Stochastic Anal. Appl.* **10**(4): 431–441.
- Klüppelberg, C., Lindner, A. & Maller, R. (2006). Continuous time volatility modelling: COGARCH versus Ornstein-Uhlenbeck models, in Y. Kabanov et al. (ed.), *From Stochastic Calculus to Mathematical Finance*, Springer, pp. 393–419.
- Kohatsu-Higa, A. (1997). High order Ito-Taylor approximations to heat kernels, *J. Math. Kyoto Univ.* **37**(1): 129–150.

- Kohatsu-Higa, A. (2001). Weak approximations. A Malliavin calculus approach, *Math. Comp.* **70**: 135–172.
- Kohatsu-Higa, A. & Ogawa, S. (1997). Weak rate of convergence for an Euler scheme of nonlinear SDEs, *Monte Carlo Methods Appl.* **3**(4): 327–345.
- Kohatsu-Higa, A. & Petterson, R. (2002). Variance reduction methods for simulation of densities on the Wiener space, *SIAM J. Numer. Anal.* **40**: 431–450.
- Kohatsu-Higa, A. & Protter, P. (1994). The Euler scheme for SDEs driven by semimartingales, in H. Kunita & H. H. Kuo (eds), *Stochastic Analysis on Infinite Dimensional Spaces*, Pitman, pp. 141–151.
- Kohler, W. E. & Boyce, W. E. (1974). A numerical analysis of some first order stochastic initial value problems, *SIAM J. Appl. Math.* **27**: 167–179.
- Komori, Y. (2007). Multi-colored rooted tree analysis of the weak order conditions of a stochastic Runge-Kutta family, *Appl. Numer. Math.* **57**: 147–165.
- Komori, Y. & Mitsui, T. (1995). Stable ROW-type weak scheme for stochastic differential equations, *Monte Carlo Methods Appl.* **1**(4): 279–300.
- Komori, Y., Mitsui, T. & Sugiura, H. (1997). Rooted tree analysis of the order conditions of ROW-type scheme for stochastic differential equations, *BIT* **37**(1): 43–66.
- Komori, Y., Saito, Y. & Mitsui, T. (1994). Some issues in discrete approximate solution for stochastic differential equations, *Comput. Math. Appl.* **28**(10-12): 269–278.
- Konakov, V. & Mammen, E. (2002). Edgeworth type expansions for Euler schemes for stochastic differential equations, *Monte Carlo Methods Appl.* **8**(3): 271–285.
- Kou, S. G. (2002). A jump diffusion model for option pricing, *Management Science* **48**: 1086–1101.
- Kozin, F. (1983). Estimation of parameters for systems driven by white noise, in K. Hennig (ed.), *Random Vibrations and Reliability*, CISM Lecture Notes, Akademie Verlag, Berlin, pp. 163–172.
- Kozlov, R. I. & Petryakov, M. G. (1986). The construction of comparison systems for stochastic differential equations and numerical methods, *Differential equations and numerical methods*, Nauka Sibirsk Otdel. Novosibirsk, pp. 45–52. (in Russian).
- Krylov, N. V. (1980). *Controlled Diffusion Processes*, Vol. 14 of *Appl. Math.*, Springer.
- Kubilius, K. & Platen, E. (2002). Rate of weak convergence of the Euler approximation for diffusion processes with jumps, *Monte Carlo Methods Appl.* **8**(1): 83–96.
- Küchler, U. & Platen, E. (2000). Strong discrete time approximation of stochastic differential equations with time delay, *Math. Comput. Simulation* **54**: 189–205.
- Küchler, U. & Platen, E. (2002). Weak discrete time approximation of stochastic differential equations with time delay, *Math. Comput. Simulation* **59**: 497–507.
- Küchler, U. & Sørensen, M. (1989). Exponential families of stochastic processes: A unifying semimartingale approach, *Internat. Statist. Rev.* **57**: 123–144.
- Küchler, U. & Sørensen, M. (1997). *Exponential Families of Stochastic Processes*, Springer.
- Kurtz, T. G. & Protter, P. (1991a). Weak limit theorems for stochastic integrals and stochastic differential equations, *Ann. Probab.* **19**(3): 1035–1070.
- Kurtz, T. G. & Protter, P. (1991b). Wong-Zakai corrections, random evolutions and simulation schemes for SDEs, in E. M. E. Meyer-Wolf & A. Schwartz (eds), *Stochastic Analysis*, Academic Press, pp. 331–346.

- Kushner, H. J. (1974). On the weak convergence of interpolated Markov chains to a diffusion, *Ann. Probab.* **2**: 40–50.
- Kushner, H. J. (1977). *Probability Methods for Applications in Stochastic Control and for Elliptic Equations*, Academic Press, New York.
- Kushner, H. J. (1984). *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory*, MIT Press, Cambridge MA.
- Kushner, H. J. (2005). Numerical approximations for nonlinear stochastic systems with delays, *Stochastics* **77**(3): 211–240.
- Kushner, H. J. & DiMasi, G. (1978). Approximation for functionals and optimal control on jump diffusion processes, *J. Math. Anal. Appl.* **63**(3): 772–800.
- Kushner, H. J. & Dupuis, P. G. (1992). *Numerical Methods for Stochastic Control Problems in Continuous Time*, Vol. 24 of *Appl. Math.*, Springer.
- Kusuoka, S. (2001). Approximation of expectation of diffusion process and mathematical finance, *Proceedings of Final Taniguchi Symposium, Nara 1998*, Vol. 31 of *Advanced Studies in Pure Mathematics*, pp. 147–165.
- Kusuoka, S. (2004). Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus, *Adv. Math. Econ.* **6**: 69–83.
- Kutoyants, Y. (1984). *Parameter Estimation for Stochastic Processes*, Helderman Verlag Berlin.
- Kuznetsov, D. F. (1998). *Some Questions in the Theory of Numerical Solution of Ito Stochastic Differential Equations*, St.-Petersburg, State Technical University Publisher. (in Russian).
- Kuznetsov, D. F. (1999). *Numerical Modelling of Stochastic Equations and Stochastic Integrals*, St.-Petersburg. (in Russian).
- Kuznetsov, D. F. (2009). *Stochastic Differential Equations: Theory and Practice of Numerical Solution*, 3rd edn, St.-Petersburg, Polytechnical Univ. Press. (in Russian).
- Lamb, H., Mattingly, J. C. & Stuart, A. M. (2003). An adaptive timestepping algorithm for stochastic differential equations, *J. Comput. Appl. Math.* **161**: 417–430.
- Lamb, H., Mattingly, J. C. & Stuart, A. M. (2007). An adaptive Euler-Maruyama scheme for SDEs: Convergence and stability, *IMA J. Numer. Anal.* **27**(3): 479–506.
- Lamberton, D. & Lapeyre, B. (2007). *Introduction to Stochastic Calculus Applied to Finance*, 2nd edn, Chapman & Hall, London. Translation from French.
- Landen, C. (2000). Bond pricing in a hidden Markov model of the short rate, *Finance Stoch.* **4**: 371–389.
- Lanska, V. (1979). Minimum contrast estimation in diffusion processes, *J. Appl. Probab.* **16**: 65–75.
- Law, A. M. & Kelton, W. D. (1991). *Simulation Modeling and Analysis*, 2nd edn, McGraw-Hill, New York.
- Le Breton, A. (1976). On continuous and discrete sampling for parameter estimation in diffusion type processes, *Math. Prog. Study* **5**: 124–144.
- Le Gland, F. (1992). Splitting-up approximation for SPDEs and SDEs with application to nonlinear filtering, *Stochastic partial differential equations and their applications*, Vol. 176 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 177–187.
- Le, T. & Platen, E. (2006). Approximating the growth optimal portfolio with a diversified world stock index, *J. Risk Finance* **7**(5): 559–574.

- L'Ecuyer, P. (1994). Uniform random number generation, *Ann. Operations Research* **53**: 77–120.
- L'Ecuyer, P. (1999). Tables of maximally equidistributed combined LFSR generators, *Math. Comp.* **68**: 261–269.
- Lehn, J., Rössler, A. & Schein, O. (2002). Adaptive schemes for the numerical solution of SDEs - a comparison, *J. Comput. Appl. Math.* **138**: 297–308.
- Leisen, D. & Reimer, M. (1996). Binomial models for option valuation - Examining and improving convergence, *Appl. Math. Finance* **3**: 319–346.
- Leobacher, G. (2006). Stratified sampling and quasi-Monte Carlo simulation of Lévy processes, *Monte Carlo Methods Appl.* **12**: 231–238.
- Leon, J., Sole, J. L., Utzet, F. & Vives, J. (2006). On Lévy processes, Malliavin Calculus and market models with jumps, *Finance Stoch.* **6**: 197–225.
- Lépingle, D. (1995). Euler scheme for reflected stochastic differential equations, *Math. Comput. Simulation* **38**(1-3): 119–126.
- Lépingle, D. & Nguyen, T. T. (2004). Approximating and simulating multivalued stochastic differential equations, *Monte Carlo Methods Appl.* **10**(2): 129–152.
- Lépingle, D. & Ribémont, B. (1991). A multistep approximation scheme for the Langevin equation, *Stochastic Process. Appl.* **37**(1): 61–69.
- Lewis, A. L. (2000). *Option Valuation Under Stochastic Volatility*, Finance Press, Newport Beach.
- Lewis, P. A. W. & Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning, *Naval Research Logistics Quarterly* **26**: 403–413.
- Li, C. W. & Liu, X. Q. (1997). Approximation of multiple stochastic integrals and its application to stochastic differential equations, *Nonlinear Anal.* **30**(2): 697–708.
- Li, C. W. & Liu, X. Q. (2000). Almost sure convergence of the numerical discretisation of stochastic jump diffusions, *Acta Appl. Math.* **62**: 225–244.
- Li, T., Abdulle, A. & Weinan E. (2008). Effectiveness of implicit methods for stiff stochastic differential equations, *Commun. Comput. Phys.* **3**(2): 295–307.
- Linkov, J. N. (1984). On the asymptotic behaviour of likelihood functions for certain problems for semi-martingales, *Teor. Sluch. Proc.* **12**(1): 141–191. (in Russian).
- Liptser, R. & Shiryaev, A. N. (1977). *Statistics of Random Processes: I. General Theory*, Vol. 5 of *Appl. Math.*, Springer.
- Liptser, R. & Shiryaev, A. N. (2001). *Statistics of Random Processes: II. Applications*, Vol. 6 of *Appl. Math.*, 2nd, revised and expanded edn, Springer.
- Liske, H. (1982). Distribution of a functional of a Wiener process, *Theory of Random Processes* **10**: 50–54. (in Russian).
- Liske, H. & Platen, E. (1987). Simulation studies on time discrete diffusion approximations, *Math. Comput. Simulation* **29**(3-4): 253–260.
- Liske, H., Platen, E. & Wagner, W. (1982). About mixed multiple Wiener integrals. Preprint P-Math-23/82, IMath, Akad. der Wiss. der DDR, Berlin.
- Liu, K. & Xia, X. (1999). On the exponential stability in mean square of neutral stochastic functional differential equations, *Systems Control Lett.* **37**: 207–215.
- Liu, M. Z., Cao, W. R. & Fan, Z. C. (2004). Convergence and stability of the semi-implicit Euler method for a linear stochastic differential delay equation, *J. Comput. Appl. Math.* **170**: 255–268.
- Liu, X. Q. & Li, C. W. (2000). Weak approximation and extrapolations of stochastic differential equations with jumps, *SIAM J. Numer. Anal.* **37**(6): 1747–1767.
- Lo, A. W. (1988). Maximum likelihood estimation of generalized Itô processes with discretely sampled data, *Econometric Theory* **4**: 231–247.

- Loewenstein, M. & Willard, G. A. (2000a). Local martingales, arbitrage, and viability: Free snacks and cheap thrills, *Econometric Theory* **16**(1): 135–161.
- Loewenstein, M. & Willard, G. A. (2000b). Rational equilibrium asset-pricing bubbles in continuous trading models, *Econometric Theory* **91**(1): 17–58.
- Løkka, A. (2004). Martingale representation of functionals of Lévy processes, *Stochastic Anal. Appl.* **22**(4): 867–892.
- Long, J. B. (1990). The numeraire portfolio, *J. Financial Economics* **26**: 29–69.
- Longstaff, F. A. (1989). A nonlinear general equilibrium model of the term structure of interest rates, *J. Financial Economics* **23**: 195–224.
- Longstaff, F. A. & Schwartz, E. S. (2001). Valuing American options by simulations: A simple least-squares approach, *Rev. Financial Studies* **14**(1): 113–147.
- Lord, R., Koekkoek, R. & van Dijk, D. (2006). A comparison of biased simulation schemes for stochastic volatility models. Tinbergen Institute Discussion Paper (TI 2006-046/4).
- Lyons, T. J. (1998). Differential equations driven by rough signals, *Rev. Mat. Iberoam.* **14**(2): 215–310.
- Lyons, T. J. & Victoir, N. (2004). Cubature on Wiener space. Stochastic analysis with applications to mathematical finance, *Proceeding of the Royal Society London*, Vol. 460, pp. 169–198.
- Ma, J., Protter, P., San Martin J. & Torres, S. (2002). Numerical method for backward stochastic differential equations, *Ann. Appl. Probab.* **12**(1): 302–316.
- Ma, J., Protter, P. & Yong, J. M. (1994). Solving forward-backward stochastic differential equations explicitly – a four step scheme, *Probab. Theory Related Fields* **98**(3): 339–359.
- Ma, J., Shen, J. & Zhao, Y. (2008). On numerical approximations of forward-backward stochastic differential equations, *SIAM J. Numer. Anal.* **46**: 26–36.
- Mackevicius, V. (1987). S^p -stability of solutions of symmetric stochastic differential equations with discontinuous driving semimartingales, *Ann. Inst. H. Poincaré Probab. Statist.* **23**(4): 575–592.
- Mackevicius, V. (1994). Second order weak approximations for Stratonovich stochastic differential equations, *Liet. Mat. Rink.* **34**(2): 226–247. (transl. *Lithuanian Math. Journal*, **34**(2), 183–200).
- Mackevicius, V. (1996). Extrapolation of approximations of solutions of stochastic differential equations, *Probability Theory and Mathematical Statistics*, World Scientific, Singapore, pp. 276–297.
- Madan, D. & Milne, F. (1991). Option pricing with (V.G.) martingale components, *Math. Finance* **1**: 39–56.
- Madan, D. & Seneta, E. (1990). The variance gamma (V.G.) model for share market returns, *J. Business* **63**: 511–524.
- Maghsoudi, Y. (1996). Mean-square efficient numerical solution of jump-diffusion stochastic differential equations, *SANKHYA A* **58**(1): 25–47.
- Maghsoudi, Y. (1998). Exact solutions and doubly efficient approximations of jump-diffusion Itô equations, *Stochastic Anal. Appl.* **16**(6): 1049–1072.
- Makarov, R. N. (2001). Statistical modelling of solutions of stochastic differential equations with reflection of trajectories from the boundary, *Russian J. Numer. Anal. Math. Modelling* **16**(3): 261–278.
- Makarov, R. N. (2003). Numerical solution of quasilinear parabolic equations and backward stochastic differential equations, *Russian J. Numer. Anal. Math. Modelling* **18**(5): 397–412.

- Malliavin, P. (1997). *Stochastic Analysis*, Springer.
- Malliavin, P. & Thalmaier, A. (2003). Numerical error for SDE: Asymptotic expansion and hyperdistribution, *C. R. Acad. Sci. Paris Sér. I Math* **336**: 851–856.
- Malliavin, P. & Thalmaier, A. (2006). *Stochastic Calculus of Variations in Mathematical Finance*, Springer.
- Maltz, F. H. & Hitzl, D. L. (1979). Variance reduction in Monte-Carlo computations using multi-dimensional Hermite polynomials, *J. Comput. Phys.* **32**: 345–376.
- Mancini, C. (2002). The European options hedge perfectly in a Poisson-Gaussian stock market model, *Appl. Math. Finance* **9**: 87–102.
- Manella, R. & Palleschi, V. (1989). Fast and precise algorithm for computer simulation of stochastic differential equations, *Phys. Rev. A* **40**: 3381–3386.
- Mao, X. (1991). Approximate solution for a class of delay stochastic differential equations, *Stochastics Stochastics Rep.* **35**: 111–123.
- Mao, X. (2007). Exponential stability of equidistant Euler-Maruyama approximations of stochastic differential delay equations, *J. Comput. Appl. Math.* **200**: 297–316.
- Mao, X., Shen, Y. & Yuan, C. (2008). Almost surely asymptotic stability of neutral stochastic differential delay equations with Markovian switching, *Stochastic Process. Appl.* **118**: 1385–1406.
- Mao, X., Yin, G. & Yuan, C. (2007). Stabilization and destabilization of hybrid systems of stochastic differential equations, *Automatica* **43**(2): 264–273.
- Mao, X. & Yuan, C. (2006). *Stochastic Differential Equations with Markovian Switching*, Imperial College Press, London.
- Mao, X., Yuan, C. & Yin, G. (2005). Numerical method for stationary distribution of stochastic differential delay equations with Markovian switching, *J. Comput. Appl. Math.* **174**: 1–27.
- Mao, X., Yuan, C. & Yin, G. (2007). Approximations of Euler-Maruyama type for stochastic differential equations with Markovian switching under non-Lipschitz conditions, *J. Comput. Appl. Math.* **205**: 936–948.
- Marcus, S. I. (1978). Modeling and analysis of stochastic differential equations driven by point processes, *IEEE Trans. Inform. Theory* **24**(2): 164–172.
- Marcus, S. I. (1981). Modeling and approximation of stochastic differential equations driven by semimartingales, *Stochastics* **4**(3): 223–245.
- Margrabe, W. (1978). The value of an option to exchange one asset for another, *The Journal of Finance* **33**(1): 177–186.
- Marsh, T. A. & Rosenfeld, E. R. (1983). Stochastic processes for interest rates and equilibrium bond prices, *J. Financial Economics* **38**: 635–650.
- Marshall, A. W. & Olkin, I. (1988). Families of multivariate distributions, *J. Amer. Statist. Assoc.* **83**: 834–841.
- Maruyama, G. (1955). Continuous Markov processes and stochastic equations, *Rend. Circ. Mat. Palermo* **4**: 48–90.
- Matsumoto, M. & Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Trans. Model. Computer Simul.* **8**(1): 3–30.
- Mauthner, S. (1998). Step size control in the numerical solution of stochastic differential equations, *J. Comput. Appl. Math.* **100**(1): 93–109.
- McLeish, D. L. (2005). *Monte Carlo Simulation and Finance*, Wiley.
- McNeil, A., Frey, R. & Embrechts, P. (2005). *Quantitative Risk Management*, Princeton University Press.

- McNeil, K. J. & Craig, I. J. D. (1988). An unconditional stable numerical scheme for nonlinear quantum optical systems - quantised limit cycle behaviour in second harmonic generation. Univ. Waikato Research Report no. 166.
- Melchior, M. & Öttinger, H. (1995). Variance reduced simulations of stochastic differential equations, *J. Chemical Physics* **103**: 9506–9509.
- Melchior, M. & Öttinger, H. (1996). Variance reduced simulations of polymer dynamics, *J. Chemical Physics* **105**: 3316–3331.
- Merton, R. C. (1973a). An intertemporal capital asset pricing model, *Econometrica* **41**: 867–888.
- Merton, R. C. (1973b). Theory of rational option pricing, *Bell J. Econ. Management Sci.* **4**: 141–183.
- Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous, *J. Financial Economics* **2**: 125–144.
- Merton, R. C. (1992). *Continuous-Time Finance*, Blackwell, Oxford.
- Metwally, S. A. K. & Atiya, A. F. (2002). Using Brownian bridge for fat simulation of jump-diffusion processes and barrier options, *J. Derivatives* **10**: 43–54.
- Mikosch, T. (1998). *Elementary Stochastic Calculus With Finance in View, Statistical Science and Applied Probability Series*, World Scientific, Singapore.
- Mikulevicius, R. (1983). On some properties of solutions of stochastic differential equations, *Liet. Mat. Rink.* **4**: 18–31.
- Mikulevicius, R. & Platen, E. (1988). Time discrete Taylor approximations for Ito processes with jump component, *Math. Nachr.* **138**: 93–104.
- Mikulevicius, R. & Platen, E. (1991). Rate of convergence of the Euler approximation for diffusion processes, *Math. Nachr.* **151**: 233–239.
- Miller, S. & Platen, E. (2005). A two-factor model for low interest rate regimes, *Asia-Pacific Financial Markets* **11**(1): 107–133.
- Miller, S. & Platen, E. (2008). Analytic pricing of contingent claims under the real world measure, *Int. J. Theor. Appl. Finance* **11**(8): 841–867.
- Miller, S. & Platen, E. (2009). Real-world pricing for a modified constant elasticity of variance model, *Appl. Math. Finance*. electronic in iFirst.
- Milstein, G. N. (1974). Approximate integration of stochastic differential equations, *Theory Probab. Appl.* **19**: 557–562.
- Milstein, G. N. (1978). A method of second order accuracy integration of stochastic differential equations, *Theory Probab. Appl.* **23**: 396–401.
- Milstein, G. N. (1985). Weak approximation of solutions of systems of stochastic differential equations, *Theory Probab. Appl.* **30**: 750–766.
- Milstein, G. N. (1988a). *Numerical Integration of Stochastic Differential Equations*, Urals Univ. Press, Sverdlovsk. (in Russian).
- Milstein, G. N. (1988b). A theorem of the order of convergence of mean square approximations of systems of stochastic differential equations, *Theory Probab. Appl.* **32**: 738–741.
- Milstein, G. N. (1995a). *Numerical Integration of Stochastic Differential Equations*, Mathematics and Its Applications, Kluwer.
- Milstein, G. N. (1995b). The solving of boundary value problems by numerical integration of stochastic equations, *Math. Comput. Simulation* **38**(1-3): 77–85.
- Milstein, G. N. (1995c). Solving the first boundary value problem of parabolic type by numerical integration of stochastic differential equations, *Theory Probab. Appl.* **40**: 657–665.

- Milstein, G. N. (1996). Application of numerical integration of stochastic equations for solving boundary value problems with Neumann boundary conditions, *Theory Probab. Appl.* **41**: 210–218.
- Milstein, G. N. (1997). Weak approximation of a diffusion process in a bounded domain, *Stochastics Stochastics Rep.* **62**: 147–200.
- Milstein, G. N. & Platen, E. (1994). The integration of stiff stochastic differential equations with stable second moments, *Technical report*, Australian National University, Canberra, Statistics Report Series. SRR 014-94.
- Milstein, G. N., Platen, E. & Schurz, H. (1998). Balanced implicit methods for stiff stochastic systems, *SIAM J. Numer. Anal.* **35**(3): 1010–1019.
- Milstein, G. N., Repin, Y. M. & Tretjakov, M. V. (2002). Numerical methods for stochastic systems preserving symplectic structure, *SIAM J. Numer. Anal.* **40**: 1583–1604.
- Milstein, G. N. & Schoenmakers, J. (2002). Monte-Carlo construction of hedging strategies against multi-asset European claims, *Stochastics Stochastics Rep.* **73**: 125–157.
- Milstein, G. N. & Tretjakov, M. V. (1994a). Mean-square numerical methods for stochastic differential equations with small noises, *SIAM J. Sci. Comput.* **18**(4): 1067–1087.
- Milstein, G. N. & Tretjakov, M. V. (1994b). Numerical solution of differential equations with colored noise, *J. Statist. Phys.* **77**(3/4): 691–715.
- Milstein, G. N. & Tretjakov, M. V. (1997). Numerical methods in the weak sense for stochastic differential equations with small noise, *SIAM J. Numer. Anal.* **34**(6): 2142–2167.
- Milstein, G. N. & Tretjakov, M. V. (2000). Numerical algorithms for semilinear parabolic equations with small parameter based on approximation of stochastic equations, *Math. Comp.* **69**(229): 237–267.
- Milstein, G. N. & Tretjakov, M. V. (2004). *Stochastic Numerics for Mathematical Physics*, Springer.
- Miltersen, K. R. & Schwartz, E. (1998). Pricing of options on commodity futures with stochastic term structure of convenience yields and interest rates, *J. Financial and Quantitative Analysis* **33**: 33–59.
- Misawa, T. (2001). A Lie algebraic approach to numerical integration of stochastic differential equations, *SIAM J. Sci. Comput.* **23**(3): 866–890.
- Mitsui, T. (1995). Stability analysis of numerical solution of stochastic differential equations, *Kokyuroku* **850**: 124–138.
- Miyahara, Y. & Novikov, A. A. (2002). Geometric Lévy process pricing model, *Proc. Steklov Inst. Math.* **2**(237): 176–191.
- Mora, C. M. (2005). Weak exponential schemes for stochastic differential equations with additive noise, *IMA J. Numer. Anal.* **25**: 486–506.
- Mordecki, E., Szepessy, A., Tempone, R. & Zouraris, G. E. (2008). Adaptive weak approximation of diffusions with jumps, *SIAM J. Numer. Anal.* **46**: 1732–1768.
- Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.
- Müller-Gronbach, T. (1996). Optimal design for approximating the path of a stochastic process, *J. Statist. Plann. Inference* **49**: 371–385.
- Müller-Gronbach, T. (2002). The optimal uniform approximation of systems of stochastic differential equations, *Ann. Appl. Probab.* **12**(2): 664–690.
- Musiela, M. & Rutkowski, M. (2005). *Martingale Methods in Financial Modelling*, Vol. 36 of *Appl. Math.*, 2nd edn, Springer.

- Nakayama, T. (2002). Approximation of BSDEs by stochastic differential equations, *J. Math. Sci. Univ. Tokyo* **9**: 257–277.
- Nelsen, R. (1999). *An Introduction to Copulas*, Springer, New York.
- Nelson, D. B. (1990). ARCH models as diffusion approximations, *J. Econometrics* **45**: 7–38.
- Newton, N. J. (1984). *Discrete Approximations for Markov Chains*, PhD thesis, Univ. London.
- Newton, N. J. (1986a). An asymptotic efficient difference formula for solving stochastic differential equations, *Stochastics* **19**: 175–206.
- Newton, N. J. (1986b). Asymptotically optimal discrete approximations for stochastic differential equations, *Theory and applications of nonlinear control systems*, North-Holland, pp. 555–567.
- Newton, N. J. (1990). An efficient approximation for stochastic differential equations on the partition of symmetrical first passage times, *Stochastics* **29**: 227–258.
- Newton, N. J. (1991). Asymptotically efficient Runge-Kutta methods for a class of Ito and Stratonovich equations, *SIAM J. Appl. Math.* **51**: 542–567.
- Newton, N. J. (1994). Variance reduction for simulated diffusions, *SIAM J. Appl. Math.* **54**(6): 1780–1805.
- Newton, N. J. (1996). Numerical methods for stochastic differential equations, *Z. Angew. Math. Mech.* **76**: 211–214. (Suppl. 3, I–XVI).
- Niederreiter, H. (1988a). Low-discrepancy and low-dispersion sequences, *J. of Number Theory* **30**: 51–70.
- Niederreiter, H. (1988b). Remarks on nonlinear pseudo random numbers, *Metrika* **35**: 321–328. was Niederreiter88.
- Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte-Carlo Methods*, SIAM, Philadelphia, PA.
- Nikitin, N. N. & Razevig, V. D. (1978). Methods of numerical modelling of stochastic differential equations and estimates of their error, *Zh. Vychisl. Mat. i. Mat. Fiz.* **18**(1): 106–117. (in Russian).
- Ninomiya, S. & Victoir, N. (2008). Weak approximation of stochastic differential equations and application to derivative pricing, *Appl. Math. Finance* **15**: 107–121.
- Novikov, A. A. (1972a). On an identity for stochastic integrals, *Theory Probab. Appl.* **17**: 717–720.
- Novikov, A. A. (1972b). Sequential estimation of the parameters of diffusion type processes, *Math. Notes* **12**: 812–818.
- Ocone, D. L. (1984). Malliavin's calculus and stochastic integral representation of functions of diffusion processes, *Stochastics* **12**(3-4): 161–185.
- Ogawa, S. (1995). Some problems in the simulation of nonlinear diffusion processes, *Math. Comput. Simulation* **38**(1-3): 217–223.
- Øksendal, B. & Sulem, A. (2005). *Applied stochastic control of jump-diffusions*, Universitext, Springer.
- Osborne, M. F. M. (1959). Brownian motion in the stock market, *Math. Oper. Res.* **7**: 145–173.
- Ozaki, T. (1992). A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: A local linearization approach, *Statistica Sinica* **2**: 113–135.
- Pagés, G. & Xiao, Y. J. (1997). Sequences with low discrepancy and pseudo-random numbers: Theoretical remarks and numerical tests, *J. Statist. Comput. & Simul.* **56**: 163–183.

- Pang, S., Deng, F. & Mao, X. (2008). Almost sure and moment exponential stability of Euler-Maruyama discretizations for hybrid stochastic differential equations, *J. Comput. Appl. Math.* **213**: 127–141.
- Panneton, F., L'Ecuyer, P. & Matsumoto, M. (2006). Improved long-period generators based on linear recurrences modulo 2, *ACM Transactions on Mathematical Software* **32**(1): 1–16.
- Pardoux, E. & Talay, D. (1985). Discretization and simulation of stochastic differential equations, *Acta Appl. Math.* **3**(1): 23–47.
- Paskov, S. & Traub, J. (1995). Faster valuation of financial derivatives, *J. Portfolio Manag.* pp. 113–120.
- Pearson, N. & Sun, T. S. (1989). A test of the Cox, Ingersoll, Ross model of the term structure of interest rates using the method of moments. Sloan School of Management, Massachusetts Institute of Technology (working paper).
- Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Scand. J. Statist.* **22**: 55–71.
- Pelsser, A. & Vorst, T. (1994). The binomial model and the Greeks, *J. Derivatives* **1**(3): 45–49.
- Petersen, W. P. (1987). Numerical simulation of Ito stochastic differential equations on supercomputers, *Random Media*, Vol. 7 of *IMA Vol. Math. Appl.*, Springer, pp. 215–228.
- Petersen, W. P. (1990). Stability and accuracy of simulations for stochastic differential equations. IPS Research Report No. 90-02, ETH Zürich.
- Petersen, W. P. (1994). Some experiments on numerical simulations of stochastic differential equations and a new algorithm, *J. Comput. Phys.* **113**(1): 75–81.
- Petersen, W. P. (1998). A general implicit splitting for stabilizing numerical simulations of Ito stochastic differential equations, *SIAM J. Numer. Anal.* **35**(4): 1439–1451.
- Petterson, R. (1995). Approximations for stochastic differential equations with reflecting convex boundaries, *Stochastic Process. Appl.* **59**: 295–308.
- Picard, J. (1984). Approximation of nonlinear filtering problems and order of convergence, *Filtering and Control of Random Processes*, Vol. 61 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 219–236.
- Pietersz, R., Pelsser, A. & van Regenmortel, M. (2004). Fast drift approximated pricing in the BGM model, *J. Comput. Finance* **8**: 93–124.
- Pivitt, B. (1999). Accessing the Intel® Random Number Generator with CDSA, *Technical report*, Intel® Platform Security Division.
- Platen, E. (1980a). Approximation of Ito integral equations, *Stochastic Differential Systems*, Vol. 25 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 172–176.
- Platen, E. (1980b). Weak convergence of approximations of Ito integral equations, *Z. Angew. Math. Mech.* **60**: 609–614.
- Platen, E. (1981). A Taylor formula for semimartingales solving a stochastic differential equation, *Stochastic Differential Systems*, Vol. 36 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 157–164.
- Platen, E. (1982a). An approximation method for a class of Itô processes with jump component, *Liet. Mat. Rink.* **22**(2): 124–136.
- Platen, E. (1982b). A generalized Taylor formula for solutions of stochastic differential equations, *SANKHYA A* **44**(2): 163–172.

- Platen, E. (1983). Approximation of first exit times of diffusions and approximate solution of parabolic equations, *Math. Nachr.* **111**: 127–146.
- Platen, E. (1984). Beiträge zur zeitdiskreten Approximation von Itôprozessen. Habilitation, Academy of Sciences, Berlin.
- Platen, E. (1985). On first exit times of diffusions, *Stochastic differential systems*, Vol. 69 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 192–195.
- Platen, E. (1987). Derivative free numerical methods for stochastic differential equations, Vol. 96 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 187–193.
- Platen, E. (1992). Higher-order weak approximation of Itô diffusions by Markov chains, *Probab. Engrg. Inform. Sci.* **6**: 391–408.
- Platen, E. (1995). On weak implicit and predictor-corrector methods, *Math. Comput. Simulation* **38**: 69–76.
- Platen, E. (1999a). An introduction to numerical methods for stochastic differential equations, *Acta Numer.* **8**: 197–246.
- Platen, E. (1999b). A short term interest rate model, *Finance Stoch.* **3**(2): 215–225.
- Platen, E. (2001). A minimal financial market model, *Trends in Mathematics*, Birkhäuser, pp. 293–301.
- Platen, E. (2002). Arbitrage in continuous complete markets, *Adv. in Appl. Probab.* **34**(3): 540–558.
- Platen, E. (2004a). A benchmark framework for risk management, *Stochastic Processes and Applications to Mathematical Finance*, Proceedings of the Ritsumeikan Intern. Symposium, World Scientific, pp. 305–335.
- Platen, E. (2004b). A class of complete benchmark models with intensity based jumps, *J. Appl. Probab.* **41**: 19–34.
- Platen, E. (2004c). Modeling the volatility and expected value of a diversified world index, *Int. J. Theor. Appl. Finance* **7**(4): 511–529.
- Platen, E. (2004d). Simulation methods for stochastic differential equations, *Encyclopedia of Actuarial Science*, Vol. 3, Wiley & Sons, pp. 1159–1164.
- Platen, E. (2005a). An alternative interest rate term structure model, *Int. J. Theor. Appl. Finance* **8**(6): 717–735.
- Platen, E. (2005b). Diversified portfolios with jumps in a benchmark framework, *Asia-Pacific Financial Markets* **11**(1): 1–22.
- Platen, E. (2006). A benchmark approach to asset management, *J. Asset Management* **6**(6): 390–405.
- Platen, E. (2008). Simulation methods for stochastic differential equations, in D. Seese, C. Weinhardt & F. Schlottmann (eds), *Handbook on Information Technology in Finance*, Springer, chapter 21, pp. 501–514.
- Platen, E. (2009). Real world pricing of long term contracts, *Technical report*, University of Technology, Sydney. QFRC Research Paper 262.
- Platen, E. & Heath, D. (2006). *A Benchmark Approach to Quantitative Finance*, Springer Finance, Springer.
- Platen, E. & Rebolledo, R. (1985). Weak convergence of semimartingales and discretization methods, *Stochastic Process. Appl.* **20**: 41–58.
- Platen, E. & Rebolledo, R. (1996). Principles for modelling financial markets, *J. Appl. Probab.* **33**: 163–172.
- Platen, E. & Rendek, R. (2009a). Exact scenario simulation for selected multi-dimensional stochastic processes, *Commun. on Stoch. Analysis* **3**(3): 443–465.

- Platen, E. & Rendek, R. (2009b). Quasi-exact approximation of hidden Markov chain filters, *Technical report*, University of Technology, Sydney. QFRC Research Paper 258.
- Platen, E. & Rendek, R. (2009c). Simulation of diversified portfolios in a continuous financial market, *Technical report*, University of Technology, Sydney. QFRC Research Paper 264.
- Platen, E. & Runggaldier, W. J. (2005). A benchmark approach to filtering in finance, *Asia-Pacific Financial Markets* **11**(1): 79–105.
- Platen, E. & Runggaldier, W. J. (2007). A benchmark approach to portfolio optimization under partial information, *Asia-Pacific Financial Markets* **14**(1-2): 25–43.
- Platen, E. & Shi, L. (2008). On the numerical stability of simulation methods for SDEs, *Technical report*, University of Technology, Sydney. QFRC Research Paper 234.
- Platen, E. & Stahl, G. (2003). A structure for general and specific market risk, *Computational Statistics* **18**(3): 355–373.
- Platen, E. & Wagner, W. (1982). On a Taylor formula for a class of Itô processes, *Probab. Math. Statist.* **3**(1): 37–51.
- Pliska, S. R. (1997). *Introduction to Mathematical Finance - Discrete Time Models*, Blackwell Publishers.
- Potters, M., Bouchaud, J.-P. & Sestovic, D. (2001). Hedged Monte-Carlo: Low variance derivative pricing with objective probabilities, *Physica A* **289**: 517–525.
- Poulsen, R. (1999). Approximate maximum likelihood estimation of discretely observed diffusion processes, *Technical report*, University of Aarhus, Centre for Analytical Finance. Working Paper No. 29.
- Prakasa Rao, B. L. S. (1999). *Statistical Inference for Diffusion Type Processes*, Arnold Publishers, Oxford University Press.
- Press, S. J. (1967). A compound events model for security prices, *J. Business* **40**: 317–335.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2002). *Numerical Recipes in C++: The Art of Scientific Computing*, 2nd edn, Cambridge University Press.
- Protter, P. (1985). Approximations of solutions of stochastic differential equations driven by semimartingales, *Ann. Probab.* **13**: 716–743.
- Protter, P. (2005). *Stochastic Integration and Differential Equations*, 2nd edn, Springer.
- Protter, P. & Talay, D. (1997). The Euler scheme for Lévy driven stochastic differential equations, *Ann. Probab.* **25**(1): 393–423.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edn, Wiley, New York.
- Rao, N. J., Borwanker, J. D. & Ramkrishna, D. (1974). Numerical solution of Itô integral equations, *SIAM J. Control Optim.* **12**(1): 125–139.
- Rascanu, A. & Tudor, C. (1995). Approximation of stochastic equations by the splitting up method, *Qualitative Problems for Differential Equations and Control Theory*, World Sci. Publishing, River Edge, NJ, pp. 277–287.
- Rathinasamy, A. & Balachandran, K. (2008a). Mean-square stability of Milstein method for linear hybrid stochastic delay differential equations, *Nonlinear Anal.* **2**: 1256–1263.

- Rathinasamy, A. & Balachandran, K. (2008b). Mean-square stability of second-order Runge-Kutta methods for multi-dimensional linear stochastic differential systems, *J. Comput. Appl. Math.* **219**: 170–197.
- Rathinasamy, A. & Balachandran, K. (2008c). Mean-square stability of semi-implicit Euler method for linear stochastic differential equations with multiple delays and Markovian switching, *Appl. Math. Comput.* **206**: 968–979.
- Raviart, P. A. & Thomas, J. M. (1983). *Introduction to the Numerical Analysis of Partial Differential Equations*, Collection of Applied Mathematics for the Master's Degree, Masson, Paris. (in French).
- Razevig, V. D. (1980). Digital modelling of multi-dimensional dynamics under random perturbations, *Automat. Remote Control* **4**: 177–186. (in Russian).
- Rebonato, R. (2002). *Modern Pricing of Interest Rate Derivatives: The Libor Market Model and Beyond*, Princeton, University Press.
- Revuz, D. & Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd edn, Springer.
- Richtmeyer, R. & Morton, K. (1967). *Difference Methods for Initial-Value Problems*, Interscience, New York.
- Ripley, B. D. (1983). *Stochastic Simulation*, Wiley, New York.
- Rodkina, A. & Schurz, H. (2005). Almost sure asymptotic stability of drift-implicit θ -methods for bilinear ordinary stochastic differential equations in \mathbb{R}^1 , *J. Comput. Appl. Math.* **180**: 13–31.
- Rogers, L. C. G. (1997). Arbitrage with fractional Brownian motion, *Math. Finance* **7**(1): 95–105.
- Rogers, L. C. G. (2002). Monte-Carlo valuation of American options, *Math. Finance* pp. 271–286.
- Rogers, L. C. G. & Talay, D. (eds) (1997). *Numerical Methods in Finance. Sessions at the Isaac Newton Institute, Cambridge, GB, 1995*, Cambridge Univ. Press.
- Römisch, W. (1983). On an approximate method for random differential equations, in J. vom Scheidt (ed.), *Problems of Stochastic Analysis in Applications*, pp. 327–337. Wiss. Beiträge der Ingenieurhochschule Zwickau, Sonderheft.
- Römisch, W. & Wakolbinger, A. (1987). On the convergence rates of approximate solutions of stochastic equations, Vol. 96 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 204–212.
- Römisch, W. & Winkler, R. (2006). Stepsize control for mean-square numerical methods for stochastic differential equations with small noise, *SIAM J. Sci. Comput.* **28**(2): 604–625.
- Ronghua, L., Hongbing, M. & Qin, C. (2006). Exponential stability of numerical solutions to SDDEs with Markovian switching, *Appl. Math. Comput.* **174**: 1302–1313.
- Ronghua, L. & Yingmin, H. (2006). Convergence and stability of numerical solutions to SDDEs with Markovian switching, *Appl. Math. Comput.* **175**: 1080–1091.
- Ronghua, L. & Zhaoguang, C. (2007). Convergence of numerical solution to SDDEs with Poisson jumps and Markovian switching, *J. Comput. Appl. Math.* **184**: 451–463.
- Rootzén, H. (1980). Limit distributions for the error in approximations of stochastic integrals, *Ann. Probab.* **8**(2): 241–251.
- Ross, S. A. (1976). The arbitrage theory of capital asset pricing, *J. Economic Theory* **13**: 341–360.
- Ross, S. M. (1990). *A Course in Simulation*, Macmillan.

- Rubenthaler, S. (2003). Numerical simulation of the solution of a stochastic differential equation driven by a Lévy process, *Stochastic Process. Appl.* **103**(2): 311–349.
- Rubinstein, M. (1994). Implied binomial trees, *J. Finance* **49**: 771–818.
- Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*, Wiley Ser. Probab. Math. Statist., Wiley, New York.
- Rümelin, W. (1982). Numerical treatment of stochastic differential equations, *SIAM J. Numer. Anal.* **19**(3): 604–613.
- Runggaldier, W. J. (2004). Estimation via stochastic filtering in financial market models, in G. Yin & Q. Zhang (eds), *Contemporary Mathematics*, Vol. 351, American Math. Society, pp. 309–318.
- Ryashko, L. B. & Schurz, H. (1997). Mean square stability analysis of some linear stochastic systems, *Dynam. Systems Appl.* **6**(2): 165–189.
- Rydén, T. & Wiktorsson, M. (2001). On the simulation of iterated Itô integrals, *Stochastic Process. Appl.* **91**: 151–168.
- Sabelfeld, K. K. (1979). On the approximate computation of Wiener integrals by Monte-Carlo method, *Zh. Vychisl. Mat. i. Mat. Fiz.* **19**: 29–43. (in Russian).
- Saito, Y. & Mitsui, T. (1993a). Simulation of stochastic differential equations, *Ann. Inst. Statist. Math.* **45**: 419–432.
- Saito, Y. & Mitsui, T. (1993b). T-stability of numerical schemes for stochastic differential equations, *World Sci. Ser. Appl. Anal.* **2**: 333–344.
- Saito, Y. & Mitsui, T. (1995). S-series in the Wong-Zakai approximation for stochastic differential equations, *Vietnam J. Math.* **23**: 303–317.
- Saito, Y. & Mitsui, T. (1996). Stability analysis of numerical schemes for stochastic differential equations, *SIAM J. Numer. Anal.* **33**(6): 2254–2267.
- Samuelides, Y. & Nahum, E. (2001). A tractable market model with jumps for pricing short-term interest rate derivatives, *Quant. Finance* **1**: 270–283.
- Samuelson, P. A. (1955). Brownian motion in the stock market. (Unpublished manuscript).
- Samuelson, P. A. (1965a). Proof that properly anticipated prices fluctuate randomly, *Industr. Manag. Rev.* **6**: 41–50.
- Samuelson, P. A. (1965b). Rational theory of warrant pricing, *Industrial Management Review* **6**(2): 13–32.
- Sandmann, K. & Sondermann, D. (1994). On the stability of lognormal interest rate models. Discussion paper No. B-263, University Bonn.
- Schein, O. & Denk, G. (1998). Numerical solution of stochastic differential-algebraic equations with applications to transient noise simulation of microelectronic circuits, *J. Comput. Appl. Math.* **100**(1): 77–92.
- Schmitz-Abe, K. & Shaw, W. (2005). Measure order of convergence without an exact solution, Euler vs Milstein scheme, *Internat. J. of Pure & Appl. Math.* **24**(3): 365–381.
- Schoenmakers, J. (2005). *Robust Libor Modelling and Pricing of Derivative Products*, Chapman & Hall/CRC.
- Schönbucher, P. J. (2003). *Credit Derivatives Pricing Models*, Wiley, Chichester.
- Schoutens, W. & Studer, M. (2003). Short term risk management using stochastic Taylor expansions under Lévy models, *Insurance: Mathematics and Economics* **33**(1): 173–188.
- Schoutens, W. & Symens, S. (2003). The pricing of exotic options by Monte-Carlo simulations in a Lévy market with stochastic volatility, *Int. J. Theor. Appl. Finance* **6**(8): 839–864.

- Schroder, M. (1989). Computing the constant elasticity of variance option pricing formula, *J. Finance* **44**(1): 211–219.
- Schurz, H. (1996a). Asymptotical mean square stability of an equilibrium point of some linear numerical solutions with multiplicative noise, *Stochastic Anal. Appl.* **14**(3): 313–354.
- Schurz, H. (1996b). Numerical regularization for SDEs: Construction of nonnegative solutions, *Dynam. Systems Appl.* **5**(3): 323–351.
- Schurz, H. (1996c). *Stability, stationarity and boundedness of some implicit numerical methods for stochastic differential equations*, PhD thesis, Humboldt University, Berlin.
- Schurz, H. (1999). Preservation of asymptotical laws through Euler methods for Ornstein-Uhlenbeck process, *Stochastic Anal. Appl.* **17**(3): 463–486.
- Schweizer, M. (1991). Option hedging for semimartingales, *Stochastic Process. Appl.* **37**: 339–363.
- Setiawaty, B. (1999). *Consistent estimation for the order of hidden Markov models*, PhD thesis, University of Adelaide, Australia.
- Seydel, R. (2006). *Tools for Computational Finance*, Universitext, 3rd edn, Springer.
- Shaw, W. (1998). *Pricing Derivatives with Mathematica*, Cambridge University Press.
- Shimizu, A. & Kawachi, T. (1984). Approximate solutions of stochastic differential equations, *Bull. Nagoya Inst. Tech.* **36**: 105–108.
- Shinozuka, M. (1971). Simulation of multivariate and multidimensional random differential processes, *J. Acoust. Soc. Amer.* **49**: 357–367.
- Shiryayev, A. N. (1984). *Probability*, Springer.
- Shiryayev, A. N. (1999). *Essentials of Stochastic Finance*, World Scientific, Singapore.
- Shoji, I. (1998). Approximation of continuous time stochastic processes by a local linearization method, *Math. Comp.* **67**(221): 287–298.
- Shoji, I. & Ozaki, T. (1997). Comparative study of estimation methods for continuous time stochastic processes, *J. Time Series Analysis* **18**: 485–506.
- Shoji, I. & Ozaki, T. (1998a). Estimation for nonlinear stochastic differential equations by a local linearization method, *Stochastic Anal. Appl.* **16**(4): 733–752.
- Shoji, I. & Ozaki, T. (1998b). A statistical method of estimation and simulation for systems of stochastic differential equations, *Biometrika* **85**(1): 240–243.
- Shreve, S. E. (2003a). *Stochastic Calculus for Finance I: Continuous Time Models*, Springer.
- Shreve, S. E. (2003b). *Stochastic Calculus for Finance II: Continuous Time Models*, Springer.
- Sickenberger, T. (2008). Mean-square convergence of stochastic multi-step methods with variable step-size, *J. Comput. Appl. Math.* **212**: 300–319.
- Sin, C. A. (1998). Complications with stochastic volatility models, *Adv. in Appl. Probab.* **30**: 256–268.
- Siopacha, M. & Teichmann, J. (2007). Weak and strong Taylor methods for numerical solutions of stochastic differential equations. Quantitative Finance Papers number [0704.0745](#), submitted.
- Situ, R. (2005). *Theory of Stochastic Differential Equations with Jumps and Applications*, Springer.
- Sklar, A. (1959). Fonction de repartition à n -dimensions et leur marges, *Publications de l'Institut Statistique de l'Université de Paris* **8**: 229–231.

- Sloan, I. H. & Wozniakowski, H. (1998). When are quasi-Monte-Carlo algorithms efficient for high dimensional integrals?, *J. Complexity* **14**(1): 1–33.
- Slominski, L. (1994). On approximation of solutions of multidimensional SDEs with reflecting boundary conditions, *Stochastic Process. Appl.* **50**(2): 197–219.
- Slominski, L. (2001). Euler approximation of solutions of SDEs with reflecting boundary, *Stochastic Process. Appl.* **94**: 317–337.
- Smith, A. M. & Gardiner, C. W. (1988). Simulation of nonlinear quantum damping using the positive representation. Univ. Waikato Research Report.
- Smith, G. D. (1985). *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, 3rd edn, Clarendon Press, Oxford.
- Smith, R. D. (2007). An almost exact simulation method for the Heston model, *J. Comput. Finance* **11**(1): 115–125.
- Sørensen, H. (2001). Discretely observed diffusions: Approximation of the continuous-time score function, *Scand. J. Statist.* **28**: 113–121.
- Sørensen, M. (1991). Likelihood methods for diffusions with jumps, in N. U. Prabhu & I. V. Basawa (eds), *Statistical Inference in Stochastic Processes*, Marcel Dekker, New York, pp. 67–105.
- Sørensen, M. (1997). Estimating functions for discretely observed diffusions: A review, in I. V. Basawa, V. P. Godambe & R. L. Taylor (eds), *Selected Proceedings of the Symposium on Estimating Functions*, Vol. 32 of *IMS Lecture Notes - Monograph Series*, Hayward: Institute of Mathematical Statistics, pp. 305–325.
- Sørensen, M. (1999). On asymptotics of estimating functions, *Brazilian J. Probab. Statist.* **13**: 111–136.
- Sørensen, M. (2000). Prediction based estimating functions, *J. Econometrics* **3**: 123–147.
- Stentoft, L. (2004). Assessing the least-squares Monte-Carlo approach to American option valuation, *Rev. Derivatives Res.* **7**: 129–168.
- Stroud, A. H. (1971). *Approximate Calculation of Multiple Integrals*, Prentice-Hall.
- Studer, M. (2001). *Stochastic Taylor Expansions and Saddlepoint Approximations for Risk Management*, PhD thesis, ETH Zurich.
- Sun, M. & Glowinski, R. (1994). Pathwise approximation and simulation for the Zakai filtering equation through operator splitting, *Calcolo* **30**(3): 219–239.
- Sussmann, H. J. (1978). On the gap between deterministic and stochastic ordinary differential equations, *Ann. Probab.* **6**(1): 19–41.
- Szepessy, A., Tempone, R. & Zouraris, G. E. (2001). Adaptive weak approximation of stochastic differential equations, *Commun. on Pure & Appl. Math.* **LIV**: 1169–1214.
- Takahashi, A. (1999). An asymptotic expansion approach to pricing financial contingent claims, *Asia-Pacific Financial Markets* **6**: 115–151.
- Takahashi, A. & Yoshida, M. (2005). Monte Carlo simulation with asymptotic method, *J. Japan Statist. Soc.* **35**: 171–203.
- Takeuchi, A. (2002). The Malliavin calculus for SDE with jumps and the partially hypoelliptic problem, *Osaka J. Math.* **39**: 523–559.
- Talay, D. (1982a). *Analyse Numérique des Équations Différentielles Stochastiques*, PhD thesis, Université de Provence, Centre Saint Charles. Thèse 3ème Cycle.
- Talay, D. (1982b). Convergence for each trajectory of an approximation scheme of SDE, *Comptes Rendus Acad. Sc. Paris, Séries I Math.* **295**(3): 249–252. (in French).

- Talay, D. (1982c). How to discretize stochastic differential equations, *Nonlinear Filtering and Stochastic Control*, Vol. 972 of *Lecture Notes in Math.*, Springer, pp. 276–292.
- Talay, D. (1983). Pathwise solution and numerical analysis of stochastic differential equations, *Stochastics* **9**(4): 275–306. in French.
- Talay, D. (1984). Efficient numerical schemes for the approximation of expectations of functionals of the solution of an SDE and applications, *Filtering and Control of Random Processes*, Vol. 61 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 294–313.
- Talay, D. (1986). Discretization of a stochastic differential equation and rough estimate of the expectations of functionals of the solution, *Modél. Math. et Anal. Numér.* **20**(1): 141–179. (in French).
- Talay, D. (1987). Classification of discretization of diffusions according to an ergodic criterion, *Stochastic Modelling and Filtering*, Vol. 91 of *Lecture Notes in Control and Inform. Sci.*, Springer, pp. 207–218.
- Talay, D. (1990). Second order discretization schemes of stochastic differential systems for the computation of the invariant law, *Stochastics Stochastics Rep.* **29**: 13–36.
- Talay, D. (1991). Approximation of upper Lyapunov exponents of bilinear stochastic differential equations, *SIAM J. Numer. Anal.* **28**(4): 1141–1164.
- Talay, D. (1995). Simulation of stochastic differential systems, in P. Krée & W. Wedig (eds), *Probabilistic Methods in Applied Physics*, Vol. 451 of *Lecture Notes in Phys.*, Springer, chapter 3, pp. 54–96.
- Talay, D. (1999). The Lyapunov exponent of the Euler scheme for stochastic differential equations, in H. Crauel & M. Gundlach (eds), *Stochastic Dynamics*, Springer, pp. 241–258.
- Talay, D. & Tubaro, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations, *Stochastic Anal. Appl.* **8**(4): 483–509.
- Talay, D. & Zheng, Z. (2003). Quantiles of the Euler scheme for diffusion processes and financial applications, *Math. Finance* **13**: 187–199.
- Tanaka, H. & Kohatsu-Higa, A. (2009). An operator approach for Markov chain weak approximations with an application to infinite activity Lévy driven SDEs, *Ann. Appl. Probab.* **19**(3): 1026–1062.
- Taraskin, A. F. (1974). On the asymptotic normality of vector-valued stochastic integrals and estimates of a multi-dimensional diffusion process, *Theory Probab. Math. Statist.* **2**: 209–224.
- Tausworthe, R. C. (1965). Random numbers generated by linear recurrence modulo two, *Math. Comp.* **19**: 201–209.
- Tavella, D. & Randall, C. (2000). *Pricing Financial Instruments. The Finite Difference Method*, Wiley, New York.
- Tebaldi, C. (2005). Hedging using simulation: A least squares approach, *J. Econom. Dynam. Control* **29**: 1287–1312.
- Teichmann, J. (2006). Calculating the Greeks by cubature formulas, *Proc. Roy. Soc. London A* **462**: 647–670.
- Tetzlaff, U. & Zschiesche, H.-U. (1984). Approximate solutions for Ito differential equations using the Taylor expansion for semigroups of operators, *Wiss. Z. Techn. Hochschule Leuna-Merseburg* **2**: 332–339. (in German).
- Thitsiklis, J. & Van Roy, B. (2001). Regression methods for pricing complex American-style options, *IEEE Trans. Neural Networks* **12**: 694–703.

- Tocino, A. (2005). Mean-square stability of second-order Runge-Kutta methods for stochastic differential equations, *J. Comput. Appl. Math.* **175**: 355–367.
- Tocino, A. & Ardanuy, R. (2002). Runge-Kutta methods for numerical solution of stochastic differential equations, *J. Comput. Appl. Math.* **138**: 219–241.
- Tocino, A. & Vigo-Aguiar, J. (2002). Weak second order conditions for stochastic Runge-Kutta methods, *SIAM J. Sci. Comput.* **24**: 507–523.
- Török, C. (1994). Numerical solution of linear stochastic differential equations, *Comput. Math. Appl.* **27**(4): 1–10.
- Tudor, C. (1989). Approximation of delay stochastic equations with constant retardation by usual Ito equations, *Rev. Roumaine Math. Pures Appl.* **34**(1): 55–64.
- Tudor, C. & Tudor, M. (1987). On approximation of solutions for stochastic delay equations, *Stud. Cerc. Mat.* **39**(3): 265–274.
- Tudor, C. & Tudor, M. (1997). Approximate solutions for multiple stochastic equations with respect to semimartingales, *Z. Anal. Anwendungen* **16**(3): 761–768.
- Turner, T., Schnell, S. & Burrage, K. (2004). Stochastic approaches for modelling in vivo reactions, *Comput. Biol. Chem.* **28**(3): 165–178.
- Unny, T. E. (1984). Numerical integration of stochastic differential equations in catchment modelling, *Water Res.* **20**: 360–368.
- van der Hoek, J. & Elliott, R. J. (2006). *Binomial Methods in Finance*, Springer Finance, Springer.
- Vanden-Eijnden, E. (2003). Numerical techniques for multi-scale dynamical systems with stochastic effect, *Commun. Math. Sci.* **1**: 385–391.
- Vasicek, O. A. (1977). An equilibrium characterization of the term structure, *J. Financial Economics* **5**: 177–188.
- Ventzel, A. D., Gladyshev, S. A. & Milstein, G. N. (1985). Piecewise constant approximation for the Monte-Carlo calculation of Wiener integrals, *Theory Probab. Appl.* **24**: 745–752.
- Victoir, N. (2004). Asymmetric cubature formulae with few points in high dimension for symmetric measures, *SIAM J. Numer. Anal.* **42**(1): 209–227.
- Wagner, W. (1987). Unbiased Monte-Carlo evaluation of certain functional integrals, *J. Comput. Phys.* **71**(1): 21–33.
- Wagner, W. (1988a). Monte-Carlo evaluation of functionals of solutions of stochastic differential equations. Variance reduction and numerical examples, *Stochastic Anal. Appl.* **6**: 447–468.
- Wagner, W. (1988b). Unbiased multi-step estimators for the Monte-Carlo evaluation of certain functionals, *J. Comput. Phys.* **79**(2): 336–352.
- Wagner, W. (1989a). Stochastische numerische Verfahren zur Berechnung von Funktionalintegralen. (in German), Habilitation, Report **02/89**, IMATH, Berlin.
- Wagner, W. (1989b). Unbiased Monte-Carlo estimators for functionals of weak solutions of stochastic differential equations, *Stochastics Stochastics Rep.* **28**(1): 1–20.
- Wagner, W. & Platen, E. (1978). Approximation of Itô integral equations. Preprint ZIMM, Akad. Wissenschaften, DDR, Berlin.
- Wang, L., Mei, C. & Xue, H. (2007). The semi-implicit Euler method for stochastic differential delay equation with jumps, *Appl. Math. Comput.* **192**: 567–578.
- Wang, P. (2008). Three-stage stochastic Runge-Kutta methods for stochastic differential equations, *J. Comput. Appl. Math.* **222**: 324–332.
- Wang, Z. & Zhang, C. (2006). An analysis of stability of Milstein method for stochastic differential equations with delay, *Comput. Math. Appl.* **51**: 1445–1452.

- Widom, H. (1969). *Lectures on Measure and Integration*, Van Nonstrand and Reinhold Company.
- Wilmott, P., Dewynne, J. & Howison, S. (1993). *Option Pricing: Mathematical Models and Computation*, Oxford Financial Press.
- Wonham, W. M. (1965). Some applications of stochastic differential equations to optimal nonlinear filtering, *SIAM J. Control Optim.* **2**: 347–369.
- Wörner, J. H. (2006). Power and multipower variation: Inference for high-frequency data, *International Conference Stochastic Finance 2004*, Springer, pp. 264–276.
- Wright, D. J. (1974). The digital simulation of stochastic differential equations, *IEEE Trans. Automat. Control* **19**: 75–76.
- Wright, D. J. (1980). Digital simulation of Poisson stochastic differential equations, *Internat. J. Systems. Sci.* **11**: 781–785.
- Wu, S., Han, D. & Meng, X. (2004). p -moment stability of stochastic differential equations with jumps, *Appl. Math. Comput.* **152**: 505–519.
- Xu, C., Qian, X. & Jiang, L. (2003). Numerical analysis on binomial tree methods for a jump-diffusion model, *J. Comput. Appl. Math.* **156**: 23–45.
- Yamada, T. (1976). Sur l'approximation des solutions d'équations différentielles stochastiques, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **36**(2): 153–164. (in French).
- Yamato, Y. (1979). Stochastic differential equations and nilpotent Lie algebras, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **47**(2): 213–229.
- Yannios, N. & Kloeden, P. E. (1996). Time-discretization solution of stochastic differential equations, in R. L. May & A. K. Easton (eds), *Computational Techniques and Applications: CTAC95*, World Scientific, Singapore, pp. 823–830.
- Yen, V. V. (1988). A stochastic Taylor formula for functionals of two-parameter semimartingales, *Acta Math. Vietnam* **13**(2): 45–54.
- Yen, V. V. (1992). A stochastic Taylor formula for two-parameter stochastic processes, *Probab. Math. Statist.* **13**(1): 149–155.
- Yuan, C. & Glover, W. (2006). Approximate solutions of stochastic differential delay equations with Markovian switching, *J. Comput. Appl. Math.* **194**: 207–226.
- Yuan, C. & Mao, X. (2004a). Convergence of the Euler-Maruyama method for stochastic differential equations with Markovian switching, *Math. Comput. Simulation* **64**: 223–235.
- Yuan, C. & Mao, X. (2004b). Robust stability and controllability of stochastic differential delay equations with Markovian switching, *Automatica* **40**: 343–354.
- Zakai, M. (1969). On the optimal filtering of diffusion processes, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **11**: 230–243.
- Zhang, H. & Gan, S. (2008). Mean square convergence of one-step methods for neutral stochastic differential delay equations, *Appl. Math. Comput.* **204**: 884–890.
- Zou, G. & Skeel, R. (2004). Robust variance reduction for random walk methods, *SIAM J. Sci. Comput.* **25**(6): 1964–1981.

Author Index

- Abdulle, A., 787, 790, **819**
Abramowitz, M., 78, **793**
Abukhaled, M. I., 786, 787, **793**
Achdou, Y., *XXVII*, 788, **793**
Acworth, P., 789, **793**
Aggoun, L., *XXVII*, 309, 434–437, 445,
 788, **805**
Ahmed, H. M., 786, **805**
Ait-Sahalia, Y., *XXIII*, 42, 390, 413,
 414, 416, 417, 788, **793**
Alabert, A., 787, **793**
Alcock, J. T., 326, 358, 574, 787, 790,
 793
Alfonsi, A., 63, 783, **793**
Allain, M. F., 785, **793**
Allen, E. J., 786, 787, 789, **793, 809**
Andersen, L., 61, 63, 107, 567, 783, 789,
 793, 794
Anderson, B. D. O., 785, 787, **802**
Andreasen, J., 567, 789, **794**
Anh, V. V., 785, **809**
Ansel, J. P., 146, 469, **794**
Ardanuy, R., 784, 787, **833**
Arnold, L., 789, **794**
Artemiev, S. S., 786–790, **794, 795**
Ash, R. B., 21, **794**
Asmussen, S., *XXVII*, 783, 786–788,
 791, 794
Atalla, M. A., 785, **794**
Athreya, K. B., 59, **794**
Atiya, A. F., 786, 789, **822**
Avellaneda, M., 598, 790, **795**
Averina, A. T., 786, 788, 789, **794, 795**
Azencott, R., 784, **795**
Babbs, S., *XXIII*, **795**
Bachelier, L., 6, 40, **795**
Bajeux-Besnainou, I., 784, **795**
Baker, C. T. H., 785, **795**
Balachandran, K., 787, 790, **828**
Balakrishnan, A. V., 788, **795**
Balakrishnan, N., 171, **814**
Bally, V., 556, 565, 598, 786, 789, 790,
 795
Baras, J. B., 440, 788, **795**
Barndorff-Nielsen, O., 14, 53, 75, 105,
 412, 414, 783, 788, **795**
Barraquand, J., 638, **795**
Barrett, R., 742, **796**
Basle, 145, 151, **796**
Bavouzet, M.-P., 791, **796**
Bayer, C., 784, **796**
Becherer, D., 146, 784, **796**
Bell, E. T., 230, **796**
Belward, J. A., 787, **799**
BenArous, G., 784, **796**
Bender, C., 784, **796**
Berkaoui, A., 63, 783, 785, **796**
Bernard, P., 787, **806**
Bernis, G., 598, 790, **796**
Berry, M., 742, **796**
Bertoin, J., 75, 783, **796**
Beskos, A., 61, 783, 784, **796**
Bhar, R., 457, 788, **796**
Bibby, B. M., 390, 397, 398, 404, 788,
 796
Bielecki, T., *XXVII*, **797**
- E. Platen, N. Bruti-Liberati, *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, Stochastic Modelling and Applied Probability 64, DOI [10.1007/978-3-642-13694-8](https://doi.org/10.1007/978-3-642-13694-8),
© Springer-Verlag Berlin Heidelberg 2010

- Billingsley, P., 723, **797**
 Binder, K., 791, **797**
 Biscay, R., 786, 787, 789, **797**, 800,
814
 Björk, T., *XXIII*, *XXVII*, 45, 56, **797**
 Black, F., 14, 37, 40–42, 47, 136, 791,
797
 Bladt, M., 390, 417, 788, **797**
 Blanchet, J., 791, **797**
 Bokor, R., 786, **799**
 Bokor, R. H., 786, 787, 790, **797**
 Borodin, A. N., *XXVII*, 132, 783, **797**
 Borwanker, J. D., 786, **827**
 Bossy, M., 63, 783, 785, 787, 790, **796**,
797
 Bouchard, B., 785, 790, **797**
 Bouchaud, J.-P., 790, **827**
 Bouleau, N., 785, 788, **797**
 Boyce, W. E., 786, **817**
 Boyle, P. P., 559, 637, 638, 703, 706,
 714, 719, 785, 788, 789, 791, **797**,
798, **814**
 Brandt, M., 390, 788, 789, **798**
 Bratley, P., 638, 716, 785, 791, **798**
 Brémaud, P., 438, 783, **798**
 Brent, R. P., 716, 785, **798**
 Breymann, W., 120, **798**
 Brigo, D., *XXVII*, 789, **798**
 Broadie, M., 61, 63, 107, 108, 601, 638,
 706, 713, 714, 783, 789, **793**, **794**,
798
 Brotherton-Ratcliffe, R., 789, **794**
 Brown, B. M., 788, **798**
 Brown, G., 791, **798**
 Bru, M.-F., 70, 89, 91, 112, 170, 783,
798
 Bruti-Liberati, N., 56, 309, 326, 332,
 347, 359, 375, 528, 572, 575, 713,
 717, 784–787, 789, 790, **798**, **799**
 Buchen, P. W., 163, **799**
 Buckwar, E., 785, 786, **795**, **799**
 Bucy, R. S., 419, 788, **814**
 Bühlmann, H., 149, 176, 784, **799**
 Burq, Z. A., 61, 783, **799**
 Burrage, K., *XXIII*, 326, 358, 375, 574,
 785–787, 789, 790, **793**, **799**, **800**,
833
 Burrage, P. M., 268, 270, 431, 784–787,
 789, 790, **799**, **800**
 Cameron, R. J., 425, 430, 785, 788, **802**
 Cao, W. R., 790, **800**, **820**
 Cao, Y., 787, 790, **800**
 Caramellino, L., 598, 790, **795**
 Carbonell, F., 786, 789, **800**
 Carr, P., 53, **800**
 Carrasco, M., 788, **800**
 Carriére, J., 789, **800**
 Carverhill, A., 638, 669, 670, 673, **802**
 Casasus, L. L., 786, **800**
 Casella, B., 784, **800**
 Castell, F., 784, 788, **800**, **801**
 Chan, K. S., 785, **801**
 Chan, N. H., *XXVII*, 788, **801**
 Chan, T. F., 742, **796**
 Chang, C. C., 638, 787, 789, 791, **801**
 Chen, N., 61, 783, 784, 789, **801**
 Cheng, R., 783, **801**
 Cheridito, P., 784, **801**
 Chernov, M., 414, 788, **800**, **801**
 Cherny, A. S., 59, **801**
 Chesney, M., *XXVII*, **814**
 Chevance, D., 785, **801**
 Chiarella, C., 457, 788, **796**, **801**
 Chib, S., 390, 788, **805**
 Chorin, A. J., 791, **801**
 Christensen, B. J., 398, 788, **801**
 Christensen, M. M., 45, 143, 784, **801**
 Clark, J. M. C., 425, 430, 601, 785, 786,
 788, **801**, **802**
 Clément, E., 789, **802**
 Clements, D. J., 785, 787, **802**
 Clewlow, L., 638, 669, 670, 673, **802**
 Colwell, D. B., 601, 606, **802**
 Cont, R., *XXIII*, *XXVI*, *XXVII*, 73, 75,
 98, 565, 783, 788, **802**
 Cortés, J. C., 786, **802**, **814**
 Courtadon, G., 42, **802**
 Cox, A. M. G., 784, **802**
 Cox, J. C., 40, 41, 181, 533, 791, **802**
 Craddock, M., 62, 75, 783, **802**
 Craig, I. J. D., 787, 790, **822**
 Crisan, D., 457, 788, **802**
 Cruzeiro, A. B., 789, **802**
 Curran, M., 641, **803**
 Cyganowski, S., 785, 788, **803**
 D'Halluin, Y., *XXVI*, **803**
 Dacunha-Castelle, D., 390, 788, **803**

- Davis, M. H. A., 6, 598, 601, 791, **803**
 Da Fonseca, J., 112, 784, **803**
 Deelstra, G., 63, 783, **803**
 Delbaen, F., *XXVII*, 40, 63, 94, 147,
 783, 784, **803**
 Dembo, A., 438, 788, **803**
 DeMiguel, V., 154, **803**
 Demmel, J., 742, **796**
 Deng, F., 787, 790, **825**
 Denk, G., 786, 787, **803**, **829**
 Derman, E., 42, 791, **797**
 Detemple, J., 638, 706, 789, 790, **798**,
 803
 Devroye, L., 716, 785, **803**
 Dewynne, J., 734, 743, 791, **834**
 de la Cruz, H., 786, **800**
 De Raedt, H., 791, **803**
 DiMasi, G., 473, 788, 791, **804**, **818**
 Ding, X., 790, **804**
 Diop, A., 63, 783, 785, **796**, **797**, **804**
 Doering, C. R., 786, **804**
 Donati-Martin, C., 71, **804**
 Dongarra, J., 742, **796**
 Doob, J. L., 20, 21, **804**
 Doss, H., 785, **804**
 Dothan, L. U., 42, **804**
 Douglas, J., 785, **804**
 Doukhan, P., 411, **804**
 Doumerc, Y., 71, **804**
 Drummond, I. T., 786, **804**
 Drummond, P. D., 787, 790, **804**
 Dsagnidse, A. A., 786, **804**
 Duane, S., 786, **804**
 Duffie, D., *XXIII*, 42, 405, 480, 481,
 638, 783, 788, **804**
 Dupuis, P. G., 723, 791, **805**, **818**
- E, W., 787, 790, **819**
 Eberlein, E., 14, 53, 75, 105, 783, **805**
 Eijkhout, V., 742, **796**
 Einstein, A., 6, **805**
 Ekeland, I., 790, **797**
 Ekström, E., 784, **805**
 El-Borai, M. M., 786, **805**
 El-Khatib, Y., 598, 790, **805**
 El-Nadi, K. E., 786, **805**
 Elerain, O., 390, 788, **805**
- Elliott, R. J., *XXVII*, 20, 54, 309,
 433–437, 445, 457, 601, 606, 788,
 791, **802**, **805**, **833**
 El Karoui, N., 783, **808**
 Embrechts, P., *XXVII*, 65, 66, 97, 783,
 805, **822**
 Engel, D., 194, 784, **805**
 Engle, R. F., 49, **805**
 Eraker, B., 390, 414, 788, **805**
 Ermakov, S. M., 637, 638, 791, **806**
 Etheridge, A., 6, **803**
 Evans, L. C., 317, **806**
- Fahrmeir, L., 785, 788, **806**
 Fan, J., 417, **793**
 Fan, Z. C., 790, **800**, **820**
 Feast, G., 783, **801**
 Feng, J. F., 786, **806**
 Fernholz, E. R., 784, **806**
 Ferrante, M., 787, **793**, **806**
 Filipović, D., 405, 783, 784, **804**, **806**
 Fineso, L., 440, 788, **795**
 Fischer, P., 323, 433, 434, 447, 457,
 781, 787, 788, 790, **805**, **806**
 Fishman, G. S., 477, 716, 785, 791, **806**
 Flannery, B. P., 64, 716, **827**
 Fleming, W. H., 40, **806**
 Fleury, G., 787, **806**
 Florens, J.-P., 788, **800**
 Florens-Zmirou, D., 390, 395, 788, **803**,
 806
 Föllmer, H., 40, 473, 475, 593, 628, 687,
 806
 Forsyth, P. A., *XXVI*, **803**
 Fosdick, L. D., 791, **806**
 Fouque, J. P., 791, **806**
 Fournie, E., 598, 638, 645, 790, 791,
 806, **807**
 Fox, B. L., 638, 716, 785, 791, **798**
 Franklin, J. N., 785, **807**
 Frey, R., *XXVII*, 65, 66, 97, 457, 783,
 788, **807**, **822**
 Friedman, A., 129, 682, **807**
 Fu, M. C., 638, **807**
 Fujisaki, M., 425, 426, 788, **807**
- Gadmer, A., 61, 246, **815**
 Gaines, J., 784, 788, **800**–**802**
 Gaines, J. G., 261, 784, 789, **807**

- Galessio, G., 784, **807**
 Gallant, A. R., 398, 414, 788, **801, 807**
 Gamba, R., 598, 790, **795**
 Gan, S., 790, **834**
 Garcia, R., 790, **803**
 Gard, T. C., 785, 787, **807**
 Gardiner, C. W., 787, 790, **831**
 Gardoñ, A., 292, 786, 787, **807**
 Garlappi, L., 154, **803**
 Gelbrich, M., 786, 789, **807**
 Geman, H., *XXIII*, 14, 53, 56, 75, 105,
 800, 808
 Geman, S., 783, **808**
 Genon-Catalot, V., 411, 788, **808**
 Gerardi, A., 786, **808**
 Ghysels, E., 414, 788, **800, 801**
 Gikhman, I. I., 785, **808**
 Giles, M. B., 789, 791, **808**
 Gillespie, D. T., 787, 790, **800**
 Gladyshev, S. A., 789, 791, **808, 833**
 Glasserman, P., *XXIII, XXVII*, 56, 57,
 62, 107, 477, 534, 601, 638, 713,
 714, 785–789, 791, **793, 798, 801,**
 808
 Gleser, L. J., 80, **808**
 Glorenne, P. Y., 786, **808**
 Glover, W., 790, **834**
 Glowinski, R., 788, **805, 831**
 Glynn, P. W., *XXVII*, 480, 481, 638,
 783, 786, 788, 791, **794, 797, 804**
 Gobet, E., 598, 785, 787, 790, **796,**
 797, 808
 Godambe, V. P., 402, 403, 788, **808**
 Golec, J., 785, 786, **809**
 Gombani, A., 457, 788, **809**
 Goodlett, S. T., 789, **809**
 Gorostiza, L. G., 785, **809**
 Gorovoi, V., 41, **809**
 Gouriéroux, C., 70, 89, 111, 112, 398,
 784, 788, **809**
 Goyal, A., 789, **798**
 Grant, D., 638, **809**
 Grasselli, M., 112, 784, **803, 809**
 Grecksch, W., 785, 786, **809**
 Greenside, H. S., 786, 789, **809**
 Greiner, A., 786, **809**
 Grorud, A., 789, **809**
 Grüne, L., 786, 787, **809**
 Guerra, M. L., 789, **809**
 Guo, S. J., 786, **810**
 Guyon, J., 556, 565, 785, 789, **810**
 Gyöngy, I., 785, 786, **810**
 Ha, Z., 789, **798**
 Hammersley, J. M., 477, 637, 791, **810**
 Han, D., 787, 790, **834**
 Handscomb, D. C., 477, 637, 791, **810**
 Hänggi, P., 787, **816**
 Hansen, L. P., 398, 416, 788, **810**
 Harris, C. J., 786, **810**
 Harrison, J. M., 146, 147, 162, **810**
 Hausenblas, E., 565, 785–787, 789, **810**
 Haussmann, U., 601, **810**
 Haworth, D. C., 789, **810**
 Heath, D., *VI, XXVII*, 26, 27, 34, 62,
 75, 77, 115, 117, 128, 129, 134,
 139, 145–148, 158, 168, 169, 176,
 405, 470, 533, 591, 596, 598, 607,
 637, 638, 645, 677, 682, 687, 735,
 781, 784, 788, 790, 791, **810, 811,**
 826
 Heidelberger, P., 791, **808**
 Helfand, E., 786, 788, 789, **809, 811**
 Herdiana, R., 790, **800**
 Hernandez, D. B., 572, 578, 787, 788,
 790, **811**
 Herz, C. S., 71, **811**
 Heston, S. L., 106, 671, 686, 707,
 709–711, 719, 784, 791, **811**
 Hewitt, J. T., 788, **798**
 Heyde, C. C., 389, 397, 398, 402–404,
 788, **808, 811**
 Higham, D. J., *XXVII*, 309, 316, 337,
 339, 572, 574, 785–790, **799, 808,**
 811, 812
 Hitzl, D. L., 638, 791, **821**
 Ho, T. S. Y., 41, **812**
 Hobson, D. G., 784, **802**
 Hoch, A., 786, **804**
 Hofmann, N., 309, 487, 528, 530, 572,
 576, 593, 638, 645, 668, 671, 687,
 715, 784–787, 789–791, **812, 816**
 Honerkamp, J., 786, **809**
 Hongbing, M., 790, **828**
 Honoré, P., 415, **812**
 Horgan, R. R., 786, **804**
 Howison, S., 734, 743, 791, **834**
 Hu, Y. Z., 784, 785, **812**

- Hull, J., 42, 638, 791, **812**
 Hulley, H., 158, 164, 166, 168, 171, 172,
 784, **812, 813**
 Hunter, C. J., 569, 789, **813**
- Iacus, S. M., 788, **813**
 Ikeda, N., *XXVII*, 21, 44, 58, 59, 278,
 783, 786, **813**
 Ingersoll, J. E., 41, 163, 533, **802, 813**
- Jäckel, P., *XXVII*, 61, 62, 477, 569,
 638, 783, 788, 789, **813, 814**
 Jackwerth, J. C., 791, **813**
 Jacobsen, M., 397, 788, **796**
 Jacod, J., 23, 75, 390, 417, 747, 783,
 785, 786, 789, **793, 813**
 Janicki, A., 785, 786, **813**
 Jansons, K. M., 786, 789, **813**
 Janssen, R., 785, 786, **813**
 Jarrow, R., *XXIII*, 702, 784, **813, 814**
 Jeanblanc, M., *XXVII*, **814**
 Jeantheau, T., 411, 788, **808**
 Jensen, B., 791, **814**
 Jentzen, A., 786, **816**
 Jiang, G. J., 788, **814**
 Jiang, L., 791, **834**
 Jimenez, J. C., 786, 787, 789, **797, 800,**
 814
 Jódar, L., 786, **802, 814**
 Johannes, M., *XXIII*, 414, **805, 814**
 Johansson, M. P., 598, 791, **803**
 Johnson, N. L., 171, **814**
 Jones, O. D., 61, 783, **799**
 Jordan, H. F., 791, **806**
 Jorion, P., *XXIII*, **814**
 Joshi, M. S., *XXVII*, 497, 569, 789,
 813, 814
 Joy, C., 638, 785, **814**
- Kabanov, Y., *XXIII*, 45, 56, **797**
 Kahl, C., 61, 326, 358, 783, 787, **814**
 Kallianpur, G., 421, 425, 426, 788, **807,**
 814
 Kalman, R. E., 419, 788, **814**
 Kalos, M. H., 638, 791, **814**
 Kampen, J., 791, **814**
 Kamrad, B., 703, **815**
 Kan, R., 42, 405, 783, **804**
 Kanagawa, S., 785, 787, 789, **815**
- Kaneko, T., 785, **815**
 Kannan, D., 788, **815**
 Karasinski, P., 42, **797**
 Karatzas, I., *XXVII*, 20, 27, 596, 610,
 784, **815**
 Kardaras, C., 784, **815**
 Kaufmann, R., 61, 246, **815**
 Kawachi, T., 786, **830**
 Kaya, O., 61, 63, 107, 108, 783, **798**
 Kazimierczyk, P., 788, **815**
 Kebaier, A., 638, 789, 791, **815**
 Keller, U., 53, 75, 783, **805**
 Kelly, J. R., 457, **815**
 Kelly, L., 120, 397, 413, 788, **798, 816**
 Kelton, W. D., 638, 640, 669, 791, **819**
 Kendall, W. S., 89, **816**
 Kessler, M., 398, 788, **816**
 Kim, C., 787, **816**
 Klauder, J. R., 785, 787, 789, 790, **816**
 Klebaner, F., *XXVII*, 198, **816**
 Klett, R., 61, 246, **815**
 Kliemann, W. H., 59, **794**
 Kloeden, P. E., *VI, XXIV, XXVII*, 27,
 61–63, 99, 104, 105, 187, 191–193,
 202, 207, 209, 211, 233, 245, 252,
 253, 259–266, 270, 273, 292, 309,
 316, 337, 339, 347, 348, 350,
 352–355, 357, 367, 390, 393, 429,
 477, 479, 491, 496, 499, 501, 504,
 520, 528, 530, 537, 538, 546, 547,
 572, 574, 586, 620, 627, 638, 642,
 715, 749, 783–790, **794, 799, 803,**
 809, 811, 812, 816, 817, 834
 Klüppelberg, C., *XXVII*, 75, 783, **805,**
 817
 Knight, J. L., 788, **814**
 Koch, G., 59, **794**
 Koekkoek, R., 63, 783, **820**
 Kohatsu-Higa, A., 598, 638, 784–787,
 789–791, **796, 806, 808, 815,**
 817, 832
 Kohler, W. E., 786, **817**
 Kohlmann, M., 601, **805**
 Kolodko, A., 791, **814**
 Komori, Y., 786, 787, 789, 790, **817**
 Konakov, V., 785, **817**
 Kopp, P. E., *XXVII*, 601, 606, **802, 805**
 Kotz, S., 171, **814**

- Kou, S. G., *XXIII*, 14, 52, 56, 75, 783, **808, 817**
- Kozin, F., 788, **817**
- Kozlov, R. I., 786, **817**
- Kreps, D. M., 146, 147, 162, **810**
- Krylov, N. V., 43, 219, 682, 748, 749, **817**
- Kubilius, K., 534, 548, 728, 786, 789, **817**
- Küchler, U., 785, 788, **817, 818**
- Kunita, H., 425, 426, 788, **807**
- Kurtz, T. G., 75, 783, 785, 786, 789, **813, 818**
- Kushner, H. J., 723, 728, 747, 785, 791, **818**
- Kusuoka, S., 784, 789, 790, **818**
- Kutoyants, Y., 389, 788, **818**
- Kuznetsov, D. F., 784, 785, **818**
- L'Ecuyer, P., 716, 785, **819, 825**
- Ladde, G. S., 785, 786, **809**
- Lagendijk, A., 791, **803**
- Lamba, H., 790, **818**
- Lamberton, D., *XXVII*, 740, 789, **802, 818**
- Landen, C., 457, 788, **818**
- Lando, D., *XXIII*, **813**
- Lanska, V., 788, **818**
- Lapeyre, B., *XXVII*, 740, **818**
- Laredo, C., 411, 788, **808**
- Larsen, K., 784, **801**
- Lasry, J. M., 598, 638, 645, 790, **806, 807**
- Lau, S. H., 559, 719, 791, **798**
- Law, A. M., 638, 640, 669, 791, **819**
- Le, T., 784, **819**
- Lebuchoux, J., 598, 790, 791, **806, 807**
- Leder, K., 791, **805**
- Lee, E. K., 787, **816**
- Lee, S. B., 41, **812**
- Lehn, J., 786, 789, **819**
- Lei, G. Y., 786, **806**
- Leisen, D., 791, **819**
- Lemor, J. P., 785, **808**
- Leobacher, G., 791, **819**
- Leon, J., 791, **819**
- Léplinge, D., 785–788, **797, 819**
- Lewis, A. L., 784, **819**
- Lewis, P. A. W., 61, 783, **819**
- Le Breton, A., 395, 788, **819**
- Le Gland, F., 788, **819**
- Li, C. W., 218, 222, 227, 509, 520, 522, 784, 789, **819, 820**
- Li, T., 787, 790, **819**
- Lindner, A., 75, 783, **817**
- Linetsky, V., 41, **809**
- Linkov, J. N., 788, **819**
- Lious, P. L., 598, 790, **806, 807**
- Liptser, R., 389, 421, 460–463, 788, **819**
- Liske, H., 784–786, **819**
- Liu, K., 787, 790, **820**
- Liu, M., 790, **804**
- Liu, M. Z., 790, **800, 820**
- Liu, X. Q., 218, 222, 227, 509, 520, 522, 784, 789, **819, 820**
- Lo, A. W., 788, **820**
- Loewenstein, M., 784, **811, 820**
- Løkka, A., 790, **820**
- Long, J. B., 146, 457, 784, **820**
- Longstaff, F. A., 42, 638, 713, 789, **820**
- Lord, R., 63, 783, **820**
- Lyons, T. G., 788, **802**
- Lyons, T. J., 261, 784, 789, **807, 820**
- Lythe, G. D., 786, 789, **813**
- Ma, J., 785, **804, 820**
- Mackevicius, V., 785, 786, 789, **820**
- Madan, D., 14, 53, 75, 105, 783, **800, 808, 820**
- Maghsoodi, Y., 347, 786, 787, **810, 820, 821**
- Makarov, R. N., 785, 787, **821**
- Maller, R., 75, 783, **817**
- Malliavin, P., *XXVII*, 598, 785, 788, 789, 791, **802, 821**
- Maltz, F. H., 638, 791, **821**
- Mammen, E., 785, **817**
- Mancini, C., 470, **821**
- Manella, R., 785, **821**
- Mao, X., 785–787, 789, 790, **808, 812, 821, 825, 834**
- Marchetti, F., 786, **808**
- Marcus, S. I., 786, **821**
- Margrabe, W., 166, **821**
- Marsh, T. A., 42, **821**
- Marshall, A. W., 66, **821**
- Martini, F., 528, 785, 789, **798, 799**
- Maruyama, G., 246, 784, **821**

- Mathé, P., 785, **812**
 Matsumoto, H., 71, **804**
 Matsumoto, M., 716, 785, **822, 825**
 Mattingly, J. C., 790, **818**
 Mauthner, S., 789, **822**
 McLeish, D. L., *XXVII*, 788, **822**
 McNeil, A., *XXVII*, 65, 66, 97, 783,
822
 McNeil, K. J., 787, 790, **822**
 Mei, C., 786, 789, **833**
 Melchior, M., 791, **822**
 Méléard, S., 75, 783, 785, 786, 789, **813**
 Meng, X., 787, 790, **834**
 Mercurio, F., *XXVII*, 789, **798**
 Merener, N., 57, 534, 786, 789, **808**
 Merton, R. C., *XXIII*, 14, 37, 38, 41,
 47, 50, 52, 158, 161, 162, 166, 414,
565, 822
 Messaoud, M., 791, **796**
 Metwally, S. A. K., 786, 789, **822**
 Meyer, P. A., 784, **812**
 Michna, Z., 786, **813**
 Mikhailov, G. A., 638, 791, **806**
 Mikosch, T., *XXVII*, **805, 822**
 Mikulevicius, R., 482, 517, 518, 525,
 534–537, 620, 728, 748, 785, 786,
789, 822
 Miller, S., 41, 158, 164, 166, 168,
 170–172, 784, **812, 822**
 Milne, F., 53, **820**
 Milstein, G. N., *XXVI*, *XXVII*, 61,
 256, 261, 278, 316, 318, 322, 324,
 352, 358, 477, 483, 499, 528, 572,
 576, 585, 586, 638, 642, 643, 645,
784–791, 808, 822, 823, 833
 Miltersen, K. R., 434, **823**
 Misawa, T., 784, **823**
 Mitsui, T., 572, 574, 786, 787, 789, 790,
799, 817, 823, 829
 Miyahara, Y., 53, **823**
 Mohammed, S.-E. A., 785, **812**
 Monfort, A., 398, 788, **809**
 Moore, J. B., *XXVII*, 309, 434–437,
 445, 788, **805**
 Mora, C. M., 786, 789, **823**
 Mordecki, E., 789, **823**
 Mortimer, I. K., 787, 790, **804**
 Morton, K., 734, 740, 791, **828**
 Mostafa, O. L., 786, **805**
 Muirhead, R. J., 71, 72, **824**
 Müller-Gronbach, T., 530, 786, 787,
789, 812, 824
 Musiela, M., *XXVII*, 163, **824**
 Nahum, E., 56, **829**
 Nakao, S., 785, **815**
 Nakayama, T., 785, **824**
 Nelsen, R., 783, **824**
 Nelson, D. B., 50, **824**
 Neuenkirch, A., 786, **816**
 Newton, N. J., 425, 430, 638, 642,
 785–788, 791, **824**
 Nguyen, T. T., 787, **819**
 Niederreiter, H., *XXVII*, 638, 716, 785,
824
 Nikitin, N. N., 786, **824**
 Nikitopoulos-Sklibosios, C., 56, 309,
347, 784, 786, 787, 798, 799
 Ninomiya, S., 789, **824**
 Nishimura, T., 716, 785, **822**
 Novikov, A. A., 53, 135, 647, 787, **823,**
824
 Ocone, D. L., 601, 606, **824**
 Ogawa, S., 786, 789, **817, 824**
 Øksendal, B., *XXVII*, 783, **824**
 Olkin, I., 66, **821**
 Ombach, J., 785, 788, **803**
 Osborne, M. F. M., 47, **825**
 Öttinger, H., 791, **822**
 Ozaki, T., 786, 788, 789, **814, 825, 830**
 Pagés, G., 716, 785, **825**
 Palleschi, V., 785, **821**
 Pan, J., *XXIII*, **804**
 Pang, S., 787, 790, **825**
 Panneton, F., 716, 785, **825**
 Papaspiliopoulos, O., 784, **796**
 Pardoux, E., 785, 787, 789, 790, **825**
 Paskov, S., 785, **825**
 Pasquali, S., 457, 788, **801**
 Pearson, N., 42, **825**
 Pearson, R. A., 785–787, **816**
 Pedersen, A. R., 390, 788, **825**
 Pelsser, A., 789, 791, **825**
 Peng, H., 417, **793**
 Penski, C., 787, **803**
 Petersen, W. P., 785–787, 789, 790,
816, 825

- Petryakov, M. G., 786, **817**
 Petterson, R., 638, 786, 791, **817, 825**
 Petzold, L., 787, 790, **800**
 Picard, J., 429, **825**
 Piccardi, M., 528, 785, 789, **798, 799**
 Pietersz, R., 789, **825**
 Pironneau, O., *XXVII*, 788, **793**
 Piterbarg, V., 783, **794**
 Pitman, J., 786, **794**
 Pivitt, B., 246, 785, **825**
 Pliska, S. R., 791, **827**
 Polson, N., 414, **805**
 Pope, S. B., 789, **810**
 Portait, R., 784, **795**
 Potters, M., 790, **827**
 Poulsen, R., 390, 398, 788, **801, 827**
 Pozo, R., 742, **796**
 Prakasa Rao, B. L. S., 389, 390, 788,
 827
 Press, S. J., 415, **827**
 Press, W. H., 64, 716, **827**
 Privault, N., 598, 790, **805**
 Prottter, P., *XXVII*, 13, 14, 23–25, 27,
 29–31, 33–35, 37, 43, 44, 58, 59,
 75, 379, 573, 783–786, 789, **802,**
 804, 813, 814, 817, 818, 820,
 827
 Qian, M. P., 786, **806**
 Qian, X., 791, **834**
 Qin, C., 790, **828**
 Rachev, S. T., 786, **807**
 Ramkrishna, D., 786, **827**
 Randall, C., 734, 744, 791, **832**
 Rao, C. R., 64, **827**
 Rao, N. J., 786, **827**
 Rascanu, A., 786, **828**
 Rathinam, M., 787, 790, **800**
 Rathinasamy, A., 787, 790, **828**
 Raviart, P. A., 740, **828**
 Razevig, V. D., 786, **824, 828**
 Rebolledo, R., 40, 747, 786, **827**
 Rebonato, R., 789, **828**
 Reimer, M., 791, **819**
 Renault, E., 398, 788, **809**
 Rendek, R., 61, 154, 158, 427, 781, 783,
 788, **827**
 Repin, Y. M., 785, **823**
 Revuz, D., 62, 89, 90, 168, 169, 783,
 828
 Ribémont, B., 787, **819**
 Richtmeyer, R., 734, 740, 791, **828**
 Riera, J. J., 787, **797**
 Rindisbacher, M., 790, **803**
 Ripley, B. D., 638, 791, **828**
 Ritchken, P., 703, **815**
 Ritter, K., 530, 786, 787, 789, **812**
 Roberts, G., 61, 783, 784, **796, 800**
 Rochet, J. C., 783, **808**
 Rodkina, A., 787, 790, **828**
 Rogers, L. C. G., 784, 785, 789, **828**
 Romine, C., 742, **796**
 Römisch, W., 788–790, **828**
 Roncoroni, A., *XXIII*, 56, **808**
 Ronghua, L., 790, **828**
 Rootzén, H., 786, **829**
 Rosa, A. M., 786, **808**
 Rosenfeld, E. R., 42, **821**
 Rosiński, J., 787, **794**
 Ross, S. A., 41, 146, 181, 533, 791, **802,**
 829
 Ross, S. M., 638, 639, 641, 669, 791,
 829
 Rössler, A., 786, 789, **819**
 Rozovski, B., 457, **802**
 Rubenthaler, S., 75, 783, **829**
 Rubinstein, M., 181, 791, **802, 829**
 Rubinstein, R. Y., 638, 791, **829**
 Rudd, A., 702, **814**
 Rümelin, W., 785, 787, **829**
 Runggaldier, W. J., *XXIII*, 45, 56, 456,
 457, 472, 473, 784, 788, 791, **796,**
 797, 801, 804, 806, 807, 809,
 827, 829
 Rutkowski, M., *XXVII*, 163, **797, 824**
 Ryashko, L. B., 787, 790, **829**
 Rydén, T., 784, **829**
 Sabelfeld, K. K., 791, **829**
 Sagsyan, K. V., 786, **804**
 Saisho, Y., 787, **815**
 Saito, Y., 572, 574, 786, 787, 790, **817,**
 829
 Salminen, P., *XXVII*, 132, 783, **797**
 Samuelides, Y., 56, **829**
 Samuelson, P. A., 14, 40, 47, 162, **829**
 Sandmann, K., 42, **829**

- Santa-Clara, P., 390, 788, 789, **798**
 Sanz-Solé, M., 787, **806**
 San Martin, J., 785, **820**
 Schachermayer, W., XXVII, 147, 783,
 784, **803, 804**
 Schäffler, S., 786, 787, **803**
 Schein, O., 786, 789, **819, 829**
 Schmitz-Abe, K., 789, **829**
 Schnell, S., XXIII, 375, 787, **833**
 Schoenmakers, J., 789–791, **814, 823,**
 829
 Scholes, M., 14, 37, 40, 47, 136, **797**
 Schönbucher, P. J., XXIII, XXVII, 56,
 167, **829**
 Schoutens, W., 784, 789, **830**
 Schrage, L., 638, 716, 785, 791, **798**
 Schroder, M., 40, **830**
 Schurz, H., 61, 245, 322, 326, 358,
 390, 393, 429, 477, 576, 585, 586,
 785–788, 790, **814, 816, 817,**
 823, 828–830
 Schwartz, E., 434, **823**
 Schwartz, E. S., 638, 713, 789, **820**
 Schweizer, M., 40, 128, 473, 475, 593,
 638, 645, 668, 671, 682, 687, 784,
 791, **806, 811–813, 830**
 Seneta, E., 53, 75, 783, **820**
 Sestovic, D., 790, **827**
 Setiawaty, B., 440, 788, **830**
 Sevilla-Peris, P., 786, **802, 814**
 Seydel, R., XXVII, 788, **830**
 Shahabuddin, P., 791, **808**
 Shardlow, T., 785, **799**
 Shaw, W., 734, 741, 744, 789, 791, **829,**
 830
 Shedler, G. S., 61, 783, **819**
 Shen, J., 785, **820**
 Shen, Y., 790, **821**
 Shephard, N., 14, 53, 75, 105, 390, 414,
 783, 788, **795, 805**
 Sheu, S.-J., 40, **806**
 Shi, L., 571, 587, 588, 781, 790, **827**
 Shimbo, K., 784, **813, 814**
 Shimizu, A., 786, **830**
 Shinozuka, M., 786, **830**
 Shirakawa, H., 40, 94, **803**
 Shiryaev, A. N., XXVII, 1, 4, 5, 23, 25,
 389, 421, 460–463, 609, 612, 615,
 631, 632, 747, 786, 788, **813, 819,**
 830
 Shkurko, I. O., 787, 790, **794**
 Shoji, I., 785, 788, 789, **830**
 Shreve, S. E., XXVII, 20, 27, 596, 610,
 815, 830
 Sickenberger, T., 787, **830**
 Sin, C. A., 784, **830**
 Singleton, K., XXIII, 788, **804**
 Siopacha, M., 789, **831**
 Situ, R., 58, **831**
 Skeel, R., 791, **834**
 Sklar, A., 783, **831**
 Skorokhod, A. V., 785, **808**
 Sloan, I. H., 785, **831**
 Slominski, L., 786, **831**
 Smereka, P., 786, **804**
 Smith, A. M., 787, 790, **831**
 Smith, G. D., 734, 791, **831**
 Smith, R. D., 61, 63, 107, 783, **831**
 Sole, J. L., 791, **819**
 Sondermann, D., 42, **829**
 Sørensen, H., 390, **831**
 Sørensen, M., 390, 393, 394, 397, 398,
 402, 411–413, 417, 788, **795–797,**
 801, 816–818, 831
 Sorini, L., 789, **809**
 Sottninen, T., 784, **796**
 Spigler, R., 572, 578, 787, 790, **811**
 Stacey, A. M., 789, **814**
 Stahl, G., 145, 151, 784, **827**
 Stegun, I. A., 78, **793**
 Stentoft, L., 789, **831**
 Stramer, O., 785, **801**
 Stricker, C., 146, 469, **794**
 Strittmatter, W., 786, **809**
 Stroud, A. H., 784, **831**
 Stroud, J. R., 789, **798**
 Stuart, A. M., 785, 787, 790, **812, 818**
 Studer, M., 194, 549, 784, 789, **830,**
 831
 Sufana, R., 70, 89, 111, 112, 784, **809**
 Sugiura, H., 786, 787, 790, **817**
 Sulem, A., XXVII, 783, **824**
 Sun, M., 788, **831**
 Sun, T. S., 42, **825**
 Sussmann, H. J., 785, **831**
 Symens, S., 789, **830**
 Szepessy, A., 789, **823, 831**

- Takahashi, A., 784, 791, **831**
 Takeuchi, A., 790, **831**
 Talay, D., 75, 316, 318, 495, 556, 565,
 783, 785–790, **795, 797, 809**,
 825, 827, 828, 832
 Talkner, P., 787, **816**
 Tan, K. S., 638, 785, **814**
 Tanaka, H., 791, **832**
 Tankov, P., *XXIII, XXVII*, 73, 75, 98,
 565, 783, 788, **802**
 Taraskin, A. F., 787, **832**
 Tauchen, G., 398, 414, 788, **801, 807**
 Tausworthe, R. C., 717, **832**
 Tavella, D., 734, 744, 791, **832**
 Tebaldi, C., 112, 784, 791, **803, 809**,
 832
 Teichmann, J., 784, 789, 791, **796, 831**,
 832
 Tempone, R., 789, **823, 831**
 Tetzlaff, U., 786, **833**
 Teukolsky, S. A., 64, 716, **827**
 Thalmaier, A., *XXVII*, 785, 788, 789,
 791, **802, 821**
 Thitsiklis, J., 789, **833**
 Thomas, J. M., 740, **828**
 Tian, T., 785, 787, 790, **800**
 Tocino, A., 784, 787, 789, 790, **833**
 Toft, K. J., 791, **798**
 Török, C., 786, **833**
 Torres, S., 785, **820**
 Touzi, N., 598, 638, 645, 785, 790, 791,
 797, 806, 807
 Toy, W., 42, 791, **797**
 Traub, J., 785, **825**
 Tretjakov, M. V., *XXVII*, 528, 784–789,
 823
 Tschitashvili, R. J., 786, **804**
 Tubaro, L., 495, 789, **832**
 Tudor, C., 785, 786, **828, 833**
 Tudor, M., 785, 786, **833**
 Tullie, T. A., 791, **806**
 Turnbull, S., *XXIII*, **813**
 Turner, T., *XXIII*, 375, 787, **833**
 Tysk, J., 784, **805**

 Unny, T. E., 786, **833**
 Uppal, R., 154, **803**
 Utzett, F., 791, **819**

 Valdes, P. A., 787, **797**
 Valkeila, E., 784, **796**
 van Dijk, D., 63, 783, **820**
 van Regenmortel, M., 789, **825**
 Vanden-Eijnden, E., 787, 790, **833**
 van der Hoek, J., *XXVII*, 457, 788, 791,
 805, 833
 van der Vorst, H., 742, **796**
 Van Roy, B., 789, **833**
 Vasicek, O. A., 3, 41, 46, **833**
 Ventzel, A. D., 789, 791, **833**
 Vetterling, W. T., 64, 716, **827**
 Vetzal, K. R., *XXVI*, **803**
 Victoir, N., 784, 789, 790, **820, 824**,
 833
 Vigo-Aguiar, J., 789, **833**
 Villafluer, L., 786, **814**
 Vives, J., 791, **819**
 Voltchkova, E., *XXVI*, **802**
 Vora, D., 638, **809**
 Vorst, T., 791, **825**

 Wadewitz, A., 786, **809**
 Wagner, W., 187, 261, 784, 785, 789,
 791, **819, 827, 833**
 Wakolbinger, A., 789, **828**
 Wang, H., 791, **805**
 Wang, L., 786, 789, **833**
 Wang, P., 787, **834**
 Wang, Z., 790, **834**
 Warin, X., 785, **808**
 Watanabe, S., *XXVII*, 21, 44, 58, 59,
 278, 783, 786, **813**
 Webber, N., *XXIII*, **795**
 Weeks, D., 638, **809**
 Weron, A., 785, 786, **813**
 White, A., 42, 638, 791, **812**
 Whitlock, P. A., 638, 791, **814**
 Widom, H., 612, **834**
 Wiktorsson, M., 784, **829**
 Willard, G. A., 784, **811, 820**
 Wilmott, P., 734, 743, 791, **834**
 Winkler, R., 786, 790, **799, 828**
 Wong, H. Y., *XXVII*, 788, **801**
 Wonham, W. M., 425, 788, **834**
 Wörner, J. H., 414, **834**
 Wozniakowski, H., 785, **831**
 Wright, D. J., 785–787, **834**
 Wright, I., 784, **817**

- Wu, K., 790, **804**
Wu, S., 787, 790, **834**
- Xia, X., 787, 790, **820**
Xiao, Y. J., 716, 785, **825**
Xu, C., 791, **834**
Xue, H., 786, 789, **833**
- Yamada, T., 785, **834**
Yamato, Y., 785, **834**
Yan, F., 785, **812**
Yannios, N., 786, **834**
Yen, V. V., 784, **834**
Yin, G., 786, 789, 790, **821**
Yingmin, H., 790, **828**
Yong, J. M., 785, **820**
Yor, M., XXVII, 14, 53, 62, 71, 75, 89,
90, 105, 168, 169, 783, **800, 804,**
808, 814, 828
- Yoshida, M., 791, **831**
Yuan, C., 785, 786, 789, 790, **821, 834**
- Zakai, M., 425, 788, **834**
Zanette, A., 598, 790, **795**
Zeitouni, O., 438, 788, **803**
Zhang, C., 790, **834**
Zhang, H., 790, **834**
Zhang, Q., 788, **815**
Zhao, X., 786, 789, **808**
Zhao, Y., 785, **820**
Zhaoguang, C., 790, **828**
Zheng, Z., 785, **832**
Zhou, G., 707, 709–711, 719, 791, **811**
Zou, G., 791, **834**
Zouraris, G. E., 789, **823, 831**
Zschiesche, H.-U., 786, **833**

Index

- absolute error, 251
- actuarial pricing formula, 149
- adapted, 4
- additive diffusion coefficient, 287
- additive noise, 45, 258, 493
- affine diffusion, 404
- affine model, 42
- almost exact
 - approximation, 63
 - simulation, 361
- American option, 700
- antithetic variates, 638
- appreciation rate, 141
- approximate
 - estimator, 396
 - GOPs, 153
 - Wonham filter, 428
- approximating probabilities, 251
- approximation
 - almost exact, 63
 - discrete-time, 247
 - finite difference, 737
 - jump-adapted, 348
 - Markov chain, 722
 - pathwise, 251
 - regular weak, 507
 - strong, 251
 - weak, 251, 478
- arbitrage
 - strong, 147
- ARCH diffusion model, 50, 113
 - multi-dimensional, 103, 115, 156
- Archimedian copula, 65
- asset-or-nothing binary, 163, 170
- asymptotically
 - p -stable, 573
 - stable, 575
- augmented sigma-algebra, 4
- Bachelier model, 40
- balanced
 - implicit Euler method, 585
 - implicit method, 322, 451
- basket option, 621
- Bayes rule, 134
- benchmark, 145
- benchmark approach, 139, 456
- benchmarked price, 145, 734
- Bessel function
 - modified, 71
- binary
 - asset-or-nothing, 163, 170
 - bond-or-nothing, 165, 171
- binomial model, 176
- binomial tree, 697
 - CRR, 698
- Black model, 41
- Black-Derman-Toy model, 42
- Black-Karasinski model, 42
- Black-Scholes
 - formula, 136
 - model, 40, 47
 - multi-asset, 49
 - multi-dimensional, 102
 - modification, 706
- bond-or-nothing binary, 165, 171

- Brownian motion, 6
- geometric, 47
- Cauchy-Schwarz inequality, 21
- CEV model, 40
- chain rule
 - classical, 34
- change of measure, 435
- Chebyshev inequality, 23
- CIR model, 41
- classical chain rule, 34, 192
- Clayton copula, 65
- coefficient function
 - Itô, 204
- commutativity condition
 - diffusion, 258, 516
 - jump, 286
- compensated
 - drift coefficient, 196
 - Poisson measure, 195, 211
 - Wagner-Platen expansion, 196, 206
 - weak Taylor scheme, 519
- compound Poisson process, 414
- concatenation operator, 197
- conditional
 - density, 5
 - expectation, 5
 - iterated, 5
 - variance, 472, 639
- conditionally Gaussian model, 460
- confidence interval, 480
- consistent, 240
- contingent claim, 470, 592
- continuation region, 128, 592
- continuous uncertainty, 139
- control variate, 670
 - general, 671
- convergence
 - order, 238
 - order of strong, 292
 - strong order, 251
 - theorem
 - jump-adapted, 535
 - Markov chain, 744
 - strong, 291
 - weak, 544
 - weak order, 519, 724
- copula, 65
 - Archimedian, 65
- Clayton, 65
- Gaussian, 65, 66
- correlation
 - property, 29
- counting process, 139, 193
- Courtadon model, 42
- covariance, 2
- covariation, 24
 - approximate, 24
 - property, 29
- Cramér-Lundberg model, 12, 52
- Crank-Nicolson method, 741
- credit worthiness, 10
- CRR binomial tree, 698
- d*-dimensional linear SDE, 48
- defaultable zero coupon bond, 166
- degree of implicitness, 239
- density
 - Gaussian, 66
- derivative-free
 - scheme, 267, 309, 334
 - strong order 1.0 scheme, 310
- deviation, 479
- difference method, 234
- differential equation
 - ordinary, 233
- diffusion
 - coefficient, 28, 248
 - additive, 287
 - multiplicative, 288
 - commutativity condition, 258, 516
 - stationary, 409
- Dirac delta function, 131
- discount rate process, 125
- discounted payoff with payoff rate, 126
- discrete-time approximation, 247
- discretization
 - spatial, 735
- distribution
 - generalized hyperbolic, 73
 - generalized inverse Gaussian, 80
 - non-central chi-square, 78
 - normal inverse Gaussian, 74
 - Student-*t*, 74
 - variance-gamma, 74
 - Wishart non-central, 80
- Diversification Theorem, 153
- diversified portfolio, 152

- drift coefficient, 28, 248
 - compensated, 196
- drift-implicit
 - Euler scheme, 316
 - simplified Euler scheme, 497
 - strong order 1.0 scheme, 318
 - strong scheme, 337
- equidistant time discretization, 246
- ergodicity, 132
- error
 - absolute, 251
 - covariance matrix, 420
 - function, 78
 - global discretization, 234
 - local discretization, 234
 - roundoff, 234
 - statistical, 478
 - strong, 368
 - systematic, 478
 - weak, 478, 699
- estimate
 - least-squares, 16
 - Monte Carlo, 478
- estimating function, 397
 - martingale, 398
- estimation
 - maximum likelihood, 389
 - parameter, 390, 413
- estimator
 - approximate, 396
 - HP, 678
 - maximum likelihood, 392, 394, 417
 - unbiased, 600
- Euler method, 234
 - balanced implicit, 585
- Euler scheme, 207, 246, 253, 276
 - fully implicit, 239
 - jump-adapted, 351
 - predictor-corrector, 576
 - simplified symmetric, 587
- Euler-Maruyama scheme, 246
- European call option, 166, 172
- event
 - driven uncertainty, 140
- exact simulation, 64
 - almost, 361
 - jump-adapted, 361
- exact solution
- almost, 105
- exercise boundary, 592
- expansion
 - compensated Wagner-Platen, 196, 206
 - general Wagner-Platen, 206
 - stochastic Taylor, 187
 - Stratonovich-Taylor, 191
 - Wagner-Platen, 187
- Expectation Maximization algorithm, 438
- expectation, conditional, 5
- explicit
 - filter, 430
 - solution, 45
 - strong order 1.5 scheme, 269
 - weak order 2.0 scheme, 491
- exponential
 - Lévy model, 14, 53
- extrapolation, 240, 495
 - Richardson, 241, 710
 - weak order 2.0, 495
 - weak order 4.0, 496
- Föllmer-Schweizer decomposition, 475
- fair price, 471
- fair zero coupon bond, 149
- Feynman-Kac formula, 124, 128
- filter, 436
 - distribution, 460
 - explicit, 430
 - finite-dimensional, 460
 - implicit, 432
 - Kalman-Bucy, 420
 - SDE, 421
- filtered probability space, 4
- filtering, 419, 456
- filtration, 4
 - natural, 4
- finite
 - difference approximation, 737
 - difference method, 734, 738
 - variation
 - property, 29
- finite-dimensional filter, 460
- finite-state jump model, 460
- first exit time, 128
- first hitting time, 19
- first jump commutativity condition, 515

- first moment estimate, 211
- first strong order balanced implicit method, 326
- Fokker-Planck equation, 130
- forward contract, 162
- forward price, 163
- fraction, 142, 474
- free lunches with vanishing risk, 147
- fully implicit
 - Euler scheme, 239
 - method, 741
- fundamental matrix, 48
- fundamental solution, 45

- gamma function
 - incomplete, 107
 - multi-dimensional, 71
- gamma process, 97
- Gaussian
 - copula, 65, 66
 - density, 66
 - elimination method, 742
 - transition density, 66
- general affine process, 96
- general market risk, 151
- general method of moments, 415
- general Wagner-Platen expansion, 206
- generalized
 - hyperbolic distribution, 73
 - inverse Gaussian distribution, 80
 - method of moment, 398
 - MMM, 123
 - volatility process, 141
- generator
 - random bit, 713, 716
 - random number
 - linear congruential, 716
- geometric Brownian motion, 47
 - multi-dimensional, 67
- geometric OU-process, 69
 - matrix, 96
- Girsanov
 - theorem, 436
 - transformation, 134, 643
- global discretization error, 234
- GOP, 144
 - approximate, 153
- Gronwall inequality, 22
- growth optimal portfolio, 115, 144, 457

- growth rate
 - long term, 145

- Heath-Platen method, 677
- hedgable part, 474
- hedge ratio, 178, 591, 595, 634
- hedging strategy, 473
- Hermite polynomial, 202, 262
- Heston model, 106, 111, 686
 - multi-dimensional, 108, 155
- Heston-Zhou trinomial tree, 708
- Heun method, 239
- Heyde-optimal, 403
- hidden Markov chain, 425
- hidden state, 420
- hierarchical set, 206, 290, 363, 519
- high intensity, 372, 378
- Ho-Lee model, 41
- Hölder inequality, 21, 22
- HP estimator, 678
- Hull-White model, 42

- implicit
 - filter, 432
 - methods, 583
 - weak Euler scheme, 498
 - weak order 2.0 scheme, 500
- incomplete
 - gamma function, 107
 - market, 458, 472
- indirect inference, 398
- indistinguishability, 57
- inequality
 - Cauchy-Schwarz, 21
 - Chebyshev, 23
 - Doob, 20
 - Gronwall, 22
 - Hölder, 21, 22
 - Jensen's, 22
 - Lyapunov, 22
 - Markov, 23
 - maximal martingale, 20
- information matrix, 393
- information set, 4
- insurance premium, 12
- integrable random variable, 4
- integral
 - Itô, 31
 - multiple stochastic, 279

- Riemann-Stieltjes, 28
- Stratonovich, 33
- symmetric, 33
- intensity, 139
 - high, 372, 378
 - measure, 13, 32
- interest rate model, 41
- interpolation, 247
- inverse Gaussian process, 97
- inverse transform method, 64, 77
- iterated conditional expectations, 5
- Itô coefficient function, 204, 363
 - compensated, 205
- Itô differential, 34
- Itô formula, 34, 36, 100
 - jump process, 36
 - multi-dimensional, 35
- Itô integral, 27, 31
 - equation, 39
 - multiple, 253
- Itô process with jumps, 44
- Jensen's inequality, 22
- jump coefficient, 376, 468
- jump commutativity condition, 286
 - first, 515
 - second, 515
- jump diffusion, 44, 413
- jump martingale, 140
- jump measure
 - Poisson, 44
- jump model
 - finite-state, 460
- jump ratio, 51
- jump size, 24
 - mark-independent, 511
- jump time, 9, 11
- jump-adapted
 - approximation, 348
 - derivative-free scheme, 355
 - drift-implicit scheme, 357
 - Euler scheme, 351, 524
 - simplified, 524
 - exact simulation, 361
 - predictor-corrector
 - Euler scheme, 531
 - scheme, 359
 - strong order γ Itô scheme, 367
 - strong order γ Taylor scheme, 363
- Taylor scheme
 - weak order 2.0, 525
- time discretization, 348, 377, 523
- weak order
 - 3.0 scheme, 527
 - weak order β Taylor scheme, 535
 - weak scheme
 - exact, 533
- Kalman-Bucy filter, 420
- knock-out-barrier option, 128
- Kolmogorov
 - backward equation, 125, 130, 598
 - forward equation, 130
 - stationary, 131
- Law of the Minimal Price, 147
- least-squares
 - estimate, 16
 - Monte Carlo, 638
- leverage effect, 41
- Lévy area, 261
- Lévy measure, 14
- Lévy process, 14, 31
 - matrix, 97
- Lévy's Theorem, 26
- Lie group symmetry, 75
- likelihood ratio, 390
- linear congruential random number generator, 716
- linear growth condition, 58
- linear SDE, 101
- linear test SDE, 572
- linearity, 29
- linearized
 - SDE, 603, 629
 - stochastic differential equation, 598
- Lipschitz condition, 58
- local discretization error, 234
- local martingale, 25
 - property, 29
- local risk minimization, 473
- lognormal
 - model, 42
- long term growth rate, 145
- Longstaff model, 42
- lookback
 - option, 627
- Lyapunov inequality, 22

- mark, 10, 11
- mark set, 13
- mark-independent jump size, 511
- market
 - activity, 119
 - incomplete, 458, 472
 - price
 - of event risk, 142
 - of risk, 141, 468
 - time, 120
- market risk
 - general, 151
 - specific, 151
- Markov chain
 - approximation, 722
 - second order, 730
 - third order, 731
 - convergence theorem, 744
 - hidden, 425
- Markov inequality, 23
- Marsaglia-Bray method, 716
- martingale, 16, 133
 - estimating function, 398
 - jump, 140
 - local, 25
 - measure
 - minimal equivalent, 593
 - property, 29
 - representation, 473, 591, 595, 596, 600, 605, 646
 - representation theorem, 607
 - strict local, 25
- matrix
 - commute, 48, 101
 - error covariance, 420
 - fundamental, 48
 - Riccati equation, 421
- maximum likelihood
 - estimation, 389
 - estimator, 392, 394, 417
- mean, 2
 - reverting, 451
- measure
 - change, 435
 - intensity, 13, 32
 - Lévy, 14
 - Poisson, 55
 - transformation, 645
- Merton model, 37, 50, 52, 55, 158, 161, 414, 562
- method
 - balanced implicit, 322, 451
 - Crank-Nicolson, 741
 - Euler, 234
 - finite difference, 734, 738
 - first strong order balanced implicit, 326
 - fully implicit, 741
 - Gaussian elimination, 742
 - Heath-Platen, 677
 - Heun, 239
 - inverse transform, 64
 - Marsaglia-Bray, 716
 - theta, 740
 - trapezoidal, 238
 - weak order
 - 1.0 predictor-corrector, 501
 - 2.0 predictor-corrector, 502
- Milstein scheme, 256, 258
- minimal
 - equivalent martingale measure, 593
 - market model, 40, 115, 168
- mixture of normals, 80
- MMM, 40
 - generalized, 123
 - multi-dimensional, 117
 - stylized, 116
- model
 - affine, 42
 - ARCH diffusion, 50, 113
 - Bachelier, 40
 - binomial, 176
 - Black, 41
 - Black-Derman-Toy, 42
 - Black-Karasinski, 42
 - Black-Scholes, 40, 47
 - CEV, 40
 - CIR, 41
 - conditionally Gaussian, 460
 - constant elasticity of variance, 40
 - Courtadon, 42
 - Cramér-Lundberg, 12, 52
 - exponential Lévy, 14, 53
 - Heston, 106, 111, 686
 - Ho-Lee, 41
 - Hull-White, 42
 - lognormal, 42

- Longstaff, 42
- Merton, 37, 50, 52, 55, 158, 414, 562
- minimal market, 40, 115, 168
- multi-factor, 458
- Ornstein-Uhlenbeck, geometric, 40
- Pearson-Sun, 42
- Platen, 42
- Sandmann-Sondermann, 42
- Vasicek, 41, 209
- modification
 - Black-Scholes, 706
- modified Bessel function, 71
- moment estimate for SDEs, 59
- Monte Carlo
 - estimate, 478
 - least-squares, 638
 - method, 477
 - simulation, 477
- multi-dimensional
 - ARCH diffusion, 156
 - Heston model, 155
 - MMM, 117
- multi-factor model, 458
- multi-index, 196, 206, 363
- multiple Itô integral, 253
 - approximate, 263
- multiple stochastic integral, 194, 197, 218
 - mixed, 283
- multiple Stratonovich integral, 253
 - approximate, 265
- multiplicative diffusion coefficient, 288
- natural filtration, 4
- non-central
 - chi-square distribution, 78, 107, 171
 - Wishart distribution, 80
- normal inverse Gaussian distribution, 74
- Novikov condition, 135, 646
- numerical instability, 244
- numerical stability, 571, 740
- numerically stable, 244
- optimal estimating equation, 403
- option
 - American, 700
 - basket, 621
 - European call, 166, 172
 - knock-out-barrier, 128
- lookback, 627
- path dependent, 638
- Optional Sampling Theorem, 20
- order of convergence, 238
- order of strong convergence, 251
- order of weak convergence, 491
- Ornstein-Uhlenbeck process, 46, 69
- orthogonal, 421
- parameter estimation, 390, 413
- partial
 - differential equation, 123
 - integro differential equation, 124
- path dependent option, 638
- pathwise approximation, 251
- payoff
 - continuous, 617
 - non-smooth, 556, 633
 - rate, 126
 - smooth, 549, 707
- Pearson-Sun model, 42
- pentanomial tree, 709
- Platen model, 42
- Platen scheme, 267
- Poisson
 - jump measure, 44
 - measure, 12, 14, 32, 55
 - compensated, 195, 211
 - process, 8
 - compensated, 17
 - compound, 10, 414
 - time transformed, 9
- portfolio, 142, 468
 - diversified, 152, 158
- predictable, 19
- predictor-corrector
 - Euler scheme, 327, 576
- price
 - benchmarked, 734
- pricing formula
 - binomial option, 181
 - Black-Scholes, 184
 - real world, 148, 177
 - risk neutral, 148
- primary security account, 140
- probability
 - measure change, 133
 - space, 1
 - filtered, 4

- probability (*cont.*)
 unnormalized conditional, 426
- process
 counting, 139, 193
 gamma, 97
 general affine, 96
 inverse Gaussian, 97
 Lévy, 14
 Ornstein-Uhlenbeck, 46, 69, 85
 multi-dimensional, 86
 Poisson, 8
 predictable, 19
 pure jump, 32, 193
 Radon-Nikodym derivative, 133
 square root, 62, 72, 259, 409
 squared Bessel, 77, 116, 210
 multi-dimensional, 89
 stationary, 2
 VG-Wishart, 98
 Wiener, 6
 Wishart, 70, 91, 94
- property
 correlation, 29
 covariation, 29
 finite variation, 29
 local martingale, 29
 martingale, 29
 regularity, 152
 supermartingale, 146
- pure jump
 process, 32, 193
 SDE, 375
- quadratic variation, 23
 approximate, 23
- Radon-Nikodym derivative, 133
- random
 bit generator, 713, 716
 measure, 12
 number, quasi, 638
 variable, integrable, 4
- real world pricing formula, 148, 177
- recovery rate, 10
- regular
 time discretization, 275
 weak approximation, 507
- regularity property, 152
- remainder set, 206, 363
- Richardson extrapolation, 241, 710
- Riemann-Stieltjes integral, 28
- risk neutral pricing formula, 148
- roundoff error, 234
- sample
 mean, 480
 space, 1
 variance, 480
- Sandmann-Sondermann model, 42
- savings bond, 172
- scale measure, 132
- scenario simulation, 233, 248, 251
- scheme
 derivative-free, 267, 309, 334
 strong order 1.0, 310
 drift-implicit
 Euler, 316
 simplified Euler, 497
 strong, 337
 strong order 1.0, 318
 Euler, 207, 246, 253, 276
 Euler-Maruyama, 246
 explicit strong order 1.5, 269
 explicit weak order 2.0, 491
 explicit weak order 3.0, 493
 implicit weak Euler, 498
 implicit weak order 2.0, 500
 jump-adapted
 derivative-free, 355
 drift-implicit, 357
 Euler, 351
 predictor-corrector, 359
 strong order γ Itô, 367
 strong order γ Taylor, 363
 Milstein, 256, 258
 Platen, 267
 predictor-corrector Euler, 327
 simplified Euler, 482
 strong order γ Itô, 331
 weak Taylor, 481
- score vector, 394
- SDE
 filter, 421
 linear, 45, 48, 101
 linearized, 603, 629
 pure jump, 375
 Stratonovich, 253
 with jumps, 52, 54

- second jump commutativity condition, 515
- second moment estimate, 212
- self-financing, 142, 468
- semimartingale, 30
 - special, 30
 - sequence of
 - approximate GOPs, 153
 - diversified portfolios, 152
 - JDMs
 - regular, 152
 - sigma-algebra, 1, 4
 - augmented, 4
 - simplified
 - Euler scheme, 482
 - symmetric predictor-corrector Euler scheme, 587
 - weak order
 - 2.0 Taylor scheme, 484
 - 3.0 Taylor scheme, 486
 - weak schemes, 712
- Sklar's theorem, 65
- smooth payoff, 707
- sparse matrix, 742
- spatial discretization, 735
- specific generalized volatility, 151
- specific market risk, 151
- square root process, 62, 72, 259, 409
- squared Bessel process, 77, 116, 210
- SR-process
 - correlated, 89
 - multi-dimensional, 72, 87
- stability region, 574
- state space, 1
- stationary
 - density, 131
 - diffusion, 409
 - independent increment, 6
 - process, 2
- statistical error, 478
- stochastic
 - differential equation, 38
 - linearized, 598
 - vector, 43
- integral
 - compensated multiple, 198
 - multiple, 194, 197, 218
- process, 1
- stationary independent increments, 14
- Taylor expansion, 187
- time change, 113
- stopping region, 592
- stopping time, 19
 - inaccessible, 19
- stratified sampling, 640
- Stratonovich equation, 191
- Stratonovich integral, 33
 - multiple, 253
- Stratonovich SDE, 253
- Stratonovich-Taylor expansion, 191
- strong
 - approximation, 251
 - arbitrage, 147
 - error, 368
 - solution, 57
- strong order
 - γ Itô scheme, 331
 - γ Taylor scheme, 290
 - 1.0 Taylor scheme, 278
 - 1.0 compensated Taylor scheme, 284
 - 1.5 Taylor scheme, 261
 - 2.0 Taylor scheme, 264, 266
- Student- t distribution, 74
- stylized MMM, 116
- sub-sigma-algebra, 5
- submartingale, 17
- subordination, 80
- supermartingale, 17, 25, 147
 - property, 146
 - strict, 17
- symmetric integral, 33
- systematic error, 478
- Taylor scheme
 - simplified
 - weak order 2.0, 484
 - weak order 3.0, 486
 - strong order γ , 290
 - strong order 1.0, 278
 - strong order 1.0 compensated, 284
 - strong order 1.5, 261
 - strong order 2.0, 264, 266
 - weak compensated, 519
 - weak order β , 519
 - weak order 2.0, 483, 508, 512
 - weak order 3.0, 485

- Taylor scheme (*cont.*)
 - weak order 4.0, 488
- theorem
 - martingale representation, 607
 - weak convergence, 746
- theta method, 740
- time change
 - stochastic, 113
- time discretization
 - equidistant, 23, 246
 - jump-adapted, 348, 377
 - regular, 275
- time set, 1
- tracking rate, 153
- transfer function, 575
- transform function, 398, 399
- transformed
 - Wiener process, 39
- transition density, 75
 - Gaussian, 66
- transition distribution, 63
- trapezoidal method, 238
- tree
 - binomial, 697
 - pentanomial, 709
 - trinomial, 704
 - Heston-Zhou , 708
- triangular solver, 742
- trinomial tree, 704
- unbiased estimator, 600
- unhedgable part, 474
- unique strong solution, 58
- uniqueness, 57
 - of strong solution, 57
- unnormalized conditional probability, 426
- variance, 2
 - conditional, 472, 639
- variance reduction, 637
 - by conditioning, 640
 - technique, 480, 638
- variance-gamma distribution, 74
- Vasicek model, 3, 41, 209
 - extended, 41
- vector
 - stochastic differential equation, 43
- VG-Wishart process, 98
- volatility, 468
- Wagner-Platen expansion, 187
- weak
 - approximation, 251, 478
 - convergence criterion, 478
 - convergence order, 491
 - convergence theorem, 746
 - error, 478, 699
 - relative, 552
- weak order
 - β Taylor scheme, 519
 - 1.0 predictor-corrector method, 501
 - 2.0 Taylor scheme, 483, 508, 512
 - 2.0 extrapolation, 495
 - 2.0 predictor-corrector method, 502
 - 3.0 Taylor scheme, 485
 - 4.0 Taylor scheme, 488
 - 4.0 extrapolation, 496
 - convergence, 519, 724
- weak scheme
 - derivative-free, 529
 - simplified, 712
- weak Taylor scheme, 481
- Wiener process, 6
 - correlated, 82
 - matrix, 83
 - multi-dimensional, 8
 - time changed, 85
 - transformed, 39
- Wishart process, 70, 91, 94
- Wonham filter
 - approximate, 428
 - problem, 425
- Zakai equation, 426
- zero coupon bond
 - defaultable, 167