

Springer Finance

Textbooks

Stéphane Crépey

Financial Modeling

A Backward Stochastic
Differential Equations Perspective

EXTRA
MATERIALS

extras.springer.com



Springer

Springer Finance

Textbooks

Editorial Board

Marco Avellaneda

Giovanni Barone-Adesi

Mark Broadie

Mark H.A. Davis

Emanuel Derman

Claudia Klüppelberg

Walter Schachermayer

Springer Finance Textbooks

Springer Finance is a programme of books addressing students, academics and practitioners working on increasingly technical approaches to the analysis of financial markets. It aims to cover a variety of topics, not only mathematical finance but foreign exchanges, term structure, risk management, portfolio theory, equity derivatives, and financial economics.

This subseries of Springer Finance consists of graduate textbooks.

For further volumes:

<http://www.springer.com/series/11355>

Stéphane Crépey

Financial Modeling

A Backward Stochastic
Differential Equations Perspective



Springer

Prof. Stéphane Crépey
Département de mathématiques,
Laboratoire Analyse & Probabilités
Université d'Evry Val d'Essonne
Evry, France

Additional material to this book can be downloaded from <http://extras.springer.com>
Password: [978-3-642-37112-7]

ISBN 978-3-642-37112-7

ISBN 978-3-642-37113-4 (eBook)

DOI 10.1007/978-3-642-37113-4

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013939614

Mathematics Subject Classification: 91G20, 91G60, 91G80

JEL Classification: G13, C63

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, 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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

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

To Irène and Camille

Preface

This is a book on financial modeling that emphasizes computational aspects. It gives a unified perspective on derivative pricing and hedging across asset classes and is addressed to all those who are interested in applications of mathematics to finance: students, quants and academics.

The book features backward stochastic differential equations (BSDEs), which are an attractive alternative to the more familiar partial differential equations (PDEs) for representing prices and Greeks of financial derivatives. First, BSDEs offer the most unified setup for presenting the financial derivatives pricing and hedging theory (as reflected by the relative compactness of the book, given its rather wide scope). Second, BSDEs are a technically very flexible and powerful mathematical tool for elaborating the theory with all the required mathematical rigor and proofs. Third, BSDEs are also useful for the numerical solution of high-dimensional nonlinear pricing problems such as the nonlinear CVA and funding issues which have become important since the great crisis [30, 80, 81].

Structure of the Book

Part I provides a course in stochastic processes, beginning at a quite elementary level in order to gently introduce the reader to the mathematical tools that are needed subsequently. Part II deals with the derivation of the pricing equations of financial claims and their explicit solutions in a few cases where these are easily obtained, although typically these equations have to be solved numerically as is done in Part III. Part IV provides two comprehensive applications of the book's approach that illustrate the versatility of simulation/regression pricing schemes for high-dimensional pricing problems. Part V provides a thorough mathematical treatment of the BSDEs and PDEs that are of fundamental importance for our approach. Finally, Part VI is an extended appendix with technical proofs, exercises and corrected problem sets.

Outline

Chapters 1–3 provide a survey of useful material from stochastic analysis. In Chap. 4 we recall the basics of financial theory which are necessary for understanding how the risk-neutral pricing equation of a generic contingent claim is derived. This chapter gives a unified view on the theory of pricing and hedging financial derivatives, using BSDEs as a main tool. We then review, in Chap. 5, benchmark models on reference derivative markets. Chapter 6 is about Monte Carlo pricing methods and Chaps. 7 and 8 deal with deterministic pricing schemes: trees in Chap. 7 and finite differences in Chap. 8.

Note that there is no hermetic frontier between deterministic and stochastic pricing schemes. In essence, all these numerical schemes are based on the idea of propagating the solution, starting from a surface of the time-space domain on which it is known (typically: the maturity of a claim), along suitable (random) “characteristics” of the problem. Here “characteristics” refers to Riemann’s method for solving hyperbolic first-order equations (see Chap. 4 of [191]). From the point of view of control theory, all these numerical schemes can be viewed as variants of Bellman’s dynamic programming principle [26]. Monte Carlo pricing schemes may thus be regarded as one-time-step multinomial trees, converging to a limiting jump diffusion when the number of space discretization points (tree branches) goes to infinity. The difference between a tree method in the usual sense and a Monte Carlo method is that a Monte Carlo computation mesh is stochastically generated and nonrecombining.

Prices of liquid financial instruments are given by the market and are determined by supply-and-demand. Liquid market prices are thus actually used by models in the “reverse-engineering” mode that consists in calibrating a model to market prices. This calibration process is the topic of Chap. 9. Once calibrated to the market, a model can be used for Greeking and/or for pricing more exotic claims (Greeking means computing risk sensitivities in order to set-up a related hedge).

Analogies and differences between simulation and deterministic pricing schemes are most clearly visible in the context of pricing by simulation claims with early exercise features (American and/or cancelable claims). Early exercisable claims can be priced by hybrid “nonlinear Monte Carlo” pricing schemes in which dynamic programming equations, similar to those used in deterministic schemes, are implemented on stochastically generated meshes. Such hybrid schemes are the topics of Chaps. 10 and 11, in diffusion and in pure jump setups, respectively. Again, this is presently becoming quite topical for the purpose of CVA computations.

Chapters 12–14 develop, within a rigorous mathematical framework, the connection between backward stochastic differential equations and partial differential equations. This is done in a jump-diffusion setting with regime switching, which covers all the models considered in the book.

Finally Chap. 15 gathers the most demanding proofs of Part V, Chap. 16 is devoted to exercises for Part I and Chap. 17 provides solved problem sets for Parts II and III.

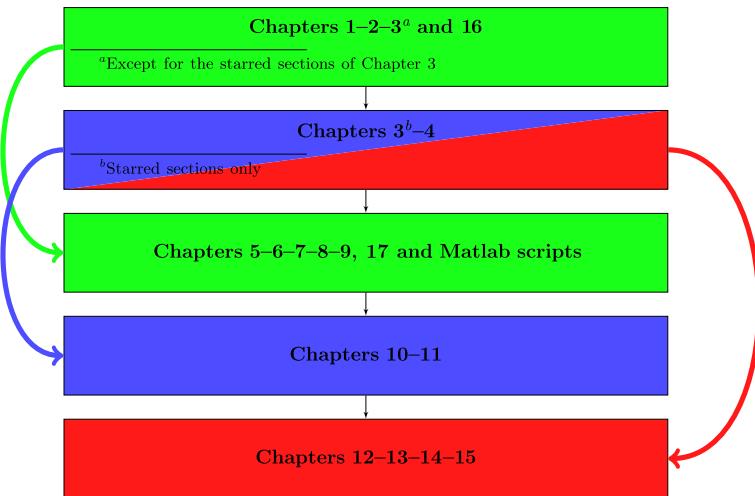


Fig. 1 Getting started with the book: roadmap of “a first and partial reading” for different audiences. *Green*: Students. *Blue*: Quants. *Red*: Academics

Roadmap

Given the dual nature of the proposed audience (scholars and quants), we have provided more background material on stochastic processes, pricing equations and numerical methods than is needed for our main purposes. Yet we have not avoided the sometimes difficult mathematical technique that is needed for deep understanding. So, for the convenience of readers, we signal sections that contain advanced material with an asterisk (*) or even a double asterisk (**) for the still more difficult portions.

Our ambition is, of course, that any reader should ultimately benefit from all parts of the book. We expect that an average reader will need two or three attempts at reading at different levels for achieving this objective. To provide additional guidance, we propose the following roadmap of what a “first and partial” reading of the book could be for three “stylized” readers (see Fig. 1 for a pictorial representation): a student (in “green” on the figure), a quant (“blue” audience) and an academic (“red”; the “blue and red” box in the chart may represent a valuable first reading for both quants and academics):

- for a graduate student (“green”), we recommend a first reading of the book at a classical quantitative and numerical finance textbook level, as follows in this order:
 - start by Chaps. 1–3 (except for the starred sections of Chap. 3), along with the accompanying (generally classical) exercises of Chap. 16,¹

¹Solutions of the exercises are available for course instructors.

- then jump to Chaps. 5 to 9, do the corrected problems of Chap. 17 and run the accompanying Matlab scripts (<http://extras.springer.com>);
- for a quant (“blue”):
 - start with the starred sections of Chap. 3, followed by Chap. 4,
 - then jump to Chaps. 10 and 11;
- for an academics or a PhD student (“red”):
 - start with the starred sections of Chap. 3, followed by Chap. 4,
 - then jump to Chaps. 12 to 14, along with the related proofs in Chap. 15.

The Role of BSDEs

Although this book isn’t exclusively dedicated to BSDEs, it features them in various contexts as a common thread for guiding readers through theoretical and computational aspects of financial modeling. For readers who are especially interested in BSDEs, we recommend:

- Section 3.5 for a mathematical introduction of BSDEs at a heuristic level,
- Chapter 4 for their general connection with hedging,
- Section 6.10 and Chap. 10 for numerical aspects and
- Part V and Chap. 15 for the related mathematics.

In Sect. 11.5 we also give a primer of CVA computations using simulation/regression techniques that are motivated by BSDE numerical schemes, even though no BSDEs appear explicitly. More on this will be found in [30], for which the present book should be a useful companion.

Bibliographic Guidelines

To conclude this preface, here are a few general references:

- on random processes and stochastic analysis, often with connections to finance (Chaps. 1–3): [149, 159, 167, 174, 180, 205, 228];
- on martingale modeling in finance (Chap. 4): [93, 114, 159, 191, 208, 245];
- on market models (Chap. 5): [43, 44, 58, 131, 146, 208, 230, 241];
- on Monte Carlo methods (Chap. 6): [133, 176, 226];
- on deterministic pricing schemes (Chaps. 7 and 8): [1, 12, 27, 104, 172, 174, 207, 248, 254, 256];
- on model calibration (Chap. 9): [71, 116, 213];
- on simulation/regression pricing schemes (Chaps. 10 and 11): [133, 136];
- on BSDEs and PDEs, especially in connection with finance (Chaps. 12–14): [96, 113, 114, 122, 164, 197, 224].

Acknowledgements²

This book grew out of various courses that I taught for many years in different quantitative finance master programs, starting with the one of my home university of Evry (MSc “M2IF”). I would like to thank all the students, who provided useful feedback through their questions and comments. The book also has a clear flavor of long run collaboration with my colleagues and friends Monique Jeanblanc, Tom Bielecki and Marek Rutkowski. A course given by Tom Bielecki at Illinois Institute of Technology was used as the basis of a first draft for Part I. Chapters 6–8 rely to some extent on the documentation of the option pricing software PREMIA developed since 1999 at INRIA and CERMICS under the direction of Bernard Lapeyre, Agnès Sulem and Antonino Zanette (see www.premia.fr); I thank Claude Martini and Katia Voltchkova for having let me use their contributions to PREMIA in preliminary versions of these chapters. Chapters 10 and 11 are joint work with my former PhD student Abdallah Rahal (see [88, 89, 229]). Thanks are due to Yann Le Cam, from BNP Paribas, and to anonymous referees for their comments on drafts of this book. Last but not least, thanks to Mark Davis who accepted to take the book in charge as Springer Finance Series editor and to Lester Senechal who helped with the final preparation for publication.

Paris, France
June 1 2013

Stéphane Crépey

²This work benefited from the support of the “Chaire Risque de Crédit” and of the “Chaire Marchés en Mutation”, Fédération Bancaire Française.

Contents

Part I An Introductory Course in Stochastic Processes

1 Some Classes of Discrete-Time Stochastic Processes	3
1.1 Discrete-Time Stochastic Processes	3
1.1.1 Conditional Expectations and Filtrations	3
1.2 Discrete-Time Markov Chains	6
1.2.1 An Introductory Example	6
1.2.2 Definitions and Examples	7
1.2.3 Chapman-Kolmogorov Equations	10
1.2.4 Long-Range Behavior	12
1.3 Discrete-Time Martingales	12
1.3.1 Definitions and Examples	12
1.3.2 Stopping Times and Optional Stopping Theorem	17
1.3.3 Doob's Decomposition	21
2 Some Classes of Continuous-Time Stochastic Processes	23
2.1 Continuous-Time Stochastic Processes	23
2.1.1 Generalities	23
2.1.2 Continuous-Time Martingales	24
2.2 The Poisson Process and Continuous-Time Markov Chains	24
2.2.1 The Poisson Process	27
2.2.2 Two-State Continuous Time Markov Chains	31
2.2.3 Birth-and-Death Processes	33
2.3 Brownian Motion	33
2.3.1 Definition and Basic Properties	34
2.3.2 Random Walk Approximation	35
2.3.3 Second Order Properties	36
2.3.4 Markov Properties	36
2.3.5 First Passage Times of a Standard Brownian Motion	38
2.3.6 Martingales Associated with Brownian Motion	39
2.3.7 First Passage Times of a Drifted Brownian Motion	42
2.3.8 Geometric Brownian Motion	43

3 Elements of Stochastic Analysis	45
3.1 Stochastic Integration	45
3.1.1 Integration with Respect to a Symmetric Random Walk	45
3.1.2 The Itô Stochastic Integral for Simple Processes	46
3.1.3 The General Itô Stochastic Integral	49
3.1.4 Stochastic Integral with Respect to a Poisson Process	51
3.1.5 Semimartingale Integration Theory (*)	51
3.2 Itô Formula	53
3.2.1 Introduction	53
3.2.2 Itô Formulas for Continuous Processes	54
3.2.3 Itô Formulas for Processes with Jumps (*)	57
3.2.4 Brackets (*)	60
3.3 Stochastic Differential Equations (SDEs)	62
3.3.1 Introduction	62
3.3.2 Diffusions	63
3.3.3 Jump-Diffusions (*)	69
3.4 Girsanov Transformations	71
3.4.1 Girsanov Transformation for Gaussian Distributions	71
3.4.2 Girsanov Transformation for Poisson Distributions	73
3.4.3 Abstract Bayes Formula	75
3.5 Feynman-Kac Formulas (*)	75
3.5.1 Linear Case	75
3.5.2 Backward Stochastic Differential Equations (BSDEs)	76
3.5.3 Nonlinear Feynman-Kac Formula	77
3.5.4 Optimal Stopping	78

Part II Pricing Equations

4 Martingale Modeling	83
4.1 General Setup	85
4.1.1 Pricing by Arbitrage	86
4.1.2 Hedging	95
4.2 Markovian Setup	102
4.2.1 Factor Processes	103
4.2.2 Markovian Reflected BSDEs and Obstacles PIDE Problems	104
4.2.3 Hedging Schemes	106
4.3 Extensions	108
4.3.1 More General Numéraires	108
4.3.2 Defaultable Derivatives	111
4.3.3 Intermittent Call Protection	119
4.4 From Theory to Practice	121
4.4.1 Model Calibration	121
4.4.2 Hedging	121
5 Benchmark Models	123
5.1 Black–Scholes and Beyond	123

5.1.1	Black–Scholes Basics	123
5.1.2	Heston Model	126
5.1.3	Merton Model	127
5.1.4	Bates Model	127
5.1.5	Log-Spot Characteristic Functions in Affine Models	127
5.2	Libor Market Model of Interest-Rate Derivatives	130
5.2.1	Black Formula	130
5.2.2	Libor Market Model	132
5.2.3	Caps and Floors	133
5.2.4	Adding Correlation	134
5.2.5	Swaptions	136
5.2.6	Model Simulation	137
5.3	One-Factor Gaussian Copula Model of Portfolio Credit Risk	138
5.3.1	Credit Derivatives	139
5.3.2	Gaussian Copula Model	140
5.4	Benchmark Models in Practice	144
5.4.1	Implied Parameters	144
5.4.2	Implied Delta-Hedging	146
5.5	Vanilla Options Fourier Transform Pricing Formulas	150
5.5.1	Fourier Calculus	150
5.5.2	Black–Scholes Type Pricing Formula	151
5.5.3	Carr–Madan Formula	153

Part III Numerical Solutions

6	Monte Carlo Methods	161
6.1	Uniform Numbers	161
6.1.1	Pseudo-Random Generators	162
6.1.2	Low-Discrepancy Sequences	164
6.2	Non-uniform Numbers	166
6.2.1	Inverse Method	166
6.2.2	Gaussian Pairs	167
6.2.3	Gaussian Vectors	169
6.3	Principles of Monte Carlo Simulation	170
6.3.1	Law of Large Numbers and Central Limit Theorem	170
6.3.2	Standard Monte Carlo Estimator and Confidence Interval .	170
6.4	Variance Reduction	171
6.4.1	Antithetic Variables	171
6.4.2	Control Variates	172
6.4.3	Importance Sampling	173
6.4.4	Efficiency Criterion	174
6.5	Quasi Monte Carlo	175
6.6	Greeking by Monte Carlo	176
6.6.1	Finite Differences	176
6.6.2	Differentiation of the Payoff	177
6.6.3	Differentiation of the Density	177

6.7	Monte Carlo Algorithms for Vanilla Options	178
6.7.1	European Call, Put or Digital Option	178
6.7.2	Call on Maximum, Put on Minimum, Exchange or Best of Options	179
6.8	Simulation of Processes	182
6.8.1	Brownian Motion	182
6.8.2	Diffusions	184
6.8.3	Adding Jumps	186
6.8.4	Monte Carlo Simulation for Processes	188
6.9	Monte Carlo Methods for Exotic Options	188
6.9.1	Lookback Options	190
6.9.2	Barrier Options	192
6.9.3	Asian Options	193
6.10	American Monte Carlo Pricing Schemes	194
6.10.1	Time-0 Price	195
6.10.2	Computing Conditional Expectations by Simulation	196
7	Tree Methods	199
7.1	Markov Chain Approximation of Jump-Diffusions	199
7.1.1	Kushner's Theorem	199
7.2	Trees for Vanilla Options	201
7.2.1	Cox–Ross–Rubinstein Binomial Tree	201
7.2.2	Other Binomial Trees	206
7.2.3	Kamrad–Ritchken Trinomial Tree	206
7.2.4	Multinomial Trees	207
7.3	Trees for Exotic Options	208
7.3.1	Barrier Options	208
7.3.2	Bermudan Options	209
7.4	Bidimensional Trees	210
7.4.1	Cox–Ross–Rubinstein Tree for Lookback Options	210
7.4.2	Kamrad–Ritchken Tree for Options on Two Assets	210
8	Finite Differences	213
8.1	Generic Pricing PIDE	213
8.1.1	Maximum Principle	214
8.1.2	Weak Solutions	215
8.2	Numerical Approximation	216
8.2.1	Finite Difference Methods	216
8.2.2	Finite Elements and Beyond	218
8.3	Finite Differences for European Vanilla Options	220
8.3.1	Localization and Discretization in Space	220
8.3.2	Theta-Schemes in Time	222
8.3.3	Adding Jumps	226
8.4	Finite Differences for American Vanilla Options	229
8.4.1	Splitting Scheme	229
8.5	Finite Differences for Bidimensional Vanilla Options	230

8.5.1	ADI Scheme	231
8.6	Finite Differences for Exotic Options	233
8.6.1	Lookback Options	233
8.6.2	Barrier Options	234
8.6.3	Asian Options	235
8.6.4	Discretely Path Dependent Options	237
9	Calibration Methods	243
9.1	The Ill-Posed Inverse Calibration Problem	243
9.1.1	Tikhonov Regularization of Nonlinear Inverse Problems	244
9.1.2	Calibration by Nonlinear Optimization	247
9.2	Extracting the Effective Volatility	247
9.2.1	Dupire Formula	248
9.2.2	The Local Volatility Calibration Problem	250
9.3	Weighted Monte Carlo	254
9.3.1	Approach by Duality	256
9.3.2	Relaxed Least Squares Approach	257
Part IV Applications		
10	Simulation/Regression Pricing Schemes in Diffusive Setups	261
10.1	Market Model	262
10.1.1	Underlying Stock	262
10.1.2	Convertible Bond	264
10.2	Pricing Equations and Their Approximation	265
10.2.1	Stochastic Pricing Equation	266
10.2.2	Markovian Case	267
10.2.3	Generic Simulation Pricing Schemes	268
10.2.4	Convergence Results	270
10.3	American and Game Options	272
10.3.1	No Call	272
10.3.2	No Protection	274
10.3.3	Numerical Experiments	275
10.4	Continuously Monitored Call Protection	277
10.4.1	Vanilla Protection	278
10.4.2	Intermittent Vanilla Protection	280
10.4.3	Numerical Experiments	282
10.5	Discretely Monitored Call Protection	283
10.5.1	“ l Last” Protection	284
10.5.2	“ l Out of the Last d ” Protection	285
10.5.3	Numerical Experiments	287
10.5.4	Conclusions	291
11	Simulation/Regression Pricing Schemes in Pure Jump Setups	293
11.1	Generic Markovian Setup	294
11.1.1	Generic Simulation Pricing Scheme	295
11.2	Homogeneous Groups Model of Portfolio Credit Risk	296

11.2.1	Hedging in the Homogeneous Groups Model	297
11.2.2	Simulation Scheme	299
11.3	Pricing and Greeking Results in the Homogeneous Groups Model	299
11.3.1	Fully Homogeneous Case	300
11.3.2	Semi-Homogeneous Case	302
11.4	Common Shocks Model of Portfolio Credit Risk	305
11.4.1	Example	308
11.4.2	Marshall-Olkin Representation	309
11.5	CVA Computations in the Common Shocks Model	310
11.5.1	Numerical Results	312
11.5.2	Conclusions	319
Part V Jump-Diffusion Setup with Regime Switching (**)		
12	Backward Stochastic Differential Equations	323
12.1	General Setup	323
12.1.1	Semimartingale Forward SDE	326
12.1.2	Semimartingale Reflected and Doubly Reflected BSDEs	328
12.2	Markovian Setup	334
12.2.1	Dynamics	336
12.2.2	Mapping with the General Set-Up	338
12.2.3	Cost Functionals	339
12.2.4	Markovian Decoupled Forward Backward SDE	340
12.2.5	Financial Interpretation	342
12.3	Study of the Markovian Forward SDE	343
12.3.1	Homogeneous Case	344
12.3.2	Inhomogeneous Case	348
12.4	Study of the Markovian BSDEs	351
12.4.1	Semigroup Properties	354
12.4.2	Stopped Problem	355
12.5	Markov Properties	358
13	Analytic Approach	359
13.1	Viscosity Solutions of Systems of PIDEs with Obstacles	359
13.2	Study of the PIDEs	362
13.2.1	Existence	362
13.2.2	Uniqueness	363
13.2.3	Approximation	365
14	Extensions	369
14.1	Discrete Dividends	369
14.1.1	Discrete Dividends on a Derivative	369
14.1.2	Discrete Dividends on Underlying Assets	371
14.2	Intermittent Call Protection	373
14.2.1	General Setup	374
14.2.2	Marked Jump-Diffusion Setup	377
14.2.3	Well-Posedness of the Markovian RIBSDE	379

14.2.4 Semigroup and Markov Properties	382
14.2.5 Viscosity Solutions Approach	384
14.2.6 Protection Before a Stopping Time Again	385
Part VI Appendix	
15 Technical Proofs (**)	391
15.1 Proofs of BSDE Results	391
15.1.1 Proof of Lemma 12.3.6	391
15.1.2 Proof of Proposition 12.4.2	392
15.1.3 Proof of Proposition 12.4.3	396
15.1.4 Proof of Proposition 12.4.7	397
15.1.5 Proof of Proposition 12.4.10	399
15.1.6 Proof of Theorem 12.5.1	400
15.1.7 Proof of Theorem 14.2.18	403
15.2 Proofs of PDE Results	405
15.2.1 Proof of Lemma 13.1.2	405
15.2.2 Proof of Theorem 13.2.1	405
15.2.3 Proof of Lemma 13.2.4	410
15.2.4 Proof of Lemma 13.2.8	416
16 Exercises	421
16.1 Discrete-Time Markov Chains	421
16.2 Discrete-Time Martingales	421
16.3 The Poisson Process and Continuous-Time Markov Chains	423
16.4 Brownian Motion	423
16.5 Stochastic Integration	424
16.6 Itô Formula	424
16.7 Stochastic Differential Equations	425
17 Corrected Problem Sets	427
17.1 Exit of a Brownian Motion from a Corridor	427
17.2 Pricing with a Regime-Switching Volatility	428
17.3 Hedging with a Regime-Switching Volatility	431
17.4 Jump-to-Ruin	434
References	441
Index	453

Part I

An Introductory Course in Stochastic Processes

The purpose of this part is to introduce a range of stochastic processes which are used as modeling tools in financial applications. It covers different classes of Markov processes: Markov chains (both discrete and continuous in time), the Poisson process, Brownian motion, diffusions and jump-diffusions. It also presents some aspects of stochastic calculus with emphasis on application to financial modeling.

This part relies significantly on the books of Lawler [180], Mikosch [205] and Shreve [245], to which the reader is referred for most of the proofs. It is written at a quite elementary level assuming only a basic knowledge of probability theory: random variables, exponential and Gaussian distributions, Bayes' formula, the law of large numbers and the central limit theorem. These are developed, for instance, in the first chapters of the book by Jacod and Protter [152]. Exercises for this part are provided in Chap. 16.

Notation The uniform distribution over a domain \mathcal{D} , the exponential distribution with parameter λ , the Poisson distribution with parameter γ and the Gaussian distribution with parameters μ and Γ (where Γ is a covariance matrix) are respectively denoted by $\mathcal{U}_{\mathcal{D}}$, \mathcal{P}_{γ} , \mathcal{E}_{λ} and $\mathcal{N}(\mu, \Gamma)$.

Throughout the book, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space. That is, Ω is a set of elementary events ω , \mathcal{F} is a σ -field of measurable events $A \subseteq \Omega$ (which thus satisfy certain closure properties: see for instance p. 7 of [152]), and $\mathbb{P}(A)$ is the probability of an event $A \in \mathcal{F}$. The expectation of a random variable (function of ω) with respect to \mathbb{P} is denoted by \mathbb{E} . By default, a random variable is \mathcal{F} -measurable; we omit any indication of dependence on ω in the notation; all inequalities between random variables are meant \mathbb{P} -almost surely; a real function of real arguments is Borel-measurable.

Chapter 1

Some Classes of Discrete-Time Stochastic Processes

1.1 Discrete-Time Stochastic Processes

1.1.1 Conditional Expectations and Filtrations

We first discuss the notions of conditional expectations and filtrations which are key to the study of stochastic processes.

Definition 1.1.1 Let ξ and $\varepsilon_1, \dots, \varepsilon_n$ be random variables. The conditional expectation $\mathbb{E}(\xi | \varepsilon_1, \dots, \varepsilon_n)$ is a random variable characterized by two properties.

- (i) The value of $\mathbb{E}(\xi | \varepsilon_1, \dots, \varepsilon_n)$ depends only on the values of $\varepsilon_1, \dots, \varepsilon_n$, i.e. we can write $\mathbb{E}(\xi | \varepsilon_1, \dots, \varepsilon_n) = \varrho(\varepsilon_1, \dots, \varepsilon_n)$ for some function ϱ . If a random variable can be written as a function of $\varepsilon_1, \dots, \varepsilon_n$, it is said to be measurable with respect to $\varepsilon_1, \dots, \varepsilon_n$.
- (ii) Suppose $A \in \mathcal{F}$ is any event that depends only on $\varepsilon_1, \dots, \varepsilon_n$. Let $\mathbb{1}_A$ denote the indicator function of A , i.e. the random variable which equals 1 if A occurs and 0 otherwise. Then

$$\mathbb{E}(\xi \mathbb{1}_A) = \mathbb{E}(\mathbb{E}(\xi | \varepsilon_1, \dots, \varepsilon_n) \mathbb{1}_A). \quad (1.1)$$

As a random variable, $\mathbb{E}(\xi | \varepsilon_1, \dots, \varepsilon_n)$ is a function of ω . Thus, the quantity $\mathbb{E}(\xi | \varepsilon_1, \dots, \varepsilon_n)(\omega)$ is a value of $\mathbb{E}(\xi | \varepsilon_1, \dots, \varepsilon_n)$. Sometimes a less formal, but more convenient notation is used: suppose $\omega \in \Omega$ is such that $\varepsilon_i(\omega) = x_i$, $i = 1, \dots, n$. Then, the notation $\mathbb{E}(\xi | \varepsilon_1 = x_1, \varepsilon_2 = x_2, \dots, \varepsilon_n = x_n)$ is used in place of $\mathbb{E}(\xi | \varepsilon_1, \dots, \varepsilon_n)(\omega)$. Likewise, for the value of the indicator random variable $\mathbb{1}_A$, where $A \in \mathcal{F}$, the notation $\mathbb{1}_{(\varepsilon_1, \dots, \varepsilon_n)(A)}(x_1, x_2, \dots, x_n)$ is used instead of $\mathbb{1}_A(\omega)$, with the understanding that $(\varepsilon_1, \dots, \varepsilon_n)(A) = \{(\varepsilon_1(\omega), \varepsilon_2(\omega), \dots, \varepsilon_n(\omega)), \omega \in A\}$.

Example 1.1.2 We illustrate the equality (1.1) with an example in which $n = 1$. Suppose that ξ and ε are discrete random variables and A is an event which involves ε . (For concreteness we may think of ε as the value of the first roll and ξ as the sum

of the first and second rolls in two rolls of a dice, and $A = \{\varepsilon \leq 2\}$). We have, using the Bayes formula in the fourth line:

$$\begin{aligned}\mathbb{E}(\mathbb{E}(\xi | \varepsilon) \mathbb{1}_A) &= \sum_x \mathbb{E}(\xi | \varepsilon = x) \mathbb{1}_{\varepsilon(A)}(x) \mathbb{P}(\varepsilon = x) \\ &= \sum_x \left[\sum_y y \mathbb{P}(\xi = y | \varepsilon = x) \right] \mathbb{1}_{\varepsilon(A)}(x) \mathbb{P}(\varepsilon = x) \\ &= \sum_y y \sum_x \mathbb{1}_{\varepsilon(A)}(x) \mathbb{P}(\xi = y, \varepsilon = x) \\ &= \sum_{y,x} (y \mathbb{1}_{\varepsilon(A)}(x)) \mathbb{P}(\xi = y, \varepsilon = x) = \mathbb{E}(\xi \mathbb{1}_A).\end{aligned}$$

For the example involving dice, taking $\varepsilon(A) = \{x \leq 2\}$

$$\begin{aligned}\mathbb{E}(\xi \mathbb{1}_A) &= \sum_{y,x} y \mathbb{1}_{\{x \leq 2\}} \mathbb{P}(\xi = y, \varepsilon = x) \\ &= \sum_y y \mathbb{P}(\xi = y, \varepsilon = 1) + \sum_y y \mathbb{P}(\xi = y, \varepsilon = 2) \\ &= \frac{1}{36}(2 + 3 + 4 + 5 + 6 + 7) + \frac{1}{36}(3 + 4 + 5 + 6 + 7 + 8) \\ &= \frac{27}{36} + \frac{33}{36} = \frac{5}{3}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\mathbb{E}(\xi | \varepsilon) \mathbb{1}_A] &= \mathbb{E}[(\varepsilon + 3.5) \mathbb{1}_{\{\varepsilon \leq 2\}}] \\ &= \sum_{x=1}^2 x \mathbb{P}(\varepsilon = x) + 3.5 \mathbb{P}(\varepsilon \leq 2) = \frac{3}{6} + 3.5 \frac{1}{3} = \frac{5}{3}.\end{aligned}$$

It will be convenient to make the notation more compact. If $\varepsilon_1, \varepsilon_2, \dots$ is a sequence of random variables we will use \mathcal{F}_n to denote the information contained in $\varepsilon_1, \dots, \varepsilon_n$, and we will write $\mathbb{E}(\xi | \mathcal{F}_n)$ for $\mathbb{E}(\xi | \varepsilon_1, \dots, \varepsilon_n)$.

We have that

$$\mathcal{F}_n \subseteq \mathcal{F}_m \subseteq \mathcal{F} \quad \text{if } 1 \leq n \leq m.$$

This is because the collection $\varepsilon_1, \dots, \varepsilon_n$ of random variables contains no more information than $\varepsilon_1, \dots, \varepsilon_n, \dots, \varepsilon_m$. A collection $\mathcal{F}_n, n = 1, 2, 3, \dots$, of σ -fields satisfying the above property is called a filtration.

1.1.1.1 Main Properties

0. Conditional expectation is a linear operation: if a, b are constants

$$\mathbb{E}(a\xi_1 + b\xi_2 | \mathcal{F}_n) = a\mathbb{E}(\xi_1 | \mathcal{F}_n) + b\mathbb{E}(\xi_2 | \mathcal{F}_n).$$

1. If ξ is measurable with respect to (i.e. is a function of) $\varepsilon_1, \dots, \varepsilon_n$ then

$$\mathbb{E}(\xi | \mathcal{F}_n) = \xi.$$

- 1'. If ξ is measurable with respect to $\varepsilon_1, \dots, \varepsilon_n$ then for any random variable χ

$$\mathbb{E}(\xi \chi | \mathcal{F}_n) = \xi \mathbb{E}(\chi | \mathcal{F}_n).$$

2. If ξ is independent of $\varepsilon_1, \dots, \varepsilon_n$, then

$$\mathbb{E}(\xi | \mathcal{F}_n) = \mathbb{E}(\xi).$$

- 2'. [See Sect. 1.4.4, rule 7 in Mikosch [205].] If ξ is independent of $\varepsilon_1, \dots, \varepsilon_n$ and χ is measurable with respect $\varepsilon_1, \dots, \varepsilon_n$, then for every function $\phi = \phi(y, z)$

$$\mathbb{E}(\phi(\xi, \chi) | \mathcal{F}_n) = \mathbb{E}_\xi \phi(\xi, \chi)$$

where $\mathbb{E}_\xi \phi(\xi, \chi)$ means that we “freeze” χ and take the expectation with respect to ξ , so $\mathbb{E}_\xi \phi(\xi, \chi) = \mathbb{E}\phi(\xi, x) |_{x=\chi}$.

3. The following property is a consequence of (1.1) if the event A is the entire sample space, so that $\mathbb{1}_A = 1$:

$$\mathbb{E}(\mathbb{E}(\xi | \mathcal{F}_n)) = \mathbb{E}(\xi).$$

- 3'. [Tower rule] If $m \leq n$, then

$$\mathbb{E}(\mathbb{E}(\xi | \mathcal{F}_n) | \mathcal{F}_m) = \mathbb{E}(\xi | \mathcal{F}_m).$$

4. [Projection; see Sect. 1.4.5 of Mikosch [205]] Let ξ be a random variable with $\mathbb{E}\xi^2 < +\infty$. The conditional expectation $\mathbb{E}(\xi | \mathcal{F}_n)$ is that random variable in $L^2(\mathcal{F}_n)$ which is closest to ξ in the mean-square sense, so

$$\mathbb{E}[\xi - \mathbb{E}(\xi | \mathcal{F}_n)]^2 = \min_{\chi \in L^2(\mathcal{F}_n)} \mathbb{E}(\xi - \chi)^2.$$

Example of Verification of the Tower Rule Let $\xi = \varepsilon_1 + \varepsilon_2 + \varepsilon_3$, where ε_i is the outcome of the i th toss of a fair coin, so that $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = 0) = 1/2$ and the ε_i are independent. Then

$$\begin{aligned} \mathbb{E}(\mathbb{E}(\xi | \mathcal{F}_2) | \mathcal{F}_1) &= \mathbb{E}(\mathbb{E}(\xi | \varepsilon_1, \varepsilon_2) | \varepsilon_1) \\ &= \mathbb{E}(\varepsilon_1 + \varepsilon_2 + \mathbb{E}\varepsilon_3 | \varepsilon_1) = \varepsilon_1 + \mathbb{E}\varepsilon_2 + \mathbb{E}\varepsilon_3 = \varepsilon_1 + 1, \end{aligned}$$

and

$$\mathbb{E}(\xi | \mathcal{F}_1) = \mathbb{E}(\xi | \varepsilon_1) = \varepsilon_1 + \mathbb{E}(\varepsilon_2 + \varepsilon_3) = \varepsilon_1 + 1/2 + 1/2 = \varepsilon_1 + 1.$$

1.2 Discrete-Time Markov Chains

1.2.1 An Introductory Example

Suppose that R_n denotes the short term interest rate prevailing on day $n \geq 0$. Suppose also that the rate R_n is a random variable which may only take two values: Low (L) and High (H), for every n . We call the possible values of R_n the states. Thus, we consider a random sequence: $R_n, n = 0, 1, 2, \dots$. Sequences like this are called discrete time stochastic processes.

Next, suppose that we have the following information available about the conditional probabilities:

$$\begin{aligned} & \mathbb{P}(R_n = j_n | R_0 = j_0, R_1 = j_1, \dots, R_{n-2} = j_{n-2}, R_{n-1} = j_{n-1}) \\ &= \mathbb{P}(R_n = j_n | R_{n-2} = j_{n-2}, R_{n-1} = j_{n-1}) \end{aligned} \quad (1.2)$$

for every $n \geq 2$ and for every sequence of states $(j_0, j_1, \dots, j_{n-2}, j_{n-1}, j_n)$, and

$$\begin{aligned} & \mathbb{P}(R_n = j_n | R_0 = j_0, R_1 = j_1, \dots, R_{n-2} = j_{n-2}, R_{n-1} = j_{n-1}) \\ &\neq \mathbb{P}(R_n = j_n | R_{n-1} = j_{n-1}) \end{aligned} \quad (1.3)$$

for **some** $n \geq 1$ and for **some** sequence of states $(j_0, j_1, \dots, j_{n-2}, j_{n-1}, j_n)$. In other words, we know that today's interest rate depends only on the values of interest rates prevailing on the two immediately preceding days (this is the condition (1.2) above). But the information contained in these two values will sometimes affect today's conditional distribution of the interest rate in a different way than the information provided only by yesterday's value of the interest rate (this is the condition (1.3) above).

The type of stochastic dependence subject to condition (1.3) is not the Markovian type of dependence (the meaning of which will soon be clear). However, due to condition (1.2) the stochastic process $R_n, n = 0, 1, 2, \dots$ can be “enlarged” (or augmented) to a so-called Markov chain that will exhibit the Markovian type of dependence.

To see this, let us note what happens when we create a new stochastic process $X_n, n = 0, 1, 2, \dots$, by enlarging the state space of the original sequence $R_n, n = 0, 1, 2, \dots$. To this end we define

$$X_n = (R_n, R_{n+1}).$$

Observe that the state space for the sequence $X_n, n = 0, 1, 2, \dots$ contains four elements: (L, L) , (L, H) , (H, L) and (H, H) . We will now examine conditional probabilities for the sequence $X_n, n = 0, 1, 2, \dots$:

$$\begin{aligned} & \mathbb{P}(X_n = i_n | X_0 = i_0, X_1 = i_1, \dots, X_{n-2} = i_{n-2}, X_{n-1} = i_{n-1}) \\ &= \mathbb{P}(R_{n+1} = j_{n+1}, R_n = j_n | R_0 = j_0, R_1 = j_1, \dots, R_{n-1} = j_{n-1}, R_n = j_n) \\ &= \mathbb{P}(R_{n+1} = j_{n+1} | R_0 = j_0, R_1 = j_1, \dots, R_{n-1} = j_{n-1}, R_n = j_n) \end{aligned}$$

which, by condition (1.2), is also equal to

$$\begin{aligned}\mathbb{P}(R_{n+1} = j_{n+1} | R_{n-1} = j_{n-1}, R_n = j_n) \\ = \mathbb{P}(R_{n+1} = j_{n+1}, R_n = j_n | R_{n-1} = j_{n-1}, R_n = j_n) \\ = \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1})\end{aligned}$$

for every $n \geq 1$ and for every sequence of states $(i_0, i_1, \dots, i_{n-1}, i_n)$. The enlarged sequence X_n exhibits the so-called Markov property.

1.2.2 Definitions and Examples

Definition 1.2.1 A random sequence X_n , $n \geq 0$, where X_n takes values in the discrete (finite or countable) set \mathcal{S} , is said to be a Markov chain with state space \mathcal{S} if it satisfies the Markov property:

$$\begin{aligned}\mathbb{P}(X_n = i_n | X_0 = i_0, X_1 = i_1, \dots, X_{n-2} = i_{n-2}, X_{n-1} = i_{n-1}) \\ = \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1})\end{aligned}\tag{1.4}$$

for every $n \geq 1$ and for every sequence of states $(i_0, i_1, \dots, i_{n-1}, i_n)$ from the set \mathcal{S} .

Every discrete time stochastic process satisfies the following property (given that the conditional probabilities are well defined):

$$\begin{aligned}\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_{n-2} = i_{n-2}, X_{n-1} = i_{n-1}, X_n = i_n) \\ = \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1 | X_0 = i_0) \times \dots \\ \times \mathbb{P}(X_n = i_n | X_0 = i_0, X_1 = i_1, \dots, X_{n-2} = i_{n-2}, X_{n-1} = i_{n-1}).\end{aligned}$$

Not every random sequence satisfies the Markov property. Sometimes a random sequence, which is not a Markov chain, can be transformed to a Markov chain by means of enlargement of the state space.

Definition 1.2.2 A random sequence X_n , $n = 0, 1, 2, \dots$, where X_n takes values in the set \mathcal{S} , is said to be a time-homogeneous Markov chain with the state space \mathcal{S} if it satisfies the Markov property (1.4) and, in addition,

$$\mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}) = q(i_{n-1}, i_n)\tag{1.5}$$

for every $n \geq 1$ and for every two of states i_{n-1}, i_n from the set \mathcal{S} , where $q : \mathcal{S} \times \mathcal{S} \rightarrow [0, 1]$ is some given function.

Time-inhomogeneous Markov chains can be transformed to time-homogeneous ones by including the time variable in the state vector, so we only consider time-homogeneous Markov chains in the sequel.

Definition 1.2.3 The (possibly infinite) matrix $Q = [q(i, j)]_{i,j \in \mathcal{S}}$ is called the (one-step) transition matrix for the Markov chain X_n .

The transition matrix for a Markov chain X_n is a stochastic matrix. That is, its rows can be interpreted as probability distributions, with nonnegative entries summing up to unity. To every pair (ϕ_0, Q) , where $\phi_0 = (\phi_0(i))_{i \in \mathcal{S}}$ is an initial probability distribution on \mathcal{S} and Q is a stochastic matrix, there corresponds some Markov chain with the state space \mathcal{S} . Such a chain can be constructed via the formula

$$\begin{aligned}\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_{n-2} = i_{n-2}, X_{n-1} = i_{n-1}, X_n = i_n) \\ = \phi_0(i_0)q(i_0, i_1)\dots q(i_{n-1}, i_n).\end{aligned}$$

In other words, the initial distribution ϕ_0 and the transition matrix Q determine a Markov chain completely by determining its finite dimensional distributions.

Remark 1.2.4 There is an obvious analogy with a difference equation:

$$x_n = ax_{n-1}, \quad n \geq 0.$$

The solution path (x_0, x_1, x_2, \dots) is uniquely determined by the initial condition x_0 and the transition rule a .

Example 1.2.5 Let ε_n , $n = 1, 2, \dots$ be i.i.d. (independent, identically distributed random variables) such that $\mathbb{P}(\varepsilon_n = -1) = p$, $\mathbb{P}(\varepsilon_n = 1) = 1 - p$. Define $X_0 = 0$ and, for $n \geq 1$,

$$X_n = X_{n-1} + \varepsilon_n.$$

The process X_n , $n \geq 0$, is a time-homogeneous Markov chain on the set $\mathcal{S} = \{\dots, -i, -i + 1, \dots, -1, 0, 1, \dots, i - 1, i, \dots\}$ of all integers, and the corresponding transition matrix Q is given by

$$q(i, i+1) = 1 - p, \quad q(i, i-1) = p, \quad i = 0, \pm 1, \pm 2, \dots$$

This is a random walk on the integers starting at zero. If $p = 1/2$, then the walk is said to be symmetric.

Example 1.2.6 Let ε_n , $n = 1, 2, \dots$ be i.i.d. such that $\mathbb{P}(\varepsilon_n = -1) = p$, $\mathbb{P}(\varepsilon_n = 1) = 1 - p$. Define $X_0 = 0$ and, for $n \geq 1$,

$$X_n = \begin{cases} -M & \text{if } X_{n-1} = -M \\ X_{n-1} + \varepsilon_n & \text{if } -M < X_{n-1} < M \\ M & \text{if } X_{n-1} = M. \end{cases}$$

The process $X_n, n \geq 0$, is a time-homogeneous Markov chain on $\mathcal{S} = \{-M, -M + 1, \dots, -1, 0, 1, \dots, M - 1, M\}$, and the corresponding transition matrix is Q given by

$$\begin{aligned} q(i, i+1) &= 1-p, & q(i, i-1) &= p, & -M < i < M \\ q(-M, -M) &= q(M, M) = 1. \end{aligned}$$

This is a random walk starting at zero, with absorbing boundaries at $-M$ and M . If $p = 1/2$, then the walk is said to be symmetric.

Example 1.2.7 Let $\varepsilon_n, n = 1, 2, \dots$ be i.i.d. such that $\mathbb{P}(\varepsilon_n = -1) = p, \mathbb{P}(\varepsilon_n = 1) = 1 - p$. Define $X_0 = 0$ and, for $n \geq 1$,

$$X_n = \begin{cases} -M + 1 & \text{if } X_{n-1} = -M \\ X_{n-1} + \varepsilon_n & \text{if } -M < X_{n-1} < M \\ M - 1 & \text{if } X_{n-1} = M. \end{cases}$$

The process $X_n, n \geq 0$, is a time-homogeneous Markov chain on $\mathcal{S} = \{-M, -M + 1, \dots, -1, 0, 1, \dots, M - 1, M\}$, and the corresponding transition matrix is Q given by

$$\begin{aligned} q(i, i+1) &= 1-p, & q(i, i-1) &= p, & -M < i < M \\ q(-M, -M+1) &= q(M, M-1) = 1. \end{aligned}$$

This is a random walk starting at zero with reflecting boundaries at $-M$ and M . If $p = 1/2$, then the walk is said to be symmetric.

Example 1.2.8 Let $\varepsilon_n, n = 0, 2, \dots$ be i.i.d. such that $\mathbb{P}(\varepsilon_n = -1) = p, \mathbb{P}(\varepsilon_n = 1) = 1 - p$. Then the stochastic process $X_n = \varepsilon_n, n \geq 0$, is a time-homogeneous Markov chain on $\mathcal{S} = \{-1, 1\}$, and the corresponding transition matrix is

$$Q = \begin{pmatrix} -1 & +1 \\ +1 & -1 \end{pmatrix} \begin{pmatrix} p & q \\ p & q \end{pmatrix}.$$

Here

$$q(i, -1) = \mathbb{P}(X_n = -1 | X_{n-1} = i) = \mathbb{P}(X_n = -1)$$

for $i = -1, 1$, and

$$q(i, 1) = \mathbb{P}(X_n = 1 | X_{n-1} = i) = \mathbb{P}(X_n = 1)$$

for $i = -1, 1$.

1.2.3 Chapman-Kolmogorov Equations

Definition 1.2.9 Given any two states $i, j \in \mathcal{S}$, the n -step transition probability $q_n(i, j)$ is defined as

$$q_n(i, j) = \mathbb{P}(X_n = j | X_0 = i)$$

for every $n \geq 0$. We define the n -step transition matrix Q_n as

$$Q_n = [q_n(i, j)]_{i,j \in \mathcal{S}}.$$

In particular, we have $q_1(i, j) = q(i, j)$ and

$$q_0(i, j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$

and thus $Q_1 = Q$ and $Q_0 = I$ (the identity matrix).

Lemma 1.2.10

- (i) We have $q_n(i, j) = \mathbb{P}(X_{k+n} = j | X_k = i)$ for $n, k \geq 0$.
- (ii) The following representation holds for the n -step transition matrix:

$$Q_n = Q^n$$

for $n \geq 0$ (by definition we have $Q^0 = I$).

Proof (i) Using the linearity of conditional expectation in the first line, the Bayes formula in the second and third ones and the Markov property in the fourth line, we have:

$$\begin{aligned} & \mathbb{P}(X_{k+n+1} = j | X_k = i) \\ &= \sum_{r \in \mathcal{S}} \mathbb{P}(X_{k+n+1} = j, X_{k+n} = r | X_k = i) \\ &= \sum_{r \in \mathcal{S}} \frac{\mathbb{P}(X_{k+n+1} = j, X_{k+n} = r, X_k = i) \mathbb{P}(X_{k+n} = r, X_k = i)}{\mathbb{P}(X_{k+n} = r, X_k = i) \mathbb{P}(X_k = i)} \\ &= \sum_{r \in \mathcal{S}} \mathbb{P}(X_{k+n+1} = j | X_{k+n} = r, X_k = i) \mathbb{P}(X_{k+n} = r | X_k = i) \\ &= \sum_{r \in \mathcal{S}} \mathbb{P}(X_{k+n+1} = j | X_{k+n} = r) \mathbb{P}(X_{k+n} = r | X_k = i). \end{aligned}$$

The proof is concluded by induction on n .

(ii) By virtue of the arguments already used in the proof of part (i), we have:

$$\begin{aligned}
q_{n+1}(i, j) &= \mathbb{P}(X_{n+1} = j \mid X_0 = i) \\
&= \sum_{r \in S} \mathbb{P}(X_{n+1} = j, X_n = r \mid X_0 = i) \\
&= \sum_{r \in S} \frac{\mathbb{P}(X_{n+1} = j, X_n = r, X_0 = i)\mathbb{P}(X_n = r, X_0 = i)}{\mathbb{P}(X_n = r, X_0 = i)\mathbb{P}(X_0 = i)} \\
&= \sum_{r \in S} \mathbb{P}(X_{n+1} = j \mid X_n = r, X_0 = i)\mathbb{P}(X_n = r \mid X_0 = i) \\
&= \sum_{r \in S} \mathbb{P}(X_{n+1} = j \mid X_n = r)\mathbb{P}(X_n = r \mid X_0 = i) \\
&= \sum_{r \in S} q(r, j)q_n(i, r),
\end{aligned}$$

by part (i). Therefore, $Q_{n+1} = Q Q_n$. The proof is concluded by induction on n . \square

Proposition 1.2.11 *The following Chapman-Kolmogorov semigroup equation is satisfied:*

$$Q_{m+n} = Q_m Q_n = Q_n Q_m$$

for every $m, n \geq 0$. Equivalently,

$$q_{m+n}(i, j) = \sum_{k \in S} q_m(i, k)q_n(k, j) = \sum_{k \in S} q_n(i, k)q_m(k, j)$$

for every $m, n \geq 0$ and every $i, j \in S$.

Proof Lemma 1.2.10 yields that $Q_{m+n} = Q^{m+n} = Q^m Q^n = Q_m Q_n$. \square

The Chapman-Kolmogorov equation provides the basis for the first step analysis:

$$Q_{n+1} = Q Q_n. \quad (1.6)$$

The last step analysis would be $Q_{n+1} = Q_n Q$. These equations can also be written as

$$\Delta Q_{n+1} = A Q_n = Q_n A,$$

where $\Delta Q_{n+1} = Q_{n+1} - Q_n$ and $A = Q - I$. Note that the diagonal elements of A are negative and that the rows sum to 0. The matrix A is called the generator for any Markov chain associated with Q .

Definition 1.2.12 The (unconditional) n -step probabilities $\phi_n(i)$ are defined as

$$\phi_n(i) = \mathbb{P}(X_n = i)$$

for every $n \geq 0$. In particular, $\phi_0(i) = \mathbb{P}(X_0 = i)$ (the initial probabilities).

We will use the notation $\phi_n = [\phi_n(i)]_{i \in \mathcal{S}}$. This is a (possibly infinite) row-vector representing the distribution of the states of the Markov process at time n .

Proposition 1.2.13 *We have*

$$\phi_n = \phi_0 Q^n$$

for every $n \geq 0$.

Proof It is straightforward:

$$\mathbb{P}(X_n = j) = \sum_{i \in \mathcal{S}} \mathbb{P}(X_0 = i) \mathbb{P}(X_n = j | X_0 = i) = \sum_{i \in \mathcal{S}} \phi_0(i) q_n(i, j).$$

We already know that the n -step transition probability $q_n(i, j)$ is the (i, j) entry of the matrix Q^n . \square

A recursive equation for the n -step transition probabilities (the conditional probabilities $\mathbb{P}(X_n = j | X_0 = i)$) is:

$$Q_{n+1} = Q_n Q, \quad n \geq 0,$$

with the initial condition $Q_0 = I$. A recursive equation for the unconditional probabilities $\mathbb{P}(X_n = j)$ is:

$$\phi_{n+1} = \phi_n Q, \quad n \geq 0,$$

with the initial condition ϕ_0 corresponding to the distribution of X_0 .

1.2.4 Long-Range Behavior

By the long-range behavior of a Markov chain we mean the behavior of the conditional probabilities Q_n and the unconditional probabilities ϕ_n for large n . In view of the fact that $\phi_n = \phi_0 Q_n = \phi_0 Q^n$, this essentially reduces to the behavior of the powers Q^n of the transition matrix for large n .

1.3 Discrete-Time Martingales

1.3.1 Definitions and Examples

In the following definition, \mathcal{F}_n denotes the information contained in a sequence $\varepsilon_1, \dots, \varepsilon_n$ of random variables. A process Y such that Y_n is measurable with respect to \mathcal{F}_n for every n is said to be adapted to the filtration $\mathbb{F} = (\mathcal{F}_n)_{n \geq 0}$. We will normally consider adapted processes only. Sometimes we abusively say that “ Y_n is

adapted to the filtration \mathcal{F}_n , Y_n is a martingale with respect to \mathcal{F}_n ”, etc. instead of “ Y is adapted to the filtration \mathbb{F} , Y is a martingale with respect to \mathbb{F} ”.

Definition 1.3.1 A stochastic process Y_n , $n \geq 0$, is a martingale with respect to a filtration \mathcal{F}_n , $n \geq 0$: if

- (i) $\mathbb{E}|Y_n| < +\infty$, for $n \geq 0$;
- (ii) $\mathbb{E}(Y_m | \mathcal{F}_n) = Y_n$, for $m \geq n$.

Condition (i) assures that the conditional expectations are well defined. Condition (ii) implies that Y_n is \mathcal{F}_n -measurable. When we say that Y_n is a martingale without reference to \mathcal{F}_n , $n \geq 0$, we understand that \mathcal{F}_n is the information contained in Y_0, \dots, Y_n .

In order to verify (ii) it is enough to show that, for all n ,

$$\mathbb{E}(Y_{n+1} | \mathcal{F}_n) = Y_n, \quad (1.7)$$

since by the tower rule

$$\mathbb{E}(Y_{n+2} | \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(Y_{n+2} | \mathcal{F}_{n+1}) | \mathcal{F}_n) = \mathbb{E}(Y_{n+1} | \mathcal{F}_n) = Y_n,$$

and so on. We also note that, for every n ,

$$\mathbb{E}Y_{n+1} = \mathbb{E}\mathbb{E}(Y_{n+1} | \mathcal{F}_n) = \mathbb{E}Y_n,$$

so that a martingale is a process with a constant mean. Because of the property (1.7), a martingale is thought of as a model of a fair game. A process, which can be thought of as a model of a favorable (unfavorable) game, is a submartingale (supermartingale) as defined below.

Definition 1.3.2 A stochastic process Y is a submartingale (supermartingale) with respect to \mathbb{F} if:

- (i) Y_n is \mathcal{F}_n -measurable and $\mathbb{E}|Y_n| < +\infty$, for $n \geq 0$;
- (ii) $\mathbb{E}(Y_m | \mathcal{F}_n) \geq (\leq) Y_n$, for $m \geq n \geq 0$.

We note that the measurability condition in item (i) is automatically satisfied for a martingale.

By the conditional Jensen inequality, a convex (respectively concave) transform of a martingale is a submartingale (respectively supermartingale) (provided it is integrable).

Example 1.3.3 (Martingales associated with a driftless random walk) Let ε_i , $i \geq 1$ be i.i.d. with $\mathbb{E}\varepsilon_i = 0$, $\mathbb{E}\varepsilon_i^2 = \sigma^2 < +\infty$. Given a constant x , we verify that

$$S_n = x + \sum_{i=1}^n \varepsilon_i$$

and

$$M_n = S_n^2 - n\sigma^2$$

are martingales with respect to \mathcal{F}_n , the information contained in $\varepsilon_1, \dots, \varepsilon_n$, or equivalently in S_0, \dots, S_n . We have:

- (i) $\mathbb{E}|S_n| \leq |x| + \sum_{i=1}^n \mathbb{E}|\varepsilon_i| < +\infty$, and
- (ii) $\mathbb{E}(S_{n+1} | \mathcal{F}_n) = S_n$.

Similarly

- (i) $\mathbb{E}|M_n| \leq \mathbb{E}S_n^2 + n\sigma^2 = x^2 + n\sigma^2 + n\sigma^2 = x^2 + 2n\sigma^2 < +\infty$,
- (ii) and

$$\begin{aligned} \mathbb{E}(M_{n+1} | \mathcal{F}_n) &= \mathbb{E}[S_{n+1}^2 - (n+1)\sigma^2 | \mathcal{F}_n] \\ &= \mathbb{E}[S_n^2 + \varepsilon_{n+1}^2 + 2S_n\varepsilon_{n+1} - (n+1)\sigma^2 | \mathcal{F}_n] \\ &= S_n^2 + \sigma^2 + 0 - (n+1)\sigma^2 = M_n. \end{aligned}$$

Example 1.3.4 (Wald's martingale) Let ε_i , $i \geq 1$, be i.i.d. with $\mathbb{E}\varepsilon_i^2 < +\infty$. Set $S_n = x + \sum_{i=1}^n \varepsilon_i$, and let

$$m(\theta) = \mathbb{E}[\exp(\theta\varepsilon_i)], \quad -\infty < \theta < +\infty,$$

be the moment generating function of ε_i . Define

$$Z_n = \frac{\exp(\theta S_n)}{[m(\theta)]^n}, \quad n \geq 0.$$

We verify that Z_n is a martingale for every θ . We have

$$\begin{aligned} \mathbb{E}|Z_n| &= \frac{\mathbb{E}[\exp(\theta S_n)]}{[m(\theta)]^n} = \frac{\exp(\theta x)[m(\theta)]^n}{[m(\theta)]^n} = \exp(\theta x) \\ \mathbb{E}(Z_{n+1} | \mathcal{F}_n) &= \mathbb{E}\left(\frac{\exp(\theta S_{n+1})}{m(\theta)^{n+1}} | \mathcal{F}_n\right) = \frac{\exp(\theta S_n)}{m(\theta)^{n+1}} \mathbb{E} \exp(\theta \varepsilon_{n+1}) = Z_n. \end{aligned}$$

Now, suppose that $x = 0$ and each ε_i is normally distributed with mean μ and variance σ^2 . The moment generating function of $\varepsilon \sim \mathcal{N}(\mu, \sigma^2)$ is

$$m(\theta) = \mathbb{E} \exp(\theta\varepsilon) = \exp\left(\theta\mu + \frac{(\theta\sigma)^2}{2}\right).$$

We have $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$ and

$$Z_n = \exp\left(-\theta n\mu - \frac{1}{2}\theta^2 n\sigma^2 + \theta S_n\right).$$

This model is related to so-called geometric Brownian motion with a drift.

Example 1.3.5 (Martingales associated with a drifted random walk) Let ε_i , $i \geq 1$, be i.i.d. random variables with $\mathbb{P}(\varepsilon_i = 1) = p$, $\mathbb{P}(\varepsilon_i = -1) = 1 - p =: q$ for some $0 < p < 1$. Set $S_n = x + \sum_{i=1}^n \varepsilon_i$. We note that $\mathbb{E}\varepsilon_i = p - q =: \mu$ and $\mathbb{V}\text{ar } \varepsilon_i = 1 - \mu^2 = 4pq$. We now verify that

$$\bar{S}_n \equiv S_n - n\mu$$

and

$$Z_n = (q/p)^{S_n}$$

are martingales. That \bar{S}_n is a martingale with respect to $\sigma(\varepsilon_1, \dots, \varepsilon_n)$ follows immediately from Example 1.3.3 by writing

$$\bar{S}_n \equiv x + \sum_{i=1}^n (\varepsilon_i - \mu).$$

We now show that Z_n is a martingale of Wald type. We have, for $i > 0$,

$$m(\theta) = \mathbb{E}[\exp(\theta\varepsilon_i)] = p \exp(\theta) + q \exp(-\theta).$$

If we choose $\theta = \ln(q/p)$, then $m(\theta) = 1$ and Wald's martingale takes the form

$$\exp[\ln(q/p)S_n] = (q/p)^{S_n} = Z_n.$$

Example 1.3.6 Consider a sequence of independent games in each of which one wins \$1 with probability p or loses \$1 with probability $1 - p$. Let ε_n , $n \geq 1$ be a sequence of i.i.d. random variables indicating the outcome of the n th game, with

$$\mathbb{P}(\varepsilon_n = 1) = p = 1 - \mathbb{P}(\varepsilon_n = -1), \quad n \geq 1.$$

We note that $\mathbb{E}\varepsilon_n = 2p - 1$. Suppose we play a betting strategy based on the past history of the game, that is, the amount ζ_{n+1} bet on the $(n+1)$ th game is

$$\zeta_{n+1} = \zeta_{n+1}(\varepsilon_1, \dots, \varepsilon_n), \quad n \geq 0$$

with $\zeta_{n+1} \geq 0$. Let Y_n , $n \geq 1$, denote our fortune after n games and set $Y_0 = 0$. Then

$$Y_{n+1} = Y_n + \zeta_{n+1}(\varepsilon_1, \dots, \varepsilon_n)\varepsilon_{n+1}, \quad n \geq 0.$$

Now denote by \mathcal{F}_n the information contained in $\varepsilon_0, \dots, \varepsilon_n$ and consider

$$\begin{aligned} \mathbb{E}(Y_{n+1} | \mathcal{F}_n) &= \mathbb{E}[Y_n + \zeta_{n+1}(\varepsilon_1, \dots, \varepsilon_n)\varepsilon_{n+1} | \mathcal{F}_n] \\ &= Y_n + \zeta_{n+1}(\varepsilon_1, \dots, \varepsilon_n)\mathbb{E}(\varepsilon_{n+1}) \\ &\begin{cases} = Y_n & \text{if } \mathbb{E}(\varepsilon_{n+1}) = 0 \quad \Leftrightarrow \quad p = \frac{1}{2} \\ \leq Y_n & \text{if } \mathbb{E}(\varepsilon_{n+1}) \leq 0 \quad \Leftrightarrow \quad p < \frac{1}{2} \\ \geq Y_n & \text{if } \mathbb{E}(\varepsilon_{n+1}) \geq 0 \quad \Leftrightarrow \quad p > \frac{1}{2} \end{cases} \end{aligned}$$

Thus when $p = \frac{1}{2}$ (respectively $< \frac{1}{2}$ or $\frac{1}{2}$), Y_n is a martingale (respectively supermartingale or submartingale). Observe that when $p = \frac{1}{2}$, no matter what betting strategy is used in the class of strategies based on the past history of the game, we have $\mathbb{E}Y_n = \mathbb{E}Y_0 = 0$ for every n .

Now recall Examples 1.3.3 and 1.3.5 above. If $p = \frac{1}{2}$ (respectively $< \frac{1}{2}$ or $\frac{1}{2}$), then the process

$$S_n = \varepsilon_1 + \cdots + \varepsilon_n, \quad n \geq 0,$$

is a martingale (respectively supermartingale or submartingale). Next, observe that $\zeta_n = \zeta_n(\varepsilon_1, \dots, \varepsilon_{n-1})$ is \mathcal{F}_{n-1} -measurable for every $n \geq 1$. Such a process is said to be predictable with respect to the filtration \mathcal{F}_n . Our fortune Y_n can be written as

$$Y_n = \sum_{k=1}^n \zeta_k(S_k - S_{k-1}), \quad n \geq 0.$$

This expression is a martingale transform of the process S_n by the process ζ_n and is the discrete counterpart of a stochastic integral $\int \zeta dS$. We know that Y_n is a martingale (and also since $\zeta \geq 0$: respectively a supermartingale, submartingale) if S_n is a martingale (respectively a supermartingale, submartingale).

Example 1.3.7 (Doubling strategy) This example is a special case of Example 1.3.6 with $p = \frac{1}{2}$ and uses the following strategy. We bet \$1 on the first game. We stop if we win. If not, we double your bet. If we win, we stop betting (i.e. set $\zeta_n = 0$ for all greater n). Otherwise, we keep doubling your bet until we eventually win. This is a very attractive betting strategy, which involves a random stopping rule: we stop when we win. Let Y_n denote our fortune after n games. Assume $Y_0 = 0$. We already know from Example 1.3.6 that Y_n is a martingale, with $\mathbb{E}Y_n = \mathbb{E}Y_0 = 0$. But in the present case we employ a randomized stopping strategy, i.e. we stop the game at the random time

$$\nu = \min\{n \geq 1 : \varepsilon_n = 1\},$$

the time at which we win. Note that $Y_\nu = 1$ on $\{\nu < +\infty\}$ and that

$$\mathbb{P}(\nu = n) = (1/2)^n, \quad n \geq 1,$$

so

$$\mathbb{P}(\nu < +\infty) = 1.$$

Therefore we win one dollar in finite time with probability one. In particular,

$$\mathbb{E}Y_\nu = 1 \neq 0 = \mathbb{E}Y_n, \quad n \geq 0.$$

This inequality occurs because ν is an unbounded stopping time, i.e. there is no finite constant C such that $\mathbb{P}(\nu \leq C) = 1$. That is why, employing this randomized

“doubling” strategy, we are guaranteed to finish the game ahead. However, consider the expected amount lost before we win (which is the expected value of the last bet)

$$\mathbb{E}(\text{amount lost}) = \sum_{n=0}^{\infty} \mathbb{P}(v = n+1)(2^n - 1) = \sum_{n=0}^{\infty} (1/2)^{n+1} (2^n - 1) = +\infty.$$

Thus, on average, infinite capital is needed in order to play a winning game, which makes the doubling strategy much less attractive.

Remark 1.3.8 Winning a positive amount with probability one is an arbitrage in the terminology of mathematical finance. The example shows us that in order to avoid arbitrages, we must put constraints on the trading strategies. This relates to the notion of admissible trading strategies, which we will examine in Sect. 4.1.1.

1.3.2 Stopping Times and Optional Stopping Theorem

The notation $\mathbb{1}(A)$ is used instead of $\mathbb{1}_A$ in this subsection.

Definition 1.3.9 A random variable v is called a stopping time with respect to \mathbb{F} if

- (i) v takes values in $\{0, 1, \dots, \infty\}$,
- (ii) for each n , $\mathbb{1}(v = n)$ is measurable with respect to \mathcal{F}_n .

Thus a stopping time is a stopping rule based only on the currently available information. Put another way, if we know which particular event from \mathcal{F}_n took place, then we know whether $v = n$ or not.

Let $v = j$ for some $j \geq 0$. Clearly, v is a stopping time. This is the most elementary example of a bounded stopping time.

Example 1.3.10 Let ε_i be i.i.d. with $\mathbb{P}(\varepsilon_i = 1) = p$, $\mathbb{P}(\varepsilon_i = -1) = 1 - p$, for some $0 < p < 1$. Set $S_n = \sum_{i=1}^n \varepsilon_i$. Let \mathcal{F}_n be the information contained in S_0, \dots, S_n (which is the same as the information contained in $\varepsilon_1, \dots, \varepsilon_n$). We consider different stopping rules.

i. Let

$$v^j = \min\{n \geq 0 : S_n = j\}$$

(meant as ∞ if $S_n \neq j$ for all $n \geq 0$). Since $\mathbb{1}(v^j = n)$ is determined by the information in \mathcal{F}_n , v^j is a stopping time with respect to \mathbb{F} .

ii. Let

$$\theta^j = v^j - 1, \quad j \neq 0.$$

Then, since $\mathbb{1}(\theta^j = n) = \mathbb{1}(v^j - 1 = n) = \mathbb{1}(v^j = n+1)$, $\mathbb{1}(\theta^j = n)$ is not \mathcal{F}_n -measurable (it is \mathcal{F}_{n+1} -measurable). Hence θ^j is not a stopping time.

iii. Let now

$$\nu^j = \max\{n \geq 0 : S_n = j\}.$$

Thus ν^j is the last time S_n visits state j . Clearly ν^j is not a stopping time.

Exercise 1.3.11 If θ is a stopping time, then $\theta^j = \min(\theta, j)$, where j is a fixed integer, is also a stopping time. Clearly $\theta^j \leq j$.

Exercise 1.3.12 If ν and θ are stopping times, then so are $\min(\nu, \theta)$ and $\max(\nu, \theta)$.

Let ν be any nonnegative integer random variable that is finite with probability one. Let X_n , $n \geq 0$ be a random sequence. Then, X_ν denotes the random variable that takes values $X_{\nu(\omega)}(\omega)$.

The following result says that we cannot beat a fair game by using a stopping rule that is a bounded stopping time.

Lemma 1.3.13 *Let M_n be a martingale and ν a stopping time. Then*

$$\mathbb{E}M_{\min(\nu, n)} = \mathbb{E}M_0, \quad n \geq 0.$$

Proof We have

$$\begin{aligned} M_{\min(\nu, n)} &= M_\nu \mathbb{1}(\nu \leq n) + M_n \mathbb{1}(\nu > n) \\ &= M_\nu \sum_{k=0}^n \mathbb{1}(\nu = k) + M_n \mathbb{1}(\nu > n) \\ &= \sum_{k=0}^n M_k \mathbb{1}(\nu = k) + M_n \mathbb{1}(\nu > n). \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{E}M_{\min(\nu, n)} &= \sum_{k=0}^n \mathbb{E}[M_k \mathbb{1}(\nu = k)] + \mathbb{E}[M_n \mathbb{1}(\nu > n)] \\ &= \sum_{k=0}^n \mathbb{E}[[\mathbb{E}(M_n | \mathcal{F}_k) \mathbb{1}(\nu = k)]] + \mathbb{E}[M_n \mathbb{1}(\nu > n)] \\ &= \sum_{k=0}^n \mathbb{E}[[\mathbb{E}(M_n \mathbb{1}(\nu = k) | \mathcal{F}_k)]] + \mathbb{E}[M_n \mathbb{1}(\nu > n)] \\ &= \sum_{k=0}^n \mathbb{E}[M_n \mathbb{1}(\nu = k)] + \mathbb{E}[M_n \mathbb{1}(\nu > n)] \\ &= \mathbb{E}[M_n \mathbb{1}(\nu \leq n)] + \mathbb{E}[M_n \mathbb{1}(\nu > n)] = \mathbb{E}M_n = \mathbb{E}M_0, \end{aligned}$$

where the second equality follows from the martingale property of M_n , the third from the fact that $\mathbb{1}(v = k)$ is measurable with respect to \mathcal{F}_k , and the fourth from the tower rule. \square

In many situations of interest the stopping time is not bounded, but is almost surely finite, as in the doubling strategy of Example 1.3.7. In this example, $\mathbb{E}Y_v = 1 \neq 0 = \mathbb{E}Y_0$. The question arises: when is $\mathbb{E}M_v = \mathbb{E}M_0$ for a stopping time that is not bounded? We have

$$M_v = M_{\min(v, n)} + M_v \mathbb{1}(v > n) - M_n \mathbb{1}(v > n).$$

Hence, using Lemma 1.3.13, we obtain for every n

$$\mathbb{E}M_v = \mathbb{E}M_0 + \mathbb{E}[M_v \mathbb{1}(v > n)] - \mathbb{E}[M_n \mathbb{1}(v > n)]. \quad (1.8)$$

This provides motivation for the following Optional Stopping Theorem (abbreviated OST, also mentioned in the literature as Doob's optional sampling theorem).

Theorem 1.3.14 *Let M be a martingale and v a stopping time. If*

- (i) $\mathbb{P}(v < +\infty) = 1$,
- (ii) $\mathbb{E}|M_v| < +\infty$,
- (iii) and

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|M_n| \mathbb{1}(v > n)] = 0, \quad (1.9)$$

then

$$\mathbb{E}M_v = \mathbb{E}M_0. \quad (1.10)$$

Proof It follows from (1.8) and (1.9) that we only need show

$$\lim_{n \rightarrow +\infty} \mathbb{E}[M_v \mathbb{1}(v > n)] = 0. \quad (1.11)$$

By (i) and (ii),

$$\begin{aligned} \mathbb{E}|M_v| &= \sum_{k=0}^{\infty} \mathbb{E}[|M_v| \mathbb{1}(v = k)] \\ &= \sum_{k=0}^n \mathbb{E}[|M_v| \mathbb{1}(v = k)] + \mathbb{E}[|M_v| \mathbb{1}(v > n)] < +\infty. \end{aligned} \quad (1.12)$$

Now (1.11) follows because we see from (1.12) that $\mathbb{E}[|M_v| \mathbb{1}(v > n)]$ is the tail of a convergent series. \square

Example 1.3.15 For the doubling strategy of Example 1.3.7 we know that (1.10) doesn't hold. We also know that for this strategy $\mathbb{P}(\nu < +\infty) = 1$ and $\mathbb{E}|Y_\nu| = 1 < +\infty$, so it must be the case that (1.9) doesn't hold. Indeed, as $n \rightarrow +\infty$,

$$\mathbb{E}[|Y_n| \mathbb{1}(\nu > n)] = |1 - 2^n| \mathbb{P}(\nu > n) = |1 - 2^n| (1/2)^n \rightarrow 1.$$

1.3.2.1 Uniform Integrability and Martingales

Here we present some conditions that imply condition (1.9), which is difficult to verify directly.

Definition 1.3.16 A sequence of random variables X_1, X_2, \dots is uniformly integrable (UI for short) if, for every $\epsilon > 0$, there exists a $\delta > 0$ such that, for every random event $A \subset \Omega$ with $\mathbb{P}(A) < \delta$, we have that

$$\mathbb{E}(|X_n| \mathbb{1}_A) < \epsilon \quad (1.13)$$

for each n .

Observe that δ must be independent of n and that (1.13) must hold for all values of n .

Example 1.3.17 Let X_1, X_2, \dots be a random sequence with $|X_n| \leq C < +\infty$ for every n . Here C doesn't depend on n , which means that the sequence is uniformly bounded. To see that the sequence is UI, fix $\epsilon > 0$, take $\delta = \frac{\epsilon}{C}$ and take any event A such that $\mathbb{P}(A) < \delta$. We have

$$\mathbb{E}(|X_n| \mathbb{1}_A) \leq C \mathbb{P}(A) < C \delta = \epsilon$$

for every n . Thus the sequence X_1, X_2, \dots is UI.

Exercise 1.3.18 Let the sequence X_n be as in the above example. Consider the sequence $S_n = \sum_{k=1}^n X_k$. Is the sequence S_n UI?

Example 1.3.19 Consider the fortune process Y_n of the doubling strategy from Example 1.3.7. We know that this process is a martingale with respect to \mathcal{F}_n , but is it a UI martingale? In order to answer this question consider the event $A_n = \{\varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_n = -1\}$. We have $\mathbb{P}(A_n) = (1/2)^n$ and $\mathbb{E}(|Y_n| \mathbb{1}_{A_n}) = (2^n - 1)/2^n$, because $|Y_n| = 2^n - 1$ if event A_n occurs. Thus, $\mathbb{E}(|Y_n| \mathbb{1}_{A_n}) = 1 - (1/2)^n$. Now, take any $\epsilon < 1$. No matter how small $\delta > 0$ is chosen, we can always find n large enough so that $\mathbb{P}(A_n) < \delta$ and $\mathbb{E}(|Y_n| \mathbb{1}_{A_n}) \geq \epsilon$. Thus, the fortune process Y_n of the doubling strategy is not a UI martingale.

Suppose now that M_0, M_1, \dots is a UI martingale and that ν is a finite stopping time so that $\mathbb{P}(\nu < +\infty) = 1$. By uniform integrability we then conclude that, since $\mathbb{P}\{\nu > n\} \rightarrow 0$,

$$\lim_{n \rightarrow \infty} \mathbb{E}(|M_n| \mathbf{1}_{\{\nu > n\}}) = 0,$$

so that condition (1.9) holds. Thus we may state a weaker version of the OST:

Theorem 1.3.20 *Let M_n be a UI martingale and ν a stopping time. Suppose that $\mathbb{P}(\nu < +\infty) = 1$ and $\mathbb{E}(|M_\nu|) < +\infty$. Then, $\mathbb{E}M_\nu = \mathbb{E}M_0$.*

Here is a useful criterion for uniform integrability. If for a sequence of random variables X_n there exists a constant $C < +\infty$ so that $\mathbb{E}X_n^2 < C$ for each n , then the sequence X_n is uniformly integrable. See p. 115 of Lawler [180] for a proof.

Example 1.3.21 Consider a driftless random walk S_n as in Example 1.3.3, assuming $\mathbb{P}(\varepsilon_i = -1) = \mathbb{P}(\varepsilon_i = 1) = 1/2$ for every $i \geq 1$. That is, we have a symmetric random walk on integers starting at 0. We know this random walk is a martingale. Now consider the process $\widehat{S}_n = \frac{S_n}{n}$. We have that $\mathbb{E}(\widehat{S}_n^2) = 1/n$ for every $n \geq 1$. The sequence \widehat{S}_n is obviously UI, since it is a bounded sequence. But the above criterion is not satisfied for the random walk S_n itself, which in fact is not UI.

1.3.3 Doob's Decomposition

A discrete-time process X is said to be predictable with respect to a filtration \mathbb{F} if X_0 is deterministic and X_n is \mathcal{F}_{n-1} -measurable for $n \geq 1$. Any deterministic process is predictable. Less trivial examples of predictable processes are given by the trading strategies ζ_n of Example 1.3.6. A finite variation process is a difference between two adapted and nondecreasing processes, starting from 0. We call drift any predictable and finite variation process.

Recall the driftless random walk of Example 1.3.3. In this example we saw that the process $M_n = S_n^2 - n\sigma^2$ is a martingale with respect to the filtration \mathcal{F}_n . The process $D_n = n\sigma^2$ is nondecreasing (in fact it is strictly increasing) and is, of course, predictable, since it is deterministic. Finally, observe that we have the following decomposition of the process S_n^2 :

$$S_n^2 = D_n + M_n.$$

Thus, we have decomposed the submartingale S_n^2 (by Jensen's inequality) into a sum of a drift and a martingale. This is a special case of the following general result known as the Doob decomposition.

Theorem 1.3.22 *Let X be a process adapted to some filtration \mathbb{F} . Assume $\mathbb{E}|X_n| < +\infty$ for every n . Then, X_n has a unique Doob decomposition*

$$X_n = D_n + M_n, \quad n \geq 0,$$

where D_n is a drift and M_n is a martingale. Furthermore, X_n is a submartingale if and only if the drift D_n is nondecreasing.

Chapter 2

Some Classes of Continuous-Time Stochastic Processes

2.1 Continuous-Time Stochastic Processes

2.1.1 Generalities

So far we have studied random processes in discrete time. We now turn to studying random processes in continuous time. Let there be given a finite time horizon $T \in (0, \infty)$, which represents the maturity of a financial claim in the financial interpretation. A family of random variables X_t , $t \in [0, T]$, on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a continuous time random (or stochastic) process. We will frequently use the notation X or, abusively, X_t to denote the process as a whole. More rigorously, X_t denotes the state at time t of our random process. That is, for every fixed t , X_t is a random variable on the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This means that $X_t(\cdot)$ is a function from Ω to the state space $\mathcal{S} : X_t(\cdot) : \Omega \rightarrow \mathcal{S}$. On the other hand, for every fixed $\omega \in \Omega$, we are dealing with a trajectory (or a sample path), denoted by $X(\omega)$, of our random process. That is, $X(\omega)$ is a function from $[0, T]$ to $\mathcal{S} : X(\omega) : [0, T] \rightarrow \mathcal{S}$.

A filtration $\mathbb{F} = \mathcal{F}_t$, $t \in [0, T]$ (i.e. a family of information sets which satisfy $\mathcal{F}_s \subseteq \mathcal{F}_t$, $s \leq t$) and the related conditional expectations are defined similarly as in discrete time. Process Y is said to be \mathbb{F} -adapted if Y_t is \mathcal{F}_t -measurable (“a function of the information contained in \mathcal{F}_t ”) for every t . The natural filtration of a process X_t , or the filtration generated by process X_t , is defined through

$$\mathcal{F}_t = \sigma(X_s, 0 \leq s \leq t)$$

= “information contained in the random variables X_s , $0 \leq s \leq t$ ”.

We will also need the concept of predictability. Although a detailed discussion of this concept for continuous time processes is beyond the scope of this book, it will be enough for us to know that, whenever a process Z_t is adapted and left-continuous, or deterministic,¹ then Z_t is predictable. In fact, the class of predictable processes

¹Borel function of time.

is generated in a certain sense by the class of all adapted and left-continuous processes.

By default henceforth, a process is defined on the time interval $[0, T]$ and \mathbb{F} -adapted; all inequalities between processes are meant $dt \otimes d\mathbb{P}$ -almost everywhere. Note that all the developments that follow, except those of Sect. 3.4 related to Girsanov transformations, can be extended to an infinite time horizon $[0, +\infty)$.

2.1.2 Continuous-Time Martingales

Since the definitions and results concerning martingales in continuous time are essentially analogous to those in discrete time, we state the following definitions and results without much elaboration (see also Sect. 1.5 of [205]).

Definition 2.1.1 The process $Y = Y_t$, $t \in [0, T]$, is said to be a martingale (resp. submartingale or supermartingale) with respect to the filtration \mathbb{F} if:

- (i) Y is \mathbb{F} -adapted and $\mathbb{E}|Y_t| < +\infty$, $t \geq 0$;
- (ii) for every $s \leq t$, we have $\mathbb{E}(Y_t | \mathcal{F}_s) = Y_s$ (resp. $\mathbb{E}(Y_t | \mathcal{F}_s) \geq Y_s$ or $\mathbb{E}(Y_t | \mathcal{F}_s) \leq Y_s$).

Definition 2.1.2 A $[0, T] \cup \{+\infty\}$ random variable τ is said to be a stopping time with respect to the continuous-time filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$ if, for each t , the indicator function $\mathbb{1}_{\{\tau \leq t\}}$ of the event $\{\tau \leq t\}$ is measurable with respect to \mathcal{F}_t .

Theorem 2.1.3 Let M be a martingale and τ a stopping time such that

- (i) $\mathbb{P}(\tau < +\infty) = 1$,
- (ii) $\mathbb{E}|M_\tau| < +\infty$,
- (iii) $\lim_{t \rightarrow +\infty} \mathbb{E}(M_t \mathbb{1}(\tau > t)) = 0$.

Then

$$\mathbb{E}M_\tau = \mathbb{E}M_0.$$

Note that (ii) and (iii) are implied by (i) in the case of a uniformly integrable martingale M_t .

2.2 The Poisson Process and Continuous-Time Markov Chains

Definition 2.2.1 A random process X is said to be a continuous time Markov chain with a discrete (finite or countable) state space \mathcal{S} if

$$\mathbb{P}(X_{s+t} = y | \mathcal{F}_s) = \mathbb{P}(X_{s+t} = y | X_s)$$

holds for any $0 \leq s, t$ and for any $y \in \mathcal{S}$. If \mathbb{F} is the natural filtration of X we can equivalently require that, for any sequence of times $0 \leq t_1 \leq t_2 \leq \dots \leq t_{n-1} \leq t_n < +\infty$ and any collection of states $x_1, x_2, \dots, x_{n-1}, x_n$, we have

$$\begin{aligned}\mathbb{P}(X_{t_n} = x_n \mid X_{t_{n-1}} = x_{n-1}, \dots, X_{t_2} = x_2, X_{t_1} = x_1) \\ = \mathbb{P}(X_{t_n} = x_n \mid X_{t_{n-1}} = x_{n-1}).\end{aligned}$$

Definition 2.2.2 A Markov chain X is time homogeneous if and only if, for all $x, y \in \mathcal{S}$ and all $s, t \geq 0$, we have

$$\mathbb{P}(X_{s+t} = y \mid X_s = x) = \mathbb{P}(X_t = y \mid X_0 = x) =: q(t; x, y).$$

Now let

$$Q(t) = (q(t; x, y))_{x, y \in \mathcal{S}}, \quad t \geq 0,$$

denote the transition probability function for a time homogeneous Markov chain X . Note that $Q(0) = I$.

Proposition 2.2.3 For every $s, t \geq 0$, the transition probability function for a time homogeneous Markov chain X satisfies

- (i) $0 \leq q(t; x, y) \leq 1, \forall x, y \in \mathcal{S}$,
- (ii) $\sum_{y \in \mathcal{S}} q(t; x, y) = 1, \forall x \in \mathcal{S}$,
- (iii) (Chapman-Kolmogorov semigroup equations)

$$q(s + t; x, y) = \sum_{z \in \mathcal{S}} q(s; x, z)q(t; z, y), \quad \forall x, y \in \mathcal{S}, \forall s, t \geq 0 \quad (2.1)$$

or, equivalently

$$Q(s + t) = Q(s)Q(t), \quad \forall s, t \geq 0.$$

Proof Making repeated use of the Bayes formula, for all x, y in \mathcal{S} and $s, t \geq 0$ we have

$$\begin{aligned}q(s + t; x, y) &= \mathbb{P}(X_{s+t} = y \mid X_0 = x) = \frac{\mathbb{P}(X_{s+t} = y, X_0 = x)}{\mathbb{P}(X_0 = x)} \\ &= \sum_{z \in \mathcal{S}} \frac{\mathbb{P}(X_{s+t} = y, X_t = z, X_0 = x)}{\mathbb{P}(X_0 = x)} \\ &= \sum_{z \in \mathcal{S}} \frac{\mathbb{P}(X_{s+t} = y \mid X_t = z, X_0 = x)\mathbb{P}(X_t = z, X_0 = x)}{\mathbb{P}(X_0 = x)} \\ &= \sum_{z \in \mathcal{S}} \mathbb{P}(X_{s+t} = y \mid X_t = z)\mathbb{P}(X_t = z \mid X_0 = x) \\ &= \sum_{z \in \mathcal{S}} q(s; z, y)q(t; x, z),\end{aligned}$$

where the Markov property of X was used in the next-to-last line. \square

Recall that, if a real continuous function $u(t)$ satisfies the equation

$$u(s+t) = u(s)u(t), \quad \forall s, t \geq 0,$$

then it is differentiable and is such that (with $\dot{u} = \frac{du}{dt}$)

$$\dot{u}(t) = au(t), \quad u(t) = e^{at}$$

for some real number a . Similarly, in the case of a continuous semigroup of transition probabilities $Q(t)$ associated with a continuous time Markov chain X , the matrix function $Q(t)$ is differentiable in $t \geq 0$ and there exists a so-called matrix generator A of X such that $Q(0) = I$ and, for $t \geq 0$,

$$\dot{Q}(t) = Q(t)A \tag{2.2}$$

(forward form). Moreover, by the time homogeneity of X , we can show that $Q(t)$ commutes with A , so that (2.2) can equivalently be written as

$$\dot{Q}(t) = A Q(t) \tag{2.3}$$

(backward form). These last two equations are called Kolmogorov equations.

Remark 2.2.4 Recall from Chap. 1 that, for a discrete time Markov chain, the n -step transition matrix Q_n satisfies the first step equation: $Q_0 = I$ and, for $n \geq 0$,

$$\Delta Q_{n+1} = A Q_n \tag{2.4}$$

or, equivalently, the last step equation: $Q_0 = I$ and, for $n \geq 0$,

$$\Delta Q_{n+1} = Q_n A, \tag{2.5}$$

where $A = Q - I$. The solution to both equations is $Q_n = Q^n = (I + A)^n$, $n \geq 0$. The forward Kolmogorov equation (2.2) is the continuous time counterpart of the last step equation (2.5). The backward Kolmogorov equation (2.3) is the continuous time counterpart of the first step equation (2.4).

The (infinite) matrix form of the unique solution to the Chapman-Kolmogorov equations is

$$Q(t) = e^{At}, \quad t \geq 0, \tag{2.6}$$

where, for any (finite or infinite) matrix, the matrix exponential is defined by

$$e^{At} = \sum_{n=0}^{\infty} \frac{(tA)^n}{n!}.$$

For any initial distribution vector $\phi_0 = (\mathbb{P}(X_0 = x), x \in \mathcal{S})$, we then have that

$$(\mathbb{P}(X_t = x), x \in \mathcal{S}) =: \phi_t = \phi_0 Q(t) = \phi_0 e^{At}.$$

Remark 2.2.5 The Chapman-Kolmogorov equation (2.1) can be equivalently written as

$$Q(s+t) - Q(t) = Q(t)(Q(s) - I), \quad \forall s, t \geq 0.$$

Now, fix t and rewrite the above, for $s > 0$,

$$\frac{Q(t+s) - Q(t)}{s} = Q(t) \frac{Q(s) - Q(0)}{s}.$$

Since the matrix function $Q(t)$ is differentiable, letting $s \rightarrow 0$ we obtain

$$\dot{Q}(t) = Q(t)\dot{Q}(0).$$

Comparing this with the forward equation (2.2), we conclude that

$$A = \dot{Q}(0). \quad (2.7)$$

2.2.1 The Poisson Process

Throughout the book the notation h represents a small time increment.

Definition 2.2.6 Let $N_0 = 0$ and let N_t denote the (random) number of occurrences of some underlying random event in the time interval $(0, t]$, $t > 0$. If N_t satisfies the following two conditions, we will call N a Poisson process with intensity (or rate) $\lambda > 0$, $\text{PP}(\lambda)$ for short:

- (i) For any sequence $0 \leq s_1 \leq t_1 \leq s_2 \leq t_2 \leq \dots \leq s_n \leq t_n < +\infty$, the random variables $N_{t_1} - N_{s_1}, N_{t_2} - N_{s_2}, \dots, N_{t_n} - N_{s_n}$ are independent.
- (ii) For every $t > 0$ we have

$$\mathbb{P}(N_{t+h} - N_t = k) = \begin{cases} \lambda h + o(h) & \text{if } k = 1 \\ o(h) & \text{if } k \geq 2 \\ 1 - \lambda h + o(h) & \text{if } k = 0, \end{cases}$$

$$\text{where } \lim_{h \rightarrow 0} \frac{o(h)}{h} = 0.$$

For any $0 \leq s \leq t$, the random variable $N_t - N_s$ denotes the number of occurrences of an underlying random event in the time interval $(s, t]$. Any random process N satisfying condition (i) of the above definition is said to be a process with independent increments. In a classical example, N_t represents the number of customers arriving at a service facility by the time t . The following important result explains the name ‘‘Poisson process’’:

Theorem 2.2.7 *Let N be a Poisson process with rate λ . Then, for any $0 \leq s, t$, we have*

$$\mathbb{P}(N_{s+t} - N_s = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad k \geq 0. \quad (2.8)$$

In other words, the increment $N_{s+t} - N_s$ is $\mathcal{P}_{\lambda t}$ -distributed.²

Proof Writing, for fixed s , $P_k(t) = \mathbb{P}(N_{s+t} - N_s = k)$, we have:

$$\begin{aligned} P_0(t+h) &= \mathbb{P}(N_{s+t+h} - N_s = 0) \\ &= \mathbb{P}(N_{s+t} - N_s = 0, N_{s+t+h} - N_{s+t} = 0) \\ &= \mathbb{P}(N_{s+t} - N_s = 0)\mathbb{P}(N_{s+t+h} - N_{s+t} = 0) \\ &= P_0(t)[1 - \lambda h + o(h)]. \end{aligned}$$

Therefore,

$$\frac{P_0(t+h) - P_0(t)}{h} = -\lambda P_0(t) + \frac{o(h)}{h} P_0(t).$$

Letting $h \rightarrow 0$ we get

$$\dot{P}_0(t) = -\lambda P_0(t), \quad t \geq 0, \quad (2.9)$$

with the initial condition

$$P_0(0) = \mathbb{P}(N_0 = 0) = 1.$$

The ordinary differential equation (2.9) has the unique solution

$$P_0(t) = e^{-\lambda t}.$$

Thus $\mathbb{P}(N_{s+t} - N_s = 0) = P_0(t) = e^{-\lambda t}$ for $t \in [0, T]$ and, for any $k \geq 1$, we have

$$\begin{aligned} P_k(t+h) &= \sum_{i=0}^k \mathbb{P}(N_{s+t} - N_s = k-i, N_{s+t+h} - N_{s+t} = i) \\ &= \sum_{i=0}^k P_{k-i}(t)\mathbb{P}(N_{s+t+h} - N_{s+t} = i) \\ &= P_k(t)\mathbb{P}(N_{s+t+h} - N_{s+t} = 0) + P_{k-1}(t)\mathbb{P}(N_{s+t+h} - N_{s+t} = 1) \\ &\quad + \sum_{i=2}^k P_{k-i}(t)\mathbb{P}(N_{s+t+h} - N_{s+t} = i) \\ &= P_k(t)[1 - \lambda h + o(h)] + P_{k-1}(t)[\lambda h + o(h)] + \sum_{i=2}^k P_{k-i}(t) o(h). \end{aligned}$$

Therefore

²Follows a Poisson distribution with parameter λt .

$$\frac{P_k(t+h) - P_k(t)}{h} = \lambda(P_{k-1}(t) - P_k(t)) + o(1) \sum_{i=2}^k P_{k-i}(t).$$

Letting $h \rightarrow 0$, we get

$$\dot{P}_k(t) = \lambda(P_{k-1}(t) - P_k(t)), \quad t \geq 0, \quad (2.10)$$

with the initial condition

$$P_k(0) = \mathbb{P}(N_0 = k) = \mathbb{P}(0 = k) = 0,$$

for $k \geq 1$. Thus, for $k = 1$, we get

$$\dot{P}_1(t) = \lambda(e^{-\lambda t} - P_1(t)), \quad t \geq 0, \quad (2.11)$$

with the initial condition

$$P_1(0) = \mathbb{P}(N_0 = 1) = 0.$$

The ordinary differential equation (2.11) has a well known solution:

$$P_1(t) = \lambda t e^{-\lambda t}, \quad t \geq 0.$$

Thus, $\mathbb{P}(N_{s+t} - N_s = 1) = P_1(t) = \frac{(\lambda t)^1 e^{-\lambda t}}{1!}$ for $t \in [0, T]$. Proceeding similarly for $k \geq 2$, we finally obtain

$$\mathbb{P}(N_{s+t} - N_s = k) = P_k(t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad t \geq 0,$$

for all $k \geq 0$. □

Corollary 2.2.8 *Let N be a Poisson process with rate λ . Then the process N is a time homogeneous Markov chain.*

Proof It is enough to verify that, for any three times $s \leq r \leq t$ and for any three integers $k \leq m \leq n$, we have

$$\mathbb{P}(N_t = n \mid N_r = m, N_s = k) = \mathbb{P}(N_t = n \mid N_r = m)$$

and that

$$\mathbb{P}(N_t = n \mid N_r = m)$$

depends only on m, n and the time differential $t - r$. Now

$$\begin{aligned} & \mathbb{P}(N_t = n \mid N_r = m, N_s = k) \\ &= \frac{\mathbb{P}(N_t - N_r = n - m, N_r - N_s = m - k, N_s = k)}{\mathbb{P}(N_r - N_s = m - k, N_s = k)} \\ &= \mathbb{P}(N_t - N_r = n - m) = \mathbb{P}(N_t - N_r = n - m \mid N_r = m) \\ &= \mathbb{P}(N_t = n \mid N_r = m). \end{aligned} \tag{2.12}$$

This proves the Markov property. From Theorem 2.2.7 we know that

$$\mathbb{P}(N_t - N_r = n - m) = \frac{(\lambda(t - r))^{n-m} e^{-\lambda(t-r)}}{(n-m)!},$$

which proves the time homogeneity, in view of (2.12). \square

In particular,

$$\mathbb{P}(N_t = n \mid N_0 = 0) = \mathbb{P}(N_t = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}.$$

Consider the random time $\tau_1 = \min\{t > 0 : N_t = 1\}$. This is the time of the first jump of the process N . Then, for all $t \geq 0$, we have

$$\mathbb{P}(\tau_1 > t) = \mathbb{P}(N_t = 0) = e^{-\lambda t},$$

and so the random time τ_1 has exponential distribution with parameter λ . More generally, the so-called sojourn times, i.e. the random times τ_n that elapse between the consecutive jumps of the Poisson process N , are i.i.d. random variables, each having exponential distribution with parameter λ .

These results tell us that the trajectories of a Poisson process with rate λ are right-continuous step functions; the height of each step is 1; the length of each step is the value of an exponential random variable with parameter λ ; the lengths of different steps are distributed independently.

The time- t transition matrix of a Poisson process is

$$\begin{aligned} Q(t) &= (\mathbb{P}(N_{s+t} = m \mid N_s = n))_{m,n \geq 0} \\ &= (\mathbb{P}(N_{s+t} - N_s = m - n))_{m,n \geq 0} = (P_{m-n}(t))_{m,n \geq 0}, \end{aligned}$$

where, in view of (2.8):

$$P_k(t) = \mathbf{I}_{k \geq 0} \frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

Thus,

$$Q(t) = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & \dots \\ 0 & P_0(t) & P_1(t) & P_2(t) & P_3(t) & P_4(t) & \dots \\ 1 & 0 & P_0(t) & P_1(t) & P_2(t) & P_3(t) & \dots \\ 2 & 0 & 0 & P_0(t) & P_1(t) & P_2(t) & \dots \\ 3 & 0 & 0 & 0 & P_0(t) & P_1(t) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Applying (2.7), the corresponding generator A is given by the following countably infinite matrix:

$$A = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & \dots \\ 0 & -\lambda & \lambda & 0 & 0 & 0 & \dots \\ 1 & 0 & -\lambda & \lambda & 0 & 0 & \dots \\ 2 & 0 & 0 & -\lambda & \lambda & 0 & \dots \\ 3 & 0 & 0 & 0 & -\lambda & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (2.13)$$

Consistent with the general form (2.2)–(2.3) of the Kolmogorov equations of a Markov process, the system of ordinary differential equations (2.9), (2.10) can be written in vector form as: $Q(0) = I$ and, for $t \geq 0$,

$$\dot{Q}(t) = A Q(t) = Q(t)A, \quad Q(t) = e^{At}.$$

For $s \leq t$ the increment $N_t - N_s$ is a $\mathcal{P}_{\lambda(t-s)}$ -random variable. We also know that this random variable is independent of all the random variables N_r , $r \leq s$. Thus, we have

$$\mathbb{E}(N_t - \lambda t | \mathcal{F}_s) = \mathbb{E}(N_t - N_s - \lambda(t-s) | \mathcal{F}_s) + \mathbb{E}(N_s - \lambda s | \mathcal{F}_s) = N_s - \lambda s.$$

This means that the process

$$M_t = N_t - \lambda t \quad (2.14)$$

is a martingale.

2.2.2 Two-State Continuous Time Markov Chains

We now consider a two-state Markov chain with the infinitesimal generator

$$A = \begin{pmatrix} 0 & 1 \\ -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

If the process is in state 0, then it waits for a random time τ_0 before it decides to jump to state 1. The random time τ_0 has an exponential distribution with parameter λ . If

the process is in state 1, then it waits for a random time τ_1 before it decides to jump to state 0. The random time τ_1 has an exponential distribution with parameter μ .

The forward Kolmogorov equation is

$$\dot{Q}(t) = Q(t)A, \quad t \geq 0, \quad Q(0) = I,$$

that is

$$\begin{aligned}\dot{q}(t; 0, 0) &= -\lambda q(t; 0, 0) + \mu q(t; 0, 1), & \dot{q}(t; 0, 1) &= \lambda q(t; 0, 0) - \mu q(t; 0, 1), \\ \dot{q}(t; 1, 0) &= -\lambda q(t; 1, 0) + \mu q(t; 1, 1), & \dot{q}(t; 1, 1) &= \lambda q(t; 1, 0) - \mu q(t; 1, 1),\end{aligned}$$

for $t \geq 0$, with the initial conditions

$$q(0; 0, 0) = q(0; 1, 1) = 1, \quad q(0; 1, 0) = q(0; 0, 1) = 0.$$

The backward Kolmogorov equation is

$$\dot{Q} = A Q(t), \quad t \geq 0, \quad Q(0) = I,$$

that is

$$\begin{aligned}\dot{q}(t; 0, 0) &= -\lambda q(t; 0, 0) + \lambda q(t; 1, 0), & \dot{q}(t; 0, 1) &= -\lambda q(t; 0, 1) + \lambda q(t; 1, 1), \\ \dot{q}(t; 1, 0) &= \mu q(t; 0, 0) - \mu q(t; 1, 0), & \dot{q}(t; 1, 1) &= \mu q(t; 0, 1) - \mu q(t; 1, 1),\end{aligned}$$

for $t \geq 0$, with the initial conditions

$$q(0; 0, 0) = q(0; 1, 1) = 1, \quad q(0; 1, 0) = q(0; 0, 1) = 0.$$

Note that the backward and the forward Kolmogorov equations yield different systems of ODEs in the $q(t; \cdot, \cdot)$.

The matrix A diagonalizes as follows (check it as an exercise):

$$A = \begin{pmatrix} 1 & -\frac{\lambda}{\mu} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -(\lambda + \mu) \end{pmatrix} \begin{pmatrix} \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \\ -\frac{\mu}{\lambda + \mu} & \frac{\mu}{\lambda + \mu} \end{pmatrix}.$$

Thus, the solution to both equations (forward and backward) is:

$$\begin{aligned}Q(t) &= e^{At} = \begin{pmatrix} 1 & -\frac{\lambda}{\mu} \\ 1 & 1 \end{pmatrix} e^{\begin{pmatrix} 0 & 0 \\ 0 & -(\lambda + \mu) \end{pmatrix}t} \begin{pmatrix} \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \\ -\frac{\mu}{\lambda + \mu} & \frac{\mu}{\lambda + \mu} \end{pmatrix} \\ &= \begin{pmatrix} 1 & -\frac{\lambda}{\mu} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{-t(\lambda + \mu)} \end{pmatrix} \begin{pmatrix} \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \\ -\frac{\mu}{\lambda + \mu} & \frac{\mu}{\lambda + \mu} \end{pmatrix}.\end{aligned}$$

That is,

$$q(t; 0, 0) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-t(\lambda + \mu)}, \quad q(t; 0, 1) = 1 - q(t; 0, 0),$$

$$q(t; 1, 1) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-t(\lambda + \mu)}, \quad q(t; 1, 0) = 1 - q(t; 1, 1),$$

for $t \geq 0$.

Observe that

$$\lim_{t \rightarrow \infty} Q(t) = \begin{pmatrix} \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \\ \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \end{pmatrix} = \begin{pmatrix} \pi \\ \pi \end{pmatrix},$$

where $\pi = (\frac{\mu}{\lambda + \mu}, \frac{\lambda}{\lambda + \mu})$. Thus, π is the unique stationary distribution for this chain.

2.2.3 Birth-and-Death Processes

The infinitesimal generator of a birth-and-death (BDP) process is the infinite matrix

$$A = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & \dots \\ -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \dots \\ \mu_1 & -\lambda_1 - \mu_1 & \lambda_1 & 0 & 0 & \dots \\ 0 & \mu_2 & -\lambda_2 - \mu_2 & \lambda_2 & 0 & \dots \\ 0 & 0 & \mu_3 & -\lambda_3 - \mu_3 & \lambda_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The constants $\mu_i \geq 0$ represent the “death” rates at various states of the process. They are intensities of “downward” transitions; note that we always have $\mu_0 = 0$. The constants $\lambda_i \geq 0$ represent the “birth” rates at various states of the process. They are intensities of “upward” transitions. Observe that the diagonal elements of the matrix A are nonpositive, and that the rows of the matrix sum to 0.

In each state i the process waits a random amount of time, τ_i , before the process “decides” to jump to either the higher state $i + 1$ or the lower state $i - 1$ (the latter outcome being possible only if $i \geq 1$). The waiting time τ_i has an exponential distribution with parameter $\mu_i + \lambda_i$. Thus, the intensity (or rate) of a jump out of state i is $\mu_i + \lambda_i$. Once the process decides to jump from state i , the probability of the jump up to $i + 1$ is $\frac{\lambda_i}{\lambda_i + \mu_i}$ and the probability of the jump down to $i - 1$ is $\frac{\mu_i}{\lambda_i + \mu_i}$.

The Poisson process $PP(\lambda)$ is a BDP process for which $\mu_i = 0$, $\lambda_i = \lambda$, $i \geq 0$.

2.3 Brownian Motion

Up to now we have been dealing with stochastic processes taking on at most a countable number of values. The one-dimensional Brownian motion process (BM for short) takes values over the entire real line; that is why it is successfully used to model certain types of random continuous motions.

2.3.1 Definition and Basic Properties

Our intention is to model a “random continuous motion” that satisfies certain desirable physical postulates, where the continuity is understood both in the time variable and the space variable. Let X_t denote the position at time t of our random process. Our postulates regarding the X_t are as follows:

- $X_0 = 0$;
- the random process has independent and time homogeneous increments. That is, for any $0 \leq s \leq t \leq u \leq v$ the random variables $X_t - X_s$ and $X_v - X_u$ are independent; in addition, for any $0 \leq s \leq t$ the distribution of $X_t - X_s$ depends only on the time differential $t - s$.

Recall that a Poisson process with rate λ satisfies the above two postulates, but is an integer-valued process. By contrast, we make the following additional postulate on X :

- the sample paths $X(\omega)$ of our random process are continuous functions from $[0, \infty)$ to the state space $\mathcal{S} = (-\infty, \infty)$.

It turns out that the above three postulates imply that, for $0 \leq s \leq t$, the distribution of the increment $X_t - X_s$ must be Gaussian:

$$X_t - X_s \sim \mathcal{N}(\mu(t-s), \sigma^2(t-s))$$

for some constants μ and $\sigma > 0$ (we assume $\sigma \neq 0$ to avoid trivialities). All this motivates the following:

Definition 2.3.1 The stochastic process X_t , $t \geq 0$, is said to be a Brownian motion with drift, or a Wiener process with drift, starting at 0, if $X_0 = 0$ and

- i. sample paths of X_t are continuous functions of t ;
- ii. X_t has independent and time homogeneous increments;
- iii. for $0 \leq s \leq t$,

$$X_t - X_s \sim \mathcal{N}(\mu(t-s), \sigma^2(t-s)).$$

It was indicated above that condition (iii) in Definition 2.3.1 is implied by conditions (i) and (ii) (assuming $X_0 = 0$). Nevertheless, it is customary to include this condition as a part of the definition of BM. Note that, in particular, we have $X_t \sim \mathcal{N}(\mu t, \sigma^2 t)$. If we change $X_0 = 0$ into $X_0 = x$ in the definition, then $X_t \sim \mathcal{N}(x + \mu t, \sigma^2 t)$. Here μ is called the drift parameter and σ^2 the variance (or diffusion) parameter.

The definition of a Brownian motion is often formulated for the case $\mu = 0$. We have chosen to give the general (i.e. $\mu \in (-\infty, \infty)$) definition above. Such a BM process is often called a Brownian motion with drift.

Definition 2.3.2 When $x = 0$, $\mu = 0$ and $\sigma^2 = 1$, the process X is said called standard Brownian motion (SBM), which we denote by W .

It can be shown that sample paths of BM, though continuous, are nowhere differentiable. It can also be shown that sample paths of BM do not have bounded variation on any finite time interval and have a self-similar (or fractal) property.

2.3.2 Random Walk Approximation

A random walk may serve as a “discrete time prototype” of BM. As a matter of fact, BM can be constructed as an appropriate limit of random walk processes. Here is how (the construction is done for the case $\mu = 0$ for simplicity). Let $\varepsilon_1, \varepsilon_2, \dots$ be i.i.d. with $\mathbb{P}(\varepsilon_1 = \pm 1) = \frac{1}{2}$, and let

$$X_t^{h,k} = k(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{[\frac{t}{h}]})$$

where $[x]$ denotes the greatest integer $\leq x$. $X_t^{h,k}$ can be interpreted as the time- t location of a particle executing a random walk with size of the step given by k and with time unit equal to h . We have

$$\begin{aligned}\mathbb{E}(X_t^{h,k}) &= 0, \\ \text{Var}(X_t^{h,k}) &= k^2 \left[\frac{t}{h} \right].\end{aligned}$$

Let $h, k \rightarrow 0$ in such a way that $\text{Var}(X_t^{h,k})$ converges to a finite, positive number (note that if we set $k = h$, then $\text{Var}(X_t^{h,k}) \rightarrow 0$). This can be accomplished by maintaining $k^2 = \sigma^2 h$ for a finite constant σ . In particular, by considering $X^n = X^{\frac{1}{n}, \frac{\sigma}{\sqrt{n}}}$ we obtain that $\text{Var}(X_t^n) \sim_{n \rightarrow +\infty} \sigma^2 t$. We then have

$$\begin{aligned}X_t^n &= \frac{\sigma}{\sqrt{n}} (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{[nt]}) \\ &= \frac{\sigma \sqrt{t} (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{[nt]})}{\sqrt{[nt]}} \frac{\sqrt{[nt]}}{\sqrt{nt}}\end{aligned}$$

for every n . As $n \rightarrow +\infty$, $\sqrt{[nt]}/\sqrt{nt} \rightarrow 1$ and thus, by the central limit theorem,

$$X_t^n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 t).$$

In addition, one can show that all joint distributions of X^n at different time points converge to multivariate normal distributions. Recall that the random variables Z_1, Z_2, \dots, Z_d are said to have a joint normal distribution if they can be represented as

$$Z_l = \sum_{j=1}^m a_{l,j} \varepsilon_j, \quad l = 1, 2, \dots, d,$$

where the ε_j are independent normal random variables and the $a_{l,j}$ are arbitrary constants.

2.3.3 Second Order Properties

For Brownian motion X , we know that $X_t \sim \mathcal{N}(\mu t, \sigma^2 t)$. We now consider the joint probability distribution of $X_{t_1}, X_{t_2}, \dots, X_{t_n}$. We have

$$\begin{aligned} X_{t_1} &= X_{t_1} - X_0 \\ X_{t_2} &= (X_{t_2} - X_{t_1}) + (X_{t_1} - X_0) \\ &\vdots \\ X_{t_n} &= (X_{t_n} - X_{t_{n-1}}) + \cdots + (X_{t_1} - X_0) \end{aligned}$$

where, by definition of Brownian motion, increments are independent normal random variables. Hence the distribution of $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ is multivariate normal with $\mathbb{E}X_{t_j} = \mu t_j$, and with covariance matrix $C = (C_{i,j})$ given by

$$C_{i,j} = \text{Cov}(X_{t_i}, X_{t_j}) = \sigma^2 \min(t_i, t_j).$$

In order to verify the last statement assume that $s < t$. Then

$$\begin{aligned} \text{Cov}(X_s, X_t) &= \mathbb{E}(X_s X_t) - (\mathbb{E}X_s)(\mathbb{E}X_t) \\ &= \mathbb{E}[X_s(X_t - X_s)] + \mathbb{E}X_s^2 - \mu^2 s t \\ &= \mathbb{E}X_s \mathbb{E}(X_t - X_s) + \sigma^2 s + \mu^2 s^2 - \mu^2 s t \\ &= \mu^2 s(t-s) + \sigma^2 s + \mu^2 s^2 - \mu^2 s t \\ &= \sigma^2 s = \sigma^2 \min(s, t). \end{aligned}$$

2.3.4 Markov Properties

Brownian motion, as a process with independent increments, is a Markov process. This can be verified as follows:

$$\begin{aligned} \mathbb{P}(X_{s+t} \leq y \mid X_s = x, X_{t_1} = x_1, \dots, X_{t_n} = x_n) \\ &= \mathbb{P}(X_{s+t} - X_s \leq y - x \mid X_s = x, X_{t_1} = x_1, \dots, X_{t_n} = x_n) \\ &= \mathbb{P}(X_{s+t} - X_s \leq y - x) = \mathbb{P}(X_{s+t} \leq y \mid X_s = x), \end{aligned}$$

where $0 \leq t_1 < t_2 < \dots < t_n < s$. Let

$$\begin{aligned} q(t; x, y) &:= \partial_y \mathbb{P}(X_{t+s} \leq y | X_s = x) = \partial_y \mathbb{P}(X_{t+s} - X_s \leq y - x) \\ &= \frac{1}{\sqrt{2\pi t}\sigma} \exp[-(y - x - \mu t)^2/2\sigma^2 t], \end{aligned}$$

where the first and second equalities follow from the second and third defining properties of Brownian motion, respectively. The function $q(t; x, y)$ is the probability density function of X_{t+s} given that $X_s = x$. It is called the transition density function of X .

Note that $q(t; x, y)$ depends on x, y as a function of $(y - x)$. Therefore BM is a spatially homogeneous process as well as a time homogeneous process. Recall that analogous properties were satisfied for a Poisson process.

Remark 2.3.3 We have the following two properties of the function $q(t; x, y)$, where we let $\partial_z = \frac{\partial}{\partial z}$, $\partial_{z^2} = \frac{\partial^2}{\partial z^2}$.

(i) For every $x \in (-\infty, \infty)$,

$$\partial_t q(t; x, y) = \mathcal{A}^* q(t; x, y), \quad \forall t \geq 0, \quad y \in (-\infty, \infty),$$

where

$$\mathcal{A}^* q(t; x, y) = -\mu \partial_y q(t; x, y) + \frac{1}{2} \sigma^2 \partial_{y^2}^2 q(t; x, y). \quad (2.15)$$

The operator \mathcal{A}^* is the adjoint infinitesimal generator of Brownian motion. Equation (2.15) is called the forward Kolmogorov equation for the transition probability density function of X (compare with the forward equation (2.2) in the case of a continuous-time Markov chain).

(ii) For every $y \in (-\infty, \infty)$,

$$\partial_t q(t; x, y) = \mathcal{A} q(t; x, y), \quad \forall t \geq 0, \quad x \in (-\infty, \infty), \quad (2.16)$$

where

$$\mathcal{A} q(t; x, y) = \mu \partial_x q(t; x, y) + \frac{1}{2} \sigma^2 \partial_{x^2}^2 q(t; x, y).$$

The operator \mathcal{A} is the infinitesimal generator of the Brownian motion. Equation (2.16) is called the backward Kolmogorov equation for the transition probability density function (compare with the backward equation (2.3) in the case of a continuous-time Markov chain).

The Brownian motion also satisfies the so-called strong Markov property, namely the property that $B_t = X_{t+\tau} - X_\tau$ is a Brownian motion independent of \mathcal{F}_τ , for every stopping time τ . Using this property the following three important features of SBM (i.e. $X = W$) can be demonstrated (see pp. 178–180 of [180]):

Reflection Principle For any $b > 0$ and for any $t > 0$,

$$\mathbb{P}(W_s \geq b \text{ for some } 0 \leq s \leq t) = 2\mathbb{P}(W_t \geq b). \quad (2.17)$$

Equivalently,

$$\mathbb{P}(\tau_b \leq t) = 2\mathbb{P}(W_t \geq b),$$

where $\tau_b = \inf\{t \geq 0 : W_t = b\}$.

Arctan Law and Recurrence For any $t > \epsilon > 0$,

$$\mathbb{P}(W_s = 0 \text{ for some } \epsilon \leq s \leq t) = \frac{2}{\pi} \operatorname{Arctan}\left(\sqrt{\frac{t}{\epsilon} - 1}\right).$$

Consequently

$$\mathbb{P}(W_s = 0 \text{ for some } \epsilon \leq s) = 1.$$

Strong Law of Large Numbers With probability 1 we have

$$\lim_{t \rightarrow \infty} \frac{W_t}{t} = 0.$$

2.3.5 First Passage Times of a Standard Brownian Motion

Let W_t be SBM and let

$$\tau_b = \min\{t \geq 0 : W_t = b\}, \quad b \neq 0.$$

Assuming first $b > 0$, let $F(t) = \mathbb{P}(\tau_b \leq t)$. We have

$$\begin{aligned} \mathbb{P}(W_t > b) &= \mathbb{P}(W_t > b, \tau_b \leq t) = \int_0^t \mathbb{P}(W_t > b \mid \tau_b = s) dF(s) \\ &= \int_0^t \mathbb{P}(W_t > b \mid W_s = b; W_u < b, u < s) dF(s) \\ &= \int_0^t \mathbb{P}(W_t - W_s > 0 \mid W_s = b; W_u < b, u < s) dF(s) \\ &= \int_0^t \mathbb{P}(W_t - W_s > 0) dF(s) = \frac{1}{2}\mathbb{P}(\tau_b \leq t), \end{aligned}$$

where the next-to-last equality follows by the independent increments property and the last one by the distributional properties of SBM. Hence

$$\begin{aligned}\mathbb{P}(\tau_b \leq t) &= 2\mathbb{P}(W_t > b) = \frac{2}{\sqrt{2\pi t}} \int_b^\infty \exp(-x^2/2t) dx \\ &= \frac{2}{\sqrt{2\pi}} \int_{b/\sqrt{t}}^\infty \exp(-y^2/2) dy,\end{aligned}$$

where in the last step we substituted $y = x/\sqrt{t}$. Thus the probability density function of τ_b is

$$\begin{aligned}q_b(t) &= \frac{d}{dt} \mathbb{P}(\tau_b \leq t) = -\frac{2}{\sqrt{2\pi}} \exp(-b^2/2t) \frac{d}{dt} \left(\frac{b}{\sqrt{t}} \right) \\ &= \frac{b}{\sqrt{2\pi}} t^{-3/2} \exp(-b^2/2t), \quad t > 0, b > 0.\end{aligned}$$

For $b < 0$, by symmetry,

$$q_b(t) = \frac{-b}{\sqrt{2\pi}} t^{-3/2} \exp(-b^2/2t).$$

Finally, the density γ_b of τ_b arises in the form of the so-called inverse Gaussian density

$$q_b(t) = \frac{|b|}{\sqrt{2\pi}} t^{-3/2} \exp(-b^2/2t), \quad b \neq 0, t > 0.$$

Consequently

$$\mathbb{P}(\tau_b < +\infty) = \lim_{t \rightarrow +\infty} \mathbb{P}(\tau_b \leq t) = \lim_{t \rightarrow +\infty} \frac{2}{\sqrt{2\pi}} \int_{b/\sqrt{t}}^\infty \exp(-y^2/2) dy = 1$$

and $\mathbb{E}\tau_b = \infty$. This shows that SBM is a null recurrent process (it behaves like a symmetric random walk on the integers, see [180]).

Let $M_t = \max_{0 \leq s \leq t} W_s$ be the maximum of SBM on the time interval $[0, t]$. We can now easily obtain the distribution of M_t :

$$\mathbb{P}(M_t > b) = \mathbb{P}\left(\max_{0 \leq s \leq t} W_s > b\right) = \mathbb{P}(\tau_b \leq t) = \frac{2}{\sqrt{2\pi}} \int_{b/\sqrt{t}}^\infty \exp(-y^2/2) dy.$$

Note that we have just demonstrated the reflection principle (2.17) for SBM.

2.3.6 Martingales Associated with Brownian Motion

Let X_t , $t \geq 0$, be $\text{BM}(\mu = 0, \sigma^2)$. Then the following processes are martingales:

- a. X_t ,
- b. $M_t = X_t^2 - \sigma^2 t$,
- c. $Z_t = \exp(\theta X_t - \frac{1}{2}\theta^2 \sigma^2 t)$, $\forall \theta \in (-\infty, +\infty)$.

To verify that the above processes are martingales, we write $X_t = X_t - X_s + X_s$, and use the increments independence property of Brownian motion. As an example,

we verify that X is a martingale. We have

$$\mathbb{E}(X_t | \mathcal{F}_s) = \mathbb{E}(X_t - X_s + X_s | \mathcal{F}_s) = \mathbb{E}(X_t - X_s) + X_s = X_s$$

and $\mathbb{E}|X_t| < +\infty$, since X_t is a normal random variable. Hence indeed X is a martingale.

Now let X be BM($\mu \neq 0, \sigma^2$). Then the following processes are martingales.

- d. $Y_t = X_t - \mu t$,
- e. $M_t = (X_t - \mu t)^2 - \sigma^2 t$,
- f. $Z_t = \exp(\theta X_t - (\theta \mu + \frac{1}{2} \theta^2 \sigma^2)t)$, $\forall \theta \in (-\infty, +\infty)$.

The processes in (c) and (f) are Wald's martingales (recall the structure of Wald's martingale in discrete time), since in both cases $Z_t = \exp(\theta X_t)/\mathbb{E}[\exp(\theta X_t)]$.

2.3.6.1 Exit Time from a Corridor

Let X be BM with parameters μ and σ^2 , and $X_0 = x$. Then, for $a < x < b$,

$$\tau_a = \min\{t \geq 0 : X_t = a\}, \quad \tau_b = \min\{t \geq 0 : X_t = b\} \quad \text{and}$$

$$\tau_{a,b} = \tau = \min\{t \geq 0 : X_t = a \text{ or } b\}$$

are \mathbb{F} -stopping times (hitting times of given levels and exit time from a corridor), where $\mathcal{F}_t = \sigma(X_s, s \leq t)$. We now use OST to compute $\mathbb{P}(X_\tau = b)$ and $\mathbb{E}\tau$. Verification of assumptions of OST can be done similarly as for random walks; see Chap. 1 and [180].

Case 1 ($\mu = 0$) In this case X_t is a martingale. By OST,

$$\mathbb{E}X_\tau = \mathbb{E}X_0 = x.$$

But

$$\mathbb{E}X_\tau = b\mathbb{P}(X_\tau = b) + a\mathbb{P}(X_\tau = a) = x.$$

Solving the last equation for $\mathbb{P}(X_\tau = b) = 1 - \mathbb{P}(X_\tau = a)$ gives

$$\mathbb{P}(X_\tau = b) = \frac{x - a}{b - a}.$$

To compute $\mathbb{E}\tau$ we use the martingale M_t in item (b) above. By OST,

$$\mathbb{E}M_\tau = \mathbb{E}X_\tau^2 - \sigma^2 \mathbb{E}\tau = x^2,$$

so that

$$\begin{aligned} \mathbb{E}\tau &= (\mathbb{E}X_\tau^2 - x^2)/\sigma^2 = \left[\left(\frac{x - a}{b - a} \right) b^2 + \left(\frac{b - x}{b - a} \right) a^2 - x^2 \right] / \sigma^2 \\ &= (x - a)(b - x)/\sigma^2. \end{aligned}$$

Case 2 ($\mu \neq 0$) To compute $\mathbb{P}(X_\tau = b)$, we apply OST to the Wald's martingale Z_t in item f above:

$$Z_t = \exp\left(\theta X_t - \left(\theta\mu + \frac{1}{2}\theta^2\sigma^2\right)t\right),$$

for $\theta = \theta^* := -2\mu/\sigma^2$. With this choice of θ ,

$$Z_t = \exp\left(-\frac{2\mu}{\sigma^2}X_t\right) = \exp(\theta^*X_t).$$

By OST,

$$\mathbb{E}Z_\tau = \mathbb{E}Z_0 = \exp(\theta^*x).$$

Solving

$$\mathbb{E}Z_\tau = \exp(\theta^*b)\mathbb{P}(X_\tau = b) + \exp(\theta^*a)\mathbb{P}(X_\tau = a) = \exp(\theta^*x)$$

for $\mathbb{P}(X_\tau = b) = 1 - \mathbb{P}(X_\tau = a)$ gives

$$\mathbb{P}(X_\tau = b) = \frac{\exp(\theta^*x) - \exp(\theta^*a)}{\exp(\theta^*b) - \exp(\theta^*a)}.$$

It now follows from the above equation that

$$\begin{aligned} \mathbb{P}(\tau_b < \infty) &= \lim_{a \rightarrow -\infty} \mathbb{P}(X_\tau = b) \\ &= \begin{cases} 1 & \text{if } \mu > 0, \text{ hence } \theta^* < 0 \\ \exp(\theta^*(x - b)) < 1 & \text{if } \mu < 0, \text{ hence } \theta^* > 0. \end{cases} \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{P}(\tau_a < \infty) &= \lim_{b \rightarrow +\infty} \mathbb{P}(X_\tau = a) = \lim_{b \rightarrow +\infty} \frac{\exp(\theta^*b) - \exp(\theta^*x)}{\exp(\theta^*b) - \exp(\theta^*a)} \\ &= \begin{cases} \exp(\theta^*(x - a)) < 1 & \text{if } \mu > 0, \text{ hence } \theta^* < 0 \\ 1 & \text{if } \mu < 0, \text{ hence } \theta^* > 0. \end{cases} \end{aligned}$$

These results show that BM with $\mu \neq 0$ is a transient process. We will now compute $\mathbb{E}\tau$. For this we use the martingale $Y_t \equiv X_t - \mu t$ in item (d) above. By OST we have $\mathbb{E}Y_\tau = \mathbb{E}X_\tau - \mu\mathbb{E}\tau = x$, so that $\mathbb{E}\tau = (\mathbb{E}X_\tau - x)/\mu$, where

$$\mathbb{E}X_\tau = a \frac{\exp(\theta^*b) - \exp(\theta^*x)}{\exp(\theta^*b) - \exp(\theta^*a)} + b \frac{\exp(\theta^*x) - \exp(\theta^*a)}{\exp(\theta^*b) - \exp(\theta^*a)}$$

and, as before, $\theta^* = -2\mu/\sigma^2$.

Suppose now that we let $a \rightarrow -\infty$ and assume that $\mu > 0$, so that $\theta^* < 0$. Then

$$\lim_{a \rightarrow -\infty} \mathbb{E}X_\tau = b,$$

which shows that

$$\mathbb{E}\tau_b = (b - x)/\mu.$$

2.3.7 First Passage Times of a Drifted Brownian Motion

Now let τ_b be the first passage time of process X to level b , where X is $\text{BM}(\mu > 0, \sigma^2)$. We assume $X_0 = x < b$, hence $\mathbb{E}\tau_b = (b - x)/\mu < +\infty$. Recall that

$$Z_t = \exp\left(\theta X_t - \left(\theta\mu + \frac{1}{2}\theta^2\sigma^2\right)t\right)$$

is a martingale for every real θ , with

$$\mathbb{E}Z_t = \mathbb{E}Z_0 = \exp(\theta x).$$

Applying OST to the martingale Z and the stopping time τ_b , we obtain

$$\mathbb{E}Z_{\tau_b} = \exp(\theta x)$$

or

$$\mathbb{E}\left[\exp\left(\theta b - \left(\theta\mu + \frac{1}{2}\theta^2\sigma^2\right)\tau_b\right)\right] = \exp(\theta x),$$

so that

$$\mathbb{E}\exp\left[-\left(\theta\mu + \frac{1}{2}\theta^2\sigma^2\right)\tau_b\right] = \exp(\theta(x - b)). \quad (2.18)$$

Let

$$\lambda = \theta\mu + \frac{1}{2}\theta^2\sigma^2.$$

We require that $\lambda > 0$, so that $\mathbb{E}\exp(-\lambda\tau_b)$ is the Laplace transform of τ_b . Solving the last equation for θ gives

$$\theta_{\pm} = \frac{-\mu \pm \sqrt{\mu^2 + 2\sigma^2\lambda}}{\sigma^2}.$$

Taking the positive root $\theta = \theta_+$ (for $\lambda > 0$) and substituting it into the right-hand side of (2.18), we obtain the Laplace transform of τ_b as

$$\mathbb{E}\exp(-\lambda\tau_b) = \exp\left[-(b - x)\left(\sqrt{\mu^2 + 2\sigma^2\lambda} - \mu\right)/\sigma^2\right].$$

This transform can be inverted explicitly to obtain the density of τ_b as

$$\gamma_b(t) = \frac{b - x}{\sigma\sqrt{2\pi t^3}} \exp\left[-\frac{(b - x - \mu t)^2}{2\sigma^2 t}\right], \quad t \geq 0.$$

The Laplace transform can be used to obtain moments of τ_b . We have

$$\frac{d}{d\lambda} \mathbb{E} \exp(-\lambda \tau_b) = \mathbb{E} \frac{d}{d\lambda} \exp(-\lambda \tau_b) = -\mathbb{E}(\tau_b \exp(-\lambda \tau_b)),$$

so that

$$\frac{d}{d\lambda} \mathbb{E} \exp(-\lambda \tau_b) \Big|_{\lambda=0} = -\mathbb{E} \tau_b.$$

Carrying out the computation on the left side of the above equation gives the result obtained previously, i.e.

$$\mathbb{E} \tau_b = (b - x)/\mu.$$

Higher moments of τ_b can be obtained by further differentiation.

2.3.8 Geometric Brownian Motion

Definition 2.3.4 Let X_t be a Brownian motion with parameters μ, σ^2 and $X_0 = x$. The process

$$S_t = e^{X_t} = \exp(x + \mu t + \sigma W_t)$$

is called Geometric Brownian motion (GBM).

Note that the state space of S_t , $t \geq 0$, is $\mathcal{S} = (0, \infty)$. Let

$$0 = t_0 < t_1 < \dots < t_n < +\infty$$

be an increasing sequence of times, and consider relative changes

$$\frac{S_{t_1} - S_{t_0}}{S_{t_0}}, \quad \frac{S_{t_2} - S_{t_1}}{S_{t_1}}, \quad \dots, \quad \frac{S_{t_n} - S_{t_{n-1}}}{S_{t_{n-1}}}.$$

These can be expressed as

$$\exp(X_{t_1} - X_{t_0}) - 1, \quad \exp(X_{t_2} - X_{t_1}) - 1, \quad \dots, \quad \exp(X_{t_n} - X_{t_{n-1}}) - 1,$$

from which we see that relative changes in disjoint time intervals for GBM are independent random variables. The process S_t is also called a lognormal process and is often used to model prices of financial assets (this is the so-called Black–Scholes model [47]; see Sect. 5.1.1). Modeling prices with GBM involves the assumption that returns are independent from period to period.

We compute $\mathbb{E} S_t$ and $\text{Var}(S_t)$. We have $X_t \sim \mathcal{N}(x + \mu t, \sigma^2 t)$, hence

$$\mathbb{E} S_t = \mathbb{E} \exp(X_t) = \exp\left(x + \mu t + \frac{1}{2}\sigma^2 t\right).$$

We can show in a similar way that

$$\mathbb{V}\text{ar}(S_t) = \exp(2x + 2\mu t + \sigma^2 t)(\exp(\sigma^2 t) - 1)$$

and also that the mean and the variance of the return are

$$\begin{aligned}\mathbb{E}\left(\frac{S_t - S_s}{S_s}\right) &= \exp\left[\mu(t-s) + \frac{1}{2}\sigma^2(t-s)\right] - 1 \\ \mathbb{V}\text{ar}\left(\frac{S_t - S_s}{S_s}\right) &= \exp[2\mu(t-s) + \sigma^2(t-s)](\exp[\sigma^2(t-s)] - 1).\end{aligned}$$

Chapter 3

Elements of Stochastic Analysis

Our purpose in this chapter is to give an overview of the basics of stochastic calculus, an important mathematical tool that is used in control engineering, in modern finance and in the insurance industry, among other fields.

3.1 Stochastic Integration

We begin this section with the study of stochastic integrals and proceed in several stages. In Sect. 3.1.1, we define and analyze stochastic “integrals” with respect to a discrete time symmetric random walk. In Sect. 3.1.2, we define and analyze stochastic integrals of random step functions with respect to the standard Brownian motion. In Sect. 3.1.3, we generalize results of Sect. 3.1.2 to stochastic integrals of general stochastic integrands with respect to the standard Brownian motion. Section 3.1.4 discusses integration with respect to a Poisson process. Finally, Sect. 3.1.5 provides a glimpse of a more general semimartingale integration theory.

3.1.1 Integration with Respect to a Symmetric Random Walk

We already constructed a discrete-time stochastic integral with respect to a symmetric random walk in Example 1.3.6. There we called it a martingale transform of the process S_n by the betting process ζ_n . For convenience, we repeat the content of the example. We have, for $n \geq 0$,

$$S_n = x + \varepsilon_1 + \cdots + \varepsilon_n,$$

where the ε_n are i.i.d. with $\mathbb{P}(\varepsilon_n = -1) = \mathbb{P}(\varepsilon_n = 1) = \frac{1}{2}$ for $n \geq 1$. We know that a symmetric random walk S_n is a martingale with respect to the filtration $\mathcal{F}_n =$

$\sigma(\varepsilon_1, \dots, \varepsilon_n) = \sigma(S_0, S_1, \dots, S_n)$. We saw that for $n \geq 0$ the fortune process Y_n could be represented as a discrete time stochastic integral

$$Y_n = \sum_{k=1}^n \zeta_k \Delta S_k,$$

where $\Delta S_k = S_k - S_{k-1}$. Recall that the process ζ_n is assumed to be predictable with respect to the filtration \mathcal{F}_n , i.e. ζ_n is \mathcal{F}_{n-1} -measurable for $n \geq 1$.

Properties Enjoyed by the Stochastic Integral $Y_n = \sum_{k=1}^n \zeta_k \Delta S_k$

- (i) We verified in Example 1.3.6 that Y_n is a martingale with respect to the filtration \mathcal{F}_n .
- (ii) $\mathbb{E}Y_n = 0$ for every $n \geq 0$. Here is why:

$$\begin{aligned} \mathbb{E}Y_n &= \mathbb{E}\left[\sum_{k=1}^n \zeta_k \Delta S_k\right] = \sum_{k=1}^n \mathbb{E}[\zeta_k \Delta S_k] \\ &= \sum_{k=1}^n \mathbb{E}[\mathbb{E}[\zeta_k \Delta S_k | \mathcal{F}_{k-1}]] = \sum_{k=1}^n \mathbb{E}[\zeta_k \mathbb{E}[\varepsilon_k | \mathcal{F}_{k-1}]] = \sum_{k=1}^n \mathbb{E}[\zeta_k \mathbb{E}[\varepsilon_k]] = 0. \end{aligned}$$

(A much faster proof is possible: due to martingale property we have $\mathbb{E}Y_n = \mathbb{E}Y_0 = 0$.)

- (iii) $\mathbb{V}ar Y_n = \mathbb{E}Y_n^2 = \sum_{k=1}^n \mathbb{E}\zeta_k^2$ for every $n \geq 1$; see p. 199 in [180].

3.1.2 The Itô Stochastic Integral for Simple Processes

Definition 3.1.1 A stochastic process Z_t , $t \geq 0$ is said to be a simple process if it satisfies the following properties:

- There exists a partition

$$\pi_n : \quad 0 = t_0 < t_1 < \dots < t_{n-1} < t_n = T$$

and a sequence of random variables Z_1, Z_2, \dots, Z_n such that

$$Z_t = \begin{cases} Z_i & \text{if } t_{i-1} \leq t < t_i, \quad i = 1, \dots, n \\ Z_n & \text{if } t = T. \end{cases}$$

- The sequence Z_i is $\mathcal{F}_{t_{i-1}}$ -adapted. I.e. Z_i is a function of the W_t , $t \leq t_{i-1}$. Moreover, $\mathbb{E}Z_i^2$ is finite.

We can now define the Itô stochastic integral¹ [151] for simple processes Z :

Definition 3.1.2 The Itô stochastic integral for simple processes Z on the interval $(0, t]$, where $t_i \leq t < t_{i+1}$, is given by the random Riemann-Stieltjes sum

$$Y_t^n = \int_0^t Z_s dW_s = \sum_{k=1}^i Z_k \Delta W_{t_k} + Z_{i+1}(W_t - W_{t_i}), \quad (3.1)$$

where $\Delta W_{t_k} = W_{t_k} - W_{t_{k-1}}$ and where, for $i = 0$, $\sum_{k=1}^0 Z_k \Delta W_{t_k} = 0$.

We can regard the standard Brownian motion W_t as a symmetric random walk continuous in time and space (see Sect. 2.3.2). Suppose now that we place bets depending only on the history of W . The bets may be placed only at a certain finite set of times $0 = t_0 < t_1 < \dots < t_{n-1} < t_n = T$. A bet Z_i placed at time t_{i-1} may only depend on the history of W up to time t_{i-1} . The game is stopped at time t . If we bet Z_i at time t_{i-1} then we receive $Z_i \Delta W_{t_i}$ at time t_i if $t_i < t$, and we receive $Z_i(W_t - W_{t_{i-1}})$ if $t_{i-1} \leq t < t_i$. Then the integral $Y_t^n = \int_0^t Z_s dW_s$ represents our fortune at time t in such a game.

The Itô stochastic integral of Z is then defined likewise on any interval $(r, t]$ with $0 \leq r \leq t$. When considered as a function of t , the Itô stochastic integral for simple processes Z , Y_t^n , is a stochastic process.

Properties Enjoyed by the Itô Stochastic Integral for Simple Processes Z

(i) The Itô stochastic integral for simple processes is a martingale. We check the two martingale conditions:

- We have $\mathbb{E}|Y_t^n| < +\infty$ for all $t \in [0, T]$. This follows from the isometry property (iii) below.
- We have $\mathbb{E}(Y_t^n | \mathcal{F}_s) = Y_s^n$ for $0 \leq s \leq t \leq T$. To demonstrate this, we first take $t_i \leq s \leq t \leq t_{i+1}$. In this case we have

$$Y_t^n = Y_s^n + Z_{i+1}(W_t - W_s).$$

Thus, since both Y_s^n and Z_{i+1} are \mathcal{F}_s -measurable (why?), we have

$$\begin{aligned} \mathbb{E}(Y_t^n | \mathcal{F}_s) &= Y_s^n + Z_{i+1}\mathbb{E}(W_t - W_s | \mathcal{F}_s) \\ &= Y_s^n + Z_{i+1}\mathbb{E}(W_t - W_s) = Y_s^n, \end{aligned}$$

where the second equality follows from the independent increments property of Brownian motion.

Exercise 3.1.3 Verify that $\mathbb{E}(Y_t^n | \mathcal{F}_s) = Y_s^n$ is true for $t_i \leq s \leq t_{i+1}$ and $t_k \leq t \leq t_{k+1}$, where $t_{i+1} \leq t_k$.

¹Kiyoshi Itô (1915–2008) was awarded the Wolf Prize in Mathematics for his contributions to stochastic analysis in 1987. He was also awarded the first Carl Friedrich Gauss Prize in 2006.

- (ii) $\mathbb{E}Y_t^n = 0$ for every t .

Exercise 3.1.4 Verify property (ii).

- (iii) **(Isometry property)** We have that

$$\text{Var } Y_t^n = \mathbb{E}[(Y_t^n)^2] = \int_0^t \mathbb{E}Z_s^2 ds, \quad t \geq 0.$$

Exercise 3.1.5 Verify property (iii).

- (iv) **(Linearity with respect to integrands)** Let Z_t and U_t be two simple processes, and let a, b be two constants. Then

$$\int_0^t (aZ_s + bU_s) dW_s = a \int_0^t Z_s dW_s + b \int_0^t U_s dW_s, \quad \forall t \geq 0.$$

This property follows immediately from the linearity property of summation.

- (v) **(Linearity on adjacent intervals)** Let $0 \leq r \leq t \leq T$. Then

$$\int_r^t Z_s dW_s = \int_0^t Z_s dW_s - \int_0^r Z_s dW_s.$$

Exercise 3.1.6 Verify property (v).

- (vi) The sample paths of the process Y_t^n are continuous. This follows since the sample paths of W_t are continuous, and we have

$$Y_t^n = Y_{t_{i-1}}^n + Z_{t_i}(W_t - W_{t_{i-1}}), \quad t_{i-1} \leq t \leq t_i.$$

Example 3.1.7 The following simple but important example is discussed in detail in Sect. 2.2.1 of [205]. Take $Z_i = W_{t_{i-1}}$. Here we have, for $t = t_i$,

$$\begin{aligned} Y_t^n &= \sum_{k=1}^i W_{t_{k-1}} \Delta W_{t_k} = \sum_{k=1}^i W_{t_{k-1}} (W_{t_k} - W_{t_{k-1}}) \\ &= \frac{1}{2} W_t^2 - \frac{1}{2} \sum_{k=1}^i (\Delta W_{t_k})^2. \end{aligned}$$

It is demonstrated on p. 98 of [205] that when the partition π_n becomes finer (i.e. $|\pi_n| := \max_{\{i=1,2,\dots,n\}}[t_i - t_{i-1}] \rightarrow 0$), the sum $\sum_{k=1}^i (\Delta W_{t_k})^2$ converges to t in the mean-square sense. This is a very important observation as we will later see.

Exercise 3.1.8 Suppose that the simple integrand process Z is deterministic, i.e. that Z_1, Z_2, \dots, Z_n are constants. Verify that, in this case, the stochastic integral Y_t^n is a random variable that is normally distributed with mean zero and variance $\int_0^t Z_s^2 ds$.

3.1.3 The General Itô Stochastic Integral

The general Itô stochastic integral for an appropriately regular integrand processes Z_t is defined as the mean-square limit of a sequence of Itô stochastic integrals for simple processes Z_t^n . The processes Z_t^n are chosen in such a way that they converge to the process Z_t in an appropriate sense. This was Itô's main idea in constructing his stochastic integral. Here are some details.

Lemma 3.1.9 *Let Z be a process satisfying the following assumptions:*

- *Z is adapted to the natural filtration of W on $[0, T]$, i.e. for every $t \in [0, T]$, the random variable Z_t is a function of the W_s , $0 \leq s \leq t$.*
- *The integral $\int_0^T \mathbb{E} Z_s^2 ds$ is finite.*

Then there exists a sequence $\{(Z_t^n, t \in [0, T]), n \geq 1\}$ of simple processes such that

$$\int_0^T \mathbb{E}[Z_s - Z_s^n]^2 ds \rightarrow 0$$

as $n \rightarrow \infty$.

Proof This is accomplished by construction of an approximating sequence $\{(Z_t^n, t \in [0, T]), n \geq 1\}$ of simple processes on increasingly finer partitions π_n . See Appendix A4 of [205] for the details. \square

Now, we already know how to evaluate the Itô stochastic integral for each simple process Z^n in the above sequence. Let Y_t^n denote the Itô stochastic integral of Z^n on the interval $[0, t]$. By general results from functional analysis it follows that there exists a process Y_t such that

$$\mathbb{E} \sup_{0 \leq t \leq T} [Y_t - Y_t^n]^2 \rightarrow 0$$

as $n \rightarrow \infty$. We say that the sequence of processes Y^n converges in mean-square to the process Y . Moreover one can prove that the limit Y doesn't depend on the choice of a sequence Z^n of simple processes approximating Z . As a consequence, we can state the following:

Definition 3.1.10 The mean-square limit process Y is called the Itô stochastic integral of Z , and it is denoted by

$$Y_t = \int_0^t Z_s dW_s, \quad t \geq 0. \tag{3.2}$$

The Itô stochastic integral of Z can likewise be defined on any interval $(r, t]$ with $0 \leq r \leq t$.

If Z is a simple process, then the Itô stochastic integral of Z is given by the Riemann-Stieltjes sum (3.1).

Properties Enjoyed by the (General) Itô Stochastic Integral All the properties enjoyed by the Itô integral of simple processes are inherited by the general Itô stochastic integral $Y = Y(Z)$. Thus,

- (i) The (general) Itô stochastic integral for processes Z is a martingale with respect to the natural filtration of W .
- (ii) $\mathbb{E}Y_t = 0$ for every t .
- (iii) (**Isometry property**) We have that

$$\mathbb{V}\text{ar } Y_t = \mathbb{E}Y_t^2 = \int_0^t \mathbb{E}Z_s^2 ds, \quad t \geq 0. \quad (3.3)$$

- (iv) (**Linearity with respect to integrands**) Let Z_t and U_t be two admissible integrands, and let a, b be two constants. Then

$$\int_0^t (aZ_s + bU_s) dW_s = a \int_0^t Z_s dW_s + b \int_0^t U_s dW_s, \quad t \geq 0.$$

- (v) (**Linearity on adjacent intervals**) Let $r \leq t$. Then

$$\int_r^t Z_s dW_s = \int_0^t Z_s dW_s - \int_0^r Z_s dW_s.$$

- (vi) The sample paths of the process Y_t are continuous.

Suppose that the integrand process Z_t is deterministic. In this case the stochastic integral Y_t is a normally distributed random variable with mean zero and variance $\int_0^t Z_s^2 ds$. Alternatively, suppose that Z_t is an arbitrary predictable and locally bounded process. Then Y_t is a well-defined local martingale.

Example 3.1.11 (Continuation of Example 3.1.7) From the above results it follows that the mean-square limit of the stochastic integrals

$$Y_t^n = \sum_{k=1}^i W_{t_{k-1}} \Delta W_{t_k} = \sum_{k=1}^i W_{t_{k-1}} (W_{t_k} - W_{t_{k-1}})$$

is the stochastic integral

$$Y_t = \int_0^t W_s dW_s, \quad t \geq 0.$$

But we have seen that

$$Y_t^n = \frac{1}{2} W_t^2 - \frac{1}{2} \sum_{k=1}^i (\Delta W_{t_k})^2$$

and that, when the partition π_n becomes finer and finer, the sum $\sum_{k=1}^i (\Delta W_{t_k})^2$ converges to t in the mean-square sense. We thus obtain the following formula:

$$Y_t = \int_0^t W_s dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t, \quad t \geq 0.$$

This result indicates that Itô stochastic calculus is different from ordinary calculus. The reader may also consult Sect. 2.4 of [205] for still other integrals, e.g., the Stratonovich integral.

3.1.4 Stochastic Integral with Respect to a Poisson Process

Let N be $\text{PP}(\lambda)$. In view of the simple structure of trajectories of a Poisson process, it is easy to define a stochastic integral with respect to such a process. However, in order that the stochastic integrals with respect to a Poisson process have nice properties, it is required that integrands be predictable and that they satisfy some mild integrability conditions. Recall that any Borel function or left-continuous process is predictable.

We define the stochastic integral of a predictable integrand Z with respect to a Poisson process N as

$$I_t := \int_0^t Z_s dN_s := \sum_{T_n \leq t} Z_{T_n}, \quad (3.4)$$

where T_n denotes the n th jump time of N .

Note that we can represent dN_t as $\sum_n \delta_{T_n}(dt)$, where T_n represents the n th jump time of N and δ_{T_n} is a Dirac mass at T_n (a random measure over the half-line). We can thus view the Poisson stochastic integral I_t as the pathwise Lebesgue-Stieltjes integral of Z against the measure dN_t .

Properties of this integral are analogous to the properties of the Itô integral discussed above, except that the process I is *not* a martingale (because process N itself is not a martingale). Letting $M_t = N_t - \lambda t$ denote the compensated martingale of N (see (2.14)), it can be verified that the process Y defined by

$$Y_t = \int_0^t Z_s dM_s := \int_0^t Z_s dN_s - \lambda \int_0^t Z_s ds \quad (3.5)$$

is a martingale.

3.1.5 Semimartingale Integration Theory (*)²

In this subsection we give a very brief account of the general semimartingale integration theory, as it is developed, for instance, in the book by Protter [228].

²For the convenience of readers, we signal advanced sections with an asterisk (*) or even with a double asterisk (**) for the more difficult parts.

Semimartingales are a class of integrators giving rise to the most flexible theory of stochastic integration. In mathematical finance another motivation for modeling prices of traded assets as semimartingales is that price processes outside this class give rise to arbitrages unless rather stringent constraints are imposed on the trading strategies [94].

Assume that the model filtration \mathbb{F} satisfies the so-called usual conditions of completeness (\mathcal{F}_0 contains all null sets of $(\Omega, \mathcal{F}, \mathbb{P})$) and right-continuity ($\mathcal{F}_{t+} = \mathcal{F}_t$, where \mathcal{F}_{t+} is the model information “right after time t ”). Then it can be shown that every semimartingale admits a càdlàg version, where càdlàg is a French acronym for “(almost surely) left limited and right continuous”. Henceforth by default all semimartingales are taken in a càdlàg version. In one of several equivalent characterizations, a semimartingale X corresponds to the sum of a local martingale M and of a finite variation process D , where:

- a local martingale M admits an increasing sequence of stopping times τ_n such that every stopped process $M_{\cdot \wedge \tau_n}$ is a uniformly integrable martingale, and
- a finite variation process is a difference between two adapted nondecreasing processes starting from 0.

Any such representation $X = D + M$ is called a Doob-Meyer decomposition of the semimartingale X . A Doob-Meyer decomposition is not generally unique. However, there is at most one such representation of a process X with D predictable. One then speaks of “the canonical Doob-Meyer decomposition of a special semimartingale X ”. In particular:

Proposition 3.1.12 *The only predictable finite variation (e.g., time-differentiable) local martingale is the null process.*

The stochastic integral of a predictable and locally bounded³ process Z with respect to a semimartingale X is defined as

$$Y_t = \int_0^t Z_s dX_s := \int_0^t Z_s dM_s + \int_0^t Z_s dD_s, \quad (3.6)$$

where $X = D + M$ is a Doob-Meyer decomposition of X and $\int_0^t Z_s dM_s$ is defined by localization of M . It is a remarkable fact that the corresponding notion of stochastic integral is independent of the Doob-Meyer decomposition of X that is used in (3.6).

Predictable and locally bounded processes notably include all left-limiting processes of the form $Z \equiv \tilde{Z}_-$, where \tilde{Z} is a semimartingale.

Proposition 3.1.13 *In the case X is a local martingale, the integral process Y is again a local martingale.*

³In particular, with almost all trajectories bounded and a bit more than that since locally bounded actually means “locally uniformly bounded”.

In the case of a continuous integrator X , it is possible (as we saw earlier in the case of a Brownian motion) to define the stochastic integral Y for non predictable integrands Z , namely for progressive integrands Z subject to suitable integrability conditions. Then, in particular, we have that

$$\int_0^t Z_s dX_s = \int_0^t Z_{s-} dX_s, \quad (3.7)$$

for every semimartingale Z .

3.2 Itô Formula

3.2.1 Introduction

Consider the function $\omega(t) = t$, so that we have

$$\int_0^t \omega(s) d\omega(s) = \frac{1}{2}\omega^2(t).$$

In fact, if $\omega(t)$ is any differentiable function of t , we have

$$\int_0^t \omega(s) d\omega(s) = \frac{1}{2}\omega^2(t) - \frac{1}{2}\omega^2(0). \quad (3.8)$$

This is just the chain rule formula:

$$\frac{d[u(\omega(s))]}{ds} = u'(\omega(s))\omega'(s), \quad (3.9)$$

which, for $u(x) = x^2$, yields

$$d(\omega^2(t)) = 2\omega(t)\omega'(t) dt = 2\omega(t) d\omega(t) \quad (3.10)$$

or, in integrated form, (3.8).

Observe next that, for an arbitrary differentiable function u (such as $u(x) = x^2$ above), the expression (3.9) can be written as

$$du(\omega(t)) = u(\omega(t+dt)) - u(\omega(t)) = u'(\omega(t)) d\omega(t).$$

On the other hand, if the function u is regular, then we have the Taylor expansion

$$u(\omega(t) + d\omega(t)) - u(\omega(t)) = u'(\omega(t)) d\omega(t) + \frac{1}{2}u''(\omega(t))(d\omega(t))^2 + \dots$$

where, as usual, $d\omega(t) = \omega(t+dt) - \omega(t)$ is the increment of the function ω on the interval $[t, t+dt]$. So, in case of a differentiable function ω , we may neglect all terms of order 2 and higher in the above Taylor expansion, since $(d\omega(t))^k = o(dt)$ for any $k \geq 2$.

3.2.1.1 What about $\int_0^t W_s dW_s$?

Now, could it be that for SBM we would have

$$\int_0^t W_s dW_s = \frac{1}{2} W_t^2? \quad (3.11)$$

Of course not! First, from the properties of the Itô integral, we know that the expectation of the left-hand side in (3.11) is zero, whereas the expectation of the right-hand side is $\frac{1}{2}t$. Second, we already saw that the true value of the stochastic integral in (3.11) is

$$\int_0^t W_s dW_s = \frac{1}{2} W_t^2 - \frac{1}{2}t.$$

Applying Taylor expansion to the function $u(W_t) = W_t^2$, we see that

$$dW_t^2 = (W_t + dW_t)^2 - W_t^2 = 2W_t dW_t + (dW_t)^2. \quad (3.12)$$

But

$$\mathbb{E}(dW_t)^2 = \mathbb{E}(W_{t+dt} - W_t)^2 = dt,$$

so that the term $(dW_t)^2$ is like dt (it is frequently written that $(dW_t)^2 = dt$). I.e. the term $(dW_t)^2$ is not $o(dt)$, and therefore it must not be neglected in (3.12). This is the reason why (3.11) is not true and the reason why a specific Itô calculus is required.

3.2.1.2 What About $\int_0^t N_{s-} dN_s$?

Could it be that, for the Poisson process N , we have

$$\int_0^t N_{s-} dN_s = \frac{1}{2} N_t^2? \quad (3.13)$$

Of course not! From the exercise 4 of Sect. 16.5, we have

$$\int_0^t N_{s-} dN_s = \frac{1}{2} (N_t^2 - N_t).$$

3.2.2 Itô Formulas for Continuous Processes

There is a general semimartingale Itô formula which leads to all of the Itô formulas in this book, but instead of stating this general result, we will only give simpler versions.

Let $u(x)$ be twice continuously differentiable function. We already know that we must not neglect the terms dW_t and $(dW_t)^2 = dt$ in the Taylor expansion of

$u(W_t + dW_t)$. However, it is known that we may neglect the higher order terms. The suitably amended Taylor expansion yields the following **simple Itô formula**:

$$du(W_t) = u'(W_t) dW_t + \frac{1}{2} u''(W_t) dt$$

or, in integral form,

$$u(W_t) - u(W_r) = \int_r^t u'(W_s) dW_s + \frac{1}{2} \int_r^t u''(W_s) ds, \quad r \leq t.$$

Before we proceed, we introduce some notation: for a function $u(t, x)$ we denote

$$\partial_t u(t, x) = \frac{\partial u(t, x)}{\partial t}, \quad \partial u(t, x) = \frac{\partial u(t, x)}{\partial x}, \quad \partial^2 u(t, x) = \frac{\partial^2 u(t, x)}{\partial^2 x}.$$

The First Extension of the Simple Itô Formula Let the function $u(t, x)$ be once continuously differentiable w.r.t. t and twice continuously differentiable w.r.t. x . Then, for $0 \leq r \leq t$,

$$u(t, W_t) - u(r, W_r) = \int_r^t \left(\partial_s u(s, W_s) + \frac{1}{2} \partial^2 u(s, W_s) \right) ds + \int_r^t \partial u(s, W_s) dW_s \quad (3.14)$$

or, in differential form for $t \in [0, T]$,

$$du(t, W_t) = \left(\partial_t u(t, W_t) + \frac{1}{2} \partial^2 u(t, W_t) \right) dt + \partial u(t, W_t) dW_t. \quad (3.15)$$

The Second Extension of the Simple Itô Formula Suppose the processes b_t and σ_t are adapted to the natural filtration of W and are such that the two integrals below are well defined. Define a new process X_t by

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \quad t \geq 0.$$

Let the function $u(t, x)$ be once continuously differentiable w.r.t. t and twice continuously differentiable w.r.t. x . Then for process X we have, for $0 \leq r \leq t$,

$$\begin{aligned} u(t, X_t) - u(r, X_r) &= \int_r^t \left(\partial_s u(s, X_s) + \partial u(s, X_s) b_s + \frac{1}{2} \sigma_s^2 \partial^2 u(s, X_s) \right) ds \\ &\quad + \int_r^t \partial u(s, X_s) \sigma_s dW_s \end{aligned} \quad (3.16)$$

or, in differential form,

$$du(t, X_t) = \left(\partial_t u(t, X_t) + \partial u(t, X_t) b_t + \frac{1}{2} \partial^2 u(t, X_t) \sigma_t^2 \right) dt + \partial u(t, X_t) \sigma_t dW_t. \quad (3.17)$$

Observe that formulas (3.14) and (3.15) are special cases of formulas (3.16) and (3.17) for the case where $b_t \equiv 0$ and $\sigma_t \equiv 1$.

3.2.2.1 Examples

Example 3.2.1 Take $u(x) = x^2$. From the simple Itô formula we get

$$W_t^2 - W_r^2 = 2 \int_r^t W_s dW_s + \int_r^t 1 ds = 2 \int_r^t W_s dW_s + (t - r), \quad 0 \leq r \leq t.$$

In particular, for $r = 0$, we get

$$\int_0^t W_s dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t.$$

Example 3.2.2 Take $u(x) = e^x$. From the simple Itô formula we get

$$e^{W_t} - e^{W_r} = \int_r^t e^{W_s} dW_s + \frac{1}{2} \int_r^t e^{W_s} ds.$$

Recall that, for a differentiable function $\omega(t)$, we have

$$de^{\omega(t)} = e^{\omega(t)} \omega'(t) dt = e^{\omega(t)} d\omega(t)$$

and thus

$$e^{\omega(t)} - e^{\omega(r)} = \int_r^t e^{\omega(s)} d\omega(s).$$

Example 3.2.3 (Itô exponential) Take $u(t, x) = e^{x - \frac{1}{2}t}$. From the first extension of the Itô formula we get

$$e^{W_t - \frac{1}{2}t} - e^{W_r - \frac{1}{2}r} = \int_r^t e^{W_s - \frac{1}{2}s} dW_s.$$

Example 3.2.4 (Geometric Brownian motion) Take $u(t, x) = e^{(b - \frac{\sigma^2}{2})t + \sigma x}$ for b real and $\sigma > 0$. From the first extension of the Itô formula we get

$$\begin{aligned} & e^{(b - \frac{1}{2}\sigma^2)t + \sigma W_t} - e^{(b - \frac{1}{2}\sigma^2)r + \sigma W_r} \\ &= \sigma \int_r^t e^{(b - \frac{1}{2}\sigma^2)s + \sigma W_s} dW_s + b \int_r^t e^{(b - \frac{1}{2}\sigma^2)s + \sigma W_s} ds. \end{aligned}$$

Thus, defining a geometric Brownian motion⁴ S_t by

$$S_t = S_0 e^{(b - \frac{1}{2}\sigma^2)t + \sigma W_t}, \quad t \geq 0,$$

⁴See Sect. 2.3.8.

we get

$$S_t - S_r = \sigma \int_r^t S_s dW_s + b \int_r^t S_s ds, \quad r \leq t,$$

or, in differential form,

$$dS_t = bS_t dt + \sigma S_t dW_t.$$

3.2.3 Itô Formulas for Processes with Jumps (*)

Suppose a function $u(t, n) : [0, \infty) \times \{0, 1, \dots\} \mapsto \mathbb{R}$ is differentiable in the first variable and such that the process $u(t, N_t)$ satisfies some mild integrability condition (e.g., u bounded). Then, we have the following **Itô formula for a Poisson process** (in integral form): for $r \leq t$,

$$u(t, N_t) - u(r, N_r) = \int_r^t \partial_s u(s, N_s) ds + \int_r^t (u(s, N_s) - u(s, N_{s-})) dN_s, \quad (3.18)$$

where (in differential form)

$$(u(s, N_s) - u(s, N_{s-})) dN_s = (u(s, N_{s-} + 1) - u(s, N_{s-})) dN_s. \quad (3.19)$$

Example 3.2.5 Taking $u(t, n) = n^2$, we get

$$N_t^2 = 2 \int_0^t N_{s-} dN_s + N_t.$$

Example 3.2.6 Taking $u(t, n) = e^n$, we get

$$e^{N_t} = 1 + (e - 1) \int_0^t e^{N_{s-}} dN_s.$$

Example 3.2.7 Taking $u(t, n) = 2^n$, we get

$$2^{N_t} = 1 + \int_0^t 2^{N_{s-}} dN_s.$$

Let A be the generator of the Poisson process N in the matrix form (2.13). Note that, for fixed t , $u(t, \cdot)$ may be considered as an infinite vector $u = (u(t, 0), u(t, 1), u(t, 2), \dots)^T$. Likewise, for fixed t , the expression Au defines a new vector $Au = (Au(t, 0), Au(t, 1), Au(t, 2), \dots)^T$. In view of the form (2.13) of A , we have that

$$Au(t, n) = \lambda(u(t, n + 1) - u(t, n)).$$

Letting $M_t = N_t - \lambda t$, it follows that the Itô formula for a Poisson process (3.19) can be rewritten:

$$du(t, N_t) = \partial_t u(t, N_t) dt + (u(t, N_{t-} + 1) - u(t, N_{t-})) dN_t \quad (3.20)$$

$$= (\partial_t u + Au)(t, N_t) dt + (u(t, N_{t-} + 1) - u(t, N_{t-})) dM_t. \quad (3.21)$$

Consequently,

$$u(t, N_t) - \int_0^t (\partial_s u(s, N_s) + Au(s, N_s)) ds \quad (3.22)$$

$$= u(0, N_0) + \int_0^t (u(s, N_{s-} + 1) - u(s, N_{s-})) dM_s, \quad (3.23)$$

which, by application of Proposition 3.1.13, is a (local) martingale.

In addition to the Poisson process N with intensity λ , now consider a standard d -variate Brownian motion W , and let $J_{(t)}$ denote a family of i.i.d. d -variate random variables with distribution denoted by $w(dy)$, all assumed to live on the same filtered probability space. Given adapted coefficients b_t (a random vector in \mathbb{R}^d), σ_t (a random matrix in $\mathbb{R}^{d \times d}$) and a predictable function $\delta_t(x)$ (a random vector in \mathbb{R}^d marked or parameterized by $x \in \mathbb{R}^d$), we consider an Itô process in the sense of a d -variate process X obeying the following dynamics:

$$dX_t = b_t dt + \sigma_t dW_t + \delta_t(J_{(t)}) dN_t, \quad t \geq 0. \quad (3.24)$$

In particular, the description of the jumps of X can be decomposed into, on the one hand, the frequency of the jumps of X (given by, on average, λ jumps of N per unit of time) and, on the other hand, the distribution w of the marks determining (also via the response function δ) the size of a jump of X in case of a jump of N .

Remark 3.2.8 By changes of measures based on Girsanov transformations, one can extend the following developments to models with a random intensity λ_t and/or a random distribution $w_t(dx)$ of jumps (or more specifically $\lambda(t, X_t)$ and/or $w(t, X_{t-}, dx)$ in the Markov setup to be introduced in Sect. 3.3.3). This allows one to design models with dependent driving noises W and N , where the latter is a point process (increasing by one at increasing random times) with intensity λ_t .

Knowing the Itô formula for continuous Itô processes, the following result is apparent from (3.24) (see also Ikeda and Watanabe [149]):

Proposition 3.2.9 (Itô formula for an Itô process) *Given a real valued, “sufficiently regular” function $u = u(t, x)$, for $t \geq 0$, we have*

$$\begin{aligned} du(t, X_t) &= \partial_t u(t, X_t) dt + \partial u(t, X_t) b_t dt + \partial u(t, X_t) \sigma_t dW_t \\ &\quad + \frac{1}{2} \partial^2 u(t, X_t) : a_t dt + \delta_t^u(X_{t-}, J_{(t)}) dN_t, \end{aligned} \quad (3.25)$$

where:

- $\partial u(t, x)$ and $\partial^2 u(t, x)$ respectively denote the row-gradient and the Hessian matrix of u with respect to x ,
- $\partial^2 u(t, x) : a(t, x)$ stands for the trace (sum of the diagonal elements) of the product of the Hessian matrix $\partial^2 u(t, x)$ with the covariance matrix $a(t, x) = \sigma(t, x)\sigma(t, x)^\top$ of X , so that

$$\partial^2 u(t, x) : a(t, x) = \sum_{1 \leq i, j, k \leq d} \sigma_t^{i,k} \sigma_t^{j,k} \partial_{x_i, x_j}^2 u(t, x),$$

- and

$$\delta_t^u(x, z) = u(x + \delta_t(z)) - u(x). \quad (3.26)$$

Moreover, letting

$$dM_t^u = \delta_t^u(X_{t-}, J_{(t)}) dN_t - \lambda \bar{\delta}_t^u(X_t) dt,$$

in which

$$\bar{\delta}_t^u(x) = \int_{\mathbb{R}^d} (u(x + \delta_t(y)) - u(x)) w(dy) \quad (3.27)$$

(assumed well-defined), one can prove the following:

Lemma 3.2.10 *Process M^u is a local martingale.*

Remark 3.2.11 In a formalism of measure-stochastic integration (see Part V and [149, 153, 167]), the compensated sum of jumps M_t^u can be represented as the stochastic integral $\int_0^t \int_{\mathbb{R}^d} \delta_t^u(X_{t-}, x) \mu(dt, dx)$ of the predictable random function $\delta_t^u(X_{t-}, x)$ with respect to the compensated random measure $\mu(dt, dx)$ of the jumps of δdN . We write for short

$$dM_t^u = \delta_t^u(X_{t-}) \cdot d\mu_t. \quad (3.28)$$

Here and later in this book $f_t(X_{t-}) \cdot d\mu_t$ is used as a shorthand for $\int_{\mathbb{R}^d} f_t(X_{t-}, x) \mu(dt, dx)$, for every matrix-valued (vector-valued in the case of $f = \delta^u$ in (3.28)) random function $f = f_t(x, y)$. The matrix integrals are performed entry-by-entry on f , so that we end up with matrices of the same dimensions as f .

In this random-measure formalism which is used in Part V (and is necessary for extension of the approach of this book to infinite activity jump measures, see Remark 12.1.1), Lemma 3.2.10 appears as a random-measure analog of Proposition 3.1.13. Also, the dynamics (3.24) of X are equivalently written in special semi-martingale form as

$$dX_t = \bar{b}_t dt + \sigma_t dW_t + \delta_t \cdot d\mu_t, \quad (3.29)$$

where $\bar{b}_t = b_t + \lambda \bar{\delta}_t$, in which

$$\bar{\delta}_t = \int_{\mathbb{R}^d} \delta_t(y) w(dy).$$

We likewise have the following special semimartingale form of the Itô formula (3.25):

$$\begin{aligned} du(t, X_t) &= \left(\partial_t u(t, X_t) + \partial u(t, X_t) b_t + \frac{1}{2} \partial^2 u(t, X_t) : a_t + \lambda \bar{\delta} u_t(X_t) \right) dt \\ &\quad + \partial u(t, X_t) \sigma_t dW_t + \delta_t^u(X_{t-}) \cdot d\mu_t. \end{aligned} \quad (3.30)$$

3.2.4 Brackets (*)

Introducing the “random generator”

$$\mathcal{A}_t^u(x) = \partial u(t, x) b_t + \frac{1}{2} \partial^2 u(t, x) : a_t + \lambda \bar{\delta}_t^u(x),$$

the Itô formula in special semimartingale form (3.30) implies, under suitable integrability conditions:

$$(dt)^{-1} \mathbb{E}(du(t, X_t) | \mathcal{F}_t) = \partial_t u(t, X_t) + \mathcal{A}_t^u(X_t) \quad (3.31)$$

(where, in the left-hand side and in analogous expressions in the sequel, what we really mean is the corresponding limit when dt goes to 0). Given another real valued function $v = v(t, x)$, let $(u, v) \mapsto \bar{\delta}^{u,v}$ and $(u, v) \mapsto \mathcal{A}^{u,v}$ represent the bilinear “carré du champ”⁵ (random) operators associated with the linear (random) operators $u \mapsto \bar{\delta}^u$ and $u \mapsto \mathcal{A}^u$, i.e.

$$\bar{\delta}_t^{u,v}(x) = \bar{\delta}_t^{uv}(x) - u(t, x) \bar{\delta}_t^v(x) - v(t, x) \bar{\delta}_t^u(x) \quad (3.32)$$

$$= \int_{\mathbb{R}^d} (u(t, x + \delta_t(y)) - u(t, x))(v(t, x + \delta_t(y)) - v(t, x)) w(dy) \quad (3.33)$$

and $\mathcal{A}^{u,v} = \mathcal{A}^{uv} - u \mathcal{A}^v - v \mathcal{A}^u$. Under square integrability conditions, one can readily show that

$$\begin{aligned} (dt)^{-1} \mathbb{C}\text{ov}(du(t, X_t), dv(t, X_t) | \mathcal{F}_t) \\ = \partial u(t, X_t) a_t (\partial v(t, X_t))^T + \lambda \bar{\delta}_t^{u,v}(X_t) = \mathcal{A}_t^{u,v}(X_t). \end{aligned} \quad (3.34)$$

Letting $Y_t = u(t, X_t)$ and $Z_t = v(t, X_t)$, the process $\mathbb{C}\text{ov}(dY_t, dZ_t | \mathcal{F}_t)$ corresponds to the so-called sharp bracket $d\langle Y, Z \rangle_t$, so that we can also write:

$$\mathcal{A}_t^{u,v}(X_t) = \frac{d\langle Y, Z \rangle_t}{dt}. \quad (3.35)$$

⁵“Squared-field operators” in English; see Sects. XV.20–26 of Dellacherie and Meyer [95].

In summary we have:

Proposition 3.2.12 *The (random) generator $u \mapsto \mathcal{A}_t^u$ of process X and its carré du champ $(u, v) \mapsto \mathcal{A}^{u,v}$ are such that, letting $Y_t = u(t, X_t)$ and $Z_t = v(t, X_t)$ for every functions u, v of t, x :*

$$(dt)^{-1} \mathbb{E}(du(t, X_t) | \mathcal{F}_t) = \partial_t u(t, X_t) + \mathcal{A}_t^u(X_t) \quad (3.36)$$

$$(dt)^{-1} \mathbb{C}\text{ov}(du(t, X_t), dv(t, X_t) | \mathcal{F}_t) = \mathcal{A}_t^{u,v}(X_t) = \frac{d\langle Y, Z \rangle_t}{dt}.$$

In particular,

$$(dt)^{-1} \mathbb{V}\text{ar}(dY_t | \mathcal{F}_t) = \mathcal{A}_t^{u,u}(X_t)$$

$$= \frac{d\langle Y \rangle_t}{dt} = \partial u(t, X_t) a_t (\partial u(t, X_t))^T + \lambda \bar{\delta}_t^{u,u}(X_t), \quad (3.37)$$

with

$$\bar{\delta}_t^{u,u}(x) = \int_{\mathbb{R}^d} (u(t, x + \delta_t(y)) - u(t, x))^2 w(dy). \quad (3.38)$$

By letting u and v range over the various coordinate mappings of X in the second line of (3.36), we obtain the following:

Corollary 3.2.13 *Taking the matrix form $\langle X \rangle = (\langle X^i, X^j \rangle)_i^j$, we have*

$$(dt)^{-1} \mathbb{E}(dX_t | \mathcal{F}_t) = b_t + \lambda \int_{\mathbb{R}^d} \delta_t(y) w(dy) \quad (3.39)$$

$$(dt)^{-1} \mathbb{C}\text{ov}(dX_t | \mathcal{F}_t) = \frac{d\langle X \rangle_t}{dt} = a_t + \lambda \int_{\mathbb{R}^d} (\delta_t \delta_t^T)(y) w(dy). \quad (3.40)$$

Observe that the sharp brackets $\langle Y, Z \rangle$ and $\langle Y \rangle$ compensate the corresponding square brackets (respectively quadratic covariation $[Y, Z]$ and quadratic variation $[Y, Y]$) defined, starting from 0 at time 0, by

$$d[Y, Z]_t = \partial u(t, X_t) a_t (\partial v(t, X_t))^T dt + \delta_t^u(X_{t-}, J_{(t)}) \delta_t^v(X_{t-}, J_{(t)}) dN_t \quad (3.41)$$

(respectively

$$d[Y, Y]_t = \partial u(t, X_t) a_t (\partial u(t, X_t))^T dt + (\delta_t^u(X_{t-}, J_{(t)}))^2 dN_t. \quad (3.42)$$

Notably, if X is a continuous Itô process, the corresponding sharp and square brackets exist and are equal. The square brackets can equivalently be defined as suitable limits of realized covariance and variance processes over time-meshes with mesh-size going to 0. They can be defined in this way for any semimartingales Y, Z .

Brackets are key in the following integration by parts formulas:

$$\begin{aligned} d(Y_t Z_t) &= Y_{t-} dZ_t + Z_{t-} dY_t + d\langle Y, Z \rangle_t + (\delta_t^u(X_{t-}) \delta_t^v(X_{t-})) \cdot d\mu_t \\ &= Y_{t-} dZ_t + Z_{t-} dY_t + d[Y, Z]_t, \end{aligned} \quad (3.43)$$

where the second identity is valid for any semimartingales Y, Z (see [71, 149, 228]).

3.3 Stochastic Differential Equations (SDEs)

3.3.1 Introduction

Consider the ordinary differential equation

$$\frac{dx(t)}{dt} = b, \quad t \geq 0, \quad x(0) = x_0. \quad (3.44)$$

The constant b may be interpreted as an infinitesimal (or instantaneous) absolute rate of change of the function $x(t)$ since, in view of (3.44), we have

$$x(t+dt) - x(t) = b dt.$$

The solution to (3.44) is

$$x(t) = x_0 + bt, \quad t \geq 0.$$

Now consider the ordinary differential equation

$$\frac{dx(t)}{dt} = bx(t), \quad t \geq 0, \quad x(0) = x_0. \quad (3.45)$$

Here, the constant b may be interpreted as an instantaneous relative rate of change of the function $x(t)$ since, in view of (3.45), we have

$$\frac{x(t+dt) - x(t)}{x(t)} = b dt.$$

The solution to (3.45) is

$$x(t) = x_0 e^{bt}, \quad t \geq 0. \quad (3.46)$$

Now imagine that both rates are perturbed by normally distributed random shocks. In the case of (3.44), this phenomenon can be modeled as

$$x(t+dt) - x(t) = b dt + \sigma(W_{t+dt} - W_t)$$

or, equivalently,

$$dx(t) = b dt + \sigma dW_t. \quad (3.47)$$

In the case of (3.45) this phenomenon can be modeled as

$$\frac{x(t+dt) - x(t)}{x(t)} = b dt + \sigma (W_{t+dt} - W_t)$$

or, equivalently,

$$dx(t) = x(t)(b dt + \sigma dW_t). \quad (3.48)$$

Equations (3.47) and (3.48) are prototypes of stochastic differential equations (SDEs); but we need to explain what is meant by a solution to an SDE, and how an SDE can be solved.

3.3.2 Diffusions

Definition 3.3.1 A Markov process X on a real interval (state space, possibly unbounded) \mathcal{S} is said to be a diffusion with drift coefficient $b(t, x)$ and diffusion coefficient $\sigma(t, x) > 0$ if:

- (i) X has continuous sample paths;
- (ii) the following relations hold as $h \rightarrow 0$, for every $t \geq 0$ and $x \in \mathcal{S}$:

$$\mathbb{E}(X_{t+h} - X_t | X_t = x) = b(t, x)h + o(h) \quad (3.49)$$

$$\mathbb{E}[(X_{t+h} - X_t)^2 | X_t = x] = \sigma^2(t, x)h + o(h). \quad (3.50)$$

The functions $b(t, x)$ and $\sigma(t, x)$ are usually assumed to be continuous. They are called the local mean function and variance functions of a diffusion. Diffusion processes behave locally like Brownian motion. Assuming (3.49), we have:

$$\begin{aligned} & \mathbb{V}\text{ar}(X_{t+h} - X_t | X_t = x) \\ &= \mathbb{E}[(X_{t+h} - X_t)^2 | X_t = x] - [\mathbb{E}(X_{t+h} - X_t | X_t = x)]^2 \\ &= \mathbb{E}[(X_{t+h} - X_t)^2 | X_t = x] + o(h). \end{aligned}$$

Therefore one can put equivalently in (3.50):

$$\mathbb{V}\text{ar}[(X_{t+h} - X_t)^2 | X_t = x] = \sigma^2(t, x)h + o(h). \quad (3.51)$$

In case the coefficients b and σ do not depend on time, we say that X is a time homogeneous diffusion.

3.3.2.1 SDEs for Diffusions

Now let ξ be a random variable, let $b(t, x)$ and $\sigma(t, x)$ be two real valued functions and suppose that a real valued process X satisfies the following three properties:

- *Property 1.* The process X is adapted with respect to the filtration generated by ξ and the SBM W .
- *Property 2.* The ordinary and Itô integrals below are well-defined for every $t \geq 0$:

$$\int_0^t b(s, X_s) ds, \quad \int_0^t \sigma(s, X_s) dW_s.$$

- *Property 3.* The equation

$$X_t = \xi + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s$$

is satisfied for all $t \geq 0$.

Definition 3.3.2 We say that a process X is a strong solution to the SDE

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t, \quad t \geq 0, \quad X_0 = \xi, \quad (3.52)$$

if the process X satisfies the Properties 1–3 above.

Solving (3.52) means determining a process X that is a strong solution to (3.52).⁶ A strong solution to the SDE (3.52) is readily seen to be a diffusion in the sense of Definition 3.3.1, so that the SDE is thus also known as the diffusion equation with drift coefficient $b(t, x)$, diffusion coefficient $\sigma(t, x)$ and initial condition ξ .

Observe that the strong solution process X for the diffusion SDE (3.52) is an Itô process in the sense of Sect. 3.2.3 (special case without jumps).

Typically, the drift and the diffusion coefficients, as well as the initial condition, are the input data for a modeler of physical phenomena. Therefore, if one attempts to model evolution of a physical phenomenon using an equation like (3.52) above, one must address questions of uniqueness and existence of strong solutions, just as one would with ordinary differential equations. I.e. one must answer the following two questions:

- What conditions on $b(t, x)$, $\sigma(t, x)$ and ξ are necessary and sufficient so that there exist solutions to (3.52)?
- Likewise, what conditions are necessary and/or sufficient for uniqueness?

We will briefly address these questions in Sect. 3.3.3.

3.3.2.2 Examples

Here are some basic examples of time-homogeneous diffusions:

- i. SBM: $b(x) = 0$, $\sigma(x) = 1$; BM: $b(x) = b$, $\sigma(x) = \sigma$.

⁶See Remark 3.3.9 regarding the concept of a weak solution.

ii. Ornstein-Uhlenbeck process: $\mathcal{S} = (-\infty, \infty)$,

$$b(x) = a(b - x), \quad \sigma(x) = \sigma, \quad \text{with } a, b, \sigma > 0.$$

This is a mean-reverting process.

iii. Mean-reverting square-root process: $\mathcal{S} = (0, \infty)$,

$$b(x) = a(b - x), \quad \sigma(x) = \sigma\sqrt{x} \quad \text{with } a, b, \sigma > 0.$$

iv. Constant elasticity of variance diffusion: $\mathcal{S} = (0, \infty)$,

$$b(x) = bx, \quad \sigma(x) = \sigma x^\gamma \quad \text{with } b, \sigma > 0, 0 \leq \gamma \leq 2.$$

v. Geometric Brownian motion S_t (GBM): $\mathcal{S} = (0, \infty)$,

$$S_t = \exp(X_t) \quad \text{where } X_t \text{ is BM.}$$

An application of the Itô formula yields

$$b(S) = S \left(b + \frac{\sigma^2}{2} \right), \quad \sigma(S) = \sigma S. \quad (3.53)$$

The formulas (3.53) can also be obtained by direct computation using:

$$\begin{aligned} \mathbb{E}(S_{t+h} - S_t | S_t = S) &= S \left[\exp \left(bh + \frac{\sigma^2}{2} h \right) - 1 \right] \\ &= S \left(1 + bh + \frac{\sigma^2}{2} h + o(h) - 1 \right) = S \left(bh + \frac{\sigma^2}{2} h \right) + o(h) \end{aligned}$$

and

$$\begin{aligned} \mathbb{V}\text{ar}(S_{t+h} - S_t | S_t = S) &= S^2 \exp(2bh + \sigma^2 h) [\exp(\sigma^2 h) - 1] \\ &= S^2 (1 + 2bh + \sigma^2 h + o(h)) (\sigma^2 h + o(h)) \\ &= \sigma^2 S^2 h + o(h). \end{aligned}$$

3.3.2.3 Solving Diffusion SDEs

An explicit formula for a strong solution to the SDE (3.52) is not generally possible (just as with ODEs and PDEs), so we will usually need to approximate solutions using numerical methods that are addressed in Sect. 6.8. Nevertheless, it is sometimes possible to guess an explicit formula for a strong solution to the SDE (3.52) and then use the Itô formula to verify that the divined formula is correct (Property 3 may be verified using the Itô formula).

We will discuss several examples of diffusion SDEs that can be explicitly solved by using Itô formulas.

Example 3.3.3 Consider the equation

$$dX_t = dW_t, \quad t \geq 0, \quad X_0 = 0. \quad (3.54)$$

Here $b(t, x) = 0$, $\sigma(t, x) = 1$. The obvious strong solution is $X = W$.

Exercise 3.3.4 Verify that the Properties 1–3 are satisfied by this solution.

Example 3.3.5 Consider the equation

$$dX_t = b dt + \sigma dW_t, \quad t \geq 0, \quad X_0 = x. \quad (3.55)$$

Here $b(t, x) = b$, $\sigma(t, x) = \sigma$. The obvious strong solution is: $X_t = x + bt + \sigma W_t$.

Exercise 3.3.6 Verify that the Properties 1–3 are satisfied by this solution.

Example 3.3.7 Consider the equation

$$dX_t = dt + 2 \operatorname{sgn}(W_t) \sqrt{X_t} dW_t, \quad t \geq 0, \quad X_0 = 0. \quad (3.56)$$

Here $b(t, x) = 1$, $\sigma(t, x, w) = 2 \operatorname{sgn}(w) \sqrt{x}$, where $\operatorname{sgn}(0) = 0$ and

$$\operatorname{sgn}(x) = \begin{cases} +1 & \text{if } x > 0 \\ -1 & \text{if } x < 0. \end{cases}$$

We verify that $X_t = W_t^2$ is a strong solution to this equation:

- *Property 1.* For every $t \geq 0$ the random variable $X_t = W_t^2$ is a function of W_t .
- *Property 2.* The integral $\int_0^t 1 ds = t$ is well defined. The integral $\int_0^t \operatorname{sgn}(W_s) \sqrt{X_s} dW_s$ is a well defined martingale, because the process $\sqrt{X}_t = |W_t|$ is adapted to the filtration of W and

$$\int_0^t \mathbb{E}[(\operatorname{sgn}(W_s) \sqrt{X_s})^2] ds = \int_0^t s ds = \frac{1}{2}t^2$$

is well defined.

- *Property 3.* Using the Itô formula we get (recall that here $X_0 = 0$)

$$\begin{aligned} X_t &= 2 \int_0^t W_s dW_s + \int_0^t 1 ds \\ &= 2 \int_0^t \operatorname{sgn}(W_s) \sqrt{X_s} dW_s + \int_0^t 1 ds, \quad t \geq 0. \end{aligned}$$

So indeed $X = W^2$ is a strong solution to (3.56).

Remark 3.3.8 Equation (3.56) is not of the form (3.52). In fact, this equation can be considered as part of the following system of SDEs for two processes X and Y :

$$\begin{aligned} dX_t &= dt + 2 \operatorname{sgn}(Y_t) \sqrt{X_t} dW_t, \quad t \geq 0, \quad X_0 = 0 \\ dY_t &= dW_t, \quad t \geq 0, \quad Y_0 = 0. \end{aligned}$$

Remark 3.3.9 A corollary of the so-called Lévy characterization theorem is that an Itô integral is SBM if and only if its sharp bracket equals t . As a consequence, the strong solution $X_t = W_t^2$ to (3.56), jointly with $B_t = \int_0^t \operatorname{sgn}(W_s) dW_s$ (which is SBM by virtue of the above-mentioned result), form a weak solution to the SDE

$$dX_t = dt + 2\sqrt{X_t} dB_t, \quad t \geq 0, \quad X_0 = 0, \quad (3.57)$$

where a weak solution means that a Brownian motion B and a process X jointly satisfying (3.57) are sought simultaneously (as opposed to the Brownian motion W being given and imposed in (3.56) for a strong solution X). In this sense the process X is a homogeneous diffusion with the drift coefficient $b(x) = b(t, x) = 1$ and the diffusion coefficient $\sigma(x) = \sigma(t, x) = 2\sqrt{x}$.⁷

Example 3.3.10 Consider the equation

$$dX_t = \frac{1}{2}X_t dt + X_t dW_t, \quad t \geq 0, \quad X_0 = 1. \quad (3.58)$$

Here $b(t, x) = x/2$ and $\sigma(t, x) = x$. Using Example 3.2.2, we easily deduce that the process $X_t = e^{W_t}$ is a strong solution to this equation.

Exercise 3.3.11 Verify that the Properties 1–3 are verified by this solution.

Example 3.3.12 More generally, consider the equation

$$dX_t = bX_t dt + \sigma X_t dW_t, \quad t \geq 0, \quad X_0 = e^y. \quad (3.59)$$

Here $b(t, x) = bx$ and $\sigma(t, x) = \sigma x$. We have seen in Example 3.2.4 that the GBM

$$X_t = e^{y+(b-\frac{1}{2}\sigma^2)t+\sigma W_t}$$

is a strong solution to (3.59). Note that the random variable $Y_t = \ln X_t$ is $\mathcal{N}(y + (b - \frac{1}{2}\sigma^2)t, \sigma^2 t)$ -normally distributed, i.e. the random variable X_t has a lognormal distribution. In particular, for $b = 0$, $\sigma = 1$ and $y = 1$, we get that

$$X_t = e^{W_t - \frac{1}{2}t} \quad (3.60)$$

⁷Weak solutions are closely related to solutions of the “martingale problem with the data b and σ ”, see the comment following Proposition 12.2.2.

satisfies

$$dX_t = X_t dW_t, \quad t \geq 0, \quad X_0 = 1 \quad (3.61)$$

and thus is a Brownian martingale, called the stochastic exponential of the Brownian motion W .

Example 3.3.13 Consider the Ornstein-Uhlenbeck equation

$$dX_t = a(b - X_t) dt + \sigma dW_t, \quad t \geq 0, \quad X_0 = x. \quad (3.62)$$

Here $b(t, x) = a(b - x)$ and $\sigma(t, x) = \sigma$. The strong solution to (3.62) is the Ornstein-Uhlenbeck (OU) process

$$X_t = xe^{-at} + b(1 - e^{-at}) + \sigma e^{-at} \int_0^t e^{as} dW_s.$$

Exercise 3.3.14 Verify that the Properties 1–3 are satisfied here.

The random variable $\int_0^t e^{as} dW_s$ has a normal distribution with mean zero and variance $\int_0^t e^{2as} ds$. Thus, for the OU process we have

$$X_t \sim \mathcal{N}\left(xe^{-at} + b(1 - e^{-at}), \frac{\sigma^2}{2a}(1 - e^{-2at})\right).$$

If $a > 0$, then for large values of t the distribution of the OU random variable is close to $\mathcal{N}(b, \frac{\sigma^2}{2a})$. This is the reason why the constant b is called the mean reversion level.

Example 3.3.15 Consider the SDE

$$dX_t = -\frac{X_t}{1-t} dt + dW_t, \quad t \in [0, 1), \quad X_0 = 0.$$

The strong solution to this equation is

$$X_t = (1-t) \int_0^t \frac{1}{1-s} dW_s, \quad t \in [0, 1).$$

It can be shown by continuity that $X_1 = 0$. Thus, the process X_t is a Gaussian process with mean and covariance functions given, for $t, s \in [0, 1]$, by:

$$m(t) := \mathbb{E}X_t = 0, \quad c(t, s) := \mathbb{C}\text{ov}(X_t, X_s) = \mathbb{E}(X_t X_s) = \min(t, s) - ts,$$

where, in order to derive the last equality, we apply the isometry property (3.3), which implies that

$$\mathbb{E}\left(\int_0^t \frac{1}{1-r} dW_r\right)^2 = \int_0^t \frac{1}{(1-r)^2} dr = \frac{t}{1-t}.$$

Recall that a Gaussian process is such that any finite linear combination of its marginals follows a Gaussian law. The process X is known as a Brownian bridge.

A Brownian bridge can also be given as

$$Y_t = W_t - tW_1, \quad t \in [0, 1]$$

(see Example 1.3.5 in [205]). Indeed, observe that processes X and Y are both Gaussian with the same mean function $m(t)$ and the same covariance function $c(t, s)$. However, the process Y is not adapted to the filtration of W .

3.3.3 Jump-Diffusions (*)

Let N be $\text{PP}(\lambda)$. In many ways SDEs driven by a Poisson process are easier to deal with than those driven by Brownian motion. As was the case with SDEs driven by Brownian motion, the Itô formula plays a fundamental role in solving SDEs driven by a Poisson process.

Example 3.3.16 The unique strong solution to the equation (for $\gamma > -1$)

$$dX_t = \gamma X_{t-} dN_t, \quad t \geq 0, \quad X_0 = 1,$$

is $X_t = (1 + \gamma)^{N_t}$. In particular, when $\gamma = 1$ we obtain $X_t = 2^{N_t}$.

Example 3.3.17 The unique strong solution to the equation

$$dX_t = bX_{t-}(dt + dN_t), \quad t \geq 0, \quad X_0 = 1,$$

is $X_t = e^{bt}(1 + b)^{N_t}$. In particular, when $b = 1$ we obtain $X_t = e^t 2^{N_t}$.

Example 3.3.18 The unique strong solution to

$$dX_t = X_{t-} dM_t, \quad t \geq 0, \quad X_0 = 1,$$

is

$$X_t = 1 + \int_0^t X_{u-} dM_u = e^{N_t \ln 2 - \lambda t} = e^{-\lambda t} 2^{N_t}.$$

Thus the process $e^{-\lambda t} 2^{N_t}$ is a martingale. Compare this with (3.60) and (3.61).

By a jump-diffusion we mean hereafter an Itô process X in the sense of Sect. 3.2.3, but for a Markovian SDE (3.24), meaning that the random coefficients b_t , σ_t and $\delta_t(x)$ of (3.24) are now given deterministically in terms of X_{t-} ⁸ as

$$b_t = b(t, X_t), \quad \sigma_t = \sigma(t, X_t), \quad \delta_t(x) = \delta(t, X_{t-}, x). \quad (3.63)$$

⁸Or of X_t , in the case of b and σ , which in view of (3.7) makes no difference in (3.24), by the continuity of t and W_t .

Equation (3.24) is thus now an SDE (implicit) in X . Well-posedness⁹ of such jump-diffusion SDEs can be studied by classical Picard iteration techniques under suitable Lipschitz and growth conditions on the coefficients (see Chap. 12). A notable feature of the solution is the so-called ***Markov property***, meaning that

$$\mathbb{E}(\phi(X_s, s \in [t, T]) | \mathcal{F}_t) = \mathbb{E}(\phi(X_s, s \in [t, T]) | X_t) \quad (3.64)$$

holds for every (possibly path-dependent) functional Φ of X that makes sense on both sides of the equality. Thus the past of X doesn't influence its future; the present of X provides all the relevant information.

Given a real valued and “sufficiently regular” function $u = u(t, x)$, by (3.30) we have the following ***special semimartingale Itô formula for a jump-diffusion***: for any $t \in [0, T]$,

$$du(t, X_t) = (\partial_t + \mathcal{A})u(t, X_t) dt + \partial u(t, X_t)\sigma_t dW_t + \delta u(t, X_{t-}) \cdot d\mu_t \quad (3.65)$$

for the compensated jump (local) martingale

$$\delta u(t, X_{t-}) \cdot d\mu_t := \delta u(t, X_{t-}, J_{(t)}) dN_t - \lambda \bar{\delta} u(t, X_t) dt$$

in which we let, for $t \in [0, T]$ and x, y in \mathbb{R}^d ,

$$\delta u(t, x, y) = u(t, x + \delta(t, x, y)) - u(t, x), \quad \bar{\delta} u(t, x) = \int_{\mathbb{R}^d} \delta u(t, x, y) w(dy)$$

(assumed well-defined), and where “the infinitesimal generator \mathcal{A} of X acts on u ” by

$$(\mathcal{A}u)(t, x) = \partial u(t, x)b(t, x) + \frac{1}{2}\partial^2 u(t, x) : a(t, x) + \lambda \bar{\delta} u(t, x). \quad (3.66)$$

Therefore, for all suitable functions u, v of (t, x) (cf. (3.36)),

$$\begin{aligned} \mathbb{E}(du(t, X_t) | \mathcal{F}_t) &= (\partial_t + \mathcal{A})u(t, X_t) dt \\ \text{Cov}(du(t, X_t), dv(t, X_t) | \mathcal{F}_t) &= (\mathcal{A}(uv) - u\mathcal{A}v - v\mathcal{A}u)(t, X_t) dt, \end{aligned} \quad (3.67)$$

in which, with $Y_t = u(t, X_t)$ and $Z_t = v(t, X_t)$,

$$\begin{aligned} &(\mathcal{A}(uv) - u\mathcal{A}v - v\mathcal{A}u)(t, X_t) \\ &= \frac{d\langle Y, Z \rangle_t}{dt} \\ &= \partial u(t, X_t)a_t(\partial v(t, X_t))^T + \lambda \int_{\mathbb{R}^d} \delta u(t, X_t, y)\delta v(t, X_t, y)w(dy). \end{aligned} \quad (3.68)$$

Also, by the Markov property of X , the conditionings with respect to \mathcal{F}_t in (3.67) can be replaced by conditionings with respect to X_t .

⁹In the strong sense.

3.4 Girsanov Transformations

Girsanov transformation is a very useful technique of converting semimartingales into martingales. It amounts to changing probabilities of random events (changing probability measures). Throughout this section we denote by \mathbb{Q} a changed measure and by $\tilde{\mathbb{E}}$ the related expectation operator, whereas the original measure and expectation are denoted, as usual, by \mathbb{P} and \mathbb{E} .

3.4.1 Girsanov Transformation for Gaussian Distributions

3.4.1.1 Gaussian Random Variables

Suppose ε is a standard normal variable, i.e. $\varepsilon \stackrel{\mathbb{P}}{\sim} \mathcal{N}(0, 1)$. Its probability density function is $g(x) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{x^2}{2})$, $x \in (-\infty, \infty)$. Now consider a function $w(x) = e^{qx - \frac{q^2}{2}}$, where q is a constant. We transform the probability measure \mathbb{P} via the function w in order to produce a new probability measure on Ω , denoted by \mathbb{Q} and defined by

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) = \mu(\omega), \quad (3.69)$$

where $\mu = w(\varepsilon)$. We call the random variable μ the density of the measure \mathbb{Q} with respect to the measure \mathbb{P} . This is equivalent to writing

$$d\mathbb{Q} = \mu d\mathbb{P} \quad (\text{implying that } \mathbb{Q} \text{ is absolutely continuous with respect to } \mathbb{P})$$

or

$$d\mathbb{P} = \mu^{-1} d\mathbb{Q} \quad (\text{implying that } \mathbb{P} \text{ is absolutely continuous with respect to } \mathbb{Q}).$$

We say that measures \mathbb{P} and \mathbb{Q} are equivalent with respect to each other.

Exercise 3.4.1 Verify that \mathbb{Q} is a probability measure.

Hint Since \mathbb{Q} is obviously nonnegative and σ -additive, verifying that \mathbb{Q} is a probability measure amounts to verifying that $\mathbb{Q}(\Omega) = 1$, where $\mathbb{Q}(\Omega) = \int_{\Omega} d\mathbb{Q}(\omega) = \int_{\Omega} \mu(\omega) d\mathbb{P}(\omega) = \mathbb{E}\mu = \int_{-\infty}^{\infty} w(x)\gamma(x) dx$.

In view of (3.69) we obtain

$$\begin{aligned} \mathbb{Q}(\varepsilon \in dx) &= w(x)\mathbb{P}(\varepsilon \in dx) \\ &= e^{qx - \frac{q^2}{2}} (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} dx \\ &= (2\pi)^{-\frac{1}{2}} e^{-\frac{(x-q)^2}{2}} dx. \end{aligned}$$

But the function $(2\pi)^{-\frac{1}{2}} e^{-\frac{(x-q)^2}{2}}$ is the density of the normal distribution $\mathcal{N}(q, 1)$. Thus the random variable ε has the normal distribution $\mathcal{N}(q, 1)$ under the measure \mathbb{Q} , which we write $\varepsilon \sim \mathcal{N}(q, 1)$.

3.4.1.2 Brownian Motion

Let q be a constant and consider the process B defined by

$$B_t = W_t - qt, \quad t \in [0, T],$$

where the process W_t is SBM under some filtered probability space (\mathbb{F}, \mathbb{P}) . Define a process μ_t by

$$\mu_t(\omega) = \exp\left(q W_t(\omega) - \frac{1}{2} q^2 t\right), \quad t \in [0, T], \quad (3.70)$$

and then define a new measure \mathbb{Q} on \mathcal{F}_T by

$$d\mathbb{Q}(\omega) = \mu_T(\omega) d\mathbb{P}(\omega).$$

A remarkable result, known as the Girsanov Theorem, states that:

- the process μ_t is a martingale with respect to the filtration \mathcal{F}_t under the measure \mathbb{P} ,
- the measure \mathbb{Q} is a probability measure,
- the process B_t is SBM under the measure \mathbb{Q} .

Remark 3.4.2 The Girsanov theorem cannot be generalized to the case $T = \infty$.

For us, the most important application of the Girsanov theorem is the following:

Example 3.4.3 (Elimination of the drift term in a continuous SDE) Consider the linear SDE

$$dX_t = b X_t dt + \sigma X_t dW_t, \quad t \geq 0, \quad X_0 = e^y. \quad (3.71)$$

From Example 3.59 we know that the strong solution to this equation is

$$X_t = e^{y + (b - \frac{1}{2}\sigma^2)t + \sigma W_t}.$$

This is not a martingale under \mathbb{P} . We introduce a process

$$B_t = W_t - qt$$

with $q = -\frac{b}{\sigma}$ and rewrite (3.71) as

$$dX_t = \sigma X_t dB_t, \quad t \geq 0, \quad X_0 = e^y. \quad (3.72)$$

By the Girsanov theorem, the process B is SBM under \mathbb{Q} . Thus the process

$$X_t = e^{y + (b - \frac{1}{2}\sigma^2)t + \sigma W_t} = e^{y - \frac{1}{2}\sigma^2 t + \sigma B_t},$$

which is the strong solution to both (3.71) and (3.72), is a martingale under \mathbb{Q} . This observation plays a crucial role in the so-called risk-neutral approach to pricing financial assets (see Chap. 4).

Exercise 3.4.4 Verify by a direct computation that the strong solution to (3.72) is a martingale under \mathbb{Q} .

Exercise 3.4.5 Write an SDE satisfied by the process μ_t of (3.70).

3.4.2 Girsanov Transformation for Poisson Distributions

3.4.2.1 Poisson Random Variables

Let $v \stackrel{\mathbb{P}}{\sim} \mathcal{P}_\lambda$, so that

$$\mathbb{P}(v = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

for $k \geq 0$ and is zero otherwise. Letting $w(k) = e^{(\lambda-\gamma)} (\frac{\gamma}{\lambda})^k = e^{(\lambda-\gamma)-k \ln \frac{\lambda}{\gamma}}$ for some $\gamma > 0$, define a new measure \mathbb{Q} on (Ω, \mathcal{F}) by

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) = \mu(\omega),$$

where $\mu = w(v)$. Note that

$$\mathbb{E}\mu = \int \mu(\omega) \mathbb{P}(d\omega) = \sum_k e^{(\lambda-\gamma)} \left(\frac{\gamma}{\lambda}\right)^k e^{-\lambda} \frac{\lambda^k}{k!} = 1,$$

which shows that \mathbb{Q} is a probability measure (with total mass equal to one). Now observe that

$$\begin{aligned} \mathbb{Q}(v = k) &= \int_{\{\omega; v(\omega)=k\}} d\mathbb{Q}(\omega) = \int_{\{\omega; v(\omega)=k\}} \mu(\omega) d\mathbb{P}(\omega) \\ &= e^{(\lambda-\gamma)-k \ln \frac{\lambda}{\gamma}} \mathbb{P}(v = k) = \frac{e^{-\gamma} \gamma^k}{k!} \end{aligned}$$

for $k \geq 0$ and is zero otherwise. Thus,

$$v \stackrel{\mathbb{Q}}{\sim} \mathcal{P}_\gamma.$$

3.4.2.2 Poisson Process

Let N_t be $\text{PP}(\lambda)$ on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Now, for $\gamma > 0$, define a process μ_t by

$$\mu_t(\omega) = e^{(\lambda-\gamma)t} \left(\frac{\gamma}{\lambda} \right)^{N_t}, \quad t \in [0, T], \quad (3.73)$$

and then define a new measure \mathbb{Q} on (Ω, \mathcal{F}_T) by

$$d\mathbb{Q}(\omega) = \mu_T(\omega) d\mathbb{P}(\omega).$$

An appropriate version the Girsanov Theorem states that:

- the process μ_t is a martingale with respect to the filtration \mathbb{F} under the measure \mathbb{P} ,
- the measure \mathbb{Q} is a probability measure,
- the process N_t is $\text{PP}(\gamma)$ under the measure \mathbb{Q} .

Example 3.4.6 (Elimination of the drift term in a linear SDE with jumps) Consider the linear SDE

$$dX_t = X_{t-}(-\gamma dt + dN_t), \quad t \geq 0, \quad X_0 = 1. \quad (3.74)$$

The process X_t is not a martingale under \mathbb{P} , but the above equation can of course be written as

$$dX_t = X_{t-} d\tilde{M}_t, \quad t \geq 0, \quad X_0 = 1, \quad (3.75)$$

where we let $\tilde{M}_t = N_t - \gamma t$. So the process X_t is a martingale under \mathbb{Q} .

Exercise 3.4.7 Verify by a direct computation that the strong solution to (3.75) is a martingale under \mathbb{Q} .

Exercise 3.4.8 Take $\lambda = 1$. Verify that the process μ of (3.73) is the strong solution to

$$d\mu_t = (\gamma - 1)\mu_{t-}(-dt + dN_t), \quad t \geq 0, \quad \mu_0 = 1.$$

Remark 3.4.9 A Girsanov transformation can be applied jointly to a pair (W, N) , where W and N are, respectively, a Brownian motion and a Poisson process defined on the same probability space, say $(\Omega, \mathcal{F}, \mathbb{P})$. If $W \sim \text{BM}(x, b, \sigma)$ and $N \sim \text{PP}(\lambda)$, then we can apply a simultaneous change of probability measure \mathbb{P} to a new measure \mathbb{Q} , so that under the new measure we have $W \sim \text{BM}(x, m, \sigma)$ and $N \sim \text{PP}(\gamma)$. See also Lemma 12.3.6.

3.4.3 Abstract Bayes Formula

Suppose that ξ is an \mathcal{F}_T -measurable and integrable random variable and let \mathbb{Q} be defined via $\frac{d\mathbb{Q}}{d\mathbb{P}} = \mu_T$, for some positive \mathbb{P} -martingale μ with unit mean under \mathbb{P} . We state the following lemma (see Proposition III.3.8 of [153] for a proof).

Lemma 3.4.10 *A process X is a \mathbb{Q} -local martingale if and only if μX is a \mathbb{P} -local martingale.*

Writing $\mathbb{E} = \mathbb{E}^{\mathbb{P}}$, $\widetilde{\mathbb{E}} = \mathbb{E}^{\mathbb{Q}}$, we obtain as a consequence:

Proposition 3.4.11 *The following Bayes formula holds for every \mathbb{Q} -integrable random variable ξ :*

$$\mu_t \widetilde{\mathbb{E}}(\xi | \mathcal{F}_t) = \mathbb{E}(\mu_T \xi | \mathcal{F}_t). \quad (3.76)$$

Proof We have a \mathbb{Q} -martingale $\widetilde{\mathbb{E}}(\xi | \mathcal{F}_t)$ and therefore, by Lemma 3.4.10, a \mathbb{P} -local martingale $\mu_t \widetilde{\mathbb{E}}(\xi | \mathcal{F}_t)$. Since $\mu_t \widetilde{\mathbb{E}}(\xi | \mathcal{F}_t)$ is a \mathbb{P} -martingale (admitting the required integrability) with terminal condition $\mu_T \xi$ at T , (3.76) results. \square

3.5 Feynman-Kac Formulas (*)

3.5.1 Linear Case

Let X be given as a jump-diffusion and let $u = u(t, x)$ be a regular function such that $u(t, X_t)$ is a local martingale. In view of the Itô formula in martingale form (3.65), we conclude from Proposition 3.1.12 that the time-differentiable local martingale

$$(\partial_t + \mathcal{A})u(t, X_t) dt = du(t, X_t) - \partial u(t, X_t) \sigma_t dW_t - \delta u(t, X_{t-}) \cdot d\mu_t$$

is constant. This in turn translates into the following partial integro-differential equation (deterministic PIDE) to be satisfied by the function u :

$$(\partial_t + \mathcal{A})u(t, x) = 0, \quad x \in \mathbb{R}^d. \quad (3.77)$$

The fundamental situation of this kind corresponds to a Doob-martingale

$$u(t, X_t) := \mathbb{E}(\phi(X_T) | X_t) = \mathbb{E}(\phi(X_T) | \mathcal{F}_t) \quad (3.78)$$

for an integrable terminal condition $\phi(X_T)$. In (3.78), the second equality, which implies the martingale property of $u(t, X_t)$, holds by virtue of the Markov property (3.64) of a jump-diffusion X . In this case the function u can typically be characterized as the unique solution to the PIDE (3.77), along with the terminal condition $u = \phi$ at time T .

More generally, given suitable running and terminal cost functions f and ϕ as well as a funding cost function r , we have

$$\begin{aligned} u(t, X_t) &:= \mathbb{E}\left(\int_t^T e^{-\int_t^s r(\zeta, X_\zeta) d\zeta} f(s, X_s) ds + e^{-\int_t^T r(s, X_s) ds} \phi(X_T) \middle| X_t\right) \\ &= \mathbb{E}\left(\int_t^T e^{-\int_t^s r(\zeta, X_\zeta) d\zeta} f(s, X_s) ds + e^{-\int_t^T r(s, X_s) ds} \phi(X_T) \middle| \mathcal{F}_t\right), \end{aligned} \quad (3.79)$$

by the Markov property of X . By an immediate extension of the previous computations, we have:

- on the one hand, the following (local) martingale that arises from the Itô formula applied to $u(t, X_t)$:

$$du(t, X_t) - (\partial_t u + \mathcal{A}u)(t, X_t) dt = \partial u(t, X_t) \sigma_t dW_t + \delta u(t, X_{t-}) \cdot d\mu_t, \quad (3.80)$$

- on the other hand, the following Doob-martingale (conditional expectation of an integrable terminal condition) that arises from (3.79):

$$du(t, X_t) + (f(t, X_t) - ru(t, X_t)) dt. \quad (3.81)$$

Subtraction of (3.80) from (3.81) yields the local martingale

$$(\partial_t u + \mathcal{A}u + f - ru)(t, X_t) dt, \quad (3.82)$$

which is therefore constant as a time-differentiable local martingale (by Proposition 3.1.12). Also, accounting for the terminal condition $u = \phi$ at time T , this translates into the following PIDE that is satisfied by the function u :

$$\begin{cases} u(T, x) = \phi(x), & x \in \mathbb{R}^d \\ (\partial_t u + \mathcal{A}u + f - ru)(t, x) = 0, & t < T, x \in \mathbb{R}^d. \end{cases} \quad (3.83)$$

The function u can then typically be characterized and computed (including numerically, if needed/possible; see Parts III and V) for the (in some sense) unique solution to (3.83).

3.5.2 Backward Stochastic Differential Equations (BSDEs)

The SDEs that were discussed in Sect. 3.3 are so-called forward SDEs, because any solution process of any such equation is supposed to satisfy the given *initial* condition. In Example 3.3.3 we considered the equation

$$dY_t = dW_t, \quad t \in [0, T], \quad Y_0 = 0, \quad (3.84)$$

which has the obvious solution $Y_t = W_t$. In the above equation the initial condition is specified—the equation is to be solved forward in time. The *backward* version of (3.84) would read

$$dY_t = dW_t, \quad t \in [0, T], \quad Y_T = \xi, \quad (3.85)$$

where ξ is some random variable. In (3.85) the terminal condition is specified—the equation is to be solved backward in time and is therefore called a backward stochastic differential equation (BSDE). It is rather clear that (3.85) is only solvable if $\xi = W_T + c$, where c is a constant, in which case we have $Y_t = W_t + c$. Note that for $c = 0$ the solution to this equation is the same as the solution to (3.84), i.e. $Y_t = W_t$. Also note that (3.85) can be written as

$$dY_t = Z_t dW_t, \quad t \in [0, T], \quad Y_T = \xi, \quad (3.86)$$

where $Z_t = 1$ (observe that the process Z is trivially adapted to the filtration \mathcal{F}_t , $t \in [0, T]$, generated by the Brownian motion W). In the case when $\xi = W_T + c$ we thus see that the BSDE (3.86) admits a solution pair (Y_t, Z_t) that is adapted to the filtration \mathbb{F} , where $Y_t = W_t + c$ and $Z_t = 1$.

In more generality, consider the BSDE (3.86), where ξ is some random variable measurable with respect \mathcal{F}_T (not necessarily $\xi = W_T + c$). If we additionally suppose that ξ is square integrable, then it can be shown that BSDE (3.86) admits a solution pair (Y_t, Z_t) that is adapted to the filtration \mathbb{F} , where $Y_t = \mathbb{E}(\xi | \mathcal{F}_t)$ and Z_t is a unique process appearing in the so-called Brownian martingale representation of ξ , i.e. $\xi = \mathbb{E}(\xi) + \int_0^T Z_s dW_s$. We then have

$$Y_t = Y_0 + \int_0^t Z_s dW_s = \mathbb{E}(\xi) + \int_0^t Z_s dW_s = \xi - \int_t^T Z_s dW_s.$$

See the preface of the book and the introduction of Chap. 4 for presentations “in words” of BSDEs, which are studied mathematically in the next subsection at a heuristic level and in Part V with all the required mathematical rigor and proofs.

3.5.3 Nonlinear Feynman-Kac Formula

In the jump-diffusion setup of Sect. 3.5, in view of (3.80) and (3.83), we can interpret the triple of processes (parameterized by $x \in \mathbb{R}^d$ in the case of V)

$$(Y_t, Z_t, V_t(x)) := (u(t, X_t), \partial u(t, X_t) \sigma(t, X_t), \delta u(t, X_{t-}, x)) \quad (3.87)$$

as a solution to the following BSDE:

$$\begin{cases} Y_T = \phi(X_T) & \text{and for } t < T, \\ -dY_t = (f(t, X_t) - r(t, X_t)Y_t) dt - Z_t dW_t - V_t \cdot d\mu_t. \end{cases} \quad (3.88)$$

In the case of a diffusion X (without jumps, so that $\lambda = 0$), there is no component V involved in the solution (or $V = 0$). The Feynman-Kac formula (3.79) for the solution u of (3.83), written in the equivalent form (as apparent from (3.88))

$$Y_t = \mathbb{E} \left(\int_t^T (f(s, X_s) - r(s, X_s) Y_s) ds + \phi(X_T) \middle| \mathcal{F}_t \right), \quad (3.89)$$

can thus also be regarded as the Feynman-Kac representation of the solution (Y, Z, V) to the BSDE (3.88).

Note that the “intrinsic” (non discounted) form (3.89) of the Feynman-Kac representation (3.79) is implicit, meaning that the right-hand side of (3.89) also depends on Y . This is, however, not a real issue in this case, as revealed by the equivalent explicit discounted representation (3.79). Now, the power of BSDEs lies precisely in the fact that this theory allows us to solve problems that are more general than the linear equations (3.83), (3.88), i.e. nonlinear problems in which the BSDE coefficient $g(t, X_t, Y_t) (= f(t, X_t) - r Y_t$ in the case of (3.83), (3.88)) depends nonlinearly on Y and possibly also on Z and V . We can thus consider the following BSDE, to be solved for the triple of processes $(Y_t, Z_t, V_t(x))$:

$$\begin{cases} Y_T = \phi(X_T) & \text{and for } t < T, \\ -dY_t = g(t, X_t, Y_t, Z_t, \widehat{V}_t) dt - Z_t dW_t - V_t \cdot d\mu_t, \end{cases} \quad (3.90)$$

where $\widehat{V}_t := \int_{\mathbb{R}^d} V_t(y) \eta(t, X_t, y) w(dy)$ for a suitable (possibly vector-valued) integration kernel η . Now let a function $u = u(t, x)$ be a solution to the following semilinear PIDE:

$$\begin{cases} u(T, x) = \phi(x), & x \in \mathbb{R}^d, \\ \partial_t u(t, x) + \mathcal{A}u(t, x) \\ \quad + g(t, x, u(t, x), \partial u(t, x)\sigma(t, x), \widehat{\delta}u(t, x)) = 0, & t < T, x \in \mathbb{R}^d, \end{cases} \quad (3.91)$$

with $\widehat{\delta}u(t, x) := \int_{\mathbb{R}^d} \delta u(t, x, y) \eta(t, x, y) w(dy)$. Straightforward extensions of the computations that led from (3.83) to (3.88) show that the triple (Y, Z, V) , given in terms of u by (3.87), satisfies the nonlinear BSDE (3.90). For this reason the formula (3.87) is sometimes called a nonlinear Feynman-Kac formula.

Remark 3.5.1 In fact, as we will see in Part V, the simplest way to rigorously solve the PIDE (3.91) is actually to go in the other direction, namely first solving the BSDE (3.90) for a triple of processes $(Y_t, Z_t, V_t(x))$ and then redoing the above computations in the reverse order so as to establish that the function u defined via $u(t, X_t) = Y_t$ satisfies (3.91).

3.5.4 Optimal Stopping

BSDEs allow us to deal not only with semilinearity, which refers to nonlinear dependence of the coefficient g with respect to (Y, Z, V) , but also with the nonlinear-

ity that may result from optimal stopping features. E.g. in (3.79) we may consider, instead of u , the function $v = v(t, x)$ such that

$$\begin{aligned} v(t, X_t) \\ := \text{esssup}_{\tau \in \mathcal{T}_t} \mathbb{E} \left(\int_t^\tau e^{-\int_s^\tau r(\xi, X_\xi) d\xi} f(s, X_s) ds + e^{-\int_t^\tau r(s, X_s) ds} \phi(X_\tau) \middle| X_t \right) \end{aligned} \quad (3.92)$$

or, in equivalent implicit intrinsic form:

$$\begin{aligned} v(t, X_t) \\ = \text{esssup}_{\tau \in \mathcal{T}_t} \mathbb{E} \left(\int_t^\tau (f(s, X_s) - r(s, X_s)v(s, X_s)) ds + \phi(X_\tau) \middle| X_t \right). \end{aligned} \quad (3.93)$$

In these dynamic programming equations, \mathcal{T}_t denotes the set of all $[t, T]$ -valued stopping times. In particular, we have $v(t, x) \geq \phi(x)$ (respectively $v(t, x) \geq u(t, x)$), as follows by considering $\tau \equiv t$ (respectively $\tau \equiv T$) in (3.92), where the conditioning with respect to X_t can be replaced by the conditioning with respect to \mathcal{F}_t , just as in (3.79) it was replaced, by the Markov property of a jump-diffusion X .

Remark 3.5.2 The set \mathcal{T}_t is uncountable, so that care is needed in taking the supremum over an uncountable family of random variables on the right-hand side of (3.92). The supremum must thus be understood in the essential supremum (“esssup”) sense.

As will be seen in Part V, we can also establish a nonlinear Feynman-Kac formula of the form (3.87) for:

- on the one hand, the solution (Y, Z, V, A) to the following reflected BSDE:

$$\begin{cases} Y_T = \phi(X_T) \quad \text{and, for } t < T: \\ -dY_t = g(t, X_t, Y_t, Z_t, \hat{V}_t) dt + dA_t - Z_t dW_t - V_t \cdot d\mu_t, \\ Y_t \geq \phi(X_t), \quad (Y_t - \phi(X_t))^+ dA_t = 0; \end{cases} \quad (3.94)$$

here A represents an additional continuous and nondecreasing process which is required for preventing the component Y_t of the solution from falling below the barrier level $\phi(X_t)$. This nondecreasing process is only allowed to increase on the event that $\{Y_t = \phi(X_t)\}$, as imposed by the minimality condition in the third line.

- on the other hand, the following obstacle problem:

$$\begin{cases} v(T, x) = \phi(x), \quad x \in \mathbb{R}^d, \\ \max(\partial_t v(t, x) + \mathcal{A}v(t, x) + g(t, x, v(t, x), \partial v(t, x)\sigma(t, x), \hat{\delta}v(t, x)), \\ \quad \phi(t, x) - v(t, x)) = 0, \quad t < T, \quad x \in \mathbb{R}^d. \end{cases} \quad (3.95)$$

The corresponding generalization of (3.93) is

$$Y_t = \operatorname{esssup}_{\tau \in \mathcal{T}_t} \mathbb{E} \left(\int_t^\tau g(s, X_s, Y_s, Z_s, \hat{V}_s) ds + \phi(X_\tau) \mid X_t \right). \quad (3.96)$$

The proof of these results involves some technicalities, in particular since the value function v of an optimal stopping problem is only of class $C^{1,1}$ (at best) on the exercise boundary $\{v = \phi\}$. This implies that v being a “solution” to (3.95) (which involves, through the generator \mathcal{A} , the second derivatives in space of v) can only be understood in a weak sense. BSDEs are useful here since there is a well established theory connecting solutions of BSDEs to the so-called viscosity class of weak solutions to nonlinear PDEs or PIDE (see Sect. 8.1.2 and Part V).

Note that BSDEs are not only useful theoretically, but also in practice, when the dimension d of X is such that the numerical solution of a nonlinear PIDE by a deterministic scheme is ruled out by Bellman’s “curse of dimensionality”. For solving such high-dimensional nonlinear problems, simulation schemes are the only possibility: BSDE-based simulation schemes (better known among practitioners in the form of the “American Monte Carlo” pricing schemes of Sect. 6.10 and Part IV) or branching particle schemes (see e.g. [135]).

Part II

Pricing Equations

Equipped with the necessary mathematical tools, we are now in a position to derive the pricing equations of financial claims. In Chap. 4 we derive such equations (both stochastic BSDEs and deterministic PDEs or PIDEs) in a general framework and, in Chap. 5, we review reference models in various asset classes.

Chapter 4

Martingale Modeling

In this chapter we show how the task of pricing and hedging financial derivatives can be reduced to that of solving related backward stochastic differential equations, called stochastic pricing equations in this book, equivalent to the deterministic pricing equations that arise in Markovian setups. The deterministic pricing equations, starting with the celebrated Black–Scholes equation, are better known to practitioners. However, these deterministic partial-differential equations, also including integral terms in models with jumps, are more “model dependent” than the stochastic pricing equations. Moreover, the deterministic pricing equations are less general since they are only available in Markovian setups. In addition, the mathematics of pricing and hedging financial derivatives is in fact simpler in terms of the stochastic pricing equations. Indeed, rigorous demonstrations based on the deterministic pricing equations unavoidably involve¹ technical notions of viscosity or Sobolev solutions.

The stochastic pricing equations belong to the class of the so-called backward stochastic differential equations (BSDEs, or reflected BSDEs in the case of products with early exercise clauses). The backward terminology refers to the fact that these equations are stated in terms of a terminal condition (random variable) ξ at a future maturity time T . Backward stochastic differential equations were first introduced by Bismut [46] in the 1970s in the case of a linear driver, and then by Peng and Pardoux [218] in the late 1980s. They have been extensively studied since then, particularly in relation to mathematical finance (see El Karoui et al. [114] for a seminal reference). To restate in financial terms the mathematical notions introduced in Sect. 3.5 and thoroughly investigated in Part V, the solution to a BSDE consists of a pair of processes (Π, Δ) , in which Π corresponds to the price of a financial derivative and Δ to its hedge. In the simplest case, a solution to a backward stochastic differential equation is obtained by invocation of a martingale representation theorem. Yet the theory of backward stochastic differential equations, properly speaking,

¹As soon as early exercise clauses and related obstacles in the deterministic equations come into the picture.

begins with the presence of an implicit driver coefficient (corresponding to the running cost of a control problem) $g = g_t(\Pi_t, \Delta_t)$ in the equation. If g is nonlinear, Picard iteration and a contraction argument are needed, beyond a martingale representation theorem, to solve a BSDE. In most real-life pricing applications, g is either exogenous (explicit) or linear; however, this is sometimes not the case, as with the nonlinear funding issues that have become stringent since the crisis [30, 80, 216]. Another important practical case of nonlinearity is not through the driver coefficient g , but via the presence of early exercise features in the case of American and game claims. In this case a third component in the solution to the pricing BSDE, the so-called reflecting process (process A throughout this book) is needed to maintain the price process Π between the early exercise payoff processes; see El Karoui et al. [112].

Well-posedness of any stochastic or deterministic pricing equation in suitable spaces of solutions is taken for granted whenever needed in this chapter, as is also the connection between stochastic and deterministic pricing equations. These topics were heuristically addressed through the Itô formula in Sect. 3.5. The interested reader is referred to Part V for more formal developments.

Outline Using BSDEs as a main tool, Sect. 4.1 summarizes the theory of risk-neutral pricing and hedging of financial derivatives that has been developed since the seminar paper of Black and Scholes [47] and almost one century after the precursory work of Louis Bachelier [14]. The central result, Proposition 4.1.15, can be informally stated as follows: under the assumption that a pricing BSDE, defined under a risk-neutral (also called pricing) probability measure \mathbb{P} over a primary market of hedging instruments, admits a solution Π , we have that Π is the minimal superhedging price up to a \mathbb{P} -local martingale cost process for the related derivative. This notion of hedge with a local martingale cost under a pricing measure establishes a connection between arbitrage prices and hedging in a rather general, possibly incomplete, market. In the case of a complete market the cost process is equal to 0.

Although theoretically pleasing, these results do not immediately lend themselves to practical computations. To make them practical, we need to specify them for a Markovian setup, which is done in Sect. 4.2. A Markovian setup is one in which the current state of the system sums up all relevant information regarding its future evolution; conditioning with respect to the information flow until today is thus equivalent to conditioning with respect to the current state. In this case, an analytic (as opposed to stochastic) approach may also be developed and more explicit hedging strategies may be computed.

Section 4.3 presents various extensions of these results. Section 4.3.1 thus generalizes the risk-neutral approach to a martingale modeling approach with respect to an arbitrary numéraire B (positive primary asset price process) which may be used for discounting other price processes, rather than the savings account (riskless asset) in the standard risk-neutral approach. This extension is particularly important for dealing with interest rate derivatives. Section 4.3.2 refines the risk-neutral martingale modeling approach of Sects. 4.1 and 4.2 to the case, important for counterparty risk modeling, of defaultable derivatives, with all cash flows stopped at the

default time τ_d of the issuer of the contract. Section 4.3.3 is an introduction to the topic of intermittent call protection, to be thoroughly investigated in Chap. 10 and Sect. 14.2. Section 4.4 makes the connection between theory and the way models are used in practice.

4.1 General Setup

Since the precursor work of Bachelier [14], the complexity of financial markets has led to modeling their evolution as random processes, a key notion in this regard being that of information. Mathematically, the proper setup for modeling random evolutions in dynamic informational contexts, is that of stochastic processes defined on filtered probability spaces. Moreover, whatever the real nature of time in finance may be, it is often more convenient mathematically to model the time as a continuous rather than a discrete variable. The evolution of a financial market model is thus given throughout most of the book in terms of stochastic processes defined on a filtered space $(\Omega, \mathcal{F}, \mathbb{F})$, where (Ω, \mathcal{F}) is a measurable space of events, and where the filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$ is given as an increasing flow of σ -fields $\mathcal{F}_t \subseteq \mathcal{F}$ on Ω , indexed by time t varying between 0 and T . In the financial interpretation, time 0 is to be understood as the current “pricing time”, and the constant $T > 0$ represents the maturity of a financial derivative.

The next step consists in assigning probabilities to events, namely defining a probability measure on the measurable space (Ω, \mathcal{F}) . Here one of the main messages of mathematical finance theory is that, in absence of arbitrage opportunities in a financial market, prices of traded assets can be represented as expectations of future cash flows discounted at the risk-free rate. Yet this doesn’t hold under the real world (also called objective, or statistical, or historical, or physical, or actuarial) probability measure, for otherwise financial assets would trade at no risk premium. It only holds under a changed measure, referred to as a risk-neutral or pricing measure. Moreover, the actuarial and pricing measures are equivalent, i.e. events which are possible under the one measure are also possible under the other. This issue of measures, which at first may seem to be a mathematical subtlety, is in fact fundamental. We thus need to deal with at least two probability measures on the events space Ω . The first, denoted hereafter by $\widehat{\mathbb{P}}$, corresponds to the historical probability measure that describes the statistical behavior of market factors. The other, denoted by \mathbb{P} , corresponds to the pricing measure which is implicit in the market prices interpreted as risk-neutral expectations of their future cash flows discounted at the risk-free rate.

Since this is a book about pricing, we will typically work under a risk-neutral measure \mathbb{P} , rather than under the historical measure $\widehat{\mathbb{P}}$. In general we will take the liberty to introduce and define a pricing model directly under \mathbb{P} . This of course doesn’t mean that analyzing a financial market under the historical measure $\widehat{\mathbb{P}}$ is not important; such analysis must actually come first in the modeling process. It means simply that this task has already been done. We thus take a pricing model

for granted, one that we suppose gives a realistic view of the financial market under consideration, up to an equivalent change of measure.

Remark 4.1.1 The validity of a risk-neutral modeling approach is subject to conditions that can be summarized by saying that one is dealing with a perfect financial market, without frictions of any kind. Regarding a financial derivative, this approach is only legitimate for a replicable claim, or at least (if the market is incomplete) for a liquidly traded derivative, so that the derivative and the primary market can be viewed as together forming an extended perfect market. Otherwise a risk-neutral pricing approach is not justified, and one should introduce risk premia explicitly into the picture. Throughout this book we work in an idealized financial market, disregarding for the sake of tractability all kinds of imperfections such as transaction costs or liquidity issues, although these are of course in reality very important aspects of trading (especially since the crisis). See [30, 80] regarding extension of the “classical” approach of this book to a multiple funding curve environment.

Regarding the model filtration \mathbb{F} , we assume for simplicity that \mathcal{F}_0 is trivial, which means that all processes have constant initial conditions at time 0. By default in this book, all processes are \mathbb{F} -adapted, which means that the value of a process at any time $t \in [0, T]$ is revealed by the information contained in \mathcal{F}_t . We also assume the so-called usual conditions, i.e. we enrich our filtration in the canonical way to make it complete and right-continuous. This implies in particular that all processes belonging to the class of so-called semimartingales may be taken in a càdlàg version. As reviewed in Sect. 3.1.5, semimartingales are a class of processes giving rise to the most flexible theory of stochastic integration and are typically used for modeling price processes. In fact, it can be shown that price processes outside this class give rise to arbitrages. We recall that “càdlàg” is a French acronym for “(almost surely) right continuous with left limits”; we saw in Sect. 3.1.5 that, in a filtration satisfying the usual conditions, every semimartingale admits a càdlàg version.

4.1.1 Pricing by Arbitrage

In this subsection we revisit in more detail the above statements concerning the equivalence between absence of arbitrage and the existence of a risk-neutral probability measure. For this we first need to introduce a primary market model and then define a financial derivative to be priced and hedged in this market model.

4.1.1.1 Primary Market Model

We consider a primary market model composed of a locally risk-free asset B , called the savings account, and of q primary risky assets. A risk-free asset (we henceforth

omit “locally” for notational simplicity) has a price process without martingale part (price process given as a finite variation and predictable special semimartingale; see Sect. 3.1.5). The savings account essentially represents an asset that is assumed to exist in the market,² whose value evolves at the short-term riskless interest rate r_t , which is the rate of a riskless loan between t and $t + dt$. The riskless discount factor β is then defined as the inverse of the savings account. Letting conventionally $\beta_0 = B_0 = 1$, we thus have, for $t \in [0, T]$,

$$\beta_t = B_t^{-1} = \exp\left(-\int_0^t r_s ds\right) \quad (4.1)$$

for a short-term interest rate process r that is bounded from below.³

We denote by P and \mathcal{D} the \mathbb{R}^q -valued price process and cumulative dividend value process of the primary risky assets. The price process P may be understood, in a sense to be made more precise shortly, as the present value of future cash flows. The dividend process \mathcal{D} accumulates all the financial cash flows that are granted to the holders of the risky assets during their lifetime. Dividends in this broad sense thus encompass actual stock dividends in the case of risky assets given as stocks, coupons in the case of bonds, or recovery of assets upon default of their issuing firm, to state only the main sources of “dividend” incomes.

Given the price process P and the dividend process \mathcal{D} , it is useful mathematically to further introduce the so-called cumulative price \widehat{P} of the primary risky assets, defined as

$$\widehat{P}_t = P_t + \beta_t^{-1} \int_{[0,t]} \beta_s d\mathcal{D}_s. \quad (4.2)$$

In the financial interpretation, the last term in (4.2) represents the current value at time t of all dividend payments of the asset over the period $[0, t]$, under the assumption that all dividends are immediately reinvested in the savings account.

Also, since the vocation of the primary market is to serve as a pool of hedging instruments for a financial derivative with maturity T , we assume that the primary assets live over $[0, T]$ (or beyond, in which case we simply “ignore” what happens to them beyond T). In particular, we assume that the primary market is “European” in the sense that it doesn’t contain assets with early exercise clauses.

We assume that the cumulative price process \widehat{P} is a locally bounded semimartingale, where the local boundedness property means the existence of a sequence of stopping times such that the process, stopped at any of these times, is uniformly bounded. Restricting attention to locally bounded semimartingales, whose choice of semimartingales has already been discussed, allows us to avoid the mathematically delicate concept of sigma martingales which would arise by trading in the primary

²This classical assumption is, of course, seriously challenged and, in fact, unrealistic since the crisis; see Remark 4.1.1.

³Typically nonnegative, but not necessarily so, as seen with the Swiss currency during the 2011 Euro sovereign debt crisis.

market were arbitrary semimartingales to be used in the modeling; see Delbaen and Schachermayer [93] (in book form, [94]) or Protter [228].

We assume that the primary market model is free of arbitrage opportunities, in the sense that the so-called “No Free Lunch with Vanishing Risk” (NFLVR) condition is satisfied.

In a broad sense, absence of arbitrage opportunities refers to the impossibility of making a profit with positive probability without risking a loss. In other words, the only nonnegative wealth process of a self-financing trading strategy, starting from 0, is the null process. Note that the mere existence of a financial market can be opposed to the assumption of absence of arbitrage opportunities. But the idea underlying the no arbitrage assumption is that, every time arbitrage opportunities arise in the market, they are “exercised” and resorbed in quick time by supply and demand, so that it is a reasonable mathematical approximation to consider that there are no such opportunities. This assumption underlies much of the financial mathematics, including early work in discrete time by Harrison and Pliska [141] and culminating in the Fundamental Theorem of Asset Pricing established in its more general continuous-time form by Delbaen and Schachermayer [93, 94].

The NFLVR condition is a specific no arbitrage condition involving wealth processes of admissible self-financing primary trading strategies. We will not reproduce here the full statement of this condition from [94], since it is rather technical and we will not use it explicitly in the sequel. It will be enough for us to recall the related notion of a trading strategy in the primary market.

Definition 4.1.2 A primary trading strategy in the primary market is an $\mathbb{R} \times \mathbb{R}^q$ -valued process (ζ^0, ζ) , where ζ^0 and the row-vector ζ respectively represent the number of units held in the savings account and in the primary risky assets, with ζ predictable and locally bounded. The related wealth process \mathcal{W} is given, for $t \in [0, T]$, by

$$\mathcal{W}_t = \zeta_t^0 B_t + \zeta_t P_t. \quad (4.3)$$

Accounting for dividends, the strategy is said to be self-financing if

$$d\mathcal{W}_t = \zeta_t^0 dB_t + \zeta_t (dP_t + dD_t) \quad (4.4)$$

or, equivalently,

$$d(\beta_t \mathcal{W}_t) = \zeta_t d(\beta_t \widehat{P}_t). \quad (4.5)$$

If, moreover, the discounted wealth process $\beta \mathcal{W}$ is bounded from below, the strategy is said to be admissible.

The equivalence between the intrinsic form (4.4) and the discounted form (4.5) of the self-financing condition is easily seen in the present context, where β is given

by (4.1). Assuming (4.4), we thus have that

$$\begin{aligned} d(\beta_t \mathcal{W}_t) &= \beta_t (d\mathcal{W}_t - r_t \mathcal{W}_t dt) \\ &= \beta_t (\zeta_t^0 dB_t + \zeta_t (dP_t + dD_t) - r_t (\zeta_t^0 B_t + \zeta_t P_t) dt) \\ &= \beta_t (\zeta_t (dP_t + dD_t) - r_t \zeta_t P_t dt) = \zeta_t (d(\beta_t P_t) + \beta_t dD_t) = \zeta_t d(\beta_t \widehat{P}_t), \end{aligned}$$

which is (4.5), and conversely.

The predictability condition on ζ in Definition 4.1.2 is a measurability assumption that is satisfied, for instance, by every deterministic or left-continuous (adapted) process ζ . Intuitively this means that the current value of ζ can be “predicted” if we know its past values. In discrete time, predictability reduces to adaptedness with respect to the “lagged” filtration $(\mathcal{F}_{i-1})_{i \geq 0}$. From a financial point of view, assuming predictability of a trading strategy is justified by the fact that effectively “implementing” a hedge always requires some small amount of time. Also note, from the mathematical point of view, that the class of local martingales is stable with respect to the stochastic integration of a predictable and locally bounded integrand ζ , as is for instance any ζ of the form ζ_- for some semimartingale $\tilde{\zeta}$ (see Proposition 3.1.13 and the associated comments).

Given the initial wealth w of a self-financing primary trading strategy (ζ^0, ζ) , the wealth process \mathcal{W}_t can also be written as

$$\beta_t \mathcal{W}_t = w + \int_0^t \zeta_s d(\beta_s \widehat{P}_s). \quad (4.6)$$

The process ζ^0 that gives the number of units held in the savings account is then uniquely determined as

$$\zeta_t^0 = \beta_t (\mathcal{W}_t - \zeta_t P_t).$$

In the sequel we restrict ourselves to self-financing trading strategies. We thus can redefine a (self-financing) primary trading strategy as a pair (w, ζ) , formed of an initial wealth $w \in \mathbb{R}$, and an \mathbb{R}^q -valued predictable locally bounded primary strategy in the risky assets ζ , with related wealth process \mathcal{W} defined by (4.6).

Remark 4.1.3 This redefinition is a convenient trick that allows us to “forget” about the risk-free funding asset and the funding issues, which become “absorbed” in the discounting at the risk-free rate. To deal with multiple-curve markets where there is no such thing as a risk-free funding asset, funding needs to be considered explicitly (see [30, 80]).

Again, the NFLVR condition is a specific “no arbitrage” condition involving wealth processes of admissible primary trading strategies. Under our assumptions, the Fundamental Theorem of Asset Pricing of Delbaen and Schachermayer [93, 94] can be summarized as follows:

Proposition 4.1.4 *The NFLVR condition is equivalent to the existence of a risk-neutral measure on the primary market, i.e. $\mathcal{M} \neq \emptyset$, where \mathcal{M} denotes the set of probability measures $\mathbb{P} \sim \widehat{\mathbb{P}}$ such that $\beta \widehat{P}$ is a \mathbb{P} -local martingale.*

The proof of the general result is very technical. The interested reader can start with the paper of Harrison and Pliska [141] for a simple result and proof in the case of a finite state market; for more general results see also Cherny and Shiryaev [66].

Since the NFLVR condition is assumed throughout this book, Proposition 4.1.4 ensures existence of a risk-neutral measure $\mathbb{P} \in \mathcal{M}$, so that $\mathcal{M} \neq \emptyset$. A companion result to Proposition 4.1.4, sometimes referred to as the second Fundamental Theorem of Asset Pricing (Proposition 4.1.4 being then more specifically designated as the first Fundamental Theorem of Asset Pricing [66]), is the characterization of uniqueness of a risk-neutral measure in Proposition 4.1.4, i.e. the characterization of the non arbitrable price processes P for which the set \mathcal{M} is reduced to a singleton. Loosely speaking, the second Fundamental Theorem of Asset Pricing states that uniqueness for a risk-neutral measure over the primary market is equivalent to the completeness of the market, which means that every European derivative in the market is replicable, i.e. perfectly hedgeable.

4.1.1.2 Contingent Claims

We now introduce a contingent claim, or financial derivative, on the primary market. A derivative is a financial claim between an investor or holder of a claim and its counterparty or issuer that involve, as made precise in Definition 4.1.7 below, some or all of the following cash flows (or payoffs):

- a bounded variation cumulative dividend process $D = (D_t)_{t \in [0, T]}$;
- terminal cash flows, consisting of:
 - a payment ξ at maturity T , where ξ denotes a real-valued random variable that is bounded from below,
 - in the case of American or game products with early exercise features, put and/or call payoff processes $L = (L_t)_{t \in [0, T]}$ and $U = (U_t)_{t \in [0, T]}$, given as real-valued càdlàg processes that are bounded from below and such that $L \leq U$ and $L_T \leq \xi \leq U_T$.

The put payoff L_t corresponds to a payment made by the issuer to the holder of the claim, in case the holder of the claim decides to terminate (or “put”) the contract at time t . The call payment U_t corresponds to a payment made by the issuer to the holder of the claim, in case the issuer of the claim decides to terminate (or “call”) the contract at time t . Moreover, there may be call protection, modeled in the form of a stopping time ϑ such that issuer calls are not allowed before time ϑ .

The terminology “derivative” comes from the fact that all the above cash flows are typically given as functions of the underlying primary asset price process P . More generally (this will be our working assumption in Sect. 4.2), the cash flows of a derivative (and then, in a Markov setup, its price Π) and the prices P of the primary

assets are all given as functions of a common set of explanatory factors X . One may then consider the issue of factor hedging the derivative with price process Π by the primary assets with price process P , building upon the common dependence of Π and P on X .

Example 4.1.5

- (i) In the simplest case of a European vanilla call/put option with strike K on $S = P^1$, the first primary risky asset, one has $D = 0$ and $\xi = (S_T - K)^{\pm}$.
- (ii) In the case of an American vanilla call/put option on an underlying stock S , one has $L_t = (S_t - K)^{\pm}$, $\xi = (S_T - K)^{\pm}$ and no early call is possible, so that $\vartheta \equiv T$.
- (iii) In the case of a vanilla⁴ convertible bond on an underlying stock S , the dividend process D consists of a cumulative flow of bond coupons, and we have

$$\xi = \bar{N} \vee S_T, \quad L_t = \bar{P} \vee S_t, \quad U_t = \bar{C} \vee S_t,$$

for nonnegative constants $\bar{P} \leq \bar{N} \leq \bar{C}$.

- (iv) In the case of a credit default swap (CDS for short) on an underlying credit name, there are no terminal cash flows in the above sense. There is only a cumulative dividend process D . From the point of view of the seller of default protection and for a notional conventionally set to one, this process is given by:

$$D_t = \int_0^t (\Lambda dJ_s + SJ_s ds),$$

where Λ and J are the fractional loss-given-default and the survival indicator process of the underlying credit name and where S is the contractual spread of the CDS.

- (v) In the case of a CDO tranche with attachment point a and detachment point b on an underlying portfolio of credit names, there are no terminal cash flows. From the point of view of the seller of default protection and for a notional conventionally set to one, the dividend process D is given by:

$$D_t = \int_0^t (\Sigma(b - a - L_s) ds - dL_s),$$

where Σ is the tranche contractual spread and the cumulative tranche loss L_t is given by

$$L_t = (\mathcal{L}_t - a)^+ - (\mathcal{L}_t - b)^+ = \min((\mathcal{L}_t - a)^+, b - a),$$

in which \mathcal{L}_t is the aggregated default loss on the portfolio at time t .

⁴Default-free and without call protection and so, in particular, $\vartheta \equiv 0$.

Remark 4.1.6 The European and American vanilla calls and puts of items (i) and (ii) will be used as standing examples throughout the book. The convertible bond of item (iii) and the credit derivatives of item (v) will be the topics of two applications in Chaps. 10 and 11.

With the credit derivatives of items (iv) and (v), the underlying consists of the default loss on reference credit names: a single credit name in the case of a CDS and a pool of credit names in the case of a CDO tranche. For a CDO tranche the underlying is not a traded asset (whereas a defaultable bond issued by the reference name can be viewed as the underlying to a CDS). However, factor hedging of the tranche by, say, its credit index is, to some extent, possible, given a common factor process X driving the tranche and the index cash flows. Note that, even if there are almost no new CDO issuances since the crisis, CDOs are still relevant in terms of risk-management as they are still present in banks' portfolios. Counterparty risk on CDOs is an important issue. Moreover, counterparty risk on portfolios of CDS gives rise to certain kinds of "giant CDOs".

We will now formally define European claims, American claims and game claims. It will soon become apparent that European claims can be considered as special cases of American claims, which are themselves included in game claims, so that we will eventually be able to reduce attention to the latter. In the following definitions, the put (or maturity) time τ and the call (or maturity) time θ represent stopping times at, respectively, the holder's and the issuer's convenience, with $\theta \geq \vartheta$ in the case of a call protection before a stopping time ϑ .

Definition 4.1.7

- (i) A European claim is a financial claim with dividend process D and with payment ξ at maturity T .
- (ii) An American claim is a financial claim with dividend process D and with payment at the terminal (put or maturity) time τ given by

$$\mathbb{1}_{\{\tau < T\}} L_\tau + \mathbb{1}_{\{\tau = T\}} \xi. \quad (4.7)$$

- (iii) A game claim is a financial claim with dividend process D , and with payment at the terminal (call, put or maturity) time $\nu = \tau \wedge \theta$ given by⁵

$$\mathbb{1}_{\{\nu = \tau < T\}} L_\tau + \mathbb{1}_{\{\theta < \tau\}} U_\theta + \mathbb{1}_{\{\nu = T\}} \xi. \quad (4.8)$$

These definitions encompass the great majority of financial derivatives encountered in practice, including not just standard options, but also convertible bonds or cancelable swaps and vanilla credit derivatives (see Example 4.1.5), to mention only a few. All cash flows in this book are seen from the point of view of the holder of the claim. The assumption that cash flows are bounded from below is satisfied by a

⁵With priority of a put over a call here, though this is immaterial in terms of pricing and hedging.

vast majority of real-life financial derivatives. From the mathematical point of view, this assumption on cash flows ensures their integrability on $\mathbb{R} \cup \{+\infty\}$.

Call protections are typically present in real-life callable products like convertible bonds (see the paragraph “Convertible Bonds” on page 113), to the effect of making those cheaper for their holder. The introduction of call protections in our definitions also allows us to consider an American claim as a game claim with call protection $\vartheta = T$. In Chap. 10 (see also Sect. 4.3.3), we will consider products with a more general intermittent form of call protection, namely call protection whenever a certain condition is satisfied, rather than call protection before a stopping time.

Bermudan products corresponding to constrained put policies might likewise be introduced. However, Bermudan products can be included in the above setup by considering a suitably adjusted put payoff process L . This will indeed be a consequence of Proposition 4.1.8(ii) below, in conjunction with our boundedness from below assumption on cash flows.

In the sequel, the statement “ Π is an arbitrage price for a derivative” is to be understood as: “(P, Π) is an arbitrage price for the augmented market consisting of the primary market and the derivative”. With regard to Remark 4.1.1, this implies that we consider a liquidly traded derivative, so that the derivative and the primary market can be viewed as an augmented perfect market.

The notion of arbitrage price process of a financial derivative, mentioned in the next result, is then nothing but the classical NFLVR notion already used above in the case of European claims (cf. Proposition 4.1.4), extended to American and game claims by Kallsen and Kühn [161].

Let \mathcal{T}_t and \mathcal{T}_t^ϑ , or simply \mathcal{T} and \mathcal{T}^ϑ in the case $t = 0$, denote the set of all $[t, T]$ -valued and $[t \vee \vartheta, T]$ -valued stopping times; let ν represent $\tau \wedge \theta$, for every $(\tau, \theta) \in \mathcal{T}_t \times \mathcal{T}_t^\vartheta$. As in Proposition 4.1.4, \mathcal{M} denotes the set of the risk-neutral probability measures over the primary market, with $\mathcal{M} \neq \emptyset$ under our standing non arbitrage assumption. The proof of the following result is based on a rather straightforward application of Theorem 2.9 in Kallsen and Kühn [161]; see Bielecki et al. [39] for the details.

Proposition 4.1.8 *For any $\mathbb{P} \in \mathcal{M}$, with \mathbb{P} -expectation denoted by \mathbb{E} :*

- (i) *The process Π defined, for $t \in [0, T]$, by:*

$$\beta_t \Pi_t = \mathbb{E} \left\{ \int_t^T \beta_s dD_s + \beta_T \xi \mid \mathcal{F}_t \right\}, \quad (4.9)$$

is an arbitrage price of the related European claim.

- (ii) *The process Π defined, for $t \in [0, T]$, by:*

$$\beta_t \Pi_t = \operatorname{esssup}_{\tau \in \mathcal{T}_t} \mathbb{E} \left\{ \int_t^\tau \beta_s dD_s + \beta_\tau (\mathbb{1}_{\{\tau < T\}} L_\tau + \mathbb{1}_{\{\tau = T\}} \xi) \mid \mathcal{F}_t \right\}, \quad (4.10)$$

is an arbitrage price of the related American claim, provided it is a semimartingale.

(iii) The process Π defined, for $t \in [0, T]$, by:

$$\begin{aligned} & \text{esssup}_{\tau \in \mathcal{T}_t} \text{essinf}_{\theta \in \mathcal{T}_t^\vartheta} \mathbb{E} \left\{ \int_t^\nu \beta_s dD_s + \beta_\nu (\mathbb{1}_{\{\nu=\tau < T\}} L_\tau + \mathbb{1}_{\{\theta < \tau\}} U_\theta + \mathbb{1}_{\{\nu=T\}} \xi) \mid \mathcal{F}_t \right\} \\ &= \beta_t \Pi_t = \text{essinf}_{\theta \in \mathcal{T}_t^\vartheta} \text{esssup}_{\tau \in \mathcal{T}_t} \mathbb{E} \left\{ \int_t^\nu \beta_s dD_s + \beta_\nu (\mathbb{1}_{\{\nu=\tau < T\}} L_\tau + \mathbb{1}_{\{\theta < \tau\}} U_\theta \right. \\ &\quad \left. + \mathbb{1}_{\{\nu=T\}} \xi) \mid \mathcal{F}_t \right\}, \end{aligned} \quad (4.11)$$

with $\nu = \tau \wedge \theta$, is an arbitrage price of the related game claim, provided (4.11) yields a well-defined semimartingale.

By virtue of our lower boundedness assumption on the cash flows, all the conditional expectations that appear in these statements are well-defined $\mathbb{R} \cup \{+\infty\}$ -valued random variables. Equation (4.9) follows directly by an application of Proposition 4.1.4 to the augmented market. The ‘‘esssup’’ and ‘‘essinf’’ which appear in (4.10) and (4.11) are necessary for $t > 0$ to give meaning to a supremum or an infimum of an uncountable set of random variables, i.e. the set of conditional expectations parameterized by the uncountable family of stopping times \mathcal{T}_t or \mathcal{T}_t^ϑ (see Remark 3.5.2). For (4.11) to define a semimartingale, a prerequisite is, of course, that equality holds between the left-hand side and the right-hand side in (4.11), whereas only ‘‘ \leq ’’ holds in general.

Arbitrage prices Π of the form (4.9), (4.10) or (4.11) are called \mathbb{P} -prices in the sequel. Given a \mathbb{P} -price Π , the discounted cumulative price of the option is defined, in analogy with the one of a primary asset in (4.2), by

$$\beta_t \widehat{\Pi}_t = \beta_t \Pi_t + \int_{[0,t]} \beta_s dD_s. \quad (4.12)$$

Under our assumptions, the discounted cumulative price process of a European option thus defined, with Π in (4.12) given by (4.9), is an $\mathbb{R} \cup \{+\infty\}$ -valued \mathbb{P} -martingale. In the American case, the discounted cumulative price, defined by (4.12) with Π from (4.10), is the Snell envelope of the payoff process defined by

$$\int_{[0,t]} \beta_s dD_s + \beta_t (\mathbb{1}_{\{t < T\}} L_t + \mathbb{1}_{\{t=T\}} \xi). \quad (4.13)$$

The Snell envelope of a given process \mathcal{L} is the smallest supermartingale $\geq \mathcal{L}$. In the game case, the discounted cumulative price (4.12), with Π given as (4.11),⁶ is the value process of the Dynkin game with lower and upper payoff processes respectively defined by the integral in (4.13) and the same integral with L replaced by U .

⁶Equation (4.11) being here assumed to yield a well-defined semimartingale.

One can thus interpret a European claim as an American claim with a fictitious put payment process L defined by $\beta L = -c$, where $-c$ is a minorant of $\int_0^T \beta_s dD_s + \beta_T \xi$. Indeed, we see from Proposition 4.1.8(ii) that exercise of the put before maturity for this specification of L is always sub-optimal for the holder of the claim. It is thus equivalent for a process Π to be an arbitrage price of the European claim with the cash flows D and ξ , or to be an arbitrage price of the American claim with the cash flows D , L and ξ , with L thus specified. Since American options are more complex to handle mathematically than European ones, interpreting a European option as an American option may seem rather convoluted. But this is in fact a simple way to establish the admissibility of tentative hedging strategies for a European option (see the comments after Definition 4.1.9).

By “financial derivative”, or “option”, we thus mean henceforth by default “game claim”, possibly with a call protection ϑ , including American claim (case $\vartheta \equiv T$; in particular European claim with L as specified above) as a special case.

4.1.2 Hedging

In this book we adopt a definition of hedging of a game option stemming from successive developments, starting from the hedging of American options examined by Karatzas [163], and subsequently followed by El Karoui and Quenez [115], Kifer [166], Ma and Cvitanic [196] and Hamadene [138].

Following the lines of Föllmer and Sondermann [123] (see also Schweizer [243, 244]), we first introduce a very large class, to be specified later, of hedges with semimartingale cost process ρ . In this approach the issuer of a financial derivative sets-up a hedge such that the corresponding wealth process \mathcal{W} reduces to a cost or hedging error ρ , after accounting for the “dividend cost” $(-D)$ and for the “terminal loss” given by $(-L)$, $(-U)$ or $(-\xi)$. The initial wealth w may then be used as a safe issuer price, up to the hedging error ρ , for the derivative at hand.

In the following definition we consider an issuer hedge starting at time 0. The adaptation of this definition to a holder hedge, and/or to a hedge starting at an arbitrary time $t \in [0, T]$, is straightforward. Also recall that we write $v = \tau \wedge \theta$.

Definition 4.1.9

- (i) A hedge with semimartingale cost process ρ for a game option is represented by a triple (w, ζ, θ) such that:
 - (w, ζ) is a (self-financing) primary trading strategy,
 - the call time θ belongs to \mathcal{T}^ϑ ,
 - for every put time τ in \mathcal{T} , the wealth process \mathcal{W} of (w, ζ) almost surely satisfies

$$\beta_v \mathcal{W}_v + \int_0^v \beta_s d\rho_s \geq \int_0^v \beta_s dD_s + \beta_v (\mathbb{1}_{\{v=\tau < T\}} L_\tau + \mathbb{1}_{\{\theta < \tau\}} U_\theta + \mathbb{1}_{\{v=T\}} \xi). \quad (4.14)$$

- (ii) In the special case of an American (in particular, European) option, so that $\vartheta \equiv T$, the set of admissible call times \mathcal{T}^ϑ is reduced to the constant time T . The previous definition thus reduces to a primary trading strategy (w, ζ) with related wealth process \mathcal{W} such that, for every put time τ in \mathcal{T} , almost surely,

$$\beta_\tau \mathcal{W}_\tau + \int_0^\tau \beta_s d\rho_s \geq \int_0^\tau \beta_s dD_s + \beta_\tau (\mathbb{1}_{\{\tau < T\}} L_\tau + \mathbb{1}_{\{\tau = T\}} \xi). \quad (4.15)$$

- (iii) In the special case of a European option, provided moreover that equality holds in (4.15) for $\tau \equiv T$, then almost surely,

$$\beta_T \mathcal{W}_T + \int_0^T \beta_s d\rho_s = \int_0^T \beta_s dD_s + \beta_T \xi. \quad (4.16)$$

In this case the strategy (w, ζ) is said to be a replicating strategy with cost ρ .

Process ρ is to be interpreted as the cumulative financing cost, i.e. the amount of cash added to (if $d\rho_t \geq 0$) or withdrawn from (if $d\rho_t \leq 0$) the hedging portfolio in order to get a perfect, but no longer self-financing, hedge. Hedges at no cost, i.e. with $\rho = 0$, are thus in effect superhedges. Hedges with a local martingale cost ρ under a particular risk-neutral measure \mathbb{P} can also be interpreted as mean-self-financing hedges in the sense of Föllmer and Sondermann [123] (see also Schweizer [243, 244]).

In relation to admissibility issues (see the end of Definition 4.1.2), note that the left-hand side of (4.14) (discounted wealth process with financing costs included) is bounded from below, for every hedge (w, ζ, θ) .

In the American case, càdlàg properties of the processes involved in (4.15) imply that satisfaction of the latter for every put time τ in \mathcal{T} is equivalent to: almost surely,

$$\beta_t \mathcal{W}_t + \int_0^t \beta_s d\rho_s \geq \int_0^t \beta_s dD_s + \beta_t (\mathbb{1}_{\{t < T\}} L_t + \mathbb{1}_{\{t = T\}} \xi), \quad t \in [0, T]. \quad (4.17)$$

In the special case of European options, requiring (4.15) for every put time τ in \mathcal{T} (or equivalently, (4.17)) may seem unjustified, since a European option can only be exercised at time $\tau = T$. Note however that in the European case, the put payoff L is defined in a very specific way (see the last paragraph of Sect. 4.1.1.2), so that the “undue” requirement that (4.17) also holds for $t < T$, and not only at $t = T$, is merely a way to encompass admissibility of the strategy (w, ζ) .

As Proposition 4.1.15 will reveal below, Definition 4.1.9 is consistent with the concept of arbitrage pricing of Proposition 4.1.8(iii) for game options, including American and European options as special cases. This general connection between pricing and hedging in the weak sense of Definition 4.1.9 will need to be completed by the analysis and discussion of various, more explicit, hedging strategies; this will be accomplished in Sect. 4.2.3.

To emphasize the very weak sense of Definition 4.1.9 and to warn the reader against any possible misunderstanding of the following results, note that “no hedge”

(thus $\zeta \equiv 0$, or in fact, more generally, any primary strategy (w, ζ)) is a hedge in the sense of this definition,⁷ with cost process essentially given by the tracking error process $e = e(\zeta)$ in the sense of the following:

Definition 4.1.10 Given a \mathbb{P} -price process Π and a primary strategy (w, ζ) , the corresponding profit-and-loss or tracking error process $(e_t)_{t \in [0, T]}$ of the issuer is given by

$$\beta_t e_t = \int_0^t (-d(\beta_s \widehat{\Pi}_s) + \zeta_s d(\beta_s \widehat{P}_s)), \quad (4.18)$$

where the cumulative prices \widehat{P} and $\widehat{\Pi}$ were defined in (4.2) and (4.12).

The tracking error corresponds to the notion of the cumulative gain of a trader hedging one short option position by the strategy (w, ζ) in the primary market. Note that the process $e = e(\zeta)$ only depends on ζ , not on w .

One theoretically pleasant property of hedges with semimartingale cost processes in Definition 4.1.9 is that they inherit from the class of semimartingales the property of invariance under equivalent changes of probability measures. However, the class of hedges with semimartingale cost processes is too large for any practical purpose, so we will now restrict our attention to hedges with a local martingale cost ρ under a particular risk-neutral measure \mathbb{P} . Henceforth in this chapter we work under a fixed risk-neutral measure \mathbb{P} , with \mathbb{P} -expectation (resp. conditional expectation) denoted by \mathbb{E} (resp. $\mathbb{E}_t \equiv \mathbb{E}(\cdot | \mathcal{F}_t)$). All the measure-dependent notions implicitly refer to the probability measure \mathbb{P} .

4.1.2.1 BSDE Modeling

We will now postulate suitable integrability and regularity conditions, embedded in the standing assumption that a related reflected backward stochastic differential equation (reflected BSDE) has a solution. BSDEs are useful to give a unified presentation of the pricing and hedging theory. Besides this pedagogical virtue, in Markovian setups BSDEs also offer a computational alternative to PDEs, which is especially important in the context of high-dimensional pricing problems for which deterministic pricing schemes are ruled out by the curse of dimensionality (see Part IV). We introduce our fundamental pricing reflected BSDE (4.19) under the risk-neutral probability measure \mathbb{P} , with the data defined in terms of those of a financial derivative. As discussed in detail in Part V, this equation is well posed under mild regularity and integrability conditions. At the level of this chapter, however, we will simply proceed under the working assumption that (4.19) has a solution. Based on this solution, we can derive explicit hedging strategies with minimal initial wealth for the related derivative. We will refer to (4.19) as the stochastic pricing

⁷For an appropriately chosen call time θ^* ; see Proposition 4.1.15 below for a precise statement.

equation of the claim under the prevailing market pricing measure \mathbb{P} , as opposed to the deterministic pricing equation (partial integro-differential equation) to be introduced in the Markovian setup of Sect. 4.2.

For the sake of notational simplicity we assume that the cumulative dividend process D of a financial derivative is time-differentiable, so $dD_t = C_t dt$ for some progressively measurable time-integrable coupon rate process C . Note for practical applications that it is also possible to deal with discrete dividends, as will be seen in Sect. 14.1. We consider the following reflected BSDE with the data $\beta, C, \xi, L, U, \vartheta$:

$$\begin{cases} \beta_t \Pi_t = \beta_T \xi + \int_t^T \beta_s C_s ds + \int_t^T \beta_s (dA_s - dM_s), & t \in [0, T] \\ L_t \leq \Pi_t \leq \overline{U}_t, & t \in [0, T] \\ \int_0^T (\Pi_s - L_s) dA_s^+ = \int_0^T (\overline{U}_s - \Pi_s) dA_s^- = 0 \end{cases} \quad (4.19)$$

where, with the convention that $0 \times \infty = 0$ in the third line,

$$\overline{U}_t = \mathbb{1}_{\{t < \vartheta\}} \infty + \mathbb{1}_{\{t \geq \vartheta\}} U_t. \quad (4.20)$$

Since $\beta_t = e^{-\int_0^t r_s ds}$, the first line of (4.19) is equivalent to

$$\Pi_t = \xi + \int_t^T (C_s - r_s \Pi_s) ds + (A_T - A_t) - (M_T - M_t), \quad t \in [0, T]. \quad (4.21)$$

In the sequel we use the following practical definition of a solution to (4.19), referring the reader to Part V for more formal definitions, including in particular the specification of the spaces for the inputs and outputs.

Definition 4.1.11 By a solution to (4.19) we mean a triple of real processes (Π, M, A) such that all conditions in (4.19) are satisfied, where:

- the value process Π is a càdlàg process,
- M is a martingale vanishing at time 0,
- A is a finite variation⁸ continuous process and A^\pm denote the Jordan components of A .

In the last point:

Definition 4.1.12 By the Jordan components of a finite variation⁹ process A we mean the terms of the unique decomposition $A = A^+ - A^-$ of A as the difference between two nondecreasing processes A^\pm , null at 0, defining mutually singular random measures on $[0, T]$.

⁸Starting from 0 by definition; see Sect. 3.1.5.

⁹Not necessarily continuous.

Thus $A_0^\pm = 0$, $dA^\pm \geq 0$ and “ $dA^+ dA^- = 0$ ”. Loosely speaking, the nondecreasing processes A^+ and A^- “cannot increase simultaneously”.

Since it involves the martingale M , the notion of a solution to (4.19) is, of course, measure-dependent, contingent on the prevailing market pricing measure \mathbb{P} which underlies its definition. The first line of (4.19) can be interpreted as giving the Doob-Meyer decomposition $\int_0^t \beta_s (dA_s - dM_s)$ of the special semimartingale

$$\beta_t \widehat{\Pi}_t := \beta_t \Pi_t + \int_0^t \beta_s C_s ds. \quad (4.22)$$

An equivalent definition of a solution to (4.19) would thus be that of a special semimartingale Π (rather than a triple of processes (Π, M, A)), such that all conditions in (4.19) are satisfied, where M and A there are to be understood as the canonical local martingale and the finite variation predictable components of the process $\int_{[0,\cdot]} \beta_t^{-1} d(\beta_t \widehat{\Pi}_t)$ (see Sect. 3.1.5). The “price” notation Π in the solution to (4.19) and the “cumulative price” notation $\widehat{\Pi}$ in (4.22) will be justified in Proposition 4.1.14.

As will be reviewed in Chap. 12, under square integrability conditions on the data, existence and uniqueness for a square integrable solution to (4.19) are essentially equivalent to the so-called Mokobodski condition, i.e. the existence of a square integrable quasimartingale¹⁰ Y such that $L \leq Y \leq \overline{U}$ on $[0, T]$. Existence and uniqueness for a solution to (4.19) thus hold when one of the barriers is a square integrable quasimartingale and, in particular, when the lower barrier is given as $(S - c)^+$ or $S \vee c$, where S is a square integrable Itô process and c is a constant in $\mathbb{R} \cup \{-\infty\}$. Moreover, one typically has $A = 0$ in the case of a European derivative. For an American or game claim, we will see below that the reflecting process A is related to the cost of the hedge for the issuer or the holder of the claim who would not behave optimally in the related optimal stopping problem. We work henceforth in this chapter under the following

Assumption 4.1.13 The stochastic pricing equation (4.19) admits a solution (Π, M, A) , with A equal to zero in the special case of a European derivative.

The following result states that the value process Π in the solution to the stochastic pricing equation (4.19) yields an arbitrage price of the related derivative, the price corresponding to the pricing measure \mathbb{P} .

Proposition 4.1.14 *The value process Π is the \mathbb{P} -price process of the derivative.*

Proof If (Π, M, A) is a solution to (4.19), then it is apparent from (4.21) that Π is a special semimartingale. Moreover, by virtue of a classical verification principle that will be established as Proposition 12.1.8, Π satisfies (4.11), which in the special cases of American (respectively European) options reduces to (4.10) (respectively (4.9)). We obtain the conclusion by an application of Proposition 4.1.8. \square

¹⁰Special semimartingale with additional integrability properties; see Sect. 12.1.2.2.

We are now ready to interpret the \mathbb{P} -price Π thus defined via (4.19) in terms of the notion of hedging introduced in Definition 4.1.9. Observe that, in view of (4.22), the tracking error $e = e(\zeta)$ of (4.18) is now a special semimartingale. Let the \mathbb{P} -local martingale $\rho = \rho(\zeta)$ be such that $\rho_0 = 0$ and $\int_0^\cdot \beta_t d\rho_t$ is the local martingale component of the special semimartingale (βe) , so that in view of (4.18) and (4.21):

$$\begin{aligned}\beta_t d\rho_t &= \beta_t dM_t - \zeta_t d(\beta_t \widehat{P}_t) \\ d(\beta_t e_t) &= \beta_t dA_t - \beta_t d\rho_t.\end{aligned}\tag{4.23}$$

Also, let the call time $\theta^* \in \mathcal{T}_t^\vartheta$ be defined by

$$\theta^* = \inf\{s \in [t \vee \vartheta, T]; \Pi_s \geq U_s\} \wedge T.\tag{4.24}$$

Using the minimality condition in the third line of (4.19) and the continuity of A^\pm , we have

$$A^- = 0 \quad \text{and} \quad A = A^+ \geq 0 \text{ on } [0, \theta^*], \quad \Pi_{\theta^*} = U_{\theta^*} \text{ on } \{\theta^* < T\}.\tag{4.25}$$

Proposition 4.1.14 below says that any primary strategy (w, ζ) is, along with the call time θ^* , an issuer hedge in the sense of Definition 4.1.9, with cost process essentially given by the tracking error process $e = e(\zeta)$ of Definition 4.1.10. As visible in the second line of (4.23), “essentially” here means up to the reflecting process A , which for $\theta = \theta^*$ reduces to A^+ and quantifies the departure of the holder from optimality in the related optimal stopping game.

The arguments underlying the following results are already present as early as, for instance, Lepeltier and Maingueneau [186]. In the specific contexts of the Cox–Ross–Rubinstein or Black–Scholes models, analogous results can also be found in Kifer [166].

Proposition 4.1.15

- (i) For any primary strategy ζ , (Π_0, ζ, θ^*) is a hedge with \mathbb{P} -local martingale cost $\rho(\zeta)$;
- (ii) Π_0 is the minimal initial wealth of a hedge with \mathbb{P} -local martingale cost.
- (iii) In the special case of a European derivative with $A = 0$, we have that (Π_0, ζ) is a replicating strategy with \mathbb{P} -local martingale cost ρ ; Π_0 is thus also the minimal initial wealth of a replicating strategy with \mathbb{P} -local martingale cost.

Proof (i) We must show that for every $\tau \in \mathcal{T}$, almost surely:

$$\begin{aligned}\Pi_0 + \int_0^{\tau \wedge \theta^*} \zeta_s d(\beta_s \widehat{P}_s) + \int_0^{\tau \wedge \theta^*} \beta_s d\rho_s \\ \geq \int_0^{\tau \wedge \theta^*} \beta_s C_s ds + \beta_{\theta^* \wedge \tau} (\mathbb{1}_{\{\tau \wedge \theta^* = \tau < T\}} L_t + \mathbb{1}_{\{\theta^* < \tau\}} U_{\theta^*} + \mathbb{1}_{\{\tau = \theta^* = T\}} \xi).\end{aligned}\tag{4.26}$$

Using (4.23), this is equivalent to:

$$\begin{aligned} \Pi_0 + \int_0^{\tau \wedge \theta^*} \beta_s dM_s \\ \geq \int_0^{\tau \wedge \theta^*} \beta_s C_s ds + \beta_{\tau \wedge \theta^*} (\mathbb{1}_{\{\tau \wedge \theta^* = \tau < T\}} L_\tau + \mathbb{1}_{\{\theta^* < \tau\}} U_{\theta^*} + \mathbb{1}_{\{\theta^* = \tau = T\}} \xi), \end{aligned} \quad (4.27)$$

where, by the first line of (4.19):

$$\Pi_0 + \int_0^{\tau \wedge \theta^*} \beta_s dM_s = \int_0^{\tau \wedge \theta^*} \beta_s C_s ds + \beta_{\theta^* \wedge \tau} \Pi_{\tau \wedge \theta^*} + \int_0^{\tau \wedge \theta^*} \beta_s dA_s.$$

But (4.27) follows from (4.25) and from the following relations, which are valid by the terminal and put conditions in (4.19):

$$\Pi_T = \xi, \quad \Pi_\tau \geq L_\tau.$$

(ii) By part (i), $(\Pi_0, \zeta = 0)$ is a hedge with initial wealth Π_0 and \mathbb{P} -local martingale cost. Moreover, for every hedge (w, ζ, θ) with \mathbb{P} -local martingale cost Q , for every $t \in [0, T]$ we have by (4.14) that:

$$\begin{aligned} w + \int_0^{t \wedge \theta} \zeta_s d(\beta_s \widehat{P}_s) + \int_0^{t \wedge \theta} \beta_s d\rho_s \\ \geq \int_0^{t \wedge \theta} \beta_s C_s ds + \beta_{t \wedge \theta} (\mathbb{1}_{\{t \wedge \theta = t < T\}} L_t + \mathbb{1}_{\{\theta < t\}} U_\theta + \mathbb{1}_{\{t = \theta = T\}} \xi). \end{aligned} \quad (4.28)$$

The left-hand side is thus a local martingale that is bounded from below, hence it is a supermartingale. Moreover, (4.28) also holds with a stopping time $\tau \in \mathcal{T}$ instead of t there. Taking expectations then yields that

$$w \geq \mathbb{E} \left\{ \int_0^{\tau \wedge \theta} \beta_s C_s ds + \beta_{\theta \wedge \tau} (\mathbb{1}_{\{\tau \wedge \theta = \tau < T\}} L_\tau + \mathbb{1}_{\{\theta < \tau\}} U_\theta + \mathbb{1}_{\{\tau = \theta = T\}} \xi) \right\}.$$

Hence $w \geq \Pi_0$ follows by (4.11).

(iii) In the special case of a European derivative, the stated results follow by setting $A = 0$ in the previous points of the proof. \square

Proposition 4.1.15 characterizes the \mathbb{P} -price (arbitrage price corresponding to the risk-neutral measure \mathbb{P}) of a derivative as the smallest initial wealth of a hedge with \mathbb{P} -local martingale cost, under the assumption that the reflected pricing BSDE (4.19) has a solution. For related results, see also Föllmer and Sondermann [123] or Schweizer [243].

The special case $\rho = 0$, in the previous results, corresponds to a suitable form of model completeness (replicability of European options as viewed in part (iii) of the proposition), in which the issuer of the option can and wishes to hedge all the

embedded risks. The case $\rho \neq 0$ corresponds either to model incompleteness or to a situation of model completeness in which the issuer wishes not to hedge all the embedded risks, in order to take some bets in specific risk directions or to limit transaction costs.

It is easy to see that one could state analogous definitions and results regarding hedging a defaultable game option starting at any date $t \in [0, T]$, rather than at time 0 above. Also, analogous definitions and results hold for holder hedges.

4.2 Markovian Setup

In order to be usable in practice, a pricing model needs to be tractable numerically. This will now be achieved by assuming that we are dealing with a Markovian stochastic pricing equation (4.19), meaning that the input data of (4.19) are given as functions of an underlying Markovian factor process X_t . The logic here is that the Markovian property of the data of (4.19) implies a Markovian property of the solution, the price process Π being then also given as a function u of (t, X_t) . Moreover, the pricing function u can be characterized analytically as the unique solution, in some sense, of a related deterministic partial integro-differential equation (see Sect. 3.5 and Part V). This deterministic pricing equation can then be solved by various means, ranging from an analytic or semi-analytic transform (Fourier or such) method (Sect. 5.5) to various kinds of numerical approximation schemes (Part III). Thus,

Definition 4.2.1 We say that the pricing BSDE (4.19) is Markovian if

- the data r, C, ξ, L and U are given by Borel functions of some \mathbb{R}^d -valued (\mathbb{F}, \mathbb{P}) -Markov factor process X , i.e.

$$\begin{aligned} r_t &= r(t, X_t), & C_t &= C(t, X_t), & \xi &= \phi(X_T) \\ L_t &= L(t, X_t), & U_t &= U(t, X_t); \end{aligned} \tag{4.29}$$

- the call protection time ϑ is the first time of entry of (t, X_t) into a given closed subset of $[0, T] \times \mathbb{R}^d$ containing $\{T\} \times \mathbb{R}^d$ (hence $\vartheta \leq T$).

In particular, the system made by specifying a forward dynamic for X , together with the BSDE (4.19), constitutes a decoupled Markovian forward-backward equation¹¹ for (X, Π, M, A) . The system is decoupled in the sense that the forward component serves as an input for the backward component (X is an input to (4.19) via (4.29)), but not conversely.

From the point of view of financial interpretation, the components of X are observable factors. Most factors are typically given as primary price processes in P .

¹¹See Definition 12.2.4 for formal statements.

Additional factors may be required for explaining some path dependence in the payoffs of the derivative at hand, or in the dynamics of the underlying assets. Conversely, some of the primary price processes may not be needed as factors, but are used for hedging purposes. Note that observability of the factor process X , in the mathematical sense of \mathbb{F} -adaptedness, is not sufficient in practice. In order for a model to be calibratable,¹² we need a constructive mapping from a collection of meaningful and directly observable economic variables to X .

4.2.1 Factor Processes

As factor process X we will use the \mathbb{R}^d -valued jump-diffusion of Sect. 3.3.3, defined as a solution to the following Markovian forward SDE: an initial condition $X_0 = x$, given as an observable or calibratable constant and, for $t \in [0, T]$,

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t + \delta(t, X_{t-}, J_{(t)}) dN_t, \quad (4.30)$$

with Markovian intensity of jumps $\lambda(t, X_t)$ of the point process N and conditional jump distribution $w(t, X_{t-}, dx)$ of $J_{(t)}$ (see Remark 3.2.8). Process X is thus Markov with the following generator (cf. (3.66)):

$$\mathcal{A}u(t, x) = \partial u(t, x)b(t, x) + \frac{1}{2}\partial^2 u(t, x) : a(t, x) + \lambda(t, x)\bar{\delta}u(t, x), \quad (4.31)$$

in which we let, for every $t \geq 0$ and x, y in \mathbb{R}^d ,

$$\delta u(t, x, y) = u(t, x + \delta(t, x, y)) - u(t, x), \quad \bar{\delta}u(t, x) = \int_{\mathbb{R}^d} \delta u(t, x, y) w(t, x, dy).$$

In the case $\lambda = 0$, the jump part vanishes in X and we are left with a diffusion. In the case $b = \sigma = 0$, X is a pure jump process; under a more specific structure on δ , λ and w (see Sect. 12.2), the jump process X is supported by a finite set which can be identified with $E = \{1, \dots, n\}$ and X is a continuous-time E -valued Markov chain X . The generator \mathcal{A} of X is then given, for every time-differentiable function u over $[0, T] \times E$ (or, equivalently, any system $u = (u^i)_{1 \leq i \leq n}$ of time-differentiable functions u^i over $[0, T]$), by:

$$\mathcal{A}u^i(t) = \lambda^i(t) \sum_{j \neq i} w^{i,j}(t)(u^j(t) - u^i(t)). \quad (4.32)$$

Remark 4.2.2 (Unbounded jump measures) For technical simplicity we do not consider the “infinite activity” case of Lévy drivers with unbounded jump intensity measure (corresponding to the product $\lambda(t, x)w(t, x, dy)$ in our finite activity setup).

¹²At least in principle, leaving alone the computational issues; these will be dealt with in Chap. 9.

However, resorting to the random measure formalism of Part V (see Remark 3.2.11) and to the corresponding formulation (3.29) of X , it is possible to deal likewise with infinite activity jump measures.

With versatile specifications ranging from continuous-time Markov chains to diffusions and (modulo the above remark) Lévy-driven models, the jump-diffusion setup (4.30) offers a flexible framework which is rich enough for most pricing applications. This setup includes, in particular, the most common forms of stochastic volatility and/or jump equity derivatives pricing models, such as the Black–Scholes model, local volatility models, the Merton model, the Heston model or the Bates model. It also covers most of the continuous-time Markov chains, including those continuous-time Markov chains, modulated by diffusions, which can be used in dynamic credit portfolio modeling [31, 35, 36, 45, 61, 70, 128, 143, 178].

As will be explained in Sect. 4.3.1, the risk-neutral modeling approach can be readily extended to a martingale modeling approach with respect to an arbitrary numéraire, rather than the savings account in the risk-neutral approach. This allows one to apply the previous models to interest rate [58] and foreign exchange [191] derivatives.

Moreover, as we will see in Sect. 4.3.2, we can also accommodate, in the risk-neutral modeling approach, defaultable derivatives with terminal payoffs of the form $\mathbb{1}_{T < \tau_d} \phi(X_T)$,¹³ where τ_d represents the default-time of a reference entity. A process X of the form (4.30) is then typically used as a pre-default factor model. This allows one to deal with equity-to-credit derivatives like convertible bonds (see the paragraph “Convertible Bonds” on page 113 and Chap. 10), and with counterparty risk more generally [30, 80, 81, 216].

4.2.2 Markovian Reflected BSDEs and Obstacles PIDE Problems

We assume a jump-diffusion factor process X of the form (4.30) and a call protection before the stopping time

$$\vartheta = \inf\{t > 0; X_t \notin \mathcal{O}\} \wedge \bar{T}, \quad (4.33)$$

for a constant $\bar{T} \in [0, T]$ and a “regular”¹⁴ open subset $\mathcal{O} \subseteq \mathbb{R}^d$. By virtue of martingale and Markov properties, we then have, as will be formally established in Proposition 13.2.5 (see also Sect. 3.5):

Proposition 4.2.3

- (i) (*Post-Protection Price*). On $[\vartheta, T]$, the \mathbb{P} -price process Π can be represented as $\Pi_t = u(t, X_t)$, where u is the unique viscosity solution to the following double

¹³Or $\mathbb{1}_{v < \tau_d} \phi(X_v)$, upon exercise at a stopping time v , for American or game claims.

¹⁴See Example 12.4.6 for a precise statement.

obstacle problem:

$$\begin{aligned} & \min(\max(\partial_t u(t, x) + \mathcal{A}u(t, x) + C(t, x) - r(t, x)u(t, x), L(t, x) - u(t, x)), \\ & U(t, x) - u(t, x)) = 0, \quad t < T, \quad x \in \mathbb{R}^d, \end{aligned} \quad (4.34)$$

with terminal condition $u(T, x) = \phi(x)$.

- (ii) (*Protection price*). On $[0, \vartheta]$, the \mathbb{P} -price process Π can be represented as $\Pi_t = \bar{u}(t, X_t)$, where the function \bar{u} is the unique viscosity solution to the following lower obstacle problem:

$$\begin{aligned} & \max(\partial_t \bar{u}(t, x) + \mathcal{A}\bar{u}(t, x) + C(t, x) - r(t, x)\bar{u}(t, x), L(t, x) - \bar{u}(t, x)) = 0, \\ & t < \bar{T}, \quad x \in \mathcal{O}, \end{aligned} \quad (4.35)$$

with boundary condition $\bar{u} = u$ on $([0, T] \times \mathbb{R}^d) \setminus ([0, \bar{T}] \times \mathcal{O})$.

In view of Proposition 4.1.15(ii), $u(0, X_0) = \Pi_0$ is therefore the minimal initial wealth of a superhedge with \mathbb{P} -local martingale cost process for the option. Moreover, in case the pricing functions u and \bar{u} are sufficiently regular for an Itô formula to be applicable,¹⁵ the martingale component M of Π satisfies, for $t \in [0, T]$ and in the “.” notation introduced in Remark 3.2.11:

$$dM_t = \partial v \sigma(t, X_t) dW_t + \delta v(t, X_{t-}) \cdot d\mu_t, \quad (4.36)$$

where the random function v equals u for $t > \vartheta$ and \bar{u} for $t \leq \vartheta$. Equations (4.34)–(4.35) constitute a cascade of two PIDEs for the pair of pricing function (u, \bar{u}) . Note the “thick” parabolic boundary $([0, T] \times \mathbb{R}^d) \setminus ([0, \bar{T}] \times \mathcal{O})$, due to the jumps of X (see Sect. 8.3.3.1 and Chap. 13), in the equation for \bar{u} in Proposition 4.34(ii).

Example 4.2.4 In the case of a continuous-time Markov chain¹⁶ X with values in $E = \{1, \dots, n\}$ and for ϑ defined by (4.33), with \mathcal{O} there given as a subset of E , (4.34), (4.35) assume the form of the following cascade of two systems of ODEs for $(u, \bar{u}) = (u^i(t), \bar{u}^i(t))_{1 \leq i \leq n}$:

$$\left\{ \begin{array}{l} u^i(T) = \phi^i(T), \quad 1 \leq i \leq n \\ \min(\max(\partial_t u^i(t) + \mathcal{A}u^i(t) + C^i(t) - r^i(t)u^i(t), \\ \quad L^i(t) - u^i(t)), U^i(t) - u^i(t)) = 0, \quad t < T, \quad 1 \leq i \leq n \\ \bar{u} = u \quad \text{on } ([0, T] \times E) \setminus ([0, \bar{T}] \times \mathcal{O}) \\ \max(\partial_t \bar{u}^i(t) + \mathcal{A}\bar{u}^i(t) + C^i(t) - r^i(t)\bar{u}^i(t), L^i(t) - \bar{u}^i(t)) = 0, \quad t < \bar{T}, \quad i \in \mathcal{O}, \end{array} \right. \quad (4.37)$$

for \mathcal{A} given by (4.32).

¹⁵Otherwise a more general but less constructive representation for M_t can be given in terms of Malliavin calculus.

¹⁶See Sect. 2.2.

4.2.3 Hedging Schemes

In view of Proposition 4.2.3, the first line of the stochastic pricing equation (4.19) (or, equivalently, (4.21)) can be rewritten as

$$-dv(t, X_t) = (C - rv)(t, X_t) dt + dA_t - \partial v \sigma(t, X_t) dW_t - \delta v(t, X_{t-}) \cdot d\mu_t \quad (4.38)$$

for the function v defined as u for $t > \vartheta$ and \bar{u} for $t \leq \vartheta$. We assume a similar structure (but without barriers) on the primary market price process P , so that $P_t = v(t, X_t)$ for some function $v(t, x)$, and

$$-dv(t, X_t) = (\mathcal{C} - rv)(t, X_t) dt - \partial v \sigma(t, X_t) dW_t - \delta v(t, X_{t-}) \cdot d\mu_t, \quad (4.39)$$

where $\mathcal{C}(t, X_t)$ represents a primary market coupon rate process. Note that v is an \mathbb{R}^q -valued function and so, in particular, the Jacobian matrix ∂v lives in $\mathbb{R}^{q \times d}$ and (4.39) lives in \mathbb{R}^q .

Building upon the common dependence of Π and P on X , we may then consider the issue of factor hedging the derivative, with price process Π , by the primary assets with price process P . The cost ρ associated with the strategy ζ and, in turn, the related tracking error e in (4.23) can thus be expressed in terms of the pricing functions v and ν and the related delta functions, as follows:

Proposition 4.2.5 *In the Markovian setup, the dynamics (4.23) of the cost process $\rho = \rho(\zeta)$ may be rewritten as*

$$d\rho_t = (\partial v \sigma(t, X_t) - \zeta_t \partial v \sigma(t, X_t)) dW_t + (\delta v(t, X_{t-}) - \zeta_t \delta v(t, X_{t-})) \cdot d\mu_t. \quad (4.40)$$

Provided $\partial v \sigma$ is left-invertible, it is thus possible to perfectly hedge the source of risk W by setting

$$\zeta_t = \partial v \sigma(\partial v \sigma)^{-1}(t, X_t). \quad (4.41)$$

In the simplest case where $d = q$ and ∂v and σ are invertible, this formula reduces to

$$\zeta_t = \partial v(\partial v)^{-1}(t, X_t). \quad (4.42)$$

Substituting this strategy into (4.40), we are left with the cost process

$$\rho = \int_0^\cdot (\delta v(t, X_{t-}) - \zeta_t \delta v(t, X_{t-})) \cdot d\mu_t, \quad (4.43)$$

with ζ defined by (4.41) or (4.42). Note that this strategy, which is perfect from the point of view of hedging W , potentially creates some jump risk via the dependence on ζ of the integrand in (4.43).

At the other extreme, in case the jump measure of X has finite support (as in the continuous-time finite Markov chain specification of Example 4.2.4), it is alternatively possible to perfectly hedge the jump risk μ , provided $\delta v(t, X_{t-})$ is left-invertible, by setting

$$\zeta_t = \delta v(\delta v)^{-1}(t, X_{t-}). \quad (4.44)$$

Substituting this strategy into (4.40), we are left with the cost process

$$\rho = \int_0^{\cdot} (\partial v \sigma(t, X_t) - \zeta_t \partial v \sigma(t, X_t)) dW_t, \quad (4.45)$$

with ζ defined by (4.44). Note, however, that this strategy potentially creates some additional risk along W , via the dependence on ζ of the integrand in (4.45).

In the context of credit derivatives (see Sect. 4.3.2), hedging against W means hedging the spread risk, whereas hedging against μ means hedging default risk. We thus see that hedging spread risk without caring about default risk, which was the main trend in the practical risk management of credit derivatives until the credit crisis of 2007–2009 (in order to spare the high cost of hedging the jump-to-default risk), can lead to leveraged default risk.

4.2.3.1 Min-Variance Hedging

Again, a perfect hedge ($\rho = 0$) is hopeless unless the jump measure of X has finite support. In the context of incomplete markets, the choice of a hedging strategy is dependent on the optimality criterion with respect to the hedging cost ρ in (4.23), (4.40). For instance, one may wish to minimize the objective, \widehat{P} -variance of $\int_0^T \beta_t d\rho_t$. But the related strategy $\widehat{\zeta}^{va}$ is intractable. In particular $\widehat{\zeta}^{va}$ typically depends on the objective model drift, a quantity difficult to estimate statistically from financial time series.

As a proxy to this strategy, we can use the strategy ζ^{va} , which minimizes the risk-neutral variance of the error. Under mild conditions, $\int_0^{\cdot} \beta dM$ and $\beta \widehat{P}$ are square integrable martingales as martingale components of the value processes in the solutions to related square integrable BSDEs. The risk-neutral min-variance hedging strategy ζ^{va} is then given by the following Galtchouk-Kunita-Watanabe decomposition of $\int_0^{\cdot} \beta dM$ with respect to $\beta \widehat{P}$ (see e.g. Corollary 1, Chapter IV.3 of Protter [228]):

$$\beta_t dM_t = \zeta_t^{va} d(\beta_t \widehat{P}_t) + \beta_t d\rho_t^{va} \quad (4.46)$$

for some \mathbb{R}^q -valued $\beta \widehat{P}$ -integrable process ζ^{va} and a real square integrable martingale $\beta_t d\rho_t^{va}$ strongly orthogonal to $\beta \widehat{P}$.

For $\omega = v$ or v , let

$$(\omega, v)(t, x) = (\partial \omega \sigma)(\partial v \sigma)^T(t, x) + \lambda(t, x) \int_{\mathbb{R}^d} (\delta \omega(\delta v)^T)(t, x, y) w(t, x, dy). \quad (4.47)$$

This results in a q -dimensional row-vector function $(v, v)(t, x)$ and a q -dimensional matrix-function $(v, v)(t, x)$. Setting, in matrix form,

$$\langle A, B \rangle = (\langle A^i, B^j \rangle)_i^j, \quad \langle A \rangle = \langle A, A \rangle,$$

we have, in view of (4.46) and (3.68):

$$\zeta_t^{va} = \frac{d\langle \Pi, P \rangle_t}{dt} \left(\frac{d\langle P \rangle_t}{dt} \right)^{-1} = (v, v)(v, v)^{-1}(t, X_{t-}), \quad (4.48)$$

where invertibility of the conditional \mathbb{P} -covariance matrix $\frac{d\langle P \rangle_t}{dt}$ is assumed.

4.3 Extensions

4.3.1 More General Numéraires

Up to now, we implicitly chose the savings account β^{-1} , assumed to be a positive finite variation process, as a numéraire, i.e. a primary asset with positive price process which is used for discounting other price processes. However, for certain applications this choice may not be available (there may not be a riskless asset in the primary market), or it may not be the best possible one (even if there is a riskless asset, the choice of another asset as a numéraire may be more convenient, as when dealing with interest rate derivatives). This motivates the extension of the previous developments to the situation where the price process B of some reference asset is a locally bounded positive semimartingale, not necessarily of finite variation. We assume for simplicity a nondividend paying asset with martingale discounted price process.¹⁷ The interpretation of B as savings account and of $\beta = B^{-1}$ as a riskless discount factor is now replaced by the interpretation of B as a numéraire, referring to the fact that other price processes will be expressed as relative (rather than discounted) prices βP . Understanding a discounted price as a relative price and a risk-neutral model as a martingale model with respect to the numéraire B under the related valuation measure \mathbb{P} , the risk-neutral modeling approach developed in the previous sections still holds under this relaxed assumption on B . In particular, the self-financing property can still be stated in the form of (4.5) (see Protter [227]), although this is not as obvious as in the special case where B was a finite variation and continuous process. Also note that the set of admissible strategies is a numéraire-dependent notion. The concept of arbitrage is thus relative to the numéraire B . The meaning of “a martingale model with respect to the numéraire B under the related valuation measure \mathbb{P} ” is that relative cumulative prices $\beta \widehat{\Pi}$ must now be modeled as \mathbb{P} -local martingales.

¹⁷See Sect. 6.9.3 for an example of the dividend-adjustment which is otherwise required in the equations.

Let a correspondence between triples of processes (Π, M, A) and (π, m, a) be defined by $M_0 = m_0 = A_0 = a_0 = 0$ and, for $t \in [0, T]$,

$$\pi_t = \beta_t \Pi_t, \quad dm_t = \beta_t dM_t, \quad da_t = \beta_t dA_t \quad \text{with } m_0 = 0 \text{ and } a_0 = 0. \quad (4.49)$$

The pricing BSDE (4.19),¹⁸ to be solved for (Π, M, A) , is equivalent to the following \mathbb{P} -BSDE with the data $(c, \chi, \ell, \bar{h}) := (\beta C, \beta_T \xi, \beta L, \beta \bar{U})$, to be solved for (π, m, a) :

$$\begin{aligned} \pi_t &= \chi + c_T - c_t + a_T - a_t - (m_T - m_t), \quad t \in [0, T] \\ \ell_t &\leq \pi_t \leq \bar{h}_t, \quad t \in [0, T] \\ \int_0^T (\pi_s - \ell_s) da_s^+ &= \int_0^T (\bar{h}_s - \pi_s) da_s^- = 0, \end{aligned} \quad (4.50)$$

which is (4.19) with input data r, C, ξ, L, \bar{U} replaced by $0, c, \chi, \ell, \bar{h}$. The conclusions of Propositions 4.1.14–4.1.15 are still valid, provided “a solution (Π, M, A) to (4.19)” therein is understood as the process (Π, M, A) defined via (4.49) in terms of a solution (π, m, a) to (4.50). The Markovian case (4.29) now corresponds to the situation where

$$c_t = c(t, X_t), \quad \chi = \psi(X_T), \quad \ell_t = \ell(t, X_t), \quad h_t = h(t, X_t) \quad (4.51)$$

for a suitable \mathbb{R}^d -valued (\mathbb{F}, \mathbb{P}) -Markov factor process X . Let X be given in the form of a jump-diffusion (4.30) under \mathbb{P} , with generator \mathcal{A} given by (4.31), and let ϑ be given by (4.33). The deterministic pricing equation formally related to the BSDE (4.50) is then written as

$$\begin{cases} u(T, x) = \psi(x), \quad x \in \mathbb{R}^q \\ \min(\max(\partial_t u + \mathcal{A}u + c, \ell - u), h - u) = 0 \quad \text{on } [0, T) \times \mathbb{R}^d \\ \bar{u} = u \quad \text{on } ([0, T] \times \mathbb{R}^d) \setminus ([0, \bar{T}] \times \mathcal{O}) \\ \max(\partial_t \bar{u} + \mathcal{A}\bar{u} + c, \ell - \bar{u}) \quad \text{on } [0, \bar{T}) \times \mathcal{O}, \end{cases} \quad (4.52)$$

to be solved for the pair (u, \bar{u}) of the no protection pricing function u and the protection pricing function \bar{u} . We then have the following analog to Proposition 4.2.3:

Proposition 4.3.1 *Under suitable conditions, the BSDE (4.50) admits a unique solution (π, m, a) and the cascade of two PIDEs (4.52) admits a unique viscosity solution (u, \bar{u}) . The connection between (π, m, a) and (u, \bar{u}) reads*

$$\pi_t = v(t, X_t),$$

where v is defined as u for $t > \vartheta$ and \bar{u} for $t \leq \vartheta$.

¹⁸In which β now refers to the discount factor associated with an arbitrary numéraire B .

Moreover, in case the pricing functions u , \bar{u} are sufficiently regular for an Itô formula to be applicable,¹⁹ we have

$$dm_t = \partial v \sigma(t, X_t) dW_t + \delta v(t, X_{t-}) \cdot d\mu_t.$$

We further assume that the primary risky price process P likewise satisfies $p := \beta P = v(t, X_t)$ for a function v such that

$$d(\beta_t \hat{P}_t) = \partial v \sigma(t, X_t) dW_t + \delta v(t, X_{t-}) \cdot d\mu_t. \quad (4.53)$$

We then have the following analog of Proposition 4.2.5:

Proposition 4.3.2 *For the game option with the data C , ξ , L , \bar{U} , $\Pi_0 = B_0 v(t, X_0)$ is the minimal initial wealth of a superhedge with \mathbb{P} -local martingale cost process. Moreover, the cost process $\rho = \rho(\zeta)$ and the tracking error process $e = e(\zeta)$ in (4.18) and (4.23) are given by*

$$\begin{aligned} d\rho_t &= (\partial v \sigma(t, X_t) - \zeta_t \partial v \sigma(t, X_t)) dW_t \\ &\quad + (\delta v(t, X_{t-}) - \zeta_t \delta v(t, X_{t-})) \cdot d\mu_t \end{aligned} \quad (4.54)$$

$$d(\beta_t e_t) = -c_t dt + \zeta_t d(\beta_t \hat{P}_t) - d\pi_t = da_t - \beta_t d\rho_t. \quad (4.55)$$

Provided $\partial v \sigma$ is left-invertible, it is thus possible to perfectly hedge against W by setting

$$\zeta_t = \partial v \sigma(\partial v \sigma)^{-1}(t, X_t). \quad (4.56)$$

In the simplest case where $d = q$ and ∂v and σ are invertible, this formula reduces to

$$\zeta_t = \partial v(\partial v)^{-1}(t, X_t). \quad (4.57)$$

Alternatively, provided $\delta v(t, X_{t-})$ is left-invertible (assuming a jump measure with finite support), it is possible to perfectly hedge against μ by setting

$$\zeta_t = \delta v(\delta v)^{-1}(t, X_{t-}). \quad (4.58)$$

Still another possibility, which minimizes the risk-neutral variance of the error, is to use the strategy ζ^{va} given by

$$\zeta_t^{va} = \frac{d\langle \pi, p \rangle_t}{dt} \left(\frac{d\langle p \rangle_t}{dt} \right)^{-1} = (v, v)(v, v)^{-1}(t, X_{t-}). \quad (4.59)$$

¹⁹Otherwise a more general but less constructive representation for m_t can be given in terms of Malliavin calculus.

4.3.1.1 Change of Numéraire

Now let there be given another numéraire, or reference asset with relative price process $\beta_t \tilde{B}_t$, assumed to be a positive \mathbb{P} -martingale. Let $\tilde{\mathbb{P}}$ be the pricing measure on (Ω, \mathcal{F}) , associated with the numéraire \tilde{B} , such that

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \mu_T, \quad (4.60)$$

with $\mu_t = \frac{B_0 \tilde{B}_t}{B_0 B_t}$. We write $\mathbb{E}^{\tilde{\mathbb{P}}} = \tilde{\mathbb{E}}$. Observe that we have, for every \mathcal{F}_t -measurable and bounded random variable χ ,

$$\tilde{\mathbb{E}}(\chi) = \mathbb{E}(\chi \mu_T) = \mathbb{E}[\chi \mathbb{E}(\mu_T | \mathcal{F}_t)] = \mathbb{E}(\chi \mu_t). \quad (4.61)$$

As a consequence, μ_t is the \mathcal{F}_t -measurable Radon-Nikodym density $\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}|_{\mathcal{F}_t}$ of $\tilde{\mathbb{P}}$ with respect to \mathbb{P} on \mathcal{F}_t , for every $t \in [0, T]$. Also, by Lemma 3.4.10, a process Θ is a $\tilde{\mathbb{P}}$ -local martingale if and only if $(\mu \Theta)$ is a \mathbb{P} -local martingale. Considering a (nondividend paying) European asset with \mathbb{P} -price process Π , we have in particular the following \mathbb{P} -martingale:

$$\frac{\Pi}{B} = \mu \frac{\tilde{B}_0 \Pi}{B_0 \tilde{B}}.$$

So $\frac{\Pi}{B}$ is a $\tilde{\mathbb{P}}$ -local martingale and (assuming it is a $\tilde{\mathbb{P}}$ -true martingale), for $t \in [0, T]$, we have that

$$\Pi_t = \tilde{B}_t \tilde{\mathbb{E}}(\tilde{B}_T^{-1} \xi | \mathcal{F}_t). \quad (4.62)$$

4.3.2 Defaultable Derivatives

We now consider an extension of the martingale modeling approach to defaultable derivatives. This is important in dealing with credit derivatives and with counterparty risk, the latter being a major issue and a main driver of investment bank P&Ls since the 2007–2009 credit crisis [8, 30, 33, 63, 80, 81, 86, 216].

Returning to risk-neutral modeling with respect to a numéraire B given as a savings account and for the riskless discount factor $\beta = B^{-1}$ of (4.1), we thus consider defaultable derivatives with terminal payoffs of the form $\mathbb{1}_{T < \tau_d} \phi(S_T)$,²⁰ where τ_d represents the default-time of a reference entity. We will follow the reduced-form intensity approach originally introduced by Lando [175] and Jarrow and Turnbull [158], subsequently generalized in many ways in the credit risk literature (see [43, 44]) and extended in particular to American and game claims in Bielecki et al. [37–40]. We give almost no proofs in this subsection. For these we refer the

²⁰Or $\mathbb{1}_{v < \tau_d} \phi(S_v)$ upon exercise at a stopping time v , in the cases of American or game claims.

interested reader to [37–39] and to Sect. 10.1 and [40] for a concrete example of a reduced-form modeling approach. The main message is that defaultable claims can be handled in essentially the same way as default-free claims, provided that the default-free discount factor process β is replaced by a credit risk adjusted discount factor α and that a fictitious dividend, continuously paid at rate γ (the so-called default intensity), is introduced to account for recovery on the claim upon default. Note in this regard that the “default-free” discount factor β in (4.1) can itself be interpreted in terms of a default risk with “intensity” or “killing” rate r_t .

4.3.2.1 Cash Flows

Given a $[0, T] \cup \{+\infty\}$ -valued stopping time τ_d representing the default time of the issuer of a claim, let

$$J_t = \mathbb{1}_{\{t < \tau_d\}}$$

represent the related survival indicator process. A defaultable game option is a game option in the sense of Definition 4.1.7(iii), with all cash flows killed at the default time τ_d . We will consider defaultable game options with call protection ϑ . For reasons analogous to those enunciated above, these encompass, as a special case, defaultable American options (case $\vartheta = T$), themselves including defaultable European options. Given a call protection $\vartheta \in \mathcal{T}$ and a time $t \in [0, T]$, let here ν represent $\tau \wedge \theta \wedge \tau_d$, for every $(\tau, \theta) \in \mathcal{T}_t \times \mathcal{T}_t^\vartheta$.

Definition 4.3.3 A defaultable game option is a game option with the ex-dividend cumulative future discounted cash flows $\beta_t \pi^t(\tau, \theta)$ given, for every time $t \in [0, T]$, issuer put time $\tau \in \mathcal{T}_t$ and holder call time $\theta \in \mathcal{T}_t^\vartheta$, by

$$\beta_t \pi^t(\tau, \theta) = \int_t^\nu \beta_s dD_s + \beta_\nu J_\nu (\mathbb{1}_{\{\nu=\tau < T\}} L_\tau + \mathbb{1}_{\{\nu < \tau\}} U_\theta + \mathbb{1}_{\{\nu=T\}} \xi),$$

where:

- the dividend process $D = (D_t)_{t \in [0, T]}$ equals

$$D_t = \int_{[0, t]} J_s C_s ds - R_s dJ_s \quad (4.63)$$

- for some coupon rate process²¹ $C = (C_t)_{t \in [0, T]}$ and some predictable recovery process $R = (R_t)_{t \in [0, T]}$;
- the put payment $L = (L_t)_{t \in [0, T]}$ and the call payment $U = (U_t)_{t \in [0, T]}$ are càdlàg processes, and the payment at maturity ξ is a random variable such that

$$L \leq U \quad \text{on } [0, T], \quad L_T \leq \xi \leq U_T.$$

²¹See Sect. 14.1 for the case of discrete coupons.

We further assume that R , L and ξ are bounded from below, as is then the cumulative discounted lower payoff. Specifically, there exists a constant c such that

$$\int_{[0,t]} \beta_s dD_s + \beta_t J_t (\mathbb{1}_{\{t < T\}} L_t + \mathbb{1}_{\{t = T\}} \xi) \geq -c, \quad t \in [0, T]. \quad (4.64)$$

Convertible Bonds The standing example of a defaultable game option is a convertible bond. Convertible bonds have two important features:

- early put and call clauses, respectively at the holder's and issuer's convenience,
- defaultability, since they are corporate bonds²² and one of the main vehicles of the so-called equity-to-credit strategies [100, 242].

We introduce:

- \bar{N} : the nominal of the bond,
 S : the price process of the stock underlying the bond,
 \bar{R} : the recovery rate process on the bond upon default of the issuer,
 Λ : the fractional loss on the underlying stock, given default of the issuer of the bond,
 κ : the bond conversion factor,
 \bar{P}, \bar{C} : the put and call nominal payments with, by assumption, $\bar{P} \leq \bar{N} \leq \bar{C}$.

Definition 4.3.4 A convertible bond is a defaultable game option with coupon rate process C , recovery process R^{cb} and payoffs L^{cb}, U^{cb}, ξ^{cb} such that

$$R_t^{cb} = (1 - \Lambda) \kappa S_{t-} \vee \bar{R}_t, \quad \xi^{cb} = \bar{N} \vee \kappa S_T \quad (4.65)$$

$$L_t^{cb} = \bar{P} \vee \kappa S_t, \quad U_t^{cb} = \bar{C} \vee \kappa S_t. \quad (4.66)$$

See [39] and [100] for detailed descriptions of convertible bond covenants and Chap. 10 for related studies.

4.3.2.2 Reduction of Filtration in the Hazard Intensity Setup

As detailed in Bielecki et al. [38], an application of Proposition 4.1.8 yields:

Proposition 4.3.5 Assume that a semimartingale Π is the value process of the Dynkin game related to a defaultable game option under the risk-neutral measure \mathbb{P} . I.e. for $t \in [0, T]$:

$$\underset{\tau \in \mathcal{T}_t}{\text{esssup}} \underset{\theta \in \mathcal{T}_t^\vartheta}{\text{essinf}} \mathbb{E}(\pi^t(\tau, \theta) | \mathcal{F}_t) = \Pi_t = \underset{\theta \in \mathcal{T}_t^\vartheta}{\text{essinf}} \underset{\tau \in \mathcal{T}_t}{\text{esssup}} \mathbb{E}(\pi^t(\tau, \theta) | \mathcal{F}_t). \quad (4.67)$$

Then Π is an arbitrage price process for the defaultable game option.

²²Since the 2011 Euro sovereign debt crisis, defaultability is also a feature of many Euro government bonds.

Looking toward applying the so-called reduced-form approach in single-name credit risk [44], we assume further that the full model filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$ is given as a reference filtration $\widetilde{\mathbb{F}} = (\widetilde{\mathcal{F}}_t)_{t \in [0, T]}$ progressively enlarged by the default time, so that $\mathbb{F} = \widetilde{\mathbb{F}} \vee \mathbb{H}$, where $\mathbb{H} = (\sigma(\tau_d \wedge t))_{t \in [0, T]}$ is the natural filtration of J . We moreover assume that the Azema supermartingale, i.e. the optional projection Q of J defined, for $t \in [0, T]$, by

$$Q_t = \mathbb{P}(\tau_d > t | \widetilde{\mathcal{F}}_t),$$

is a positive, continuous and finite variation process.

Since Q is continuous, τ_d is a totally inaccessible \mathbb{F} -stopping time [95]. Moreover, τ_d avoids $\widetilde{\mathbb{F}}$ -stopping times in the sense that $\mathbb{P}(\tau_d = \tau) = 0$ for every $\widetilde{\mathbb{F}}$ -stopping time τ [67].

The (\mathcal{H}) - (or immersion) hypothesis means that all $\widetilde{\mathbb{F}}$ -local martingales are \mathbb{F} -local martingales, whereas τ_d being an $\widetilde{\mathbb{F}}$ -pseudo-stopping time means that all $\widetilde{\mathbb{F}}$ -local martingales stopped at τ_d are \mathbb{F} -local martingales [212]. Assuming that Q is continuous, the further assumption that Q has finite variation in fact implies that Q is nonincreasing, which lies somewhere between further assuming the (stronger) immersion hypothesis and assuming that τ_d is an $\widetilde{\mathbb{F}}$ -pseudo-stopping time.

We assume for simplicity that Q is time-differentiable, and we define the default hazard intensity γ , the credit-risk-adjusted-interest-rate \tilde{r} and the credit-risk-adjusted-discount-factor α , respectively, by:

$$\gamma_t = -\frac{d \ln Q_t}{dt}, \quad \tilde{r}_t = r_t + \gamma_t, \quad \alpha_t = \beta_t \exp\left(-\int_0^t \gamma_s ds\right) = \exp\left(-\int_0^t \tilde{r}_s ds\right).$$

Under the previous assumptions, the compensated jump-to-default process

$$H_t = \mathbb{1}_{\tau_d \leq t} - \int_0^{t \wedge \tau_d} \gamma_s ds, \quad t \in [0, T] \tag{4.68}$$

is an \mathbb{F} -martingale. The quantities $\tilde{\tau}$ and $\tilde{\Pi}$ introduced in the next lemma (see Bi-elecki et al. [38] for proof and references) are called the pre-default values of τ and Π .

Lemma 4.3.6

- (i) For any \mathbb{F} -adapted, respectively \mathbb{F} -predictable process Π over $[0, T]$, there exists a unique $\widetilde{\mathbb{F}}$ -adapted, respectively $\widetilde{\mathbb{F}}$ -predictable process $\tilde{\Pi}$ over $[0, T]$, such that $J\Pi = J\tilde{\Pi}$, respectively $J_-\Pi = J_-\tilde{\Pi}$, over $[0, T]$.
- (ii) For any $\tau \in \mathcal{T}$, there exists a $[0, T]$ -valued $\widetilde{\mathbb{F}}$ -stopping time $\tilde{\tau}$ such that $\tau \wedge \tau_d = \tilde{\tau} \wedge \tau_d$.

In view of the structure of the payoffs π in (4.63), we assume without loss of generality that C, L, U are $\widetilde{\mathbb{F}}$ -adapted, ξ is $\widetilde{\mathcal{F}}_T$ -measurable, R is $\widetilde{\mathbb{F}}$ -predictable and ϑ, θ, τ are $\widetilde{\mathbb{F}}$ -stopping times. For any $t \in [0, T]$, let henceforth in this subsection

\mathcal{T}_t , or \mathcal{T} in the case $t = 0$, denote the set of $[t, T]$ -valued $\widetilde{\mathbb{F}}$ - (rather than \mathbb{F} -before) stopping times; also, let v denote $\tau \wedge \theta$ (rather than $\tau \wedge \theta \wedge \tau_d$ above), for every $t \in [0, T]$ and $\tau, \theta \in \mathcal{T}_t$. The next lemma (see Bielecki et al. [38] for proof and references) shows that the computation of conditional expectations of cash flows $\pi^t(\tau, \theta)$ with respect to \mathcal{F}_t can be reduced to the computation of conditional expectations of $\widetilde{\mathbb{F}}$ -equivalent cash flows $\tilde{\pi}^t(\tau, \theta)$ with respect to $\widetilde{\mathcal{F}}_t$.

Lemma 4.3.7 *For any stopping times $(\tau, \theta) \in \mathcal{T}_t \times \mathcal{T}_t^\vartheta$, we have*

$$\mathbb{E}(\pi^t(\tau, \theta) | \mathcal{F}_t) = J_t \mathbb{E}(\tilde{\pi}^t(\tau, \theta) | \widetilde{\mathcal{F}}_t),$$

where $\tilde{\pi}^t(\tau, \theta)$ is given, with $v = \tau \wedge \theta$, by

$$\alpha_t \tilde{\pi}^t(\tau, \theta) = \int_t^v \alpha_s f_s ds + \alpha_v (\mathbb{1}_{\{v=\tau < T\}} L_\tau + \mathbb{1}_{\{\theta < \tau\}} U_\theta + \mathbb{1}_{\{v=T\}} \xi), \quad (4.69)$$

where we have set $f = C + \gamma R$.

The following result, which is a consequence of Lemma 4.3.7, reduces the pricing task in the original market, which is subject to the default risk in which cash flows are discounted according to the discount factor β , to pricing in a fictitious default-free market in which cash flows are discounted according to the credit risk adjusted discount factor α .

Proposition 4.3.8 (see Bielecki et al. [38]) *If an $\widetilde{\mathbb{F}}$ -semimartingale $\tilde{\Pi}$ solves the $\widetilde{\mathbb{F}}$ -Dynkin game with payoff $\tilde{\pi}$ in the sense that, for every $t \in [0, T]$,*

$$\begin{aligned} \underset{\tau \in \mathcal{T}_t}{\text{esssup}} \underset{\theta \in \mathcal{T}_t^\vartheta}{\text{essinf}} \mathbb{E}(\tilde{\pi}^t(\theta, \tau) | \widetilde{\mathcal{F}}_t) &= \tilde{\Pi}_t \\ &= \underset{\theta \in \mathcal{T}_t^\vartheta}{\text{essinf}} \underset{\tau \in \mathcal{T}_t}{\text{esssup}} \mathbb{E}(\tilde{\pi}^t(\tau, \theta) | \widetilde{\mathcal{F}}_t), \end{aligned}$$

then $\Pi := J \tilde{\Pi}$ is an \mathbb{F} -semimartingale solving the \mathbb{F} -Dynkin game with payoff π .

Hence, by Proposition 4.3.5, Π is an arbitrage price for the option, with pre-default price process $\tilde{\Pi}$.

4.3.2.3 Pre-default BSDE Modeling

The next step consists in modeling $\tilde{\Pi}$ as the value process of a solution $(\tilde{\Pi}, \tilde{M}, \tilde{A})$, assumed to exist, to the following doubly reflected BSDE with the data

$$\alpha, f = C + \gamma R, \quad \xi, L, \overline{U} = \mathbb{1}_{\{\cdot < \vartheta\}} \infty + \mathbb{1}_{\{\cdot \geq \vartheta\}} U,$$

where a solution to (4.70) below is defined in analogy with Definition 4.1.11:

$$\begin{aligned} \alpha_t \tilde{\Pi}_t &= \alpha_T \xi + \int_t^T \alpha_s (f_s ds + d\tilde{A}_s - d\tilde{M}_s), \quad t \in [0, T], \\ L_t \leq \tilde{\Pi}_t \leq \bar{U}_t, \quad t &\in [0, T], \\ \int_0^T (\tilde{\Pi}_s - L_s) d\tilde{A}_s^+ &= \int_0^T (\bar{U}_s - \tilde{\Pi}_s) d\tilde{A}_s^- = 0. \end{aligned} \tag{4.70}$$

Hence, by Proposition 4.1.14, the $\tilde{\mathbb{F}}$ -semimartingale $\tilde{\Pi}$ solves the $\tilde{\mathbb{F}}$ -Dynkin game with payoff $\tilde{\pi}$. Thus, by Proposition 4.3.8, $\Pi := J\tilde{\Pi}$ is an arbitrage price for the option, with related pre-default price process $\tilde{\Pi}$. For $t \in [0, T]$, we define

$$\Pi_t = \mathbb{1}_{\{t < \tau_d\}} \tilde{\Pi}_t, \quad \beta_t \hat{\Pi}_t = \beta_t \Pi_t + \int_{[0, t]} \beta_s dD_s, \tag{4.71}$$

where D_t is as in (4.63). We also define M by $M_0 = 0$ and, for $t \in [0, T]$,

$$\int_{[0, t]} \beta_s dM_s = \beta_t \hat{\Pi}_t + \int_0^t \beta_s J_s dA_s. \tag{4.72}$$

The following result allows us to interpret (4.72) as the canonical decomposition of the \mathbb{F} -special semimartingale $\beta \hat{\Pi}$, the discounted cumulative price of the option as defined in (4.22). In particular, M is the canonical \mathbb{F} -local martingale component of $\int_{[0, \cdot]} \beta_t^{-1} d(\beta_t \hat{\Pi}_t)$.

Lemma 4.3.9 *The process M defined by (4.72) is an \mathbb{F} -local martingale stopped at τ_d .*

Proof By (4.70) we have, for every $t \in [0, T]$,

$$\int_0^t \alpha_s d\tilde{M}_s = \alpha_t \tilde{\Pi}_t - \tilde{\Pi}_0 + \int_0^t \alpha_s d\tilde{A}_s + \int_0^t \alpha_s (C_s + \gamma_s R_s) ds.$$

By standard computations analogous to those involved in the proof of Lemma 4.3.7 (see [38]), we thus obtain, for every $0 \leq t \leq s \leq T$,

$$\mathbb{E}\left(\beta_t^{-1} \int_t^s \beta_r dM_r \mid \mathcal{F}_t\right) = J_t \mathbb{E}\left(\alpha_t^{-1} \int_t^s \alpha_r d\tilde{M}_r \mid \tilde{\mathcal{F}}_t\right) = 0. \quad \square$$

Let

$$\theta^* = \inf\{s \in [\vartheta, T]; \tilde{\Pi}_s \geq U_s\} \wedge T. \tag{4.73}$$

For any primary strategy ζ , let the \mathbb{F} -local martingale $\rho = \rho(\zeta)$ still be defined by the first line of (4.23), with M now given as in (4.72). Proposition 4.3.10 can be regarded as an extension of Proposition 4.1.15 to the defaultable case, in which two

filtrations are involved. Note that our assumptions here are made with respect to the filtration $\tilde{\mathbb{F}}$, with respect to which the BSDE (4.70) is defined, whereas conclusions are drawn with respect to the full model filtration \mathbb{F} .

Proposition 4.3.10 (see Bielecki et al. [37, 38])

- (i) For any hedging strategy ζ , (Π_0, ζ, θ^*) is a hedge with (\mathbb{F}, \mathbb{P}) -local martingale cost ρ .
- (ii) Π_0 is the minimal initial wealth of a hedge with (\mathbb{F}, \mathbb{P}) -local martingale cost.
- (iii) In the special case of a European derivative with $\tilde{A} = 0$, (Π_0, ζ) is a replicating strategy with (\mathbb{F}, \mathbb{P}) -local martingale cost ρ , and Π_0 is the minimal initial wealth of a replicating strategy with (\mathbb{F}, \mathbb{P}) -local martingale cost.

Analysis of Hedging Strategies Our analysis of hedging strategies will rely on the following lemma that yields the dynamics of the price process of a game option or, more precisely, of the \mathbb{F} -local martingale component M of process $\int_{[0,\cdot]} \beta_t^{-1} d(\beta_t \widehat{\Pi}_t)$. The compensated jump-to-default \mathbb{F} -martingale H_t was introduced in (4.68).

Lemma 4.3.11 The \mathbb{F} -local martingale M defined in (4.72) satisfies, for $t \in [0, T \wedge \tau_d]$:

$$dM_t = d\tilde{M}_t + \Delta \widehat{\Pi}_t dH_t \quad (4.74)$$

with $\Delta \widehat{\Pi}_t := R_t - \tilde{\Pi}_{t-}$.

Proof (We give but a sketch; see Bielecki et al. [38] for the detail.) This follows by computations similar to those in the proof of Kusuoka's Theorem 2.3 in [171] (where the (\mathcal{H}) -hypothesis and a more specific Brownian reference filtration are assumed), using in particular the avoidance property that $\mathbb{P}(\tau_d = \tau) = 0$ for every $\tilde{\mathbb{F}}$ -stopping time τ . \square

In analogy with the structure of the payoffs of a defaultable derivative, we assume henceforth that the primary market dividend vector-process \mathcal{D} is given by

$$\mathcal{D}_t = \int_{[0,t]} J_s \mathcal{C}_s ds - \mathcal{R}_s dJ_s$$

for suitable coupon rate and recovery processes \mathcal{C} and \mathcal{R} . We also assume that $P = J \tilde{P}$, without loss of generality for the application of hedging a defaultable derivative. In particular, any value of the primary market at τ_d is embedded in the recovery \mathcal{R} in \mathcal{D} . We further define, along with the usual primary cumulative price \tilde{P} , the pre-default cumulative price, by letting for $t \in [0, T]$:

$$\overline{P}_t = \tilde{P}_t + \alpha_t^{-1} \int_0^t \alpha_s g_s ds,$$

where we have set $g = \mathcal{C} + \gamma \mathcal{R}$. The following decomposition is the analog, for the primary market, of decomposition (4.74) for a game option.

Lemma 4.3.12 (see Bielecki et al. [37]) *Process $\alpha \bar{P}$ is an $\widetilde{\mathbb{F}}$ -local martingale, and we have, for $t \in [0, T \wedge \tau_d]$:*

$$\beta_t^{-1} d(\beta_t \widehat{P}_t) = \alpha_t^{-1} d(\alpha_t \bar{P}_t) + \Delta \widehat{P}_t dH_t, \quad (4.75)$$

where $\Delta \widehat{P}_t := \mathcal{R}_t - \widetilde{P}_{t-}$.

Substitution of (4.74) and (4.75) into the hedging cost ρ (first line of (4.23)) of the strategy (Π_0, ζ, θ^*) yields the following decomposition of the hedging cost ρ :

Proposition 4.3.13 *Under the previous assumptions, for every primary strategy ζ , the related cost $\rho = \rho(\zeta)$ in Proposition 4.3.10 satisfies, for $t \in [0, T \wedge \tau_d]$,*

$$d\rho_t = dM_t - \zeta_t \beta_t^{-1} d(\beta_t \widehat{P}_t) = [d\widetilde{M}_t - \zeta_t \alpha_t^{-1} d(\alpha_t \bar{P}_t)] + [\Delta \widehat{P}_t - \zeta_t \Delta \widehat{P}_t] dH_t. \quad (4.76)$$

4.3.2.4 Pre-default Markovian Setup

We now assume that the pre-default pricing BSDE (4.70) is Markovian, in the sense that the pre-default input data $\tilde{r} = r + \gamma$, $f = C + \gamma R$, ξ , L , U of (4.70) are given as Borel functions of an $(\widetilde{\mathbb{F}}, \mathbb{P})$ -Markov factor process X , i.e.

$$\begin{aligned} \tilde{r}_t &= \tilde{r}(t, X_t), & f_t &= f(t, X_t), & \xi &= \phi(X_T) \\ L_t &= L(t, X_t), & U_t &= U(t, X_t). \end{aligned}$$

We assume more specifically that the pre-default factor process X is defined by (4.30) with respect to the filtration $\widetilde{\mathbb{F}} = \mathbb{F}^{W, \mu}$ with related generator \mathcal{A} , and that ϑ is defined by (4.33). We can then introduce the pre-default pricing PIDE cascade related to the pre-default pricing BSDE (4.70), to be solved for the pair (u, \bar{u}) of the pre-default no protection pricing function u and the pre-default protection pricing function \bar{u} . Thus (cf. (4.34)–(4.35) or (4.52)):

$$\begin{cases} u(T, x) = \phi(x), & x \in \mathbb{R}^d \\ \min(\max(\partial_t u + \mathcal{A}u + f - \tilde{r}u, L - u), U - u) = 0 & \text{on } [0, T) \times \mathbb{R}^d \\ \bar{u} = u & \text{on } ([0, T] \times \mathbb{R}^d) \setminus ([0, \bar{T}) \times \mathcal{O}) \\ \max(\partial_t \bar{u} + \mathcal{A}\bar{u} + f - \tilde{r}\bar{u}, L - \bar{u}) & \text{on } [0, \bar{T}) \times \mathcal{O}. \end{cases} \quad (4.77)$$

We then have, as before, by application of the results of Chaps. 12 and 13:

Proposition 4.3.14 Under mild conditions, the cascade of the two PIDEs (4.77) is well-posed in the sense of viscosity solutions, and its solution (u, \bar{u}) is related to the solution $(\tilde{\Pi}, \tilde{M}, \tilde{A})$ of (4.70), for $t \in [0, T]$, by:

$$\tilde{\Pi}_t = v(t, X_t), \quad (4.78)$$

where v is defined as u for $t > \vartheta$ and \bar{u} for $t \leq \vartheta$.

Moreover, in case the pricing functions u and \bar{u} are sufficiently regular for an Itô formula to be applicable,²³ we have for $t \in [0, T]$:

$$d\tilde{M}_t = \partial v \sigma(t, X_t) dW_t + \delta v(t, X_{t-}) \cdot d\mu_t. \quad (4.79)$$

Accordingly, the first line of (4.70) assumes the following form:

$$-dv(t, X_t) = (f - \tilde{r}v)(t, X_t) dt + d\tilde{A}_t - \partial v \sigma(t, X_t) dW_t - \delta v(t, X_{t-}) \cdot d\mu_t. \quad (4.80)$$

We assume the same structure (without the barriers) on a primary market pre-default price process \tilde{P} , so that $\tilde{P}_t = v(t, X_t)$ where, setting $g(t, x) = \mathcal{C}(t, x) + \gamma(t, x)\mathcal{R}(t, x)$:

$$-dv(t, X_t) = (g - \tilde{r}v)(t, X_t) dt - \partial v \sigma(t, X_t) dW_t - \delta v(t, X_{t-}) \cdot d\mu_t. \quad (4.81)$$

Substituting (4.80) and (4.81) into (4.76) yields:

Proposition 4.3.15 For $t \in [0, T \wedge \tau_d]$,

$$\begin{aligned} d\rho_t = & [\partial v \sigma(t, X_t) - \zeta_t \partial v \sigma(t, X_t)] dW_t + [\delta v(t, X_{t-}) - \zeta_t \delta v(t, X_{t-})] \cdot d\mu_t \\ & + [\Delta v_t - \zeta_t \Delta v_t] dH_t, \end{aligned} \quad (4.82)$$

where we have set $\Delta v_t = R_t - v(t, X_{t-})$ and $\Delta v_t = \mathcal{R}_t - v(t, X_{t-})$.

As in Sect. 4.2.3 (see also Bielecki et al. [40]), this decomposition of the hedging cost ρ can then be used for devising practical hedging schemes of a defaultable game option, such as superhedging ($\rho = 0$), hedging only the market (spread) risk W , hedging only the default risk H or min-variance hedging.

4.3.3 Intermittent Call Protection

We now want to consider callable products with a more general form of intermittent call protection, i.e. call protection whenever a certain condition is satisfied, rather

²³Otherwise a more general but less constructive representation for \tilde{M}_t can be given in terms of Malliavin calculus.

than simply call protection before a certain stopping time. This will be stated in terms of an effective call payoff process \bar{U} of the following form:

$$\bar{U}_t = \Omega_t^c \infty + \Omega_t U_t \quad (4.83)$$

for given càdlàg event-processes²⁴ Ω_t , $\Omega_t^c = 1 - \Omega_t$. The interpretation of (4.83) is that a call is possible whenever $\Omega_t = 1$; otherwise call protection is in force. Note that (4.20) corresponds to the special case $\Omega_t = \mathbb{1}_{\{t \geq \vartheta\}}$ in (4.83). The identification between:

- the arbitrage, or infimal superhedging \mathbb{P} -price process of a game option with intermittent call protection, and
- the value process Π of a solution (Π, M, A) (assumed to exist) to the BSDE (4.19), in which \bar{U} is given by (4.83),

can be established by a straightforward adaptation of the arguments developed in Sect. 4.1.

In the jump-diffusion factor process model X defined by (4.30), assuming

$$\Omega_t = \Omega(t, X_t, H_t) \quad (4.84)$$

for a suitably augmented finite-dimensional Markovian factor process (X_t, H_t) and a related Boolean function Ω of (t, X, H) , it is expected that one should then have $\Pi_t = u(t, X_t, H_t)$ on $[0, T]$ for a suitable pricing function u . Under suitable technical conditions, this is precisely what emerges from the results of Sect. 14.2, in the case of a call protection discretely monitored at the dates of a finite time grid $\mathfrak{T} = \{T_0, T_1, \dots, T_m\}$. As standing examples of such discretely monitored call protections, we mention the following clauses, which are commonly found in convertible bond contracts on an underlying stock S . Let S_t be given as X_t^1 , the first component of our factor process X_t .

Example 4.3.16 Given a constant trigger level \bar{S} and a constant integer i :

- (i) call possible whenever $S_t \geq \bar{S}$ at the last i monitoring times T_l ; call protection otherwise.

Or, more generally, given a further integer $j \geq i$:

- (ii) call possible whenever $S_t \geq \bar{S}$ on at least i of the last j monitoring times T_l ; call protection otherwise.

This will be investigated further in Chap. 10, using the results of Sect. 14.2.

²⁴Boolean-valued processes.

4.4 From Theory to Practice

4.4.1 Model Calibration

In practical applications we can think of \mathbb{P} as “the pricing measure chosen by the market” to price a contingent claim. For hedging purposes or in order to implement bets on specific risk factors, and also for pricing exotic or structured products, traders need to know the market pricing measure \mathbb{P} .

In practice, the measure \mathbb{P} is typically estimated by calibration of a model to market data.²⁵ Indeed there are two sets of constraints that the market pricing measure \mathbb{P} must satisfy. First, \mathbb{P} must satisfy structural requirements stemming from its equivalence with the objective probability measure $\hat{\mathbb{P}}$. Any process must thus have the same trajectory properties (such as continuity or lack of it) under the objective and under an equivalent pricing measure. Second, the cross-section $\Pi_t^\pm(T, K)$ of the market prices of European vanillas quoted at any pricing time t on an underlying S must satisfy

$$\Pi_t^\pm(T, K) = \beta_t^{-1} \mathbb{E}_t \beta_T (S_T - K)^\pm, \quad (T, K) \in \text{obs}_t, \quad (4.85)$$

where obs_t is the set of the most liquid options on S (European vanilla at-the-money or slightly out-of-the-money calls and puts) quoted in the market at time t .

Constraints of type (4.85) are called calibration constraints. A model is said to fit the market smile, at a given time t , if it satisfies the calibration constraints (4.85). Accounting also for synchronization and noise issues in the market data, one commonly relaxes the calibration equality constraints (4.85) into inequality constraints within the bid-ask spread. Quite a few classes of models can fit the smile within the bid-ask spread, provided their parameters are suitably calibrated. A further requirement can be to fit the smile dynamics priced by the market. This corresponds to additional calibration constraints associated with market prices of exotic options.

Finally, for being usable in practice, a pricing model needs to be constructive and implementable in real time. Concretely this leads to work with a low-dimensional “Markovian proxy” X for the “true” (if any) market factor process.

4.4.2 Hedging

A model calibrated to the market can be used for hedging purposes, and for dealing with more exotic products. When a bank sells a derivative, it immediately sets up a hedge composed of liquid instruments, such as the asset(s) underlying the derivative, and/or further vanilla derivatives. But for feasibility, as well as for transaction

²⁵Model calibration will be the topic of Chap. 9.

cost issues (note that transaction costs are not considered in our setup), the bank is restricted to piecewise constant hedging strategies ζ^h such that

$$\zeta_t^h = \zeta_{t_i}^h \quad \text{for } t_i < t \leq t_{i+1}, \quad (4.86)$$

where $(t_i)_{0 \leq i \leq n}$ is a time-grid over $[0, T]$. In practice n may vary from one (static hedging) to the number of days or weeks between 0 and T . Since derivative payoffs are typically nonlinear, in order to get a good hedge the composition of the hedging portfolio must be updated at a high enough frequency.

In an idealized, complete market model, a continuously rebalanced delta-hedge ζ^* provides a perfect hedge to an option's seller (profit-and-loss identically equal to 0 or, in the case of an American option, super-hedge in case of sub-optimal exercise by the option holder). By contrast, in the real-world there are many reasons why a practical strategy ζ^h typically leads to negative P&Ls under some scenarios:

- Hedge slippage: the hedge ζ^h is rebalanced at discrete times, whereas a hedge must be rebalanced in continuous time to provide (in the complete market case) a perfect hedge.
- Model misspecification: note that hedging ratios are typically model-dependent, even among models calibrated to the same data set.
- Market incompleteness.
- Transaction costs and all kinds of costs, illiquidities and market imperfections, which are not accounted for in our formalism.
- Arbitrages which may occur in the market and are excluded by assumption in the standard pricing and hedging theory of this book.

Chapter 5

Benchmark Models

In this chapter we give a very succinct primer of basic models for reference derivative markets: equity derivatives (Black–Scholes model and stochastic volatility or/and jump extensions), interest rate derivatives (Libor market model) and credit derivatives (one-factor Gaussian copula model).

5.1 Black–Scholes and Beyond

We consider a primary market composed of the savings account $B = \beta^{-1}$ and of an underlying S , which may represent a stock, an index, a futures price, an exchange rate, a forward interest or swap rate, the value of a commodity, or any quantity reasonably modeled in the form of a nonnegative jump-diffusion. The riskless interest rate r in the economy and a constant yield dividend q on S are assumed to be constant,¹ so that, in particular, $B_t = e^{rt}$. Recall from Chap. 4 that, by arbitrage, the discounted wealth process of any admissible self-financing trading strategy in S and B must be a local martingale under a risk-neutral probability measure \mathbb{P} . In particular, the process $\beta_t S_t e^{qt} = S_t e^{-\kappa t}$, where $\kappa = r - q$, must be a \mathbb{P} -local martingale.

5.1.1 Black–Scholes Basics

Consistent with this requirement, the risk-neutral form of the Black–Scholes–Merton model [47, 202] postulates the following diffusion for S , driven by a standard \mathbb{P} -Brownian motion W :

$$dS_t = S_t(\kappa dt + \sigma dW_t), \quad (5.1)$$

for a constant volatility parameter σ . Or, explicitly:

¹The exact interpretation of r and q depends on the nature of the underlying S .

$$S_t = S_0 e^{bt + \sigma W_t}, \quad (5.2)$$

where $b = \kappa - \frac{1}{2}\sigma^2$.

Equivalent to (5.1), in terms of $F_t^T = S_t e^{\kappa(T-t)}$, we have:

$$dF_t^T = \sigma F_t^T dW_t. \quad (5.3)$$

The T -forward price F_t^T is thus a \mathbb{P} -Brownian martingale with constant volatility σ . In terms of the cumulative stock price \widehat{S}_t , from (5.3) we get:

$$d(\beta_t \widehat{S}_t) = d(\beta_t S_t) + \beta_t q S_t dt = e^{-qt} d(\beta_t S_t e^{qt}) = \beta_t \sigma S_t dW_t. \quad (5.4)$$

The \mathbb{P} -price process of a European vanilla option with integrable payoff $\phi(S_T)$ at T , say ϕ measurable and bounded for simplicity, is in turn given, for $t \in [0, T]$, by

$$\Pi_t = e^{-r(T-t)} \mathbb{E}_t \phi(S_T), \quad (5.5)$$

so that the discounted price $(e^{-rt} \Pi_t)$ is a \mathbb{P} -martingale. Moreover, by the Markov property of the risk-neutral Black–Scholes stock S , we have that

$$\mathbb{E}_t \phi(S_T) = \mathbb{E}(\phi(S_T) | S_t)$$

and therefore $\Pi_t = v(t, S_t)$ for a deterministic pricing function v . Assuming that v is sufficiently regular, an application of the Itô formula (3.30) yields:

$$e^{rt} d(e^{-rt} v(t, S_t)) = (\partial_t v + \mathcal{A}_S v - rv)(t, S_t) dt + \sigma S_t \partial_S v(t, S_t) dW_t,$$

where $\mathcal{A}_S v = \kappa S \partial_S v + \frac{\sigma^2 S^2}{2} \partial_{S^2}^2 v$. Since $e^{-rt} v(t, S_t) = e^{-rt} \Pi_t$ is a martingale, we have by Proposition 3.1.12 that

$$\partial_t v + \mathcal{A}_S v = rv.$$

Accounting for the terminal condition $\Pi_T = \phi(S_T)$, this leads to the following Black–Scholes pricing PDE:

$$\begin{cases} v(T, S) = \phi(S), & S \in (0, +\infty), \\ \partial_t v + \kappa S \partial_S v + \frac{1}{2} \sigma^2 S^2 \partial_{S^2}^2 v - rv = 0 & \text{in } [0, T) \times (0, +\infty). \end{cases} \quad (5.6)$$

Conversely, for sufficiently regular and bounded ϕ , the PDE (5.6) is known to have a unique classical solution v bounded on $[0, T] \times \mathbb{R}_+$ [129]. We then have, by application of the Itô formula:

$$\beta_T \phi(S_T) = \beta_t v(t, S_t) + \int_t^T \beta_s \partial_S v(s, S_s) \sigma S_s dW_s,$$

where, by (5.4),

$$\beta_s \partial_S v(s, S_s) \sigma S_s dW_s = \partial_S v(s, S_s) d(\beta_s \widehat{S}_s).$$

Thus

$$\beta_T \phi(S_T) = \beta_t v(t, S_t) + \int_t^T \partial_S v(u, S_u) d(\beta_u \widehat{S}_u). \quad (5.7)$$

The stochastic integral on the right-hand side is a bounded \mathbb{P} -local martingale and therefore a \mathbb{P} -martingale. We thus have that:

- $\Pi_t = v(t, S_t)$, which is obtained by taking conditional expectations in (5.7).
- The self-financing hedging strategy defined, for $t \in [0, T]$, by

$$\zeta_t^{bs} = \partial_S v(t, S_t) \quad (5.8)$$

units of stock S and $\beta_t(v(t, S_t) - S_t \partial_S v(t, S_t))$ units of the riskless asset $B_t = e^{rt}$, replicates the option payoff $\phi(S_T)$, starting from the wealth $v(0, S_0) = \Pi_0$ at time 0.

This provides a direct derivation, in the simple setup of the Black–Scholes model, of the outputs of Proposition 4.2.5.

Remark 5.1.1 Replicability of contingent claims in the Black–Scholes model explains why the Black–Scholes price doesn't depend on the physical drift, say μ , of S even though μ may be responsible for fat tails or skewness in the physical returns of S .

In order to illustrate the use of Proposition 4.2.5, we now consider the case of a primary market defined, as before, by the savings account B and a T -forward contract on S (instead of S above) with strike \mathcal{K} . The arbitrage price process Θ of the T -forward contract on S is thus given, for $t \in [0, T]$, by:

$$\Theta_t = S_t e^{-q(T-t)} - \mathcal{K} e^{-r(T-t)} = \Theta(t, S_t)$$

so that $\partial_S \Theta(t, S_t) = e^{-q(T-t)}$. By application of Proposition 4.2.5, a perfect replication strategy for the option with initial wealth $v(0, S_0)$ at time 0 is defined, for $t \in [0, T]$, by

$$\tilde{\zeta}_t^{bs} = \partial_S v(t, S_t) = e^{q(T-t)} \partial_S v(t, S_t) \quad (5.9)$$

units of T -forward contracts on S , and therefore $\beta_t(v(t, S_t) - \tilde{\zeta}_t^{bs} \Theta_t)$ units of the riskless asset B .

5.1.1.1 Black–Scholes Formulas

In the special case of a European call option with payoff function $\phi(S) = (S - K)^+$, we can check by verification in (5.6) that the Black–Scholes call pricing function Π^{bs} and the delta function $\Delta^{bs} = \partial_S \Pi^{bs}$ are given by the so-called Black–Scholes formulas:

$$\begin{aligned}\Pi^{bs}(t, S, T, K; r, q, \sigma) &= Se^{-q\tau} \mathcal{N}(d_+^{bs}) - Ke^{-r\tau} \mathcal{N}(d_-^{bs}), \\ \Delta^{bs}(t, S, T, K; r, q, \sigma) &= e^{-q\tau} \mathcal{N}(d_+^{bs}),\end{aligned}\quad (5.10)$$

where $\tau = T - t$, \mathcal{N} is the Gaussian cumulative distribution function and

$$d_{\pm}^{bs} = d_{\pm}^{bs}(t, S, T, K; r, q, \sigma) = \frac{\ln(\frac{S}{K}) + \kappa\tau}{\sigma\sqrt{\tau}} \pm \frac{1}{2}\sigma\sqrt{\tau}. \quad (5.11)$$

The argument $(t, S, T, K; r, q, \sigma)$ will be abbreviated, when no confusion might arise, by (t, S, T, K) or (t, S) , or it may sometimes even (as in d_{\pm}^{bs} above) be omitted. The above formulas admit straightforward extensions to the case where r , q and σ are time-integrable functions: simply replace $r\tau$, $q\tau$ and $\sigma\sqrt{\tau}$, respectively, by $\int_t^T r(u) du$, $\int_t^T q(u) du$, and Σ with $\Sigma^2 = \int_t^T \sigma^2(u) du$, in (5.10)–(5.11). We also have the corresponding formulas deduced by call-put parity for a put option.

Remark 5.1.2 The Black–Scholes formulas (5.10) can also easily be derived by a direct computation based on (5.5), like formulas (5.63) for the price and (5.66) for the delta in more general models below. Setting $t = 0$ and using the same notation as in the proof of Proposition 5.5.2, elementary log-normal density computations yield in the case of the Black–Scholes model:

$$P_0 = \mathbb{P}(S_T > K) = \mathcal{N}(d_+^{bs}(0, S)), \quad \tilde{P}_0 = \tilde{\mathbb{P}}(S_T > K) = \mathcal{N}(d_-^{bs}(0, S)).$$

As will be explained in Sect. 5.4, the Black–Scholes model is strongly misspecified. This leads us to consider various extensions of this model, adding stochastic volatility or/and jumps into the picture.

5.1.2 Heston Model

The best known stochastic volatility model is the Heston model [144], which postulates affine dynamics for the instantaneous variance process V_t . So, under a risk-neutral measure \mathbb{P} ,

$$\begin{cases} dV_t = \lambda(\theta - V_t) dt + \eta\sqrt{V_t} dB_t, \\ dS_t = S_t(\kappa dt + \sqrt{V_t} dW_t), \end{cases} \quad (5.12)$$

where:

- W and B are two \mathbb{P} -Brownian motions with correlation ρ ,
- λ is the speed of mean-reversion of the instantaneous variance V_t ,
- θ is the long-term variance mean, so that $\theta = \lim_{t \rightarrow \infty} \mathbb{E} V_t$,
- η is the volatility of the volatility parameter.

Remark 5.1.3 Despite the fact that the Heston SDE is non-Lipschitz, it can be shown to have a unique strong solution (see for instance [159]).

5.1.3 Merton Model

The Merton model [203] is obtained by adding an independent compound Poisson jump process to the Black–Scholes model. So, under a risk-neutral measure \mathbb{P} :

$$\frac{dS_t}{S_{t-}} = \kappa dt + \sigma dW_t + J_{(t)} dN_t - \lambda \bar{J} dt \quad (5.13)$$

where:

- $(N_t)_{t \geq 0}$ is a Poisson process with jump intensity λ and with ordered jump times denoted by T_l ,
- the $J_{(t)}$ are i.i.d. jump sizes such that $j_{(t)} := \ln(1 + J_{(t)})$ is $\mathcal{N}(\alpha, \beta)$ -distributed, so that

$$\bar{j} := \mathbb{E} j_{(t)} = \alpha, \quad \mathbb{E} j_{(t)}^2 = \alpha^2 + \beta, \quad \bar{J} := \mathbb{E} J_{(t)} = e^{\alpha + \frac{\beta}{2}} - 1,$$

- W, N and the $J_{(t)}$ are independent.

By application of the Itô formula (3.25), the model can be rewritten explicitly in terms of the log-spot $X_t = \ln(S_t)$ as:

$$dX_t = a dt + \sigma dW_t + j_{(t)} dN_t, \quad (5.14)$$

where $a = b - \lambda \bar{J}$, in which $b = \kappa - \frac{1}{2}\sigma^2$. Thus, using here and henceforth a simple subscript “ \cdot_l ” as a shorthand for “ $\cdot_{(T_l)}$ ”, $X_T = x + aT + \sigma W_T + \sum_{l=1}^{N_T} j_l$, where $x = \ln(S_0)$. Therefore

$$S_T = S_0 e^{aT + \sigma W_T} \prod_{l=1}^{N_T} (1 + J_l). \quad (5.15)$$

5.1.4 Bates Model

The Bates model [24] is the following combination of the Heston and the Merton models:

$$\begin{cases} dV_t = \lambda(\theta - V_t) dt + \eta \sqrt{V_t} dB_t, \\ \frac{dS_t}{S_{t-}} = (\kappa - \lambda \bar{J}) dt + \sqrt{V_t} dW_t + J_{(t)} dN_t. \end{cases} \quad (5.16)$$

5.1.5 Log-Spot Characteristic Functions in Affine Models

The risk-neutral log-spot characteristic function

$$\Phi_T(u) = \mathbb{E}[\exp(iuX_T)] = \mathbb{E}[S_T^{iu}],$$

where $X_t = \ln(S_t)$ and $i^2 = -1$, is explicitly known in the above models, which all belong to the class of affine jump-diffusions [103]. Vanilla option prices and Greeks can then be computed by the Fourier transform techniques of Sect. 5.5.

Proposition 5.1.4 *Let $x = X_0 = \ln(S_0)$, $v = V_0$ and*

$$\begin{aligned}\Phi_T^\kappa(u) &= \exp[iu(x + \kappa T)], \\ \Phi_T^j(u) &= \exp[-\lambda T(iu(e^{\alpha+\frac{\beta}{2}} - 1) - (e^{iu\alpha-u^2\frac{\beta}{2}} - 1))].\end{aligned}$$

In the Black–Scholes, Merton, Heston and Bates models we have:

$$\begin{aligned}\Phi_T^{bs}(u) &= \Phi_T^\kappa(u) \exp\left[-\frac{1}{2}u(i+u)\sigma^2T\right], & \Phi_T^{me}(u) &= \Phi_T^{bs}(u)\Phi_T^j(u), \\ \Phi_T^{he}(u) &= \Phi_T^\kappa(u) \exp[C(u, T)\theta + D(u, T)v], & \Phi_T^{ba}(u) &= \Phi_T^{he}(u)\Phi_T^j(u),\end{aligned}\tag{5.17}$$

where

$$C(u, T) = \lambda \left[Ty_- - \frac{2}{\eta^2} \ln\left(\frac{1 - ge^{-pT}}{1 - g}\right) \right], \quad D(u, T) = \frac{1 - e^{-pT}}{1 - ge^{-pT}} y_-, \tag{5.18}$$

in which

$$p = \sqrt{y^2 - 4wz}, \quad y_\pm = \frac{y \pm p}{\eta^2}, \quad g = \frac{y_-}{y_+}$$

for

$$w = -\frac{1}{2}u(i+u) = \frac{1}{2}ui(ui-1) = -\frac{u^2}{2} - \frac{ui}{2}, \quad y = \lambda - \rho\eta i u, \quad z = \frac{\eta^2}{2}.$$

Proof In the case of the Heston model, we introduce $F_t = S_t e^{-\kappa t}$ and we compute $\Phi_T^{he}(u)$ as

$$\mathbb{E}S_T^{ui} = e^{\kappa uiT} \mathbb{E}F_T^{ui} = e^{\kappa uiT} \Phi(0, F_0, V_0),$$

with $\Phi(t, F, V) := \mathbb{E}(F_T^{ui} | F_t = F, V_t = V)$. In particular, $\Phi(T, F, V) = F^{ui}$. We then seek an explicit solution Φ of the form

$$\Phi(t, F, V) = F^{ui} \exp[C(u, T-t)\theta + D(u, T-t)V],$$

for suitable coefficients $C = C(u, \tau)$, $D = D(u, \tau)$ such that, in particular,

$$C(u, 0)\theta + D(u, 0)V = 0.$$

Since $\Phi(t, F_t, V_t)$ is local martingale, by an application of the Itô formula (3.30) and Proposition 3.1.12 we obtain that

$$\partial_t \Phi + \mathcal{A}_{F,V} \Phi = 0,$$

where

$$\mathcal{A}_{F,V} = \frac{1}{2} V F^2 \partial_{F^2} + \frac{1}{2} \eta^2 V \partial_{V^2} + \rho \eta V F \partial_{FV} - \lambda (V - \theta) \partial_V.$$

This yields the following equations for (C, D) : for every real u , $C(u, 0) = D(u, 0) = 0$ and for $\tau > 0$:

$$\begin{aligned} -\theta \partial_\tau C - V \partial_\tau D + \frac{1}{2} V (ui)(ui - 1) + \frac{1}{2} \eta^2 V D^2 + \eta \rho V (ui) D \\ - \lambda (V - \theta) D = 0 \end{aligned} \tag{5.19}$$

or, equivalently,

$$\begin{cases} \partial_\tau C = \lambda D, \\ \partial_\tau D = w + \frac{1}{2} \eta^2 D^2 + \rho \eta ui D - \lambda D = w + z D^2 - y D = z(D - y_+)(D - y_-), \end{cases}$$

in which the second line is a Riccati equation in D . The reader may verify that the solution $C(u, \tau)$, $D(u, \tau)$ is given by (5.18) (with T replaced by τ). This proves the result in the case of the Heston model.

In the case of the Bates model, we introduce the process

$$L_t = \prod_{l=1}^{N_t} (1 + J_l) e^{-\lambda \bar{J}_t}, \quad dL_t = L_{t-d} \left(\sum_{l=1}^{N_t} J_l - \lambda \bar{J}_t \right) \tag{5.20}$$

with generator

$$\mathcal{A}_L \varphi(L) = -\lambda \bar{J}_L \partial_L \varphi + \lambda (\mathbb{E} \varphi((1 + J_1)L) - \varphi(L)).$$

Denoting the stock in the Heston model by S^{he} , by (5.12) and (5.20) we have that

$$d(S_t^{he} L_t) = S_t^{he} dL_t + L_{t-d} dS_t^{he} = S_t^{he} L_{t-d} \left(\kappa dt + \sqrt{V_t} dW_t + d \left(\sum_{l=1}^{N_t} J_l - \lambda \bar{J}_t \right) \right).$$

Hence, by the uniqueness result mentioned in Remark 5.1.3, $S^{he} L$ coincides with the stock S^{ba} in the Bates model (5.16). Thus, by independence, $\Phi_T^{ba}(u) = \Phi_T^{he}(u) \Phi_T^L(u)$. It remains to prove that $\Phi_T^L(u) := \mathbb{E} L_T^{ui} = \Phi_T^j(u)$. Note that

$$\mathcal{A}_L L^{ui} = -\lambda \bar{J}_L \partial_L L^{ui} + \lambda L^{ui} (\mathbb{E}[(1 + J_1)^{ui}] - 1).$$

By the Itô formula (3.30) we thus have, with “ $\dot{\equiv}$ ” standing for “equality up to a local martingale”:

$$dL_t^{ui} \dot{\equiv} L_t^{ui} \delta \lambda dt,$$

with

$$\delta = -\bar{J} ui + (\mathbb{E}[(1 + J_1)^{ui}] - 1) = -ui(e^{\alpha + \frac{\beta}{2}} - 1) + e^{ui\alpha - \frac{u^2\beta}{2}} - 1.$$

It follows that

$$\mathbb{E}L_t^{ui} = L_0^{ui} e^{\delta \lambda t},$$

i.e. $\mathbb{E}L_t^{ui} = \Phi_t^j(u)$.

Finally, setting $\lambda = \rho = 0$ and letting $\eta \rightarrow 0+$ yields

$$y = 0, \quad y_{\pm} = \frac{\pm p}{\eta^2}, \quad p = \sqrt{-2w}\eta, \quad 1 - g = 2,$$

$$D(u, T) \rightarrow \frac{pT}{1-g} y_- = \frac{-p^2 T}{2\eta^2} = wT,$$

and $C(u, T)\theta + D(u, T)v$ reduces to $-\frac{1}{2}u(i+u)vT$. Thus $\Phi_T^{he}(u)$ reduces to $\Phi_T^{bs}(u)$ for $\sigma = \sqrt{v}$. We can likewise check that, for $\alpha = \beta = 0$, $\Phi_T^{me}(u)$ reduces to $\Phi_T^{bs}(u)$. \square

5.2 Libor Market Model of Interest-Rate Derivatives

5.2.1 Black Formula

The Black formula extends the Black–Scholes formula to the case of stochastic interest rates. It is derived within a martingale pricing model with respect to the numéraire defined by the price process B^T of a T -discount bond. We let $\beta_t^T = 1/B_t^T$ and we call T -forward-neutral the measure \mathbb{P}^T associated with the numéraire B^T . From Sect. 4.3.1, the T -forward value

$$F_t^T = \beta_t^T S_t e^{-\int_t^T q(u) du} \tag{5.21}$$

of S , where $q(t)$ denotes a deterministic dividend yield on S , must be a \mathbb{P}^T -local martingale. Consistent with this arbitrage requirement, the Black T -forward-neutral model postulates that F_t^T is a lognormal process with a deterministic volatility process $\sigma(t)$ under \mathbb{P}^T . Since $B_T^T = 1$, the general numéraire pricing formula (4.62), applied with $\tilde{B} = B_t^T$ and $\tilde{\mathbb{P}} = \mathbb{P}^T$ yields

$$\Pi_t = B_t^T \mathbb{E}^{\mathbb{P}^T} (\xi | \mathcal{F}_t). \tag{5.22}$$

Substituting $\xi = (S_T - K)^+ = (F_T^T - K)^+$ into (5.23) and using the Black–Scholes formulas (5.10)–(5.11) with $r = q = 0$ to compute the expectation on the right-hand side of (5.22), we obtain:

$$\Pi_t := \Pi_t^{bl} = B_t^T \pi^{bl}(t, F_t^T, T, K; \sigma),$$

$$\delta_t^{bl} := \partial_F \pi^{bl}(t, F_t^T, T, K; \sigma) = \delta^{bl}(t, F_t^T, T, K; \sigma),$$

with

$$\pi^{bl}(t, F, T, K; \sigma) = F\mathcal{N}(d_+^{bl}) - K\mathcal{N}(d_-^{bl}), \quad \delta^{bl}(t, F, T, K; \sigma) = \mathcal{N}(d_+^{bl}), \quad (5.23)$$

where

$$d_{\pm}^{bl} = \frac{\ln(F/K)}{\Sigma} \pm \frac{1}{2}\Sigma \quad \text{for } \Sigma^2 = \int_t^T \sigma^2(u) du. \quad (5.24)$$

We also have the corresponding formulas deduced by call-put parity for a put option. The argument $(t, F, T, K; \sigma)$ will be abbreviated, when no confusion can arise, by (t, F, T, K) , or (t, F) , or sometimes even (as in d_{\pm}^{bl} above) omitted.

The delta δ_t^{bl} is a key ingredient of any hedging scheme for a vanilla call option in the Black model. First consider a primary market defined by the numéraire (T -discount bond) B_t^T , which is used as funding asset, along with a T -forward contract on S with strike price K . The relative price process $\beta_t^T \Theta_t$ of the T -forward contract on S is given by

$$\beta_t^T \Theta_t = \beta_t^T (S_t e^{-\int_t^T q(u) du} - K B_t^T) = F_t^T - K = \Theta(t, F_t^T).$$

So $\partial_F(\beta_t^T \Theta_t)(t, F_t^T) = 1$. By application of (4.57), a perfect replication strategy with initial wealth Π_0^{bl} at time 0 is thus given by

$$\tilde{\xi}_t^{bl} = \delta^{bl}(t, F_t^T) (\partial_F \beta_t^T \Theta_t)^{-1} = \delta^{bl}(t, F_t^T)$$

units of T -forward contracts on S and therefore

$$\beta_t^T (B_t^T \pi^{bl}(t, F_t^T) - \delta^{bl}(t, F_t^T) B_t^T (F_t^T - K)) = -K\mathcal{N}(d_-^{bl}(t, F_t^T)) + \delta^{bl}(t, F_t^T) K$$

units of the T -discount bond.

We now consider a primary market defined by the T -discount bond B_t^T and the stock $S_t = B_t^T F_t^T e^{\int_t^T q(u) du}$. By application of (4.57), with

$$\beta_t^T S_t = F_t^T e^{\int_t^T q(u) du}, \quad \partial_F(\beta_t^T S_t)(t, F_t^T) = e^{\int_t^T q(u) du},$$

a perfect replication strategy with initial wealth Π_0^{bl} for the European call option is defined, for every time $t \in [0, T]$, by

$$\xi_t^{bl} = e^{-\int_t^T q(u) du} \delta^{bl}(t, F_t^T)$$

units of S and therefore $(-K\mathcal{N}(d_-^{bl}(t, F_t^T)))$ units of the T -discount bond.

As a reality check note that, in the case of deterministic interest rates with B_t^T given as $e^{-\int_t^T r(u) du}$ for some function $r(t)$, so that $B_t^T = e^{-\int_0^T r(u) du} B_t$ and $\mathbb{P}^T = \mathbb{P}$, we have

$$F_t^T = S_t e^{\int_t^T (r(u) - q(u)) du}, \quad d_{\pm}^{bl}(t, F_t^T) = d_{\pm}^{bs}(t, S_t)$$

and we obtain, consistent with (5.9),

$$\tilde{\xi}_t^{bl} = \mathcal{N}(d_+^{bl}(t, F_t^T)) = \mathcal{N}(d_+^{bs}(t, S_t)) = \tilde{\xi}_t^{bs}.$$

The F^T -component of the replication strategy in the T -forward-neutral Black model with factor F^T and primary assets F^T and B^T is thus the same as the one in the risk-neutral Black–Scholes model with factor S and primary assets F^T and B . We likewise obtain, consistent with (5.8),

$$\xi_t^{bl} = e^{-\int_t^T q(u) du} \mathcal{N}(d_+^{bl}(t, F_t^T)) = e^{-\int_t^T q(u) du} \mathcal{N}(d_+^{bs}(t, S_t)) \xi_t^{bs},$$

so that the S -component of the replication strategy in the T -forward-neutral Black model with factor F^T and primary assets S and B^T is the same as the one in the Black–Scholes risk-neutral model with factor S and primary assets S and B .

5.2.2 Libor Market Model

Libor stands for London Inter Bank Offered Rates, which are daily published rates based on the interest rates at which a group of banks in London claim they would agree to borrow money from each other when needed.

Remark 5.2.1 Libor are of primary influence as underlyings to most vanilla interest-rate derivatives such as FRA, IRS, cap/floor and swaptions. However, the interbank loan market has been very severely impacted since the great crisis and the ensuing liquidity squeeze. Consequently, in parallel to the drying up of the interbank loan market, on the one hand, Libor got disconnected from OIS rates (overnight or money market rates and related derivatives); on the other hand, as more and more trades get collateralized, their effective funding rate is the corresponding collateral repo rate, which is typically indexed on OIS. This creates a situation where the price of an interest-rate product (even the simplest flow instrument such as an FRA) involves (at least) two curves, an OIS and a Libor curve, and the related convexity adjustment. In addition, the situation where an underlying to financial derivatives has become in a sense arbitrarily fixed by a panel of key players in the derivative market poses insider issues, as illustrated by the ongoing Libor affairs. We will not deal with any of these topics here and refer interested readers to [83, 84, 121, 201].

Libor are simply compounded interest rates. Given a tenor structure $0 \leq t_1 < \dots < t_{n+1}$ and the pricing time $t_0 = 0$, for $i = 1, \dots, n$ we write

$$h_i = t_{i+1} - t_i, \quad L_t^i = L_t(t_i, t_{i+1}), \quad B_t^i = B_t^{t_{i+1}}, \quad \mathbb{P}^i = \mathbb{P}^{t_{i+1}}, \quad \mathbb{E}^i = \mathbb{E}^{\mathbb{P}^i}.$$

Also, letting $B_t^0 = B_t^{t_1}$, for $i = 1, \dots, n$ and $t \leq t_i$, we thus have

$$h_i L_t^i = \frac{B_t^{i-1} - B_t^i}{B_t^i}, \quad B_t^i = \frac{1}{1 + h_i L_t^i} B_t^{i-1}. \quad (5.25)$$

Since $B_{t_{i+1}}^i = 1$, it follows by induction that, for every $1 \leq i \leq l + 1 \leq n + 1$,

$$B_{t_i}^l = \prod_{k=i}^l \frac{1}{1 + h_k L_{t_i}^k}. \quad (5.26)$$

The quantity $B_t^{i-1} - B_t^i$ is the difference between prices of traded assets, so that the ratio $h_i L_t^i = \frac{B_t^{i-1} - B_t^i}{B_t^i}$ must be a local martingale under the pricing measure \mathbb{P}^i . Consistent with this arbitrage requirement, the log-normal Libor Market Model (LMM, or BGM model in reference to Brace, Gatarek and Musiela [54]) postulates the following Black dynamics for the forward Libor L^i :

$$dL_t^i = \sigma_i(t) L_t^i dW_t^i, \quad \text{or} \quad L_t^i = L_0^i \exp\left(\int_0^t \sigma_i(s) dW_s^i - \frac{1}{2} \int_0^t \sigma_i^2(s) ds\right) \quad (5.27)$$

for some \mathbb{P}^i -Brownian motion W^i and some deterministic volatility function $\sigma_i(t)$. Since the (forward) Libor process L^i is stopped at t_i , the function $\sigma_i(t)$ vanishes after t_i . In particular, the variance of $\ln L_{t_{i+1}}^i$ is equal to

$$\Sigma_i^2(t) := \int_0^{t_i} \sigma_i^2(s) ds.$$

Remark 5.2.2 As is Black–Scholes on equities, the above log-normal interest rate derivatives model is of course misspecified. The market standard is to use smiled extensions of this basic setup in the form of SABR models, which are hybrid local volatility/stochastic volatility models with explicit asymptotics for the implied volatility and flexible smile dynamics [58, 137, 231].

5.2.3 Caps and Floors

A caplet is a call option with arrear settlement on a Libor. Considering the caplet with maturity t_i and strike K , the payoff to the caplet holder at t_{i+1} is

$$\Pi_{t_{i+1}}^i = h_i (L_{t_i}^i - K)^+.$$

Working in the numéraire B^i , by (5.22) we thus have, for $t \leq t_i$,

$$\Pi_t^i = h_i B_t^i \mathbb{E}_t^i (L_{t_i}^i - K)^+.$$

Hence by the log-normality postulated in (5.27):

$$\Pi_t^i = h_i B_t^i [L_t^i \mathcal{N}(d_+^i(t, L_t^i)) - K \mathcal{N}(d_-^i(t, L_t^i))], \quad \delta_t^i := \mathcal{N}(d_+^i(t, L_t^i)), \quad (5.28)$$

with

$$d_{\pm}^i(t, L_t^i) = \frac{\ln(\frac{L_t^i}{K})}{\Sigma_i} + \frac{\Sigma_i}{2} \quad \text{for } \Sigma_i^2(t) = \int_t^{t_i} \sigma_i^2(s) ds.$$

In view of Proposition 4.3.2 and the subsequent formulas, δ_t^i is the key to the hedging of a caplet in the log-normal Libor market model. So, in a primary market defined by the numéraire B^i and a forward swap over the period (t_i, t_{i+1}) , we can show, by application of (4.57), that a perfect replication strategy for the caplet with initial wealth Π_0^i is defined, for $t \in [0, t_i]$, by

$$\zeta_t = \delta_t^i$$

units of the forward swap over the period (t_i, t_{i+1}) and the ensuing number of t_{i+1} -discount bonds through the self-financing condition on the replicating portfolio.

We further get the price and delta of a cap (portfolio of caplets) with maturity t_1 and tenor structure t_1, \dots, t_{n+1} as

$$\Pi_t = \sum_{i=1}^n \Pi_t^i, \quad \delta_t = \sum_{i=1}^n \delta_t^i,$$

where Π_t^i and δ_t^i were defined in (5.28). The analogous formulas for a floor are obtained by cap/floor parity. Due to the additive structure of the payoffs, the prices of caps and floors depend only on the marginal laws of the L^i at tenor dates t_i , not on the Libor correlation.

5.2.4 Adding Correlation

As opposed to caps and floors, swaptions are sensitive to the correlation structure of L .² We now define a correlation structure between the L^i by expressing their joint dynamics under the so-called terminal measure \mathbb{P}^n . By definition of \mathbb{P}^i , for $t \leq t_i$ we have (cf. (4.61)):

$$v_t^i := \frac{d\mathbb{P}^{i-1}}{d\mathbb{P}^i} \Big|_{\mathcal{F}_t} = \mathbb{E}^i \left[\frac{d\mathbb{P}^{i-1}}{d\mathbb{P}^i} \mid \mathcal{F}_t \right] = \frac{B_0^i B_t^{i-1}}{B_0^{i-1} B_t^i} = \frac{B_0^i}{B_0^{i-1}} (1 + h_i L_t^i). \quad (5.29)$$

We write, in vector-form,

$$dL_t^i = s_i(t) L_t^i d\mathbb{W}_t^i = \sigma_i(t) L_t^i dW_t^{i,i} \quad (5.30)$$

for the row-vector volatility function $s_i(t) = (0, \dots, 0, \sigma_i(t), 0, \dots, 0)$ and for an n -dimensional \mathbb{P}^i -Brownian motion $\mathbb{W}^i = (W^{l,i})_{1 \leq l \leq n}$ with correlation matrix $\rho =$

²Swaptions are also very sensitive to the term-structure of the volatility, whereas caps and floors are only sensitive to the integrated variance; see Rebonato [230].

$(\rho_{l,i})_{1 \leq i, l \leq n}$. In order to derive the \mathbb{P}^n -dynamics of the L^i , we proceed by backward induction over i , starting from $dL_t^n = \sigma_n(t)L_t^n dW_t^{n,n}$. We next define \mathbb{W}_t^{n-1} as $\mathbb{W}_t^n - \rho \int_0^t \mu_u^\top du$, for some row-vector process μ . By the Girsanov theorem, in order that \mathbb{W}^{n-1} thus defined to be a \mathbb{P}^{n-1} -Brownian motion on $[0, t_n]$, it is sufficient that μ satisfies $d\nu_t^n = \mu_t \nu_t^n d\mathbb{W}_t^n$, where ν_t^n is the \mathcal{F}_t -measurable Radon-Nikodym density of \mathbb{P}^{n-1} with respect to \mathbb{P}^n on \mathcal{F}_t . Now, by (5.29)–(5.30), we have

$$d\nu_t^n = \nu_t^n \frac{h_n L_t^n s_n(t)}{1 + h_n L_t^n} d\mathbb{W}_t^n.$$

We thus set $\mu_t = \frac{h_n L_t^n s_n(t)}{1 + h_n L_t^n}$. Hence

$$\begin{aligned} dL_t^{n-1} &= s_{n-1}(t)L_t^{n-1} \left(d\mathbb{W}_t^n - \frac{h_n L_t^n \rho s_n^\top(t)}{1 + h_n L_t^n} dt \right) \\ &= \sigma_{n-1}(t)L_t^{n-1} dW_t^{n-1,n} - \frac{h_n L_t^n \sigma_n(t) \rho_{n,n-1}}{1 + h_n L_t^n} \sigma_{n-1}(t)L_t^{n-1} dt. \end{aligned}$$

For every $i = 1, \dots, n$, we likewise define

$$dL_t^i = \sigma_i(t)L_t^i dW_t^{i,n} - \sum_{l=i+1}^n \frac{h_l L_t^l \sigma_l(t) \rho_{l,i}}{1 + h_l L_t^l} \sigma_i(t)L_t^i dt \quad (5.31)$$

or, in log-returns,

$$d\ln(L_t^i) = \sigma_i(t) dW_t^{i,n} - \left(\sum_{l=i+1}^n \frac{h_l L_t^l \sigma_l(t) \rho_{l,i}}{1 + h_l L_t^l} \sigma_i(t) + \frac{1}{2} \sigma_i(t)^2 \right) dt. \quad (5.32)$$

Then L^i is a \mathbb{P}^i -martingale for every $i \leq n$. Note that for $i < n$, L^i has a non vanishing \mathbb{P}^n -drift that depends on the L^l for $l > i$.

5.2.4.1 Correlation Structures

A first possibility is to set ρ as an historical estimate of the correlation matrix of the Libor. But an historical correlation matrix is difficult to estimate in practice. A generally preferred alternative (see Chap. 9) is to calibrate a parametric form of ρ to market quotes of swaptions. Various parameterizations of ρ are classically used in this calibration, such as $\rho_{i,l} = \exp(-\gamma|i-l|)$, or

$$\rho_{i,l} = \rho_\infty + (1 - \rho_\infty) \exp[-|t_i - t_l| \gamma(t_i, t_l)], \quad (5.33)$$

with $\gamma(t_i, t_l) = \gamma_1 - \gamma_2 \max(t_i, t_l)$, or

$$\rho_{i,l} = \exp \left[-\frac{|i-l|}{n-1} (-\ln \nu_\infty + \eta_1 \varphi(i, l, n) + \eta_2 \psi(i, l, n)) \right], \quad (5.34)$$

with

$$\varphi(i, l, n) = \frac{i^2 + l^2 + il - 3ni - 3nl + 3i + 3l + 2n^2 - n - 4}{(n-2)(n-3)}$$

$$\psi(i, l, n) = \frac{i^2 + l^2 + il - ni - nl - 3i - 3l + 3n + 2}{(n-2)(n-3)}.$$

Note, however, that the formula (5.33) can fail to define a correlation matrix for some values of its parameters. The formula (5.34) is known to produce a correlation matrix provided $0 \leq \eta_2 \leq 3\eta_1$ and $0 \leq \eta_1 + \eta_2 \leq -\ln v_\infty$ (see [58]).

5.2.5 Swaptions

An interest rate swap with tenor dates t_1, \dots, t_{n+1} and strike K is a contract with a cash flow $h_i(L_{t_i}^i - K)$ at each time t_{i+1} , for $i = 1, \dots, n$. By (5.22), the value of the swap at time $t \leq t_1$ is given by

$$\sum_{i=1}^n h_i B_t^i \mathbb{E}_t^i (L_{t_i}^i - K) = \sum_{i=1}^n h_i B_t^i (L_t^i - K).$$

The related swaption is an option to enter this swap at the maturity time t_1 . The swaption's payoff is given by the positive part of

$$\sum_{i=1}^n B_{t_1}^i h_i (L_{t_1}^i - K)^+$$

at t_1 , i.e.

$$\left(\sum_{i=1}^n h_i B_{t_1}^i \right) (S_{t_1} - K)^+,$$

where the forward swap rate S_t is defined, for $t \leq t_1$, by

$$S_t = \frac{B_t^0 - B_t^n}{\sum_{i=1}^n h_i B_t^i}.$$

Hence

$$\sum_{i=1}^n h_i B_{t_1}^i L_{t_1}^i = B_{t_1}^0 - B_{t_1}^n = \sum_{i=1}^n h_i B_{t_1}^i S_{t_1}$$

by (5.25). The process S_t is thus a local martingale under the pricing measure \mathbb{Q} associated with the numéraire $N_t = \sum_{i=1}^n h_i B_t^i$. Consistent with this requirement, the swap market model (also known as the Libor Swap Model) consists in modeling the swap rate S in Black form under \mathbb{Q} ; exact Black formulas for swaptions prices and deltas result.

In the Libor market model, the swap rate S_{t_1} is not lognormal under \mathbb{Q} . But it is numerically close to \mathbb{Q} -lognormality for an integrated squared variance Σ_0^2 given by the following Rebonato's formula (see p. 248 of Brigo and Mercurio [58]):

$$\Sigma_0^2 = \frac{1}{S_0^2} \sum_{i,l=1}^n w^i w^l L_0^i L_0^l \rho_{i,l} \int_0^{t_1} \sigma_i(t) \sigma_l(t) dt \quad (5.35)$$

for weights w^l proportional to the $h_l B_0^l$. One thus has the following approximate time-0 price-and-delta formulas for a swaption in the LMM:

$$\Pi_0 = N_0(S_0 \mathcal{N}(d_+) - K \mathcal{N}(d_-)), \quad \Delta_0 = \mathcal{N}(d_+),$$

with

$$d_{\pm} = \frac{\ln(\frac{S_0}{K})}{\Sigma_0} \pm \frac{\Sigma_0}{2},$$

where Σ_0 is given by (5.35).

5.2.6 Model Simulation

Interest rate derivatives cashflows are typically given as functions ϕ of the $L_{t_j}^i$. One thus has for a cap, by (4.62) and (5.26),

$$\begin{aligned} \Pi_0 &= B_0^n \sum_{i=1}^n \mathbb{E}^n \left[\frac{h_i (L_{t_i}^i - K)^+}{B_{t_{i+1}}^n} \right] \\ &= B_0^n \mathbb{E}^n \left[\sum_{i=1}^n h_i (L_{t_i}^i - K)^+ \prod_{l=i+1}^n (1 + h_l L_{t_{l+1}}^l) \right]. \end{aligned}$$

To properly discount each cash flow $h_i (L_{t_i}^i - K)^+$ under \mathbb{P}^n , one thus needs to know the values of the $L_{t_{i+1}}^l$ with $l > i$.

For pricing and Greeking by Monte Carlo in the LMM, we can simulate the forward Libor under \mathbb{P}^n by an Euler scheme on a time-grid refining the tenor structure (see Sect. 6.8.2.1).

Remark 5.2.3 Hull and White report in [147] that, for standard tenor lengths such as $h_i = 3$ m to 6 m, discretizing (5.32) by an Euler scheme at tenor dates is accurate enough. This at least holds provided t_0 is close enough to t_1 ; otherwise, of course, the discretization grid must be refined between t_0 and t_1 .

Starting from L_0 , deduced from $B_0 = (B_0^i)_{0 \leq i \leq n}$ by (5.26), we thus obtain for every $l = 0, \dots, n-1$ and $i = l+1, \dots, n$:

Table 5.1 Trajectory simulated in a one-factor LMM

t	0	t_1	t_2	t_3	t_4
$\sqrt{h_l} \varepsilon_l$	-0.371379	1.81768	-0.204069	0.512108	
L^1	5 %	4.698 %			
L^2	5 %	4.699 %	6.135 %		
L^3	5 %	4.701 %	6.138 %	5.918 %	
L^4	5 %	4.702 %	6.142 %	5.923 %	6.36 %

$$L_{t_{l+1}}^i = L_{t_l}^i \exp \left[\sigma_i(t_l) \sqrt{h_l} (\Lambda \varepsilon^l)_i - \left(\sum_{k=i+1}^n \frac{h_k L_{t_l}^k \sigma_k(t_l) \rho_{k,i}}{1 + h_k L_{t_l}^k} \sigma_i(t_l) + \frac{1}{2} \sigma_i^2(t_l) \right) h_l \right],$$

where ε^l represents a vector of n independent Gaussian random variables, and the matrix Λ is a square-root of ρ , so that $\Lambda \Lambda^\top = \rho$ (a square-root Λ of ρ can be obtained, as in item (i) below, by Cholesky decomposition of ρ ; see Sect. 6.2.3).

Example 5.2.4

- (i) In the case $n = 2$:

$$\begin{aligned} \mathbb{W}^2 &= \begin{pmatrix} W^{1,2} \\ W^{2,2} \end{pmatrix}, \\ \frac{d\langle W^{1,2}, W^{2,2} \rangle_t}{dt} &= \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \Lambda \Lambda^\top \quad \text{with } \Lambda = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}. \end{aligned}$$

- (ii) In the case of a rank-one correlation structure of the form (5.41) that we will encounter in a different context in Sect. 5.3.2, one standard univariate Gaussian draw ε_l per time step is enough to simulate a model trajectory. Table 5.1 shows one trajectory of the Libor thus simulated for $n = 4$, $\sigma_i = 15\%$, $h_i = 0.5$, and $L_0^i = 5\%$ (flat initial term structure at the level 5 %).

5.3 One-Factor Gaussian Copula Model of Portfolio Credit Risk

We now move to portfolio credit derivatives. The market model in this regard is still the static one-factor Gaussian copula model,³ first introduced in Li [188].

Remark 5.3.1 It was said that the Gaussian copula model was responsible for huge losses on CDOs in the 2007–2009 credit crisis. Of course, from the theoretical point of view the shortcomings of the model are quite obvious; in particular this is a static quotation device, rather than a full-flesh dynamic model (even though it is possi-

³In combination with stochastic recoveries since the 2007–2009 credit crisis.

ble to dynamize this setup by the introduction of a relevant filtration [85, 120]). However, the biggest losses on securitization products in the credit crisis occurred not on synthetic CDOs that were actively risk managed by banks, as explained in Sect. 5.4.2 with Gaussian copula deltas, but on cash CDOs (often in the form of ABS held by nonbanking institutions) that were simply not risk-managed. It is thus not the Gaussian copula model which is so much to blame here, but rather the regulation, which made it possible for banks to externalize such risks to institutions not committed to hedge them.

One considers d reference entities (firms) with respective default times and loss-given-defaults denoted by τ_l and Λ_l , for $l = 1, \dots, d$. The cumulative portfolio loss at time t is given by

$$\mathcal{L}_t = \sum_{l=1}^d \Lambda_l \mathbb{1}_{\tau_l \leq t}.$$

5.3.1 Credit Derivatives

5.3.1.1 Single-Name CDSs

A single-name credit default swap (CDS for short) on name l has cumulative discounted cash flows given by, from the point of view of the seller of default protection:

$$\int_0^T \beta_t (S J_t^l dt + \Lambda_l dJ_t^l),$$

where J^l is the survival indicator process of name l . The related price process, for $t \in [0, T]$, is written

$$\mathbb{E}_t \int_t^T \beta_s (S J^l ds + \Lambda_l dJ_s^l).$$

The contractual CDS spread S is typically set such that the CDS is entered at no cost at inception so that, assuming deterministic interest rates $r_t = r(t)$,

$$S = \frac{\Lambda_l \int_0^T \beta_t dF_l(t)}{\int_0^T \beta_t (1 - F_l(t)) dt},$$

where $F_l(t) = \mathbb{P}(\tau_l \leq t)$.

5.3.1.2 CDO Tranches

A single tranche Collateralized Debt Obligation (CDO) with attachment point a , detachment point b and maturity T is an option with the following cumulative discounted cash flows, from the point of view of a seller of default protection:

$$\int_0^T \beta_t [\Sigma(b - a - L_t) dt - dL_t],$$

where

$$L_t = (\mathcal{L}_t - a)^+ - (\mathcal{L}_t - b)^+ = \min((\mathcal{L}_t - a)^+, b - a)$$

is the cumulative tranche loss and Σ is the tranche contractual spread. A CDO tranche can thus be interpreted as a call-spread with strikes a and b on the portfolio loss \mathcal{L}_t . For instance, on the DJ iTraxx market, which is a family of CDS indices for Europe and Asia, CDO tranches were liquidly quoted until May 2009 for (a, b) in $(0\%, 3\%)$, $(3\%, 6\%)$, $(6\%, 9\%)$, $(9\%, 12\%)$ and $(12\%, 22\%)$. Since the crisis only the indices are still quoted, no longer the tranches. However CDOs are still relevant in terms of risk-management as they are still present in bank portfolios (see Remark 4.1.6). The tranches $(0\%, 3\%)$ and $(9\%, 12\%)$ are respectively called the equity and the senior tranche, whereas intermediate tranches are known as the mezzanine tranches.

The price process of a tranche is given, for $t \in [0, T]$, by

$$\mathbb{E}_t \int_t^T \beta_s [\Sigma(b - a - L_s) ds - dL_s]. \quad (5.36)$$

The contractual tranche spread Σ is typically set such that the tranche is entered at no cost at inception, so that

$$\Sigma = \frac{\mathbb{E} \int_0^T \beta_t dL_t}{\mathbb{E} \int_0^T \beta_t (b - a - L_t) dt}.$$

Assuming deterministic interest rates $r_t = r(t)$, the values of either leg of the CDO only depends on the expected tranche losses $\mathbb{E}L_t$, $t \in [0, T]$. Indeed, one thus has for the fees leg:

$$\mathbb{E} \int_0^T \beta_t (b - a - L_t) dt = \int_0^T \beta_t (b - a - \mathbb{E}L_t) dt. \quad (5.37)$$

For the protection leg, observe that $d(\beta_t L_t) = \beta_t (dL_t - r_t L_t dt)$ and $L_0 = 0$. Hence Fubini's theorem yields

$$\mathbb{E} \int_0^T \beta_t dL_t = \beta_T \mathbb{E}L_T + \int_0^T r_t \beta_t \mathbb{E}L_t dt. \quad (5.38)$$

5.3.2 Gaussian Copula Model

A copula function C is the joint cumulative distribution function of an \mathbb{R}^d -valued random vector with uniform marginals on $[0, 1]$. Thus, in particular, $C(1, \dots, 1, u)$,

$1, \dots, 1) = u$ for every $u \in [0, 1]$. Sklar's theorem states that for every joint multivariate cumulative distribution function $F(t_1, \dots, t_d)$ with marginal cumulative distribution functions $F_1(t_1), \dots, F_d(t_d)$, there exists a copula function C such that, for any t_1, \dots, t_d ,

$$F(t_1, \dots, t_d) = C[F_1(t_1), \dots, F_d(t_d)]. \quad (5.39)$$

In the one-factor Gaussian copula model [179, 188], it is postulated that, on top of some marginal cumulative distribution functions F_l , the dependence structure between the τ_l is defined by the one-factor Gaussian copula

$$C_\rho(u_1, \dots, u_d) = \mathcal{N}_\rho[\mathcal{N}^{-1}(u_1), \dots, \mathcal{N}^{-1}(u_d)], \quad (5.40)$$

where \mathcal{N} and \mathcal{N}_ρ respectively represent the standard univariate Gaussian cumulative distribution function and the d -variate Gaussian cumulative distribution function with covariance matrix

$$\begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & 1 & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}. \quad (5.41)$$

Let the τ_l follow a Gaussian copula model. For $l = 1, \dots, d$ and $t \geq 0$, we define

$$X_l = \mathcal{N}^{-1}(F_l(\tau_l)), \quad x_t^l = \mathcal{N}^{-1}(F_l(t)).$$

Thus $\tau_l = F_l^{-1}(\mathcal{N}(X_l))$, and $\tau_l \leq t$ if and only if $X_l \leq x_t^l$. Using (5.39)–(5.40), it follows that, for every t_1, \dots, t_d ,

$$\begin{aligned} \mathbb{P}(X_1 \leq x_{t_1}^1, \dots, X_d \leq x_{t_d}^d) &= \mathbb{P}(\tau_1 \leq t_1, \dots, \tau_d \leq t_d) = F(t_1, \dots, t_d) \\ &= \mathcal{N}_\rho(x_{t_1}^1, \dots, x_{t_d}^d). \end{aligned}$$

Therefore $X = (X_1, \dots, X_d)$ is a Gaussian vector with covariance matrix given by (5.41). We thus have the following representation in law: for $l = 1, \dots, d$,

$$X_l = \sqrt{\rho}Y + \sqrt{1 - \rho}Y_l \quad (5.42)$$

for an independent standard Gaussian common factor Y and independent Gaussian random variables Y_l . The X_l and therefore the τ_l are conditionally independent given Y . For $l = 1, \dots, d$, let

$$p_{\tau_l}^{l|y} := \mathbb{P}(\tau_l \leq t | Y = y) = \mathbb{P}(X_l \leq x_{t_l}^l | Y = y) = \mathcal{N}\left(\frac{x_{t_l}^l - \sqrt{\rho}y}{\sqrt{1 - \rho}}\right). \quad (5.43)$$

By conditional independence

$$F(t_1, \dots, t_d) = \mathbb{E}[\mathbb{P}[X_1 \leq x_{t_1}^1, \dots, X_d \leq x_{t_d}^d | Y]] = \int_{-\infty}^{\infty} \prod_{l=1}^d p_t^{l|y} g(y) dy,$$

where g is the standard Gaussian density. The moment generating function $\Psi_{\mathcal{L}_t}(u) = \mathbb{E}[e^{u\mathcal{L}_t}]$ of the portfolio loss \mathcal{L}_t is given by (with $\Psi_{\mathcal{L}_t}^Y(u) = \mathbb{E}[e^{u\mathcal{L}_t} | Y]$)

$$\begin{aligned} \Psi_{\mathcal{L}_t}(u) &= \mathbb{E}_Y \Psi_{\mathcal{L}_t}^Y(u) = \mathbb{E} \left[\prod_{l=1}^d \mathbb{E}[e^{u\Lambda_l \mathbb{1}_{\tau_l \leq t}} | Y] \right] \\ &= \int_{-\infty}^{\infty} \prod_{l=1}^d (1 - p_t^{l|y} + p_t^{l|y} e^{u\Lambda_l}) g(y) dy. \end{aligned} \quad (5.44)$$

5.3.2.1 Exact CDO Pricing Schemes

We assume in this subsection that the loss-given-defaults Λ_l are commensurate or, more specifically and without loss of generality, natural numbers. With $q_t^k = \mathbb{P}(\mathcal{L}_t = k)$ and $\Lambda = \sum_{l=1}^d \Lambda_l$ we thus have:

$$\Psi_{\mathcal{L}_t}(u) = \sum_{k=0}^{\Lambda} q_t^k e^{uk}, \quad \mathbb{E}L_t = \sum_{k=0}^{\Lambda} ((k-a)^+ \wedge (b-a)) q_t^k. \quad (5.45)$$

The expected tranche loss $\mathbb{E}L_t$ is thus a function of the portfolio loss distribution $q_t = (q_t^k)_{0 \leq k \leq \Lambda}$. This distribution q_t can be computed by numerical inversion of the Laplace transform $\Psi_{\mathcal{L}_t}(u)$. Choosing $(\Lambda + 1)$ as a power of 2, this inversion can be done in time $O(\Lambda \ln \Lambda)$ by the fast Fourier transform (FFT) (see end of Sect. 5.5.3).

Alternatively, the portfolio loss probability distribution q_t can be computed recursively as follows. Let $q_t^{k|y}$ and $q_t^{k|y}(i)$ respectively represent $\mathbb{P}(\mathcal{L}_t = k | y)$ and the conditional probability that the aggregated loss over the i first names of the portfolio equals k , so that $q_t^{k|y} = q_t^{k|y}(d)$.

Lemma 5.3.2 *We have $q_t^{k|y}(0) = \mathbb{1}_{k=0}$ and, for every $l = 1, \dots, d$ and $k = 0, \dots, \Lambda$:*

$$q_t^{k|y}(l) = p_t^{l|y} q_t^{k-\Lambda_l|y}(l-1) + (1 - p_t^{l|y}) q_t^{k|y}(l-1). \quad (5.46)$$

Proof Let $q_t^{k,\epsilon|y}(i)$, for $\epsilon = 0, 1$, represent the conditional probability that the aggregated loss over the i first names of the portfolio equals k and $\mathbb{1}_{\tau_i \leq t} = \epsilon$. We have

$$\begin{aligned}
q_t^{k|y}(l) &= q_t^{k, I_t^l=1|y}(l) + q_t^{k, I_t^l=0|y}(l) \\
&= q_t^{k-\Lambda_l, I_t^l=1|y}(l-1) + q_t^{k, I_t^l=0|y}(l-1) \\
&= q_t^{k-\Lambda_l|y}(l-1)p_t^{l|y} + q_t^{k|y}(l-1)(1-p_t^{l|y}),
\end{aligned}$$

where the first two identities are elementary and the last one follows by conditional independence with respect to Y . \square

Once the conditional loss distribution $q_t^{k|y} = q_t^{k|y}(d)$ has been recursively computed using (5.46), by numerical integration we recover

$$q_t^k = \int_{-\infty}^{\infty} q_t^{k|y} g(y) dy.$$

Recursive relations analogous to (5.46) can also be derived for the sensitivities of the tranches with respect to the input data $F_l(t)$. Following Andersen and Sidenius [7], we first compute $\partial_{F_l(t)} \mathbb{E}(L_t | Y)$ from

$$(\partial_{p_t^{l|y}} \mathbb{E}(L_t | Y)) (\partial_{F_l(t)} p_t^{l|y}) = (\partial_{p_t^{l|y}} \mathbb{E}(L_t | Y)) \frac{\partial_{x_t^l} p_t^{l|y}}{\partial_{x_t^l} F_l(t)}. \quad (5.47)$$

Now, by (5.43),

$$\partial_{x_t^l} p_t^{l|y} = \frac{1}{\sqrt{2\pi(1-\rho)}} \exp\left(-\frac{(x_t^l - \sqrt{\rho}Y)^2}{2(1-\rho)}\right), \quad \partial_{x_t^l} F_l(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_t^l)^2}{2}\right).$$

Moreover, by (5.45),

$$\partial_{p_t^{l|y}} \mathbb{E}(L_t | Y) = \sum_{k=0}^A ((k-a)^+ \wedge (b-a)) \partial_{p_t^{l|y}} q_t^{k|y}, \quad (5.48)$$

in which, by (5.46),

$$\partial_{p_t^{l|y}} q_t^{k|y} = q_t^{k-\Lambda_l|y}(\{l\}) - q_t^{k|y}(\{l\}),$$

where $q_t^{k|y}(\{l\})$ denotes the conditional distribution of the aggregated loss over the portfolio deprived of name l , which can be computed recursively along the same lines as (5.46). We finally recover the unconditional sensitivity $\partial_{F_l(t)} \mathbb{E} L_t$ by numerical integration with respect to the Gaussian density $g(y)$.

5.3.2.2 Approximate CDO Pricing Schemes

Fast approximate schemes may also be used to compute the portfolio loss distribution q_t . Moreover, these approximate schemes do not require the assumption of commensurate losses.

We refer the reader to El Karoui and Jiao [110, 111] for an efficient, easy to implement and mathematically justified approach based on Gauss-Poisson approximations of the portfolio conditional loss distributions. Related yet heuristic and harder to implement saddle-point methods (see [259]) are based on the following inverse Laplace transform representation for $q_t^{x|y} = \frac{\mathbb{P}(\mathcal{L}_t \in dx | Y=y)}{dx}$, in a weak sense as explained below:

$$q_t^{x|y} = \frac{1}{2\pi i} \int_{\eta-i\infty}^{\eta+i\infty} \Psi_{\mathcal{L}_t}^y(u) e^{-ux} du, \quad (5.49)$$

where $\Psi_{\mathcal{L}_t}^y(u) = \mathbb{E}(e^{u\mathcal{L}_t} | Y=y)$. The integration is parallel to the imaginary axis in the complex plane, with $\eta > 0$. By (5.49) in the weak sense, we mean that⁴

$$\mathbb{E}[\varphi(\mathcal{L}_t) | Y] = \frac{1}{2\pi i} \int_{\eta-i\infty}^{\eta+i\infty} \left(\int_0^\infty e^{-ux} \varphi(x) dx \right) \Psi_{\mathcal{L}_t}^Y(u) du, \quad (5.50)$$

for every test-function φ . Noting that $\partial_x[(ux + 1)e^{-ux}] = -u^2 xe^{-ux}$, so $\int_0^\infty xe^{-ux} dx = \frac{1}{u^2}$, we obtain by application of (5.50) to $\varphi(\mathcal{L}_t) = (\mathcal{L}_t - a)^+$:

$$\mathbb{E}[(\mathcal{L}_t - a)^+ | Y] = \frac{1}{2\pi i} \int_{u=\eta-i\infty}^{\eta+i\infty} \Psi_{\mathcal{L}_t}^Y(u) e^{-ua} u^{-2} du. \quad (5.51)$$

Saddle-point methods are then based on the approximation of $\Psi_{\mathcal{L}_t}^Y(u) e^{-ua}$ in (5.51) by suitable Taylor expansions around a well chosen point u^* , so that the resulting integral can be computed explicitly. Depending on the expansion point u^* and the order of the expansion, we obtain a whole family of approximate pricing schemes. In the simplest case we recover the large portfolio approximation of Vasicek [252].

A last possibility for computing the portfolio loss distribution q_t , or the value of a CDO tranche directly, is to proceed by Monte Carlo simulation. But simulation methods are much slower on these problems than any of the previous procedures: Gauss–Poisson or saddle-point approximations, or even, assuming commensurate A_I , exact fast Fourier transform or recursive schemes. Note that the integrals in all these algorithms, all of which involve the Gaussian kernel $g(y) dy$, can be computed efficiently by Gauss–Hermite quadrature.

5.4 Benchmark Models in Practice

5.4.1 Implied Parameters

The reader must not misunderstand the meaning of the previous “benchmark models”. In fact, the Black–Scholes or the one-factor Gaussian copula pricing formulas

⁴See also (5.60).

are essentially used by traders for conveying information about the relative value of different options in the market. The idea is to express prices in a unit of measurement, implied volatility or implied correlation, that is less sensitive to the strike and maturity of an option than its money-value. These formulas are thus no more than “wrong formulas into which to put a wrong number (the implied volatility of an option or the implied correlation of a CDO tranche) to get the right result (an option market price of a CDO tranche market spread)”.

5.4.1.1 Black–Scholes Implied Volatility

Black–Scholes formulas are effectively used in the reverse-engineering mode for determining, given a European vanilla price observed on the market, the corresponding value of the Black–Scholes volatility consistent with that option price. Given values of r and q inferred at time t , from riskless bonds for r and from call-put parity for q , the Black–Scholes implied volatility of a European vanilla option is the value σ_t such that

$$\Pi^{bs}(t, S_t, T, K; r, q, \sigma_t) = \Pi_t^{ma}(T, K), \quad (5.52)$$

where $\Pi_t^{ma}(T, K)$ denotes the market price of the option at time t . Since the Black–Scholes price of a vanilla option is increasing in σ , (5.52) yields a unique σ_t provided the market price lies within the related arbitrage bounds. Moreover, (5.52) is easy to solve numerically, by dichotomy or Newton–Raphson zero search. Proceeding in this way for a range of strikes K and a fixed maturity T , one typically obtains:

- a symmetrical smile on foreign exchange derivative markets;
- a negative skew on equity derivative markets;
- a smirk on interest rate derivative markets.

One can also observe a positive skew on “negative beta” derivative markets (see Fig. 5.1 in the case of options on gold futures).

5.4.1.2 Gaussian Copula Implied Correlation

The one-factor Gaussian copula model is also used in the reverse-engineering mode for quoting CDO tranches in terms of their implied correlations. A preliminary step consists in inferring the cumulative distribution functions F_l from the respective CDS markets, based on the following pricing equations (assuming deterministic interest rates; see Sect. 5.3.1.1):

$$S_t^l(T) \int_t^T \beta_s (1 - F_l(s)) ds - \Lambda_l \int_t^T \beta_s dF_l(s) = 0, \quad (5.53)$$

where $S_t^l(T)$ denotes the fair spread at time t of a CDS with maturity T on name l of the pool. This relation allows us to bootstrap the function F_l (for $s \geq t$) from the

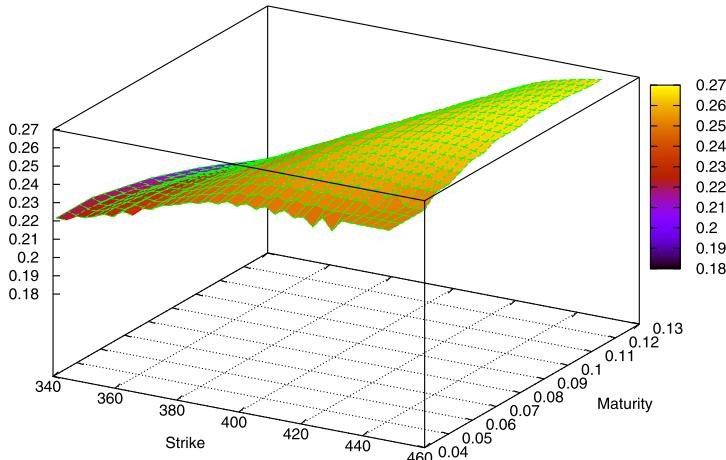


Fig. 5.1 “Golden smile”: an implied volatility surface on gold futures

market CDS curve observed on name l at time t . Then, given $F = (F_l(s))_{1 \leq l \leq d}$ for $s \geq t$:

- the compound implied correlation of a tranche is defined as the value of the correlation $\hat{\rho}_t$ in a one-factor Gaussian copula model such that

$$\Sigma^{gc}(t, F, T, a, b; \hat{\rho}_t) = \Sigma_t^{ma}(T, a, b),$$

where $\Sigma_t^{ma}(T, a, b)$ denotes the market tranche spread at time t and Σ^{gc} is the Gaussian copula CDO tranches pricing function;

- the base implied correlation of a tranche is defined as the value of the correlation ρ_t in a one-factor Gaussian copula model such that

$$\Sigma^{gc}(t, F, T, 0, b; \rho_t) = \Sigma_t^{ma}(T, 0, b),$$

where $\Sigma_t^{ma}(T, 0, b)$ denotes a synthetic market spread reconstructed from the observed market spreads of the tranches with detachment point b and below (see O’Kane and Livesey [215]).

The base implied correlation is more stable numerically than the compound implied correlation, because the function Σ^{gc} is decreasing in its argument ρ for $a = 0$, but not for $a > 0$ (see [215]).

5.4.2 Implied Delta-Hedging

We now explain how models as simple as Black–Scholes or the Gaussian copula are commonly used in the implied mode for delta-hedging.

Remark 5.4.1 Apart from serving as benchmark models on respective derivative markets, the Black–Scholes model and the Gaussian copula setup have of course little in common. However, it is possible to draw some analogies (and note some obvious differences) between the theory and practice of delta-hedging in the two models; see [73, 79].

5.4.2.1 Implied Delta-Hedging with the Black–Scholes Model

We consider the problem of delta-hedging a European option with maturity T and payoff ξ on an underlying S . A bank sells the option at price Π_0 at time 0 and must pay the payoff ξ at time T . The hedging strategy of the bank consists in rebalancing, at every step of a time-grid $t_i = ih$, $0 \leq i \leq n - 1$, a self-financing hedge in the underlying and in the savings account. We assume that dividends on S are paid and kept as new stock shares falling at yield q in the hedging portfolio. We are thus considering a hedging strategy of the form $\zeta_t = \zeta_{t_i} e^{q(t-t_i)}$ on $(t_i, t_{i+1}]$, for $i = 0, \dots, n - 1$. So, recalling (5.4),

$$\zeta_t d(\beta_t \widehat{S}_t) = \zeta_{t_i} e^{-qt_i} d(\beta_t S_t e^{qt}). \quad (5.54)$$

By (4.18), the discounted P&L of the trader at T is (assuming a nondividend-paying option, so that $\Pi = \widehat{\Pi}$) therefore

$$\beta_T e_T = \Pi_0 + \sum_{i=0}^{n-1} e^{-qi\hbar} \zeta_{i\hbar} (\beta_{(i+1)\hbar} S_{(i+1)\hbar} e^{q(i+1)\hbar} - \beta_{i\hbar} S_{i\hbar} e^{qi\hbar}) - \beta_T \xi = \sum_{i=0}^{n-1} \beta_{i\hbar} \delta_i e, \quad (5.55)$$

where

$$\beta_{i\hbar} \delta_i e = -(\beta_{(i+1)\hbar} \Pi_{(i+1)\hbar} - \beta_{i\hbar} \Pi_{i\hbar}) + e^{-qi\hbar} \zeta_{i\hbar} (e^{-\kappa(i+1)\hbar} S_{(i+1)\hbar} - e^{-\kappa i\hbar} S_{i\hbar}), \quad (5.56)$$

in which $\kappa = r - q$. This is valid for every discrete-time hedging scheme $(\zeta_{t_i})_{0 \leq i \leq n-1}$. The Black–Scholes implied delta hedging scheme corresponds to the following specification:

$$\zeta_{t_i} = \Delta^{bs}(t_i, S_{t_i}, T, K; \sigma_{t_i}),$$

where σ_{t_i} denotes the Black–Scholes implied volatility of the option at time t_i .

5.4.2.2 Implied Delta-Hedging with the Gaussian Copula Model

Our next aim is to hedge the spread risk of a CDO tranche between two successive default times of the reference entities. Note that we do not aim at hedging defaults in this approach. More precisely, our goal will be to hedge homogeneous bumps on homogeneous time intervals of the underlying CDS curves, using CDS index

contracts as hedging instruments. A CDS index contract covers default risk on all names in a credit pool. CDS index contracts may be considered as kinds of averages of individual CDSs, and they can be priced essentially like the latter, using a relation of the form (5.53).

We thus rebalance, at every time step h (typically taken of the order of one week for credit derivatives), a hedging position in a primary market consisting of the savings account $B_t = \beta_t^{-1}$, and of q CDS index contracts with increasing maturities T_j , $j = 1, \dots, q$ where $T_{q-1} < T \leq T_q$. Also, let $T_0 = 0$. Considering a bank that bought credit protection through a tranche and is short ζ_t^j units in the CDS index contract with maturity T_j , the discounted P&L increment of the hedged position on a time interval $(t, t + h)$ is written, assuming no defaults of the underlying names on $(t, t + h)$, as:

$$\begin{aligned}\beta_t \delta e &= -\beta_t \delta e^\star + \sum_j \zeta_t^j \beta_t \delta e^j \\ &= (\beta_{t+h} D_{t+h} - \beta_t D_t) - \Sigma_0^* (\beta_{t+h} F_{t+h} - \beta_t F_t) - \beta_t \Sigma_0^* h \\ &\quad - \sum_j \zeta_t^j (\Lambda \beta_{t+h} D_{t+h}^j - \Sigma_t^j \beta_{t+h} F_{t+h}^j - \beta_t \Sigma_t^j h),\end{aligned}\tag{5.57}$$

where:

- δe^\star (respectively δe^j) is the increment of the P&L on a unit position on the tranche (respectively on the CDS index contract with maturity T_j),
- Σ_0^* is the contractual spread of the tranche, and Σ_t^j is the spread of the CDS index contract with maturity T_j at the current time t ,
- D and F , respectively D^j and F^j , denote the value processes of the default and fees legs of the tranche, respectively of the CDS index contract with maturity T_j , and
- Λ is a common and constant loss-given-default on the credit index contracts.

Note that the δe^j only depend on the value of the CDS index contracts at time $t + h$. This is due to the fact that the CDS index contracts used for hedging at time t are new CDSs, freshly emitted at t . So their value at time t is equal to zero, by definition of the fair spread Σ_t^j . The Gaussian copula implied delta hedge then consists in setting a row-vector of CDS index hedging positions $\zeta_t = (\zeta_t^j)_j$ in (5.57), such that

$$\Delta_t^\star = \zeta_t \Delta_t,\tag{5.58}$$

where Δ_t^\star (respectively Δ_t) represents the row-vector of the sensitivities of the tranche (respectively the matrix of the sensitivities of the CDS index contracts) with respect to homogeneous bumps of one basis point on the time interval $[T_{j-1}, T_j]$ of the underlying CDS curves, for $j = 1, \dots, q$. In (5.58):

- the Δ_t^j are computed by assessing, using a relation analogous to (5.53), the impact of the j^{th} bump on the underlying CDS curves, on the default and fees legs

- of the i^{th} CDS index contract. Note that the Δ_i^j are not sensitive to the dependence structure (copula function) of (τ_1, \dots, τ_d) , and that the matrix Δ is lower triangular;
- the $\Delta^{*,j}$ are computed by assessing the impact on the default and fees legs of the tranche of the j^{th} bump on the underlying CDS curves.

More precisely, in the second item:

- we bootstrap the marginal cumulative distribution functions F_l from the underlying CDS spread curves using (5.53), and we compute the base implied correlation ρ_t of the tranche;
- we compute the associated values D and F of the default and fees legs of the tranche.

Then, for $j = 1, \dots, q$:

- we recalibrate as in item (i) every marginal cumulative distribution function \tilde{F}_l to the corresponding CDS curve bumped by +1bp on the time interval $[T_{j-1}, T_j]$;
- we compute the values \tilde{D} and \tilde{F} of the default and fees legs of the tranche in the Gaussian copula model with marginal cumulative distribution functions \tilde{F}_l and with correlation parameter ρ_t ;
- we set

$$\Delta^{*,j} = \Sigma_0^*(\tilde{F} - F) - (\tilde{D} - D).$$

We finally solve the triangular system (5.58) for the row-vector ζ_t .

5.4.2.3 Sticky Deltas

The previous Black–Scholes (resp. Gaussian copula) implied delta hedges are only partial hedges, which do not fairly account for the volatility or (resp. correlation, leaving alone the jump-to-default) risk.

To improve on the Black–Scholes implied delta, we can alternatively use a finite difference Black–Scholes implied delta of the form

$$(2\alpha S_t)^{-1}(\Pi^{bs}(t, (1+\alpha)S_t, T, K; r, q, \tilde{\sigma}_t) - \Pi^{bs}(t, (1-\alpha)S_t, T, K; r, q, \sigma_t)),$$

in which the Black–Scholes volatility $\tilde{\sigma}_t$ is a suitable update of σ_t , accounting for the correlation between stock returns and implied volatility changes. For instance, the sticky delta rule stipulates that the implied volatility surface evolves deterministically, when parameterized in terms of the time-to-maturity ($T - t$) and the moneyness $\ln(\frac{K}{S_t})$ (see Balland [15]).

Likewise, to improve over the Gaussian copula implied delta, we can use an updated correlation parameter $\tilde{\rho}_t$ in item (iv) above in order to account for the correlation between spread moves and implied correlation changes. A sticky delta rule thus stipulates that the implied correlation smile evolves deterministically when parameterized in terms of the time to maturity ($T - t$) and of a suitable notion of moneyness of a tranche.

5.5 Vanilla Options Fourier Transform Pricing Formulas

In many models, semi-explicit Fourier pricing formulas are available for vanilla options. This is notably the case in all affine jump-diffusion (AJD) models [103]. These formulas are only semi-explicit in the sense that they involve deterministic integrals that must be valued numerically. However, these valuations can be done very quickly and accurately (often using fast Fourier transform algorithms) so that, in particular, we can use these formulas in the context of model calibration (see Chap. 9).

5.5.1 Fourier Calculus

The Fourier transform $\mathcal{F}f$ of an absolutely integrable function f from \mathbb{R} to itself is defined, for every $u \in \mathbb{R}$, by

$$\mathcal{F}f(u) = \int_{-\infty}^{\infty} e^{iux} f(x) dx.$$

The differentiation operator translates into multiplication by iu in the Fourier space, so $\mathcal{F}f'(u) = (-iu)\mathcal{F}f(u)$. The inverse Fourier transform formula states that for $x \in \mathbb{R}$,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \mathcal{F}f(u) du. \quad (5.59)$$

Moreover the Fourier transform $\mathcal{F}f$ may be extended to complex values of its argument, resulting in the so-called complex Fourier transform of f , for u in suitable strips of analyticity of $\mathcal{F}f$ parallel to the real axis. So, for sufficiently small $\eta > 0$,

$$f(x) = \frac{1}{2\pi} \int_{-\eta i - \infty}^{-\eta i + \infty} e^{-iux} \mathcal{F}f(u) du. \quad (5.60)$$

In case f represents the density of the law of a real random variable X , the Fourier transform of f coincides with the characteristic function $\Phi(u) = \mathbb{E}[\exp(iuX)]$ of X . Then (5.60) implies that

$$\begin{aligned} G(x) &= \mathbb{P}(X > x) \\ &= \frac{1}{2\pi} \int_{-\eta i - \infty}^{-\eta i + \infty} \Phi(u) \left(\int_x^{\infty} e^{-iuy} dy \right) du = \frac{1}{2\pi i} \int_{-\eta i - \infty}^{-\eta i + \infty} \frac{e^{-iux}}{u} \Phi(u) du. \end{aligned}$$

Letting $\eta \rightarrow 0+$, we can then show, by application of the Cauchy residue formula [187, 250], that

$$G(x) = \frac{1}{2} + \frac{1}{2\pi} \lim_{\varepsilon \rightarrow 0+} \int_{|u|>\varepsilon} \frac{e^{-iux} \Phi(u)}{iu} du, \quad (5.61)$$

where the integral is defined in the sense of principal values.

Remark 5.5.1 Here is a heuristic argument to prove (5.61). On the one hand, the Cauchy residue formula yields, for $\eta > 0$,

$$\frac{1}{2\pi i} \int_{-\eta i - \infty}^{-\eta i + \infty} \frac{e^{-iux}}{u} \Phi(u) du - \frac{1}{2\pi i} \int_{\eta i - \infty}^{\eta i + \infty} \frac{e^{-iux}}{u} \Phi(u) du = 1.$$

On the other hand, the principal value of the integral in the second line below is such that

$$\begin{aligned} & \lim_{\eta \rightarrow 0+} \left(\int_{-\eta i - \infty}^{-\eta i + \infty} \frac{e^{-iux}}{u} \Phi(u) du + \int_{\eta i - \infty}^{\eta i + \infty} \frac{e^{-iux}}{u} \Phi(u) du \right) \\ &= 2 \text{ p.v.} \int_{\mathbb{R}} \frac{e^{-iux} \Phi(u)}{u} du = 2 \lim_{\varepsilon \rightarrow 0+} \int_{|u|>\varepsilon} \frac{e^{-iux} \Phi(u)}{u} du. \end{aligned}$$

The formula (5.61) for $G(x) = \frac{1}{2\pi i} \int_{-\eta i - \infty}^{-\eta i + \infty} \frac{e^{-iux}}{u} \Phi(u) du$ results.

Since $\frac{e^{iux} \Phi(-u)}{-iu}$ is the complex conjugate of $\frac{e^{-iux} \Phi(u)}{iu}$, we have

$$\lim_{\varepsilon \rightarrow 0+} \int_{|u|>\varepsilon} \frac{e^{-iux} \Phi(u)}{iu} du = 2 \lim_{\varepsilon \rightarrow 0+} \int_{\varepsilon}^{\infty} \Re e \left[\frac{e^{-iux} \Phi(u)}{iu} \right] du.$$

So, finally,

$$G(x) = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \Re e \left[\frac{ie^{-iux} \Phi(u)}{iu} \right] du. \quad (5.62)$$

5.5.2 Black–Scholes Type Pricing Formula

We now consider the problem of valuing a European vanilla call of maturity T and strike K on a positive underlying $S = e^X$. For simplicity we assume a constant risk-free interest rate r , a constant dividend yield q on S , and we write $\kappa = r - q$ as usual.

Proposition 5.5.2 A European vanilla call price at time 0, $\Pi_0 = \mathbb{E}e^{-rT}(S_T - K)^+$ can be represented as

$$\Pi_0 = S_0 e^{-qT} \tilde{P}_0 - K e^{-rT} P_0, \quad (5.63)$$

where P_0 and \tilde{P}_0 are defined in terms of the characteristic function $\Phi_T(u) = \mathbb{E}[\exp(iuX_T)]$ (assumed to exist) as

$$\begin{aligned} P_0 &= \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re e \left[\frac{e^{-iuk} \Phi_T(u)}{iu} \right] du, \\ \widetilde{P}_0 &= \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re e \left[\frac{e^{-iuk} \Phi_T(u-i)}{iu \Phi_T(-i)} \right] du, \end{aligned} \quad (5.64)$$

with $\Phi_T(-i) = \mathbb{E} S_T = S_0 e^{\kappa T}$ in P_0 .

Proof We write

$$\Pi_0 = \mathbb{E}(e^{-rT} e^{X_T} \mathbb{1}_{\{X_T > k\}}) - K e^{-rT} \mathbb{P}(X_T > k),$$

where $k = \ln K$. By (5.62),

$$P_0 = \mathbb{P}(X_T > k) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re e \left[\frac{e^{-iuk} \Phi_T(u)}{iu} \right] du.$$

We introduce the probability measure $\widetilde{\mathbb{P}}$ equivalent to \mathbb{P} such that

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = \frac{e^{X_T}}{\mathbb{E}(e^{X_T})} = \frac{S_T}{S_0 e^{\kappa T}}.$$

Thus

$$\begin{aligned} \mathbb{E}(e^{-rT} e^{X_T} \mathbb{1}_{\{X_T > k\}}) &= \int_{x \in \mathbb{R}} e^{-rT} e^x \mathbb{1}_{\{x > k\}} \mathbb{P}(X_T \in dx) \\ &= \mathbb{E}(e^{-rT} e^{X_T}) \widetilde{\mathbb{P}}(X_T > k) = S_0 e^{-qT} \widetilde{\mathbb{P}}(X_T > k). \end{aligned}$$

The characteristic function of X_T under $\widetilde{\mathbb{P}}$ is given by

$$\widetilde{\mathbb{E}}(e^{iux_T}) = \int_{x \in \mathbb{R}} e^{iux} \widetilde{\mathbb{P}}(X_T \in dx) = \frac{\mathbb{E}(e^{X_T} e^{iux_T})}{\mathbb{E}(e^{X_T})} = \frac{\Phi_T(u-i)}{\Phi_T(-i)}.$$

Hence, by (5.62),

$$\widetilde{P}_0 = \widetilde{\mathbb{P}}(X_T > k) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re e \left[\frac{e^{-iuk} \Phi_T(u-i)}{iu \Phi_T(-i)} \right] du. \quad \square$$

In many models the characteristic function Φ_T can be computed explicitly. This is, for instance, the case in all AJD models. Knowing Φ_T , the Fourier pricing formula (5.63) enables one to compute Π_0 numerically. For implementation details, including the use of Gaussian quadratures methods which can be efficiently used here, or for numerical issues related to multivalued complex integrands, the reader is referred to Kahl and Jackel [160].

5.5.2.1 Delta in Homogeneous Models

By homogeneous models, we mean models in which the European vanilla option prices are degree one homogeneous with respect to the pair (S_0, K) (for given values of the remaining parameters and risk factors in the model), i.e.

$$\Pi(0, \alpha S_0, T, \alpha K) = \alpha \Pi(0, S_0, T, K), \quad \alpha > 0,$$

or, equivalent to this,

$$S_0 \partial_S \Pi(0, S_0, T, K) + K \partial_K \Pi(0, S_0, T, K) = \Pi(0, S_0, T, K). \quad (5.65)$$

Note that, in general,

$$\partial_K \Pi_0 = e^{-rT} \partial_K \mathbb{E}(S_T - K)^+ = -e^{-rT} \mathbb{E} \mathbb{1}_{\{S_T > K\}} = -e^{-rT} \mathbb{P}(S_T > K) = -e^{-rT} P_0.$$

So, in a homogeneous model,

$$\begin{aligned} S_0 \partial_S \Pi(0, S_0, T, K) &= \Pi(0, S_0, T, K) - K \partial_K \Pi(0, S_0, T, K) \\ &= \Pi_0 + K e^{-rT} S_0 = S_0 e^{-qT} \tilde{P}_0 \end{aligned}$$

by (5.63). Hence

$$\Delta_0 = \partial_S \Pi(0, S_0, T, K) = e^{-qT} \tilde{P}_0. \quad (5.66)$$

5.5.3 Carr–Madan Formula

Fast Fourier transform (FFT) algorithms cannot be used to evaluate the integrals in (5.64), due to the singularity of the integrands at $u = 0$. We now present an alternative formula, due to Carr and Madan [62], that is amenable to valuation by FFT. A further advantage of this formula is that it allows us to compute simultaneously the prices of a whole family of options with various strikes at time 0 (as is required for calibration purposes). The idea is to compute the Fourier transform of the call price, viewed as a function $\Pi_0(k)$ of the log-strike $k = \ln K$. The function $\Pi_0(k)$ is not integrable, since $\lim_{k \rightarrow -\infty} \Pi_0(k) = e^{-rT} \mathbb{E} S_T = S_0 e^{-qT} > 0$. However, we obtain an integrable function by setting $\Pi_0^\alpha(k) = e^{\alpha k} \Pi_0(k)$ for a fixed $\alpha > 0$. Letting q_T denote the density (assumed to exist) of $X_T = \ln(S_T)$, for every real u we compute

$$\begin{aligned} \mathcal{F}\Pi_0^\alpha(u) &= \int_{-\infty}^{\infty} e^{iku} \Pi_0^\alpha(k) dk = \int_{-\infty}^{\infty} e^{(\alpha+iu)k} \Pi_0(k) dk \\ &= e^{-rT} \int_{-\infty}^{\infty} e^{(\alpha+iu)k} \left(\int_{-\infty}^{\infty} (e^x - e^k)^+ q_T(x) dx \right) dk \end{aligned}$$

$$\begin{aligned}
&= e^{-rT} \int_{-\infty}^{\infty} q_T(x) \left(\int_{-\infty}^x e^{(\alpha+iu)k} (e^x - e^k) dk \right) dx \\
&= e^{-rT} \int_{-\infty}^{\infty} q_T(x) \left(\frac{e^x}{\alpha+iu} [e^{(\alpha+iu)k}]_{-\infty}^x - \frac{1}{\alpha+iu+1} [e^{(\alpha+iu+1)k}]_{-\infty}^x \right) dx \\
&= e^{-rT} \int_{-\infty}^{\infty} q_T(x) \left(\frac{e^x e^{(\alpha+iu)x}}{\alpha+iu} - \frac{e^{(\alpha+iu+1)x}}{\alpha+iu+1} \right) dx,
\end{aligned}$$

where the last equality follows from the fact that $\lim_{k \rightarrow -\infty} e^{(v+iu)k} = 0$, for every $v > 0$ and $u \in \mathbb{R}$. Therefore⁵

$$\begin{aligned}
\mathcal{F}\Pi_0^\alpha(u) &= e^{-rT} \int_{-\infty}^{\infty} q_T(x) \frac{e^{(\alpha+iu+1)x}}{(\alpha+iu)(\alpha+iu+1)} dx \\
&= \frac{e^{-rT} \Phi_T(u - (\alpha+1)i)}{(\alpha+iu)(\alpha+iu+1)}. \tag{5.67}
\end{aligned}$$

By the inverse Fourier transform formula (5.59) we then have, for every real k ,

$$\begin{aligned}
\Pi_0(k) &= e^{-\alpha k} \Pi_0^\alpha(k) = \frac{e^{-\alpha k}}{2\pi} \int_{-\infty}^{\infty} e^{-iku} \mathcal{F}\Pi_0^\alpha(u) du \\
&= \frac{e^{-\alpha k}}{\pi} \Re e \left[\int_0^{\infty} e^{-iku} \mathcal{F}\Pi_0^\alpha(u) du \right], \tag{5.68}
\end{aligned}$$

where the last equality holds because the function

$$\mathbb{R} \ni k \mapsto \Pi_0^\alpha(k)$$

is real-valued, which implies that $\mathcal{F}\Pi_0^\alpha(-u) = \overline{\mathcal{F}\Pi_0^\alpha(u)}$ for every $u \in \mathbb{R}$. We then approximate, by numerical integration based on Simpson's rule:⁶

$$\Pi_0(k) \approx \frac{e^{-\alpha k}}{\pi} \Re e \left[\sum_{j=0}^{N-1} e^{-iku_j} \mathcal{F}\Pi_0^\alpha(u_j) w_j \right], \tag{5.69}$$

with N even in Simpson's rule and where, for $j = 0, \dots, N-1$:

$$u_j = jh, \quad w_j = \frac{h}{3} (3 + (-1)^{j+1} - \mathbb{1}_{\{j=0 \text{ or } N-1\}}).$$

Moreover, for k of the form $k_n = \underline{k} + \frac{2\pi n}{Nh}$, where \underline{k} will be fixed in the end, we have:

$$ku_j = k_n u_j = \underline{k} u_j + \frac{2\pi n}{N} j.$$

⁵ Assuming $\Phi_T(-(\alpha+1)i) = \mathbb{E} S_T^{\alpha+1} < +\infty$.

⁶Simpson's integration rule gives a good accuracy for a relatively small value of N .

Substituting this into (5.69) yields

$$\Pi_0(k_n) \approx \frac{e^{-\alpha k_n}}{\pi} \Re e \left[\sum_{j=0}^{N-1} e^{-2\pi i \frac{jn}{N}} e^{-ik u_j} \mathcal{F}\Pi_0^\alpha(u_j) w_j \right], \quad 0 \leq n \leq N-1. \quad (5.70)$$

The discrete Fourier transform $(F\varphi_n)_{0 \leq n \leq N-1}$ of a vector $\varphi = (\varphi_j)_{0 \leq j \leq N-1}$ is given by

$$F\varphi_n = \sum_{j=0}^{N-1} e^{-2\pi i \frac{jn}{N}} \varphi_j, \quad 0 \leq n \leq N-1. \quad (5.71)$$

In the sum in (5.70), we thus recognize the discrete Fourier transform of

$$\varphi = (e^{-ik u_j} \mathcal{F}\Pi_0^\alpha(u_j) w_j)_{0 \leq j \leq N-1}.$$

Choosing $k = \ln(S_0) + \kappa T - \frac{\pi}{h}$, so that

$$k_n = \ln(S_0) + \kappa T - \frac{\pi}{h} + \frac{2\pi}{h} \frac{n}{N}, \quad n = 0, \dots, N-1,$$

we can thus price a call for N values of the strike K around the T -forward value of the stock $F_0 = S_0 e^{\kappa T}$ by computing the discrete Fourier transform of φ . For N given as a power of 2, the N sums in (5.70) can be computed in time $O(N \ln N)$ by FFT.

Part III

Numerical Solutions

We have seen in the previous chapter that, in affine models, semi-explicit formulas are available for the prices and Greeks of European vanilla options. In Chap. 9 we will dwell upon the usefulness of fast pricing formulas (their necessity in a sense) at the stage of model calibration, which typically involves intensive pricing of vanilla options. We can also mention in this regard the SABR model [137], in which explicit asymptotic formulas are available for the Black–Scholes implied volatility; these formulas explain why SABR (as a smiled extension of the log-normal Libor market model of Sect. 5.2.2) has become the standard of the interest rate derivative market (see [231]).

However, as far as pricing exotics or dealing with less standard models is concerned, the pricing BSDEs/PIDEs have to be solved numerically. In this part we deal with the numerical solution of the pricing equations by stochastic simulation methods (Chap. 6) or, provided the dimension of the model is not too large, by deterministic schemes like trees (Chap. 7) or finite differences (Chap. 8).

For the sake of concreteness and simplicity, we present reference numerical schemes in simple models, most often in Black–Scholes, although numerical solutions are typically not obligatory in this case since many explicit formulas are available. The explicit formulas are useful to assess the accuracy of the numerical schemes. The numerical schemes themselves are generic so that it is rather straightforward to extend them to arbitrary jump-diffusions. Moreover, regarding numerical schemes, we think that it is important to provide completely solved examples in definite, although simple, setups.

Accuracy Versus Computational Cost A typical benchmark of numerical accuracy in computational finance may vary from an order of 10^{-2} % (“one basis point” or bp) to 1 % in relative terms, depending on the application. As for computation times, the benchmark also varies greatly with the application, but as far as “real time” option pricing is concerned, “instantaneous” pricing is the target, and more than a few minutes is prohibitive. Now, a solution within a 1 bp normalized error by a finite differences ADI method, which is the industry standard today as far as deterministic methods are concerned (see Sect. 8.5), requires about

Table III.1 Compared computational costs of a Monte Carlo pricing scheme (m simulation runs) versus a deterministic pricing scheme (m_1 mesh points per space dimension, i.e. $m = m_1^d$ space mesh points), in space dimension d ; n is the number of points of a time-grid which is used in both schemes

	CPU time	Accuracy	Memory cost
MC	$O(nm)$	$O(n^{-1} + m^{-\frac{1}{2}})$	$O(1)$
PDE	$O(nm)$	$O(n^{-1} + m^{-\frac{2}{d}})$	$O(m)$

Table III.2 Deterministic versus stochastic pricing schemes: which one is preferred?

		European problem	American or control problems
$d \leq 3$	PDE or MC	PDE	
$d > 3$	MC		Nonlinear MC

300 grid points per space dimension. This yields a computation time in $O(300^d)$, i.e. a computation time ranging on present day computers from a few milliseconds for $d = 1$ to a few minutes for $d = 3$. In practice, this limits the range of applicability of deterministic pricing schemes to problems in space dimension ≤ 3 , unless sophisticated sparse grid or grid refinement techniques are used. This is an effect of Bellman's curse of dimensionality, referring to the fact that the computational cost of numerical integration grows exponentially with space dimension d as m_1^d , where m_1 is the number of discretization points in a generic dimension of the state space.

Table III.1 provides a crude comparison of the computational costs of typical Monte Carlo and deterministic schemes (Monte Carlo algorithm with time discretization of the underlying factor process, cf. Sect. 6.8.4, versus an ADI PDE method of Sect. 8.5). A rough conclusion¹ is that deterministic methods are more efficient in space dimension ≤ 3 , but they are often harder to implement. In dimension > 3 Monte Carlo methods are preferred when available. This leads to the dictionary of Table III.2 regarding the type of numerical method to be used, depending on the space dimension and on the nature of a pricing problem. In the upper left cell of the Table, the choice between PDE and Monte Carlo should be dictated by the relative interest of performance as compared with implementation cost and risk, both generally higher with deterministic schemes. In the lower right cell, "Nonlinear MC" refers to nonlinear simulation schemes, which can be used for solving nonlinear control problems. The latter include, in particular, American or more general game option (such as convertible bond) pricing problems, or pricing in the uncertain volatility model of Avellaneda, Levy, and Parás [13] in which a stochastic volatility is only known to remain in a positive interval $[\underline{\sigma}, \bar{\sigma}]$ (which also relates to the notion of second-order BSDEs of Soner, Touzi, and

¹We do not detail the constants involved in these estimates.

Zhang [247]). “Nonlinear MC” schemes include, in particular, nonlinear simulation/regression schemes such as those of Sect. 6.10 and Chap. 10. They also include purely forward particle simulation schemes (not covered in this book) based on the theory of branching Markov processes, developed from seminal contributions by the Japanese school (Ikeda, Nagasawa and Watanabe [148], Fujita [130]), by McKean [198] and by Dynkin [107, 108], Le Gall [183] and others since the 1990s. The field has seen new vitality in recent years in connection with numerical applications, including those related to finance (Henry-Labordère [142]; Guyon and Henry-Labordère [135]).

Note that for high-dimensional problems, an important issue is parallel computing, which is today broadly available via graphic card programming and allows reduction in computation times by factors which may vary from, say, 2 to 1000. In particular, parallel computing is very efficient for Monte Carlo applications.

Markovian Dimension Versus Martingale Order of Multiplicity The space dimension d above refers to the Markovian dimension of a pricing problem, in the sense of the dimension of the state space of a related factor process, which may differ from the nominal dimension of a pricing model. For instance, when pricing an Asian option in the Black–Scholes model (see Sect. 8.6.3), the nominal dimension of the model is one, but the Markovian dimension of the pricing problem is two (unless specific degenerate cases are considered, as in Sect. 8.6.3.1). Also, in the case of models with jumps such as portfolio credit risk models, the notion of dimension d of the state space of the factor process is not always so well defined. In the case of a Markov chain model, it seems reasonable to define d as the logarithm of the dimension of the matrix generator.

A competing notion for the dimension of a pricing problem is its multiplicity μ , which is the minimal size of a family of martingales with the representation property (see [92]) or, in financial terms, the minimal size of a family of replicating assets. In rough terms, this corresponds to the number of independent sources of noise in the model. It is often the case that $\mu = d$. However, this need not be so. We thus have:

- $\mu = 1 < 2 = d$ in the problem of pricing an Asian option in the Black–Scholes model (see Sect. 8.6.3),
- $d = 1 < 2 = \mu$ in the problem of pricing a vanilla option in the jump-diffusion Merton model (see Sect. 5.1.3),
- $d = n$ and $\mu = 2^n$ in the case of a general continuous-time Markov chain bottom-up model of portfolio credit risk with n obligors, but we have $\mu = n$ if simultaneous defaults are excluded (see [35]).

The multiplicity μ ² of a model is the relevant notion of dimension as far as a Monte Carlo pricing scheme is concerned. Thus observe that the complexity of

²Multiplied by the number of dates in the time-grid, in case of a time-discretized problem.

a Monte Carlo method is linear in μ (there is a linear dependence in μ hidden in the $O(nm)$ in the upper left cell of Table III.1), whereas that of a deterministic method is exponential in d . This gives another perspective on the fact that simulation methods are better suited than deterministic ones for facing the curse of dimensionality.

Chapter 6

Monte Carlo Methods

This chapter is about Monte Carlo pricing methods. The term “Monte Carlo” for computational methods involving simulated random numbers was introduced in Metropolis and Ulam [204]. Like deterministic pricing schemes, simulation pricing schemes can be used in any Markovian (or, of course, static one-period) setup. In the case of European claims, simulation pricing schemes reduce to the well known Monte Carlo loops. For products with early exercise features, or for more general control problems, numerical schemes by simulation are available too, yet these are more sophisticated; see e.g. Sect. 6.10 and Part IV.

Monte Carlo methods are attractive by their genericity:

- genericity of their theoretical properties (such as the confidence interval they provide for the solution) that are insensitive to the dimension of a problem or to the irregularity of a payoff function—at least for genuine pseudo Monte Carlo methods, as opposed to the quasi Monte Carlo methods that we will also consider in this chapter,
- genericity of implementation,
- ease of parallelization.

But (pseudo) Monte Carlo methods are slow, only converging at the rate $\sigma m^{-\frac{1}{2}}$ where m is the number of simulation runs and σ is the standard deviation of the sampled payoff. To accelerate the convergence, various variance reduction techniques can be used to transform a given payoff into another one with less variance. An alternative to variance reduction is quasi Monte Carlo, which converges faster in practice than pseudo Monte Carlo (at least in low dimension). However quasi Monte Carlo estimates do not come with confidence intervals, and their performance can be strongly altered in high dimension.

6.1 Uniform Numbers

To sample “random” vectors in \mathbb{R}^d with specific distributions, one first draws “uniform random” vectors u_j over $[0, 1]^d$. One then transforms the u_j into x_j with the

desired distribution. “Uniform random vectors” u_j over $[0, 1]^d$ may be obtained either:

- by sampling “i.i.d. random variables” over $[0, 1]$ with a pseudo-random generator, and arranging them into sequences of length d ,
- or by using a quasi-random generator (low-discrepancy sequence) in dimension d .

The use of quotation marks above indicates that simulated random numbers only “look” random and uniform (and independent, in the case of pseudo-random numbers). However, whenever a simulated sequences is generated deterministically on a computer, one can always find a statistical test for uniformity or independence that a simulated sequence will fail to pass. And, of course, the quality of a generator puts an upper bound on the quality of any simulation pricing scheme using it.

By a d -variate uniform, respectively Gaussian, random draw we henceforth mean a uniform point over $[0, 1]^d$, respectively a d -variate centered and normalized Gaussian vector, with $d = 1$ by default.

6.1.1 Pseudo-Random Generators

Pseudo-random generators are used to simulate independent uniform variables over $[0, 1]$. L’Ecuyer [181, 182] formally defines a pseudo-random number generator as a quintuple $\mathcal{G} = (s, S, T, U, G)$ where S is a finite set of states, $s \in S$ is the initial state, the mapping $T : S \rightarrow S$ is the transition function, and $G : S \rightarrow U$ is the output function from S to a finite set U of outputs. Since S is finite, the sequence of states is ultimately periodic. The period is the smallest positive integer m such that $s_{j+m} = s_j$ for all sufficiently large j . The following properties are required for a good pseudo-random number generator:

- i. *Large period length*: At least 2^{60} , say.
- ii. *Good equidistribution properties and statistical independence of successive pseudo-random draws*: The generator should pass statistical tests of uniformity and independence—general tests such as chi-square or Kolmogorov-Smirnov tests, and more specific tests such as equidistribution tests, serial tests, gap tests, partition tests.
- iii. *Little intrinsic structure*: Successive values produced by some generators produce undesirable lattice structures.
- iv. *Efficiency, fast generation, not requiring too much memory space*: Especially if we use many generators together or in parallel.
- v. *Repeatability (fixing a given seed)*: Very useful for practical applications. Otherwise one can use the time of execution of the program given by the computer clock to initialize the generator.
- vi. *Portability*: The generator should produce the same sequence on different computers or with different compilers.

- vii. *Unpredictability*: one should not be able to predict the next generated value from the previous ones.¹

Note that, from the point of view of period length, built-in library generators may behave poorly. The function `rand()` in the C++ standard library thus returns a pseudo-random integer in the range 0 to `RAND_MAX`, where the value of the constant `RAND_MAX` may vary between implementations, but can only be assumed to be at least 32767.

6.1.1.1 Pseudo-Random Uniform Numbers

Linear schemes are the simplest method for constructing pseudo-random numbers. They use a linear recurrence relation to compute the next value from the previous one. The j th random number is given by

$$u_j = \frac{U_j}{c} \in [0, 1], \quad \text{where } U_j = (aU_{j-1} + b) \bmod c,$$

for well chosen fixed integers $a > 0$, b and $c > 0$, starting from a given seed U_0 . For instance, in the Fortran IMSL Library,

$$a = 397204094, \quad b = 0 \quad \text{and} \quad c = 2^{31} - 1.$$

Such generators are very simple but prone to produce a lot of regularity in sequences along with a lattice structure. Yet it is possible to combine any of these with another, using for instance the shuffling procedure of Bayes and Durham, obtaining thereby a longer period generator with better properties.

6.1.1.2 Rejection-Acceptance Method

The rejection-acceptance method allows us to draw pseudo-uniform points in an arbitrary subset of \mathbb{R}^d , starting from pseudo-uniform points in a larger set, based on the following

Proposition 6.1.1 Suppose the u_j are i.i.d. uniform points over a Lebesgue set D of \mathbb{R}^d . Let there be given a subset Δ of D with $\frac{\lambda(\Delta)}{\lambda(D)} = \alpha \in (0, 1)$, where λ represents Lebesgue measure over \mathbb{R}^d . Let $0' = 0$ and, for $j \geq 1$,

$$j' = \inf\{l > (j-1)'; u_l \in \Delta\}.$$

Then the $v_j := u_{j'}$ are i.i.d. uniformly distributed over Δ . For every $j \geq 1$, the average acceptance time $\mathbb{E}(j' - (j-1)')$ equals α^{-1} .

¹However, this is less important in finance than for other applications such as cryptography.

Proof We have $\mathbb{P}(1' = l) = (1 - \alpha)^{l-1} \alpha$, so that

$$\mathbb{E}1' = \sum_{l \geq 1} l \mathbb{P}(1' = l) = \sum_{l \geq 1} l (1 - \alpha)^{l-1} \alpha = \alpha^{-1}.$$

For any sequence of subsets Δ_j of Δ , we show by induction that, for every $m \geq 0$,

$$\mathbb{P}(v_j \in \Delta_j, j = 1 \dots m) = \prod_{j=1}^m \frac{\lambda(\Delta_j)}{\lambda(\Delta)}.$$

For $m = 0$ this is trivially satisfied. Moreover, for $m \geq 1$ we have that

$$\begin{aligned} & \mathbb{P}(v_j \in \Delta_j, j = 1 \dots m) \\ &= \sum_{l \geq m-1, j \geq 1} \mathbb{P}(v_1 \in \Delta_1, \dots, v_{m-1} \in \Delta_{m-1}, (m-1)' = l, m' = l + j, u_{l+j} \in \Delta_m) \\ &= \sum_{l \geq m-1} \mathbb{P}(v_1 \in \Delta_1, \dots, v_{m-1} \in \Delta_{m-1}, (m-1)' = l) \sum_{j \geq 1} (1 - \alpha)^{j-1} \frac{\lambda(\Delta_m)}{\lambda(\Delta)} \\ &= \left(\sum_{l \geq m-1} \mathbb{P}(v_j \in \Delta_j, \dots, v_{m-1} \in \Delta_{m-1}, (m-1)' = l) \right) \frac{\lambda(\Delta_m)}{\lambda(\Delta)} \\ &= \mathbb{P}(v_j \in \Delta_j, j = 1 \dots m-1) \frac{\lambda(\Delta_m)}{\lambda(\Delta)}. \end{aligned}$$
□

6.1.2 Low-Discrepancy Sequences

Quasi-random numbers, or successive points u_j of low-discrepancy sequences [206, 209, 210], are not independent. But they have good equidistribution properties on $[0, 1]^d$, implying good convergence properties of $\frac{1}{m} \sum_{j=1}^m \psi(u_j)$ to $\int_{[0,1]^d} \psi(u) du$ as $m \rightarrow \infty$.

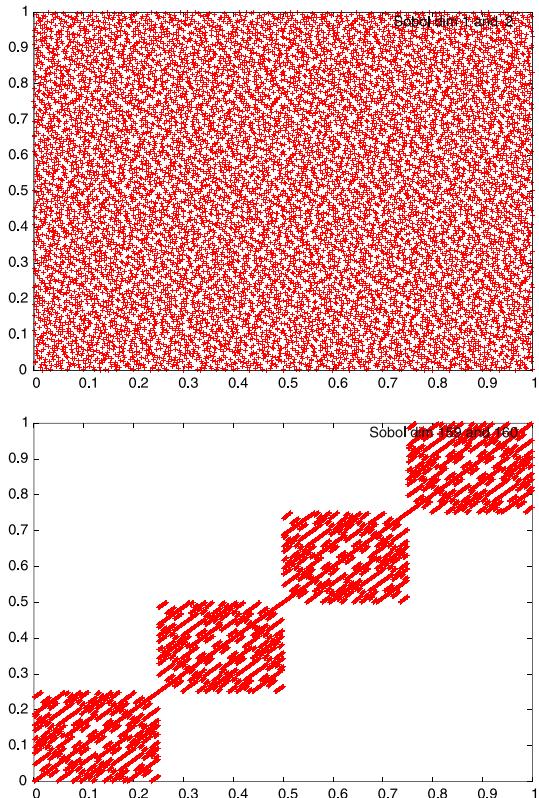
The discrepancy can be viewed as a quantitative measure for the deviation from the uniform distribution. Let $[0, x]$ denote $\{y \in [0, 1]^d, y \leq x\}$, where $y \leq x$ is defined componentwise. Given a $[0, 1]^d$ -valued sequence $u = (u_j)_{j \geq 1}$ and $x \in [0, 1]^d$, let

$$D_m(u, x) = \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{[0,x]}(u_j) - \prod_{l=1}^d x_l.$$

Definition 6.1.2

- (i) A sequence u is said to be equidistributed on $[0, 1]^d$ if $\lim_{m \rightarrow \infty} D_m(u, x) = 0$, for every $x \in [0, 1]^d$.

Fig. 6.1 Projection on the first two, respectively last two, coordinates of the first 10^4 points of the Sobol sequence in dimension $d = 160$. If a $d(\geq 2)$ -dimensional sequence is uniformly distributed in $[0, 1]^d$, then a two-dimensional sequence, formed by pairing two coordinates, should be uniformly distributed in the unit square. The appearance of nonuniformity in such projections is an indication of potential problems in using a high-dimensional quasi-random sequence (see [206])



(ii) The value $D_m(u)$ defined by

$$D_m(u) = \sup_{x \in [0,1]^d} |D_m(u, x)|$$

is called the discrepancy of u at rank m .

(iii) The sequence u is said to have low-discrepancy, or to be quasi-random, if $D_m(u) = O\left(\frac{(\ln m)^d}{m}\right)$.

Using the law of the iterated logarithm, one can show that a pseudo-random sequence has a discrepancy $O\left(\left(\frac{\ln \ln m}{m}\right)^{\frac{1}{2}}\right)$ and is therefore not quasi-random. With respect to a lattice, quasi-random numbers have the advantage that points can be added incrementally. Low discrepancy sequences perform very well in low dimension. But, for large dimension d , the theoretical bound $\frac{(\ln m)^d}{m}$ may be meaningful only for extremely large values of m . This is essentially because, in dimension d , a lattice can only be refined by increasing the number of points by a factor 2^d . The Sobol sequence [246] is one of the most popular quasi-random sequences in financial applications (see Fig. 6.1). Its construction is based on number theory and can be implemented very efficiently using bitwise XOR (“exclusive or”) operations.

6.2 Non-uniform Numbers

6.2.1 Inverse Method

The inverse simulation method is based on the following elementary but fundamental:

Lemma 6.2.1 *For every real random variable X with cumulative distribution function F_X , for uniform U we have,*

$$F_X^{-1}(U) \stackrel{\mathcal{L}}{=} X, \quad F_X(X) \stackrel{\mathcal{L}}{=} U, \quad (6.1)$$

where F_X^{-1} denotes the generalized inverse of the nondecreasing function F_X .

Proof Letting $Y = F_X^{-1}(U)$ we have, for every x ,

$$\mathbb{P}(Y \leq x) = \mathbb{P}(U \leq F_X(x)) = F_X(x),$$

as U is uniform and $F_X(x) \in [0, 1]$. Thus X and Y have the same cumulative distribution function and so they have the same law. \square

As an application of this lemma, we obtain an exponential random variable with parameter λ through $E = -\lambda^{-1} \ln(1 - U)$ or, equally in law, $(-\lambda^{-1} \ln U)$, with U uniform.

A \mathcal{P}_λ -random variable P is such that $\mathbb{P}(P = n) = e^{-\lambda} \frac{\lambda^n}{n!}$. To simulate P , we can draw a uniform number U over $[0, 1]$ and set $P = v$ such that

$$\sum_{n=0}^v e^{-\lambda} \frac{\lambda^n}{n!} \leq U < \sum_{n=0}^{v+1} e^{-\lambda} \frac{\lambda^n}{n!}.$$

Also, it is well known that the number of clients at time T in a queue with i.i.d. exponential inter-arrival times of parameter λ , is $\mathcal{P}_{\lambda T}$ -distributed. So, to simulate P , another possibility is to draw independently $\epsilon_j = -\frac{1}{\lambda} \ln u_j$ until $\sum_{l=1}^j \epsilon_l > T = 1$ and to set, with the convention $\max \emptyset = 0$,

$$P = \max \left\{ j \geq 1; \sum_{l=1}^j \epsilon_l < 1 \right\} = \max \left\{ j \geq 1; \prod_{l=1}^j u_l > e^{-\lambda} \right\}.$$

The inverse Gaussian cumulative distribution function \mathcal{N}^{-1} is not known explicitly. To simulate Gaussian variables by inversion, one can use Moro's numerical approximation for \mathcal{N}^{-1} . However, this entails a simulation bias.

6.2.2 Gaussian Pairs

Next come two exact methods for generating a standard Gaussian pair starting from uniform numbers as input data. Note that to use these methods with quasi-random numbers, we must generate the coordinates of the underlying uniform points “independently”, e.g. from two different one-dimensional quasi-random sequences, or as the two coordinates of a point from a two-dimensional quasi-random sequence, but not as two successive values of a one-dimensional quasi-random sequence.

6.2.2.1 Box-Müller Method

An exact method for simulating a Gaussian pair is the Box-Müller transformation which is based on the following:

Lemma 6.2.2 *If (U, V) is bivariate uniform, then (X, Y) defined by*

$$\begin{aligned} X &= \sqrt{-2 \ln U} \cos(2\pi V) \\ Y &= \sqrt{-2 \ln U} \sin(2\pi V) \end{aligned}$$

is bivariate Gaussian.

Proof We have that $(-2 \ln(U))$ and $(2\pi V)$ are $\mathcal{E}_{\frac{1}{2}}$ - and $\mathcal{U}_{[0, 2\pi]}$ -distributed. So, for every test-function φ ,

$$\begin{aligned} \mathbb{E}\varphi(X, Y) &= \int_0^\infty \int_0^{2\pi} \varphi(\rho \cos \theta, \rho \sin \theta) \frac{1}{2} e^{-\frac{\rho^2}{2}} d(\rho^2) \frac{d\theta}{2\pi} \\ &= \int_0^\infty \int_0^{2\pi} \varphi(\rho \cos \theta, \rho \sin \theta) \rho d\rho e^{-\frac{\rho^2}{2}} \frac{d\theta}{2\pi} \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x, y) e^{-\frac{x^2+y^2}{2}} \frac{dx dy}{2\pi}. \end{aligned}$$

□

6.2.2.2 Marsaglia Method

Lemma 6.2.3 (Change of variables formula for densities) *Let real random variables X and Y be such that $Y = f \circ X$, where f denotes a diffeomorphism between two subsets of \mathbb{R}^d . Let $g = f^{-1}$ and Jg denote the Jacobian matrix $(\partial_{y_j} g_i)_{\substack{1 \leq j \leq d \\ 1 \leq i \leq d}}$. Assuming X has a density p_X , then so does Y , and the density p_Y of Y at $y = f(x)$ is given by*

$$p_Y(y) = |\det(Jg)(y)| p_X(x) \quad (6.2)$$

(and $p_Y(y) = 0$ for y outside the image set of f).

Proof This is seen by change of variables $x = g(y)$ in $\mathbb{E}\varphi(Y) = \int \varphi[f(x)] p_X(x) dx$, where φ denotes an arbitrary test-function on \mathbb{R}^d . □

Example 6.2.4 Applying the formula (6.2) to $Y = S_t$ in Black–Scholes and $X = \sigma W_t = g(S_t)$, with $g(S) = \ln(\frac{S}{S_0}) - bt$ where $b = \kappa - \frac{\sigma^2}{2}$, we recover the lognormal density

$$p_{S_t}(S) = \frac{1}{\sigma S \sqrt{2\pi t}} e^{-\frac{g(S)^2}{2\sigma^2 t}} = \frac{1}{S} p_{(\sigma W_t)}(g(S)).$$

In the Marsaglia method of simulation of a Gaussian pair that is provided by the next lemma, the uniform point (U, V) on the unit disk can be simulated by rejection-acceptance, using uniform points on the square $[0, 1]^2$ as input data (with an acceptance rate of $\frac{\pi}{4}$).

Lemma 6.2.5 *If (U, V) is uniform on the unit disk D , then (X, Y) defined by*

$$\begin{aligned} X &= \sqrt{\frac{-2 \ln(\rho^2)}{\rho^2}} U \\ Y &= \sqrt{\frac{-2 \ln(\rho^2)}{\rho^2}} V, \end{aligned}$$

where $\rho^2 = U^2 + V^2$, is bivariate Gaussian.

Proof Let (ρ, θ) denote the polar coordinates of (U, V) . Using the transformation

$$D \ni (U, V) \xrightarrow{f} \left(\rho^2, \frac{\theta}{2\pi} \right) \in [0, 1]^2,$$

the formula (6.2) yields that the density of $(\rho^2, \frac{\theta}{2\pi})$ at a point (α, β) of $[0, 1]^2$ is given by the density $\frac{1}{\pi}$ of (U, V) at the inverse image

$$f^{-1}(\alpha, \beta) = (\sqrt{\alpha} \cos(2\pi\beta), \sqrt{\alpha} \sin(2\pi\beta))$$

of (α, β) under f , multiplied by the determinant of the matrix

$$Jf^{-1}(\alpha, \beta) = \begin{pmatrix} \frac{1}{2\sqrt{\alpha}} \cos(2\pi\beta) & -2\pi\sqrt{\alpha} \sin(2\pi\beta) \\ \frac{1}{2\sqrt{\alpha}} \sin(2\pi\beta) & 2\pi\sqrt{\alpha} \cos(2\pi\beta) \end{pmatrix}.$$

The determinant is π , so that the density of $(\rho^2, \frac{\theta}{2\pi})$ is equal to $\frac{\pi}{\pi} = 1$ uniformly over the square $[0, 1]^2$, meaning that $(\rho^2, \frac{\theta}{2\pi}) \sim \mathcal{U}_{[0,1]^2}$. We deduce by Box–Müller that

$$\sqrt{-2 \ln(\rho^2)} \begin{pmatrix} \cos(2\pi \frac{\theta}{2\pi}) \\ \sin(2\pi \frac{\theta}{2\pi}) \end{pmatrix} = \sqrt{\frac{-2 \ln(\rho^2)}{\rho^2}} \begin{pmatrix} \rho \cos(\theta) \\ \rho \sin(\theta) \end{pmatrix} = \sqrt{\frac{-2 \ln(\rho^2)}{\rho^2}} \begin{pmatrix} U \\ V \end{pmatrix}$$

is bivariate Gaussian. \square

6.2.3 Gaussian Vectors

A d -variate Gaussian vector X with zero mean and covariance matrix Γ can be simulated as follows:

- compute a square root of Γ , namely a matrix Σ such that $\Gamma = \Sigma \Sigma^T$;
- generate a d -variate Gaussian vector ε .

Then $X = \Sigma \varepsilon$ is $\mathcal{N}(0, \Gamma)$ -distributed. A square root $\Sigma = \Sigma^{cho}$ of Γ may be computed as the lower triangular matrix obtained by Cholesky decomposition of Γ . So, for $p = 1, \dots, d$:

$$\begin{aligned}\Sigma_{p,p} &= \sqrt{\Gamma_{p,p} - \sum_{r=1}^{p-1} \Sigma_{p,r}^2} \\ \Sigma_{q,p} &= \frac{\Gamma_{p,q} - \sum_{r=1}^{p-1} \Sigma_{p,r} \Sigma_{q,r}}{\Sigma_{p,p}} \quad \text{for } q = p+1, \dots, d.\end{aligned}\tag{6.3}$$

Alternatively, we can perform a spectral decomposition of Γ (as in a Principal Component Analysis), setting $\Sigma = \Sigma^{pca} = P \Lambda^{\frac{1}{2}}$, where $P \Lambda P^T = \Gamma$ with P orthonormal and Λ diagonal. A spectral decomposition is unique up to a reordering of the PCA factors² along with the eigenvalues of Γ .

Example 6.2.6 In the two-dimensional case with

$$\Gamma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

the Cholesky decomposition of Γ yields

$$\Sigma^{cho} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{pmatrix},$$

whereas the spectral decomposition of Γ reads

$$\Sigma^{pca} = \begin{pmatrix} \sigma_1 \sqrt{\frac{1+\rho}{2}} & \sigma_1 \sqrt{\frac{1-\rho}{2}} \\ \sigma_2 \sqrt{\frac{1+\rho}{2}} & -\sigma_2 \sqrt{\frac{1-\rho}{2}} \end{pmatrix}.$$

²Eigenfactors of Γ , columns of P .

6.3 Principles of Monte Carlo Simulation

We want to estimate $\Theta = \mathbb{E}[\phi(X)]$, where ϕ is some function on $\mathcal{D} \subseteq \mathbb{R}^d$ and X is a \mathcal{D} -valued random variable. Note that Θ can be expressed in integral form as

$$\Theta = \int_{\mathcal{D}} \phi(x) d\mathbb{P}^X(x).$$

Monte Carlo simulation is a general method for evaluating an integral as an expected value, based on the strong law of large numbers and the central limit theorem. It provides an unbiased estimate, and the error on the estimate is controlled within a confidence interval.

6.3.1 Law of Large Numbers and Central Limit Theorem

For x_j i.i.d. to X with $\mathbb{E}|\phi(X)| < +\infty$, we have

$$\frac{1}{m} \sum_{j=1}^m \phi(x_j) \xrightarrow{a.s.} \Theta \quad \text{as } m \rightarrow \infty.$$

If, moreover, $\sigma^2 = \text{Var}[\phi(X)] < +\infty$, then the normalized error converges in law to the Gaussian distribution:

$$\frac{\sqrt{m}}{\sigma} \left(\frac{1}{m} \sum_{j=1}^m \phi(x_j) - \Theta \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{as } m \rightarrow \infty.$$

6.3.2 Standard Monte Carlo Estimator and Confidence Interval

An unbiased estimate with m trials for Θ is thus given by

$$\Theta_m = \frac{1}{m} \sum_{j=1}^m \phi(x_j).$$

The variance of this estimate is $\frac{\sigma^2}{m}$, independent of the dimension d and with unbiased estimate (squared standard error)

$$\sigma_m^2 = \frac{1}{m-1} \left[\frac{1}{m} \sum_{j=1}^m \phi^2(x_j) - \Theta_m^2 \right].$$

The speed of convergence of Θ_m to Θ is $\frac{\sigma}{\sqrt{m}}$, which is estimated by σ_m . A confidence interval I at the threshold (confidence level) $1 - 2\alpha$ for Θ , i.e. so that $\mathbb{P}(\Theta \in I) = 1 - 2\alpha$, is given by

$$I = [\Theta_m - z_\alpha \sigma_m, \Theta_m + z_\alpha \sigma_m],$$

with $z_\alpha = \mathcal{N}^{-1}(1 - \alpha)$. For instance, if the threshold is set at 95 %, then $\alpha = 2.5\%$ and $z_\alpha \approx 1.96$. A natural stopping criterion consists in exiting a Monte Carlo loop when $z_\alpha \sigma_m$ becomes less than some fraction ϵ (e.g. $\epsilon = 10 \text{ bp} = 10^{-3}$) of Θ_m , so that one knows Θ up to a relative error equal to η , at the level of confidence $1 - \alpha$.

We briefly summarize some advantages and disadvantages of the Monte Carlo method.

- *Advantages:*

- We can implement this method very easily if we are able to simulate the random variable X .
- The estimate is unbiased and we can build a confidence interval justified by the central limit theorem.
- The properties of the method are not altered by irregularity of the function ϕ or dimensionality of the factor X .
- *Disadvantage:* Convergence is slow and therefore computing time can be very large.

6.4 Variance Reduction

The main disadvantage of the standard Monte Carlo simulation is its convergence rate in $\sigma m^{-\frac{1}{2}}$. So, to improve the accuracy by a factor of 10, we must increase the number m of simulations by a factor of 100. Alternatively, and this is the strategy of variance reduction, we may try to rewrite Θ in terms of a new random variable with less variance than $\phi(X)$.

6.4.1 Antithetic Variables

The idea of the antithetic variables method is to introduce some “good” correlation between the terms of the estimate. To fix the ideas, we consider simulation by the inverse method, based on uniform numbers u_j on $[0, 1]$ (assuming $d = 1$). In the antithetic variables method, we use each u_j twice, via $x_j = F_X^{-1}(u_j)$ and $\bar{x}_j = F_X^{-1}(1 - u_j)$. These two variables have the same law, but they are not independent. An unbiased estimate Θ is provided by

$$\overline{\Theta}_m = \frac{1}{2m} \sum_{j=1}^m (\phi(x_j) + \phi(\bar{x}_j)).$$

The variance of $\overline{\Theta}_m$ is

$$\bar{\sigma}_m^2 = \frac{1}{2m} (\sigma^2 + \text{Cov}(\phi(X), \phi(\overline{X}))),$$

with $\overline{X} = F_X^{-1}(1 - U)$. The following result gives a simple condition ensuring variance reduction by this method.

Proposition 6.4.1 *If ϕ is a monotone function, then $\bar{\sigma}_m^2 \leq \frac{1}{2} \frac{\sigma^2}{m}$.*

Proof Introducing an independent copy V of U , by the monotonicity of $\psi = \phi \circ F_X^{-1}$ we have:

$$(\psi(U) - \psi(V))(\psi(1 - U) - \psi(1 - V)) \leq 0,$$

so that

$$\mathbb{E}[\psi(U)\psi(1 - U) + \psi(V)\psi(1 - V)] - \mathbb{E}[\psi(U)\psi(1 - V) + \psi(V)\psi(1 - U)] \leq 0$$

and

$$\text{Cov}(\phi(X), \phi(\overline{X})) = \text{Cov}(\psi(U), \psi(1 - U)) \leq 0.$$

□

6.4.2 Control Variates

The idea of control variates is to introduce another payoff function ψ and/or factor model Y , for which we have an explicit price, and to estimate the correction term by Monte Carlo. Then we decompose

$$\Theta = \mathbb{E}[\phi(X)] = \mathbb{E}[\phi(X) - \psi(Y)] + \mathbb{E}[\psi(Y)],$$

where $\mathbb{E}[\psi(Y)]$ is known. An unbiased estimate with m trials of Θ is defined by

$$\widehat{\Theta}_m = \mathbb{E}[\psi(Y)] + \frac{1}{m} \sum_{j=1}^m (\phi(x_j) - \psi(y_j)),$$

with x_j i.i.d. for the law of X and y_j i.i.d. for the law of Y . The variance of $\widehat{\Theta}_m$ is given by

$$\begin{aligned} \widehat{\sigma}_m^2 &= \frac{1}{m} \text{Var}[\phi(X) - \psi(Y)] \\ &= \frac{1}{m} (\sigma^2 + \text{Var}[\psi(Y)] - 2 \text{Cov}(\phi(X), \psi(Y))). \end{aligned}$$

Variance reduction holds if the original random variable $\phi(X)$ and the control variate $\psi(Y)$ have a sufficient positive correlation.

6.4.3 Importance Sampling

The basic idea of importance sampling is to concentrate the sample distribution in the part of the space which contributes most to the payoff. To this end we introduce a changed probability measure \mathbb{Q} , with related expectation denoted by $\tilde{\mathbb{E}}$, such that

$$\text{supp}(\mathbb{P}) \cap \text{supp}(\phi(X)) \subseteq \text{supp}(\mathbb{Q}) \subseteq \text{supp}(\mathbb{P}) \quad (6.4)$$

or, equivalently in terms of the Radon-Nikodym density of \mathbb{Q} with respect to \mathbb{P} :³

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \mu,$$

for some nonnegative random variable μ with unit mean under \mathbb{P} and such that $\frac{\phi(X)}{\mu}$ is integrable under \mathbb{Q} , where for $\omega \notin \text{supp}(\mathbb{Q})$ the ratio $\frac{\phi(X)}{\mu}(\omega)$ is understood as 0 in $\tilde{\mathbb{E}}[\frac{\phi(X)}{\mu}] = \int_{\Omega} \frac{\phi(X)}{\mu}(\omega) \mathbb{Q}(d\omega)$. Note that we don't restrict ourselves to equivalent measures \mathbb{Q} , but open the door to more general measures \mathbb{Q} satisfying (6.4) in which the right-hand side inclusion expresses the property that \mathbb{Q} is absolutely continuous with respect to \mathbb{P} , but in which the left-hand side is less restrictive than \mathbb{P} absolutely continuous with respect to \mathbb{Q} . This is crucial, since the whole idea of importance sampling is precisely to concentrate the sample distribution in the part of the space which contributes most to the payoff, in particular by the use of measures \mathbb{Q} for which the left-hand side inclusion is strict in (6.4).

Given a measure \mathbb{Q} satisfying (6.4) or the equivalent conditions in terms of μ , it follows much like as in (4.62) (case $t = 0$ there) that

$$\Theta = \mathbb{E}[\phi(X)] = \tilde{\mathbb{E}}\left[\frac{\phi(X)}{\mu}\right]. \quad (6.5)$$

The Monte Carlo estimate of Θ related to (6.5) is

$$\tilde{\Theta}_m = \frac{1}{m} \sum_{j=1}^m \frac{\phi(\tilde{x}_j)}{\mu_j},$$

where the \tilde{x}_j are i.i.d. for the law of X under \mathbb{Q} .

Remark 6.4.2 As compared with control variate an inconvenience of importance sampling is that one must simulate X under the changed measure \mathbb{Q} . Since in practice X is typically given as the value at T of some process defined in SDE form, this concretely means that a prerequisite for using this method is to implement the change of measure in the SDE that defines X . This is done in practice by means of Girsanov transforms, which are used to rewrite the SDE in terms of fundamental martingales (Brownian motions and/or compensated jump measures) with respect to \mathbb{Q} ; see Sects. 3.4 and 12.3.2.

³See Sects. 3.4 and 4.3.1.1.

The goal of the exercise is then to pick an admissible measure \mathbb{Q} minimizing the \mathbb{Q} -variance of $\tilde{\Theta}_m$. Since

$$\tilde{\mathbb{E}}\tilde{\Theta}_m = \tilde{\mathbb{E}}\left[\frac{\phi(X)}{\mu}\right] = \Theta,$$

which doesn't depend on μ , this is equivalent to minimizing

$$\tilde{\mathbb{E}}\tilde{\Theta}_m^2 = \frac{1}{m}\tilde{\mathbb{E}}\left(\frac{\phi(X)}{\mu}\right)^2.$$

The minimum variance (null in the case of a nonnegative payoff $\phi(X)$) is reached for $\mu \propto |\phi(X)|$, so (since an admissible μ must have unit mean under \mathbb{P})

$$\mu = \mu^\sharp := \frac{|\phi(X)|}{\mathbb{E}|\phi(X)|}. \quad (6.6)$$

But usually μ^\sharp cannot be computed explicitly. Note in particular that, for $\phi > 0$, the number Θ that we are looking for sits in the denominator of the right-hand side of (6.6). In practice we use

$$\mu^\flat = \frac{|\psi(Y)|}{\mathbb{E}|\psi(Y)|},$$

for approximate payoff functions ψ and/or factor model Y obeying the same rationale as for control variate in Sect. 6.4.2. So $\phi(X)$ and $\psi(Y)$ should have as large a positive correlation as possible and $\mathbb{E}|\psi(Y)|$ should be computable explicitly.

6.4.4 Efficiency Criterion

We now introduce a heuristic criterion to compare the efficiency of various simulation schemes, with or without variance reduction. This criterion takes into account not only the accuracy (variance), but also the computation time required by the simulation for each scheme. The efficiency of a method \tilde{M} with respect to a method M is defined as

$$\mathcal{E} = \lim_{m, \tilde{m} \rightarrow \infty} \frac{\sigma_m}{\tilde{\sigma}_{\tilde{m}}} \sqrt{\frac{t_m}{\tilde{t}_{\tilde{m}}}},$$

where σ_m and t_m (respectively $\tilde{\sigma}_{\tilde{m}}$ and $\tilde{t}_{\tilde{m}}$) are the standard error and the computation times of method M based on m simulation runs (respectively \tilde{M} based on \tilde{m} simulation runs). Method \tilde{M} is considered to be more efficient than method M if $\mathcal{E} \geq 1$. For instance, $\mathcal{E} = 3$ means that for a given computation time method \tilde{M} is three times more accurate than method M , or that for a given accuracy method \tilde{M} is nine times faster than method M . Assuming computation times proportional to the sample sizes (so that $t_m = km$, where k is a factor which expresses the complexity

of the algorithm for method M , and likewise $\tilde{t}_{\tilde{m}} = \tilde{k}\tilde{m}$ for method \tilde{M}), then we have

$$\mathcal{E} \sim_{m, \tilde{m} \rightarrow \infty} \frac{\sigma}{\tilde{\sigma}} \sqrt{\frac{k}{\tilde{k}}}.$$

Efficiency in the sense of this criterion \mathcal{E} is thus asymptotically independent of the sample size.

6.5 Quasi Monte Carlo

Quasi Monte Carlo consists in estimating

$$\Theta = \mathbb{E}[\psi(U)] = \int_{[0,1]^d} \psi(u) du$$

by $\frac{1}{m} \sum_{j=1}^m \psi(u_j)$, where u is a d -variate low-discrepancy sequence. Unlike genuine Monte Carlo, Quasi Monte Carlo doesn't provide a confidence interval. The empirical variance of the sample is not meaningful because successive terms of the sequence are not independent. This is due to the construction of low-discrepancy sequences.

However, one has the following deterministic inequality. The definition of the Hardy-Krause variation of ψ which appears in its statement is rather technical. In dimension one, its value for a regular function ψ coincides with the usual notion $\int_{[0,1]} |\psi'(u)| du$.

Theorem 6.5.1 (Koksma-Hlawka inequality) *For every sequence u of points in \mathbb{R}^d and for every $m \geq 1$, one has:*

$$\left| \frac{1}{m} \sum_{j=1}^m \psi(u_j) - \int_{[0,1]^d} \psi(u) du \right| \leq V(\psi) D_m(u),$$

where $V(\psi)$ denotes the Hardy-Krause variation of ψ .

The Koksma-Hlawka inequality gives an a priori deterministic bound for the error in the approximation of $\int_{[0,1]^d} \psi(x) dx$ by $\frac{1}{m} \sum_{j=1}^m \psi(u_j)$. This error is expressed in terms of the discrepancy of the sequence u and of the variation of the function ψ . Through this inequality we understand the interest in having sequences with discrepancy D_m as small as possible. But it is often difficult to calculate or even to estimate the variation of ψ . Moreover, since for large dimensions d the asymptotic bound $\frac{(\ln m)^d}{m}$ of a low-discrepancy sequence may only be meaningful for very large values of m , and because $\frac{(\ln m)^d}{m}$ increases exponentially with d , we see that for large d the bound in the Koksma-Hlawka inequality gives no relevant information for realistic sample sizes m . Fortunately, the effective dimension of a problem,

in the sense of the number of risk factors that explain most of its variance, is often much lower than its nominal dimension d .

We benefit most from a low-discrepancy sequence by assigning the main risk factors of the problem, ordered by decreasing amount of explained variance, to the successive components of the points of a multivariate low-discrepancy sequence. Thus, even though we use a low-discrepancy sequence in the nominal dimension d of the problem, which may be high, the circumstance that the first coordinates of the quasi-random points are assigned to the main risk factors of the problem avoids many of the drawbacks generally associated with high-dimensional low-discrepancy sequences (see Fig. 6.1).

6.6 Greeking by Monte Carlo

Price sensitivities, or Greeks, are a key issue in financial modeling. Indeed, unless exotic products are considered, derivative prices are created by supply-and-demand in the market. As we will see in Chap. 9, liquid market prices are in fact used rather than those produced by models, in a reverse-engineering mode, as calibration input data. Greeks, on the contrary, can only be computed within models. Now, in many cases, Greeks, like prices, can also be put in the form $\Theta = \mathbb{E}[\phi(X)]$, so that the Monte Carlo pricing techniques we have seen so far can also be used for Greeking. This is briefly illustrated in this section in the problem of computing, by Monte Carlo, the delta of an option in the Black–Scholes model. We thus want to compute $\Delta_0 = \partial_s \mathbb{E}[\phi(S_T^s)]$, where S_T^s is the value at maturity of an underlying S with initial condition s at time 0. One obvious method consists in repricing the payoff by Monte Carlo for a perturbed initial condition in order to get a finite difference estimate for Δ_0 . But such a resimulation procedure is costly and biased. In many cases, direct approaches without resimulation are possible:

- by differentiation of the payoff, provided the latter is smooth enough;
- by differentiation of the transition probability density $p_T(s, S)$ of S , provided the latter exists and is smooth enough.

In general these two approaches rely, respectively, on the theory of stochastic flows [60, 133] and on Malliavin calculus [124, 125]. However, in the Black–Scholes model the related computations are, as we will now see, elementary.

6.6.1 Finite Differences

For fixed $\alpha > 0$, one can approximate Δ_0 at the order one of consistency⁴ by the uncentered finite difference

$$\frac{1}{\alpha s} (\mathbb{E}[\phi(S_T^{(1+\alpha)s})] - \mathbb{E}[\phi(S_T^s)])$$

⁴See Sect. 8.2.1.2.

or, at an improved order two of consistency, by the centered finite difference

$$\frac{1}{2\alpha s} (\mathbb{E}\phi(S_T^{(1+\alpha)s}) - \mathbb{E}\phi(S_T^{(1-\alpha)s})). \quad (6.7)$$

The expectations in (6.7) can be estimated by Monte Carlo. In order to decrease the variance, common random numbers should be used to estimate both expectations in (6.7). In the Black–Scholes model, $S_T = s \exp(bT + \sigma W_T)$, we thereby obtain the following estimate for Δ_0 :

$$\frac{1}{2\alpha sm} \sum_{j=1}^m (\phi((1+\alpha)s e^{bT+\sigma\sqrt{T}\varepsilon_j}) - \phi((1-\alpha)s e^{bT+\sigma\sqrt{T}\varepsilon_j})),$$

where the ε_j are independent Gaussian draws.

6.6.2 Differentiation of the Payoff

Alternatively, in case ϕ is sufficiently regular and S_T^s is differentiable with respect to s , Δ_0 may be computed, assuming commutation of the expectation and differentiation operators, by

$$\Delta_0 = \partial_s \mathbb{E}\phi(S_T^s) = \mathbb{E}\partial_s \phi(S_T^s) = \mathbb{E}[\phi'(S_T^s)\partial_s S_T^s] \quad (6.8)$$

(see [132, 133]). In multiplicative models such as Black–Scholes or more general homogeneous models for which $\frac{S_T}{s}$ doesn't depend on s , one has $\partial_s S_T^s = \frac{S_T^s}{s}$, so that

$$\Delta_0 = \frac{1}{s} \mathbb{E}[\phi'(S_T^s) S_T^s].$$

Note that ϕ here needs to be differentiable, at least outside a $\mathbb{P}^{S_T^s}$ -null set. This is, for instance, the case with a vanilla option since ϕ' (in the sense of distributions) is then defined as a step function and S_T^s has no atoms in Black–Scholes. But this is not the case for a step function ϕ , since ϕ' is then only defined as a Dirac mass at the point of discontinuity of ϕ . This method is thus applicable to the computation of the delta of a call option, but not the delta of a digital option (or gamma of a vanilla call option).

6.6.3 Differentiation of the Density

Assume S admits a transition probability density $p_T(s, S)$ from $(0, s)$ to (T, S) , differentiable in its first argument s (the initial condition at time 0 of the process S_t).

Then, under mild regularity conditions,

$$\Delta_0 = \int_{S>0} \phi(S) \partial_1 p_T(s, S) dS = \int_{S>0} \phi(S) \frac{\partial_1 p_T(s, S)}{p_T(s, S)} p_T(s, S) dS,$$

so that

$$\Delta_0 = \mathbb{E}[\phi(S_T^s) \partial_1 \ln(p_T(s, S_T^s))].$$

In the Black–Scholes model, we saw in Example 6.2.4 that

$$p_T(s, S) = \frac{1}{S\sqrt{2\pi T}\sigma} e^{-\frac{(\ln(\frac{S}{s}) - bT)^2}{2\sigma^2 T}}, \quad \partial_1 \ln(p_T(s, S_T^s)) = \frac{W_T}{s\sigma T},$$

so that

$$\Delta_0 = \mathbb{E}\left[\phi(S_T^s) \frac{W_T}{s\sigma T}\right].$$

6.7 Monte Carlo Algorithms for Vanilla Options

6.7.1 European Call, Put or Digital Option

Monte Carlo provides price and delta estimates with a confidence interval. The Quasi Monte Carlo method only provides price and delta estimates, without a confidence interval. The option price and delta at time 0 are

$$\Pi_0 = \mathbb{E}[e^{-rT} \phi(S_T)], \quad \Delta_0 = \partial_s \mathbb{E}[e^{-rT} \phi(S_T)].$$

The corresponding Monte Carlo estimates are written as

$$\begin{aligned} \widehat{\Pi}_0 &= \frac{1}{m} e^{-rT} \sum_{j=1}^m \pi^j, \\ \widehat{\Delta}_0 &= \frac{1}{m} e^{-rT} \sum_{j=1}^m \delta^j \quad \text{with } \delta^j = \partial_s \pi^j. \end{aligned}$$

The expressions of π^j and δ^j are detailed for each option with strike K below, assuming a Black–Scholes underlying.

- **Put:** The payoff is $(K - S_T)^+$, hence

$$\pi^j = (K - S_T^j)^+, \quad \delta^j = \begin{cases} -\partial_s S_T^j = -\frac{S_T^j}{s} & \text{if } \pi^j \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

- **Call** (see Fig. 6.2): The payoff is $(S_T - K)^+$. The call-put parity relations for the price and delta are:

$$\Pi_0^+ = \Pi_0^- + se^{-qt} - Ke^{-rT}, \quad \Delta_0^+ = \Delta_0^- + e^{-qt},$$

where Π_0^\pm and Δ_0^\pm denote the call/put prices and deltas at time 0. These relations may be used for the call, in order to limit the variance, if the call is in-the-money.

- **Digital option:** The payoff is $R\mathbb{1}_{\{S_T \geq K\}}$ for some rebate R , hence

$$\pi^j = R\mathbb{1}_{\{S_T^j \geq K\}}.$$

A Monte Carlo finite difference estimate of the delta is obtained with

$$\delta^j = \frac{1}{2\alpha s} [\phi((1+\alpha)S_T^j) - \phi((1-\alpha)S_T^j)].$$

6.7.1.1 Adding Jumps

We will now add jumps in S , postulating the risk-neutral Merton model of Sect. 5.1.3:

$$S_T = se^{aT + \sigma W_T} \prod_{l=1}^{N_T} (1 + J_l),$$

where $a = b - \lambda \bar{J}$, where N_t is a Poisson process with intensity λ and where the J_l are i.i.d. such that $\ln(1 + J_l)$ is $\mathcal{N}(\alpha, \beta)$ -distributed. Essentially nothing changes except for the fact that S_T^j of Sect. 6.7.1 is now given by

$$S_T^j = se^{aT + \sigma \sqrt{T} \varepsilon_j} \prod_{l=1}^{N_T^j} \exp(\alpha + \sqrt{\beta} \varepsilon_j^l),$$

where

- N_T^j is an independent $\mathcal{P}_{\lambda T}$ -draw;
- ε_j and the ε_j^l are independent Gaussian draws.

6.7.2 Call on Maximum, Put on Minimum, Exchange or Best of Options

Assume underlying assets evolve according to the following risk-neutral bivariate Black–Scholes model: $S_0^1 = s_1$, $S_0^2 = s_2$ and, for $t \in [0, T]$,

$$\begin{cases} dS_t^1 = S_t^1 (\kappa_1 dt + \sigma_1 dW_t^1) \\ dS_t^2 = S_t^2 (\kappa_2 dt + \sigma_2 dW_t^2), \end{cases}$$

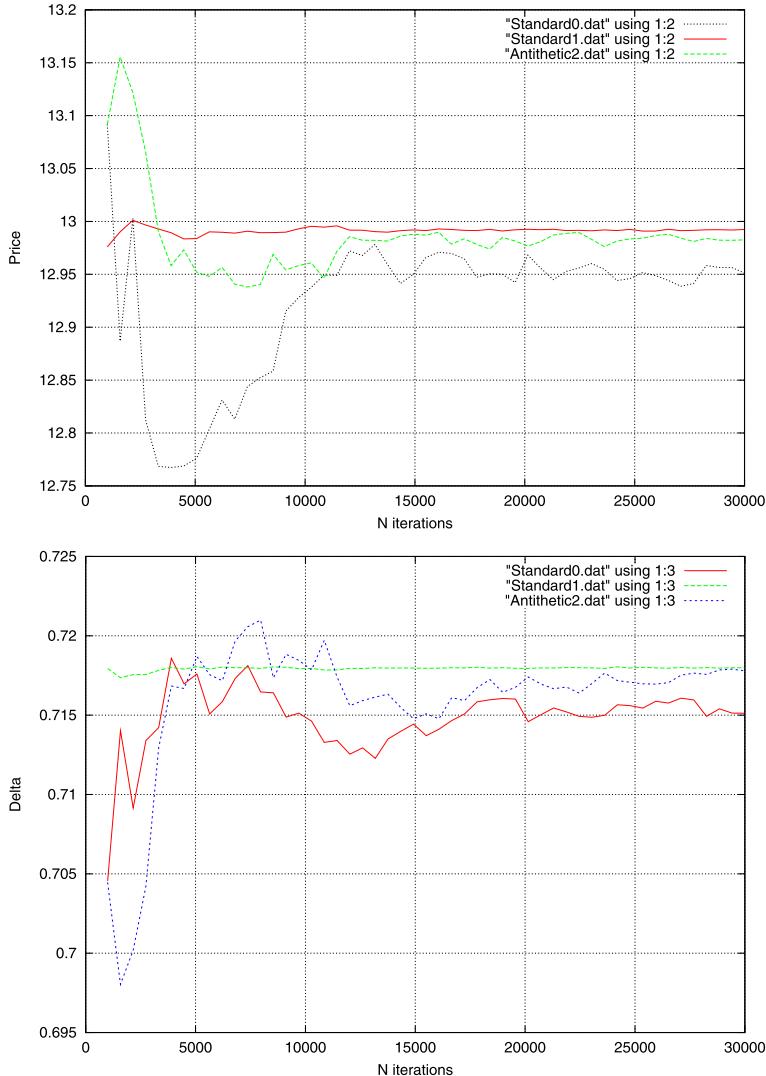


Fig. 6.2 European vanilla call priced in the Black–Scholes model by Monte Carlo, first using L’Ecuyer’s pseudo-random numbers generator, then with antithetic variables and finally by Quasi Monte Carlo based on the one-dimensional Sobol sequence

with $\kappa_l = r - q_l$, where W^1 and W^2 are two Brownian motions with correlation ρ . So, in terms of a third Brownian motion \tilde{W}^2 independent from W^1 :

$$\begin{cases} S_T^1 = s_1 \exp(b_1 T + \sigma_{1,1} W_T^1) \\ S_T^2 = s_2 \exp(b_2 T + \sigma_{2,1} W_T^1 + \sigma_{2,2} \tilde{W}_T^2), \end{cases}$$

with

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \kappa_1 - \frac{\sigma_1^2}{2} \\ \kappa_2 - \frac{\sigma_2^2}{2} \end{pmatrix}, \quad \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{pmatrix}.$$

Remark 6.7.1 The latter matrix arises by Cholesky decomposition (6.3) of the covariance matrix $h^{-1} \text{Cov}_t \begin{pmatrix} \sigma_1 dW_t^1 \\ \sigma_2 dW_t^2 \end{pmatrix}$ (see Corollary 3.2.13).

The price of an option at time 0 is

$$\Pi_0 = \mathbb{E}[e^{-rT} \phi(S_T^1, S_T^2)],$$

where ϕ denotes a payoff function. The deltas at time 0 are given by

$$\Delta_0^1 = \partial_{s_1} \mathbb{E}[e^{-rT} \phi(S_T^1, S_T^2)], \quad \Delta_0^2 = \partial_{s_2} \mathbb{E}[e^{-rT} \phi(S_T^1, S_T^2)].$$

The price and delta estimates are written as

$$\widehat{\Pi}_0 = \frac{1}{m} e^{-rT} \sum_{j=1}^m \pi^j, \quad \widehat{\Delta}_0^l = \frac{1}{m} e^{-rT} \sum_{j=1}^m \partial_{s_l} \pi^j = \frac{1}{m} e^{-rT} \sum_{j=1}^m \delta_l^j,$$

for $l = 1, 2$. The values for π^j and δ_l^j are detailed below for each option with strike K .

- **Put on the Minimum:** The payoff is $(K - \min(S_1, S_2))^+$, so that

$$\begin{aligned} \pi^j &= (K - \min(S_T^{1,j}, S_T^{2,j}))^+ \\ \delta_1^j &= \begin{cases} -\exp(b_1 T + \sigma_{1,1} W_T^1) & \text{if } \pi^j > 0 \text{ and } S_T^{1,j} \leq S_T^{2,j} \\ 0 & \text{otherwise} \end{cases} \\ \delta_2^j &= \begin{cases} -\exp(b_2 T + \sigma_{2,1} W_T^1 + \sigma_{2,2} \tilde{W}_T^2) & \text{if } \pi^j > 0 \text{ and } S_T^{1,j} \geq S_T^{2,j} \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

- **Call on the Maximum:** The payoff is $(\max(S_1, S_2) - K)^+$, so that

$$\begin{aligned} \pi^j &= (\max(S_T^{1,j}, S_T^{2,j}) - K)^+ \\ \delta_1^j &= \begin{cases} \exp(b_1 T + \sigma_{1,1} W_T^1) & \text{if } \pi^j > 0 \text{ and } S_T^{1,j} \geq S_T^{2,j} \\ 0 & \text{otherwise} \end{cases} \\ \delta_2^j &= \begin{cases} \exp(b_2 T + \sigma_{2,1} W_T^1 + \sigma_{2,2} \tilde{W}_T^2) & \text{if } \pi^j > 0 \text{ and } S_T^{1,j} \leq S_T^{2,j} \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

- **Exchange Option:** The payoff is $(S_1 - K S_2)^+$, so that

$$\begin{aligned}\pi^j &= (S_T^1 - K S_T^2)^+ \\ \delta_1^j &= \begin{cases} \exp(b_1 T + \sigma_{1,1} W_T^1) & \text{if } \pi^j > 0 \\ 0 & \text{otherwise} \end{cases} \\ \delta_2^j &= \begin{cases} -K \exp(b_2 T + \sigma_{2,1} W_T^1 + \sigma_{2,2} \tilde{W}_T^2) & \text{if } \pi^j > 0 \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

- **Best of Option:** The payoff is $[\max(S_1 - K_1, S_2 - K_2)]^+$, so that

$$\begin{aligned}\pi^j &= [\max(S_T^1 - K_1, S_T^2 - K_2)]^+ \\ \delta_1^j &= \begin{cases} \exp(b_1 T + \sigma_{1,1} W_T^1) & \text{if } \pi^j > 0 \text{ and } S_T^{1,j} - K_1 \geq S_T^{2,j} - K_2 \\ 0 & \text{otherwise.} \end{cases} \\ \delta_2^j &= \begin{cases} \exp(b_2 T + \sigma_{2,1} W_T^1 + \sigma_{2,2} \tilde{W}_T^2) & \text{if } \pi^j > 0 \text{ and } S_T^{1,j} - K_1 \leq S_T^{2,j} - K_2 \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

6.8 Simulation of Processes

Until now we were able to simulate S_T directly, without time-discretization of the associated SDE. However, to deal with more complex models, or even with simple models with path-dependent payoffs, we need to simulate the whole trajectory of the underlying between the pricing time 0 and the maturity T . Simulating trajectories is also necessary for testing the performances of a hedging scheme. In this case, the “path-dependent payoff” that interests us consists of the profit-and-loss at maturity of a delta-hedged option position.

6.8.1 Brownian Motion

Recall from Sect. 2.3.1 that a standard Brownian motion W is a continuous process starting from 0 at time 0 with the properties that, for $0 \leq s < t$, the increment $W_t - W_s$ is independent of \mathcal{F}_s^W and is normally distributed with mean zero and variance $t - s$. By the Cholesky decomposition (6.3) (case $d = 3$ there) of the covariance matrix of (W_s, W_t, W_T) , we can represent in law

$$\begin{cases} W_T = \sqrt{T} \varepsilon_T \\ W_t = \sqrt{t} \left(\sqrt{\frac{t}{T}} \varepsilon_T + \sqrt{1 - \frac{t}{T}} \varepsilon_t \right) \\ W_s = \sqrt{s} \left(\sqrt{\frac{s}{T}} \varepsilon_T + \rho \varepsilon_t + \sqrt{1 - \frac{s}{T} - \rho^2} \varepsilon_s \right), \end{cases}$$

for $(\varepsilon_s, \varepsilon_t, \varepsilon_T)$ standard trivariate Gaussian and for ρ in the last line defined by $s = \sqrt{st}(\sqrt{\frac{ts}{T^2}} + \sqrt{1 - \frac{t}{T}}\rho)$. Therefore, in particular,

$$\mathcal{L}(W_t | W_T) = \mathcal{N}\left(\frac{t}{T}W_T, \frac{t(T-t)}{T}\right) \quad (6.9)$$

and

$$\text{Cov}(W_s, W_t | W_T) = \sqrt{ts}\sqrt{1 - \frac{t}{T}}\rho = \sqrt{ts}\left(\sqrt{\frac{s}{t}} - \sqrt{\frac{ts}{T^2}}\right) = s\left(1 - \frac{t}{T}\right). \quad (6.10)$$

We now present two approaches for simulating a Brownian path on a time-grid $t_0 = 0 < t_1 < \dots < t_n = T$. Let $h_i = t_{i+1} - t_i$ for $i = 0 \dots n - 1$, so that $h_i = h = \frac{T}{n}$ in the case of a uniform time-grid.

6.8.1.1 Forward Simulation

The forward simulation of W is defined by $W_0 = 0$ and, for $0 \leq i \leq n - 1$,

$$W_{t_{i+1}} = W_{t_i} + \sqrt{h_i}\varepsilon_i$$

for independent Gaussian draws ε_i . Note that independent simulation of each W_{t_i} as $\sqrt{t_i}\varepsilon_i$ would not give the right variance for $(W_{t_{i+1}} - W_{t_i})$.

6.8.1.2 Backward Simulation

The backward simulation of W is based on the following Brownian Bridge property, which follows from (6.9):

$$\mathcal{L}(W_{\frac{t+s}{2}} | W_s = x, W_t = y) = \mathcal{N}\left(\frac{x+y}{2}, \frac{t-s}{4}\right).$$

For this algorithm one must choose n as a power of 2. The first step is directly from 0 to T , letting $W_T = \sqrt{T}\varepsilon_T$. Intermediates steps are then filled in by taking successive subdivisions of the time intervals into halves. For every $v \geq 0$, assuming we have already simulated $W_{\frac{iT}{2^v}}$ for $0 \leq i \leq 2^v$, for every $0 \leq i \leq 2^v - 1$ we simulate

$$W_{\frac{(2i+1)T}{2^{v+1}}} = \frac{1}{2}(W_{\frac{iT}{2^v}} + W_{\frac{(i+1)T}{2^v}}) + \sqrt{\frac{T}{2^{(v+2)}}}\varepsilon_{\frac{(2i+1)T}{2^{v+1}}}$$

for independent Gaussian draws ε . This algorithm can also be adapted to a nonuniform time mesh by using the law of the Brownian Bridge at any time u between s and t :

$$\mathcal{L}(W_u | W_s = x, W_t = y) = \mathcal{N}\left(\frac{t-u}{t-s}x + \frac{u-s}{t-s}y, \frac{(t-u)(u-s)}{t-s}\right).$$

Both the forward and the backward simulations of W require a vector of n independent Gaussian draws per trajectory. This vector can be computed from either n successive values of a univariate pseudo-random sequence, or from an n -variate quasi-random point. Regarding the backward simulation scheme for W , note that the values of W successively determined on each trajectory are drawn by order of decreasing variance. In case an n -variate quasi-random generator is used for generating the ε , this decreasing variance property means that the first components of every simulated quasi-random point, which are also “the more uniform ones” (recall Fig. 6.1), are used for simulating the main directions of risk of the trajectory of W .

6.8.2 Diffusions

In the case of a Brownian SDE with constant coefficients, X_t is a function of W_t . A path of X can then be simulated exactly (without time-discretization error) as the corresponding function of a path of W . This is, for instance, the case with the risk-neutral Black–Scholes model stated in log-returns variable $X = \ln(S)$. As soon as an SDE has non-constant coefficients (and cannot simply be transformed into an SDE with constant coefficients), the simplest way to simulate it is by discretizing time and using the results of the constant coefficient case locally on small time intervals. However, this entails a time-discretization error.

Let us thus consider a d -dimensional diffusion

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t \quad (6.11)$$

(jumps will be added later). As is well known and reviewed in Sect. 3.3.3, this SDE admits a unique strong solution X over $[0, T]$ under suitable Lipschitz and growth conditions on the coefficients. The two best-known time-discretization schemes for (6.11) are the Euler and the Milstein schemes.

6.8.2.1 Euler Scheme

The Euler scheme is defined by $\widehat{X}_0 = X_0$ and, for every $i = 0, \dots, n - 1$,

$$\widehat{X}_{t_{i+1}} = \widehat{X}_{t_i} + b(t_i, \widehat{X}_{t_i}) h_i + \sigma(t_i, \widehat{X}_{t_i})(W_{t_{i+1}} - W_{t_i}).$$

Simulation is obtained with a forward algorithm by $\widehat{X}_0 = X_0$ and, for $i = 0, \dots, n - 1$,

$$\widehat{X}_{t_{i+1}} = \widehat{X}_{t_i} + b(t_i, \widehat{X}_{t_i}) h_i + \sigma(t_i, \widehat{X}_{t_i}) \sqrt{h_i} \varepsilon_i,$$

for i.i.d. standard Gaussian draws ε_i . Let $h_i = h$ for notational simplicity.

Proposition 6.8.1 (see [176, 226])

(i) *We have*

$$\mathbb{E} \left(\sup_{0 \leq i \leq n} |X_{ih} - \widehat{X}_{ih}|^2 \right) = O(h)$$

$$\sup_{0 \leq i \leq n} |X_{ih} - \widehat{X}_{ih}| = o(h^{\frac{1}{2}-\alpha}) \quad \text{almost surely, for every } \alpha > 0.$$

(ii) **Convergence in law.** For b, σ and ϕ regular enough, we have

$$|\mathbb{E}\phi(X_T) - \mathbb{E}\phi(\widehat{X}_T)| = O(h).$$

Thus L^2 -convergence is of order $h^{\frac{1}{2}}$, almost sure convergence is of order $h^{\frac{1}{2}}$ (essentially) and convergence in law is linear in h in regular cases. For pricing applications, the most relevant notion of convergence is that of convergence in law.

Continuous Euler scheme The continuous Euler scheme is the continuous-time approximation scheme $(\overline{X}_t)_{t \geq 0}$ defined by interpolation of the Euler scheme by a Brownian Bridge between (t_i, \widehat{X}_{t_i}) and $(t_{i+1}, \widehat{X}_{t_{i+1}})$, for $i = 0, \dots, n-1$. So, on $[t_i, t_{i+1}]$,

$$\overline{X}_t = \widehat{X}_{t_i} + b(t_i, \widehat{X}_{t_i})(t - t_i) + \sigma(t_i, \widehat{X}_{t_i})B_t^i,$$

where B^i is a Brownian Bridge on $[t_i, t_{i+1}]$ such that $\overline{X} = \widehat{X}$ at t_i and t_{i+1} .

6.8.2.2 Milstein Scheme

Assuming a one-dimensional and time-homogeneous diffusion, the Milstein Scheme for X appears as follows: $\widetilde{X}_0 = X_0$ and, for every $i = 0, \dots, n-1$,

$$\begin{aligned} \widetilde{X}_{t_{i+1}} &= \widetilde{X}_{t_i} + \left(b(\widetilde{X}_{t_i}) - \frac{1}{2}\sigma'(\widetilde{X}_{t_i})\sigma(\widetilde{X}_{t_i}) \right)h_i + \sigma(\widetilde{X}_{t_i})(W_{t_{i+1}} - W_{t_i}) \\ &\quad + \frac{1}{2}\sigma'(\widetilde{X}_{t_i})\sigma(\widetilde{X}_{t_i})(W_{t_{i+1}} - W_{t_i})^2. \end{aligned}$$

Simulation is obtained with a forward algorithm by

$$\widetilde{X}_{t_{i+1}} = \widetilde{X}_{t_i} + \left(b(\widetilde{X}_{t_i}) - \frac{1}{2}\sigma'(\widetilde{X}_{t_i})\sigma(\widetilde{X}_{t_i}) \right)h_i + \sigma(\widetilde{X}_{t_i})\sqrt{h_i}\varepsilon_i + \frac{1}{2}\sigma'(\widetilde{X}_{t_i})\sigma(\widetilde{X}_{t_i})h_i\varepsilon_i^2$$

for independent Gaussian draws ε_i . Let $h_i = h$ for notational simplicity. The following results show that convergence in law of the Milstein Scheme is again linear in h in regular cases. But the rates of L^2 - and almost sure convergences are improved with respect to the Euler scheme.

Proposition 6.8.2 (see [176, 226])

(i) *Assuming b and σ sufficiently enough, we have*

$$\begin{aligned} \mathbb{E}\left(\sup_{0 \leq i \leq n} |X_{ih} - \tilde{X}_{ih}|^2\right) &= O(h^2) \\ \sup_{0 \leq i \leq n} |X_{ih} - \tilde{X}_{ih}| &= o(h^{1-\alpha}) \quad \text{almost surely, for every } \alpha > 0. \end{aligned}$$

(ii) **Convergence in law.** Same statement as for the Euler scheme.

Example 6.8.3 We consider the risk-neutral Heston model

$$\begin{cases} dV_t = \lambda(\theta - V_t) dt + \eta\sqrt{V_t} dB_t \\ dS_t = S_t(\kappa dt + \sqrt{V_t} dW_t) \end{cases} \quad (6.12)$$

with $d\langle W, B \rangle = \rho dt$. The Heston SDE is not Lipschitz, but one can show existence and uniqueness for a strong solution (V, S) of (6.12). A commonly used time-discretization scheme for (6.12) consists of a Milstein scheme for V and an Euler scheme for $X = \ln(S)$, so: $\tilde{V}_0 = V_0$, $\hat{X}_0 = x = \ln(S_0)$, and, for every $i = 0, \dots, n-1$,

$$\begin{cases} \tilde{V}_{t_{i+1}} - \tilde{V}_{t_i} = \left(\frac{\eta^2}{4} - \lambda(\tilde{V}_{t_i} - \theta)\right)h_i + \eta\sqrt{\tilde{V}_{t_i}^+ h_i} \varepsilon_i + \frac{h_i \eta^2 \varepsilon_i^2}{4} \\ \hat{X}_{t_{i+1}} - \hat{X}_{t_i} = \left(\kappa - \frac{\tilde{V}_{t_i}}{2}\right)h_i + \sqrt{\tilde{V}_{t_i}^+ h_i} (\rho \varepsilon_i + \sqrt{1 - \rho^2} \tilde{\varepsilon}_i), \end{cases} \quad (6.13)$$

for i.i.d. standard Gaussian pairs $(\varepsilon_i, \tilde{\varepsilon}_i)$.

Note that for well-definedness of the scheme we took the positive part of $\tilde{V}_{t_i}^+$ under the square roots in (6.13) (another possibility would have been to use $|\tilde{V}_{t_i}|$), which induces a specific simulation bias. The Milstein scheme \tilde{V} has a better trajectoryal convergence to V than the Euler scheme \hat{V} and so \tilde{V} is less prone to take negative values than \hat{V} ; hence a less biased scheme results. Time-discretization of the Heston model, and of affine processes more generally, is a delicate issue; see [133].

6.8.3 Adding Jumps

We now consider the \mathbb{R}^d -valued jump-diffusion

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t + \delta(t, X_{t-}, J_{(t)}) dN_t, \quad (6.14)$$

with jump intensity $\lambda(t, X_t)$ of a point process N and distribution $w(t, X_{t-}, dx)$ of $J_{(t)}$. As reviewed in Sect. 3.3.3 (see also Chap. 12), (6.14) admits a unique strong

solution X over $[0, T]$ under suitable Lipschitz and growth conditions on the coefficients.

6.8.3.1 Poisson Process

First, we consider the case of a Poisson process N_t with constant intensity λ . To simulate a path of N_t over $[0, T]$, one possibility is to simulate the successive i.i.d. \mathcal{E}_λ -sojourn times of N_t until time T (see Sect. 2.2.1). Alternatively, we can first simulate $N_T \sim \mathcal{P}_{\lambda T}$ (see Sect. 6.2.1) and then use the property that, conditionally on N_T , the jump times of the process on $[0, T]$ follow the order statistics of N_T i.i.d. $\mathcal{U}_{[0, T]}$ -random variables (see e.g. [71]).

6.8.3.2 Euler Scheme

To simulate (6.14) at the times $0 < t_1 < \dots < t_n = T$, we set $\widehat{X}_0 = X_0$ and, for $i = 0, \dots, n - 1$:

- i. we simulate

$$\check{X}_{t_{i+1}} = \widehat{X}_{t_i} + b(t_i, \widehat{X}_{t_i})h_i + \sigma(t_i, \widehat{X}_{t_i})\sqrt{h_i}\varepsilon_i,$$

- ii'. we compute $\widehat{X}_{t_{i+1}}$ by adding to $\check{X}_{t_{i+1}}$, with probability $1 - e^{-\lambda(t_i, \widehat{X}_{t_i})h_i} \approx \lambda(t_i, \widehat{X}_{t_i})h_i$ (for small h_i), a jump term “equal to $\delta(t_i, \widehat{X}_{t_i}, x)$ with probability $w(t_i, \widehat{X}_{t_i}, dx)$ ”.

Alternatively, in case λ does not depend on X , to simulate each path of X , we can first simulate the ordered jump times T_l of N , add these to the time-grid (t_i) and run the above algorithm on the resulting enlarged time-grid (still denoted by (t_i) for simplicity), with (ii) replaced by

- ii''. If $t_{i+1} = T_l$ for some l , then we simulate

$$\widehat{X}_{T_l} = \check{X}_{T_l} + \delta(T_l, \check{X}_{T_l}, J_l) \quad \text{where } J_l \sim w(T_l, \check{X}_{T_l}, dx),$$

otherwise we set $\widehat{X}_{t_{i+1}} = \check{X}_{t_{i+1}}$.

6.8.3.3 Continuous Euler Scheme

This is the continuous-time approximation scheme defined, using interpolation with the alternative Euler scheme at the previous section, by a Brownian Bridge between (t_i, \widehat{X}_{t_i}) and $(t_{i+1}, \check{X}_{t_{i+1}})$, for every i .

6.8.4 Monte Carlo Simulation for Processes

In the case of Monte Carlo simulation for processes, the error can be decomposed into

$$\begin{aligned} \mathbb{E}[\phi(X_T)] - \frac{1}{m} \sum_{j=1}^m \phi(\hat{X}_T^j) &= (\mathbb{E}[\phi(X_T)] - \mathbb{E}[\phi(\hat{X}_T)]) \\ &\quad + \left(\mathbb{E}[\phi(\hat{X}_T)] - \frac{1}{m} \sum_{j=1}^m \phi(\hat{X}_T^j) \right), \end{aligned} \quad (6.15)$$

where the two terms on the right-hand side are respectively referred to as the time-discretization error and the Monte Carlo or simulation error:

- for usual time-discretization schemes such as the Euler or the Milstein scheme, convergence in law is linear in h , so the time-discretization error is $O(h)$;
- the (pseudo) Monte Carlo error is $O(\widehat{\sigma} m^{-\frac{1}{2}})$, where $\widehat{\sigma}$ represents the standard deviation of the approximate payoff $\phi(\hat{X}_T)$.

The overall error is $O(h) + O(m^{-\frac{1}{2}})$, in comparison with $O(h) + O(m_1^{-2})$ in the case of a typical finite difference numerical scheme with a generic number m_1 of mesh points per space dimension (see Chap. 8). Taking m as m_1^d in order to balance the computation times allocated to the two methods (stochastic Monte Carlo and deterministic finite differences), we conclude that Monte Carlo is more accurate and therefore more efficient for $d > 4$, and less efficient for $d < 4$. This conclusion is consistent with the related discussion in the introduction to this part of the book. Also note that in order to balance the two terms of the error in (6.15), we should take m of order of n^2 . This can be implemented incrementally with backward time-discretization schemes (see Sect. 6.8.1.2 in the case of the Brownian motion W).

6.9 Monte Carlo Methods for Exotic Options

A nice feature of Monte Carlo is that it can easily cope with path dependence. However, specific treatments must be applied in order to preserve convergence rates. A recurrent idea in this regard is use of the continuous-time Euler scheme in order to try to recover the lost information about “what happens between the points of the time-grid” of a time-discretization scheme. Let $M_t = \sup_{0 \leq s \leq t} W_s$ denote the running supremum of the Brownian motion and let

$$W_t^\lambda = W_t + \lambda t, \quad M_t^\lambda = \sup_{0 \leq s \leq t} W_s^\lambda \quad (6.16)$$

for any real λ .

Lemma 6.9.1

- (i) The bivariate process (W, M) admits the following transition probability density between times 0 and t :

$$q_t(w, m) = \mathbb{1}_{m \geq w^+} \frac{2(2m - w)}{\sqrt{2\pi t^3}} \exp\left[-\frac{(2m - w)^2}{2t}\right]; \quad (6.17)$$

- (ii) the random variable

$$Z_t^\lambda = (2M_t^\lambda - W_t^\lambda)^2 - (W_t^\lambda)^2$$

is, conditionally on W_t^λ , $\mathcal{E}_{\frac{1}{2t}}$ -distributed;

- (iii) the conditional cumulative distribution function and inverse cumulative distribution function of M_t^λ , given $W_t^\lambda = w$, are written as

$$\begin{aligned} F_t(m | w) &= \left(1 - \exp\left[-\frac{2}{t}m(m - w)\right]\right), \quad m \geq w^+, \\ F_t^{-1}(u | w) &= \frac{1}{2}\left(w + \sqrt{w^2 - 2t \ln(1-u)}\right), \quad 0 \leq u \leq 1. \end{aligned} \quad (6.18)$$

Proof (i) The formula (6.17) is obtained by crossed differentiation with respect to x and y in the following “mirror formula”, which is valid for every $x \in \mathbb{R}$ and $y \geq x^+$ (see [164, 233]):

$$\mathbb{P}(W_t \geq 2y - x) = \mathbb{P}(W_t \leq x, M_t \geq y).$$

- (ii) By the Girsanov formulas of Sect. 3.4.1.2 we have, for all real x, y :

$$\begin{aligned} \mathbb{P}(W_t^\lambda \leq x) &= \mu \mathbb{P}(W \leq x), \\ \mathbb{P}(W_t^\lambda \leq x, M_t^\lambda \leq y) &= \mu \mathbb{P}(W \leq x, M_t \leq y) \end{aligned} \quad (6.19)$$

for some random weight μ , depending on λ but common to both equations in (6.19). Therefore

$$\mathbb{P}(M_t^\lambda \leq y | W_t^\lambda \leq x) = \mathbb{P}(M_t \leq y | W_t \leq x).$$

This shows that the law of M_t^λ conditional on W_t^λ doesn't depend on λ . We may thus restrict our attention to the case $\lambda = 0$. Denoting $Z_t = (2M_t - W_t)^2 - (W_t)^2$ and introducing the one-to-one mapping

$$[x^+, +\infty) \ni y \mapsto z = (2y - x)^2 - x^2 \in \mathbb{R}_+,$$

we have that

$$\mathbb{P}(Z_t \in dz | W_t = x) = \mathbb{P}(M_t \in dy | W_t = x)$$

and therefore, by (6.17),

$$\mathbb{P}(Z_t \in dz \mid W_t = x) = e^{\frac{-z}{2t}} \frac{dz}{2t}.$$

Part (iii) follows directly from (ii) by Lemma 6.2.1. \square

6.9.1 Lookback Options

We consider a lookback option with payoff $\phi(X_T, M_T)$, where X is given in the form of the following one-dimensional diffusion:

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

and $M_t = \sup_{0 \leq s \leq t} X_s$. The following result, based on Lemma 6.9.1, gives a way of simulating the pair (\bar{X}, \bar{M}) , where $\bar{M}_t = \sup_{[0,t]} \bar{X}$ and \bar{X} is the continuous-time Euler scheme for X . Let $\hat{X} = (\hat{X}_{t_i})_{0 \leq i \leq n}$ denote the (discrete time) Euler scheme for X and let $\tilde{M}_i = \sup_{t_i \leq t \leq t_{i+1}} \bar{X}_t$ for every $i = 0, \dots, n - 1$.

Proposition 6.9.2 *We have*

$$\mathcal{L}((\tilde{M}_i)_{0 \leq i \leq n-1} \mid \hat{X}) = \mathcal{L}((\hat{M}_i)_{0 \leq i \leq n-1}),$$

where, for every i ,

$$\hat{M}_i := \frac{1}{2}(\hat{X}_{t_i} + \hat{X}_{t_{i+1}} + \sqrt{(\hat{X}_{t_i} - \hat{X}_{t_{i+1}})^2 - 2\sigma(t_i, \hat{X}_{t_i})^2 h_i \ln(1 - U_i)}) \quad (6.20)$$

for independent uniforms U_i .

Proof Letting $\lambda_i = \frac{b(t_i, \hat{X}_{t_i})}{\sigma(t_i, \hat{X}_{t_i})}$, in the notation of (6.16) we have, for $t \in [t_i, t_{i+1}]$:

$$\mathcal{L}\left(\frac{\bar{X}_t - \hat{X}_{t_i}}{\sigma(t_i, \hat{X}_{t_i})} \mid \hat{X}\right) = \mathcal{L}\left(W_t^{\lambda_i} - W_{t_i}^{\lambda_i} \mid W_{t_{i+1}}^{\lambda_i} - W_{t_i}^{\lambda_i} = \frac{\hat{X}_{t_{i+1}} - \hat{X}_{t_i}}{\sigma(t_i, \hat{X}_{t_i})}\right)$$

and (for every i and also jointly in all i , by independence between the \tilde{M}_i given \hat{X})

$$\mathcal{L}\left(\frac{\tilde{M}_i - \hat{X}_{t_i}}{\sigma(t_i, \hat{X}_{t_i})} \mid \hat{X}\right) = \mathcal{L}\left(\sup_{t \in [t_i, t_{i+1}]} W_t^{\lambda_i} - W_{t_i}^{\lambda_i} \mid W_{t_{i+1}}^{\lambda_i} - W_{t_i}^{\lambda_i} = \frac{\hat{X}_{t_{i+1}} - \hat{X}_{t_i}}{\sigma(t_i, \hat{X}_{t_i})}\right).$$

By application of Lemma 6.9.1(ii) to the drifted Brownian motions $W_{t_i+}^{\lambda_i} - W_{t_i}^{\lambda_i}$ on $[0, h_i]$, and also using Lemma 6.2.1, the law of the \tilde{M}_i given \hat{X} is then the same as that of the

$$\hat{X}_{t_i} + \sigma(t_i, \hat{X}_{t_i}) F_{h_i}^{-1}\left(U_i \mid \frac{\hat{X}_{t_{i+1}} - \hat{X}_{t_i}}{\sigma(t_i, \hat{X}_{t_i})}\right) = \hat{M}_i, \quad (6.21)$$

as follows from the expression of F^{-1} in (6.18). \square

In summary, to simulate a pair (\bar{X}_T, \bar{M}_T) , which can then be substituted into a Monte Carlo loop for pricing a lookback option with payoff $\phi(X_T, M_T)$:

- i. we simulate a trajectory \hat{X} of the Euler scheme for X , using n independent uniforms u_i ;
- ii. given this trajectory \hat{X} , we simulate related \hat{M}_i by (6.20), using n new independent uniforms U_i .

Then we set $\bar{X}_T = \hat{X}_T$, $\bar{M}_T = \max_i \hat{M}_i$. If quasi-random numbers are used, we must employ a $2n$ -dimensional low-discrepancy sequence in (i), (ii) (but the use of high-dimensional low-discrepancy sequences must be considered with caution).

6.9.1.1 Black–Scholes Case

In the special case of a Black–Scholes model S_t (see Andersen and Brotherton-Ratcliffe [6]), the Euler discretization is exact provided one works in the log-returns variable $X_t = \ln(S_t)$. In this case one can then take n equal to one in (i)–(ii) above. Letting $M_T = \max_{t \in [0, T]} S_t$, the time-0 price and delta of a lookback option with payoff ϕ are

$$\Pi_0 = \mathbb{E}[e^{-rT} \phi(S_T, M_T)], \quad \Delta_0 = \partial_s \mathbb{E}[e^{-rT} \phi(S_T, M_T)],$$

with related estimates

$$\widehat{\Pi}_0 = \frac{1}{m} e^{-rT} \sum_{j=1}^m \pi^j, \quad \widehat{\Delta}_0 = \frac{1}{m} e^{-rT} \sum_{j=1}^m \delta_s \pi^j = \frac{1}{m} e^{-rT} \sum_{j=1}^m \delta^j.$$

- **Fixed Strike Lookback Call.** The payoff is $(M_T - K)^+$, so

$$\pi^j = (M_T^j - K)^+, \quad \delta^j = \begin{cases} \partial_s M_T^j = \frac{M_T^j}{s} & \text{if } \pi^j \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

- **Floating Strike Lookback Put.** The payoff is $(M_T - S_T)$, so

$$\pi^j = (M_T^j - S_T^j), \quad \delta^j = \partial_s M_T^j - \partial_s S_T^j = \frac{M_T^j - S_T^j}{s} = \frac{\pi^j}{s}.$$

Simulation of the Maximum M_T At run j :

- i. S_T^j is generated as $e^{X_T^j}$, with $X_T^j = x + bT + \sigma \sqrt{T} \varepsilon_j$ for an independent Gaussian draw ε_j ;
- ii. M_T^j is generated as $s e^{\sigma F_T^{-1}(u_j | \frac{X_T^j - x}{\sigma})}$ for an independent uniform draw u_j .

6.9.2 Barrier Options

With a barrier option the right to exercise the payoff at maturity depends on additional events such as the underlying having crossed or reached certain levels on $[0, T]$. Such options were created as a way to provide the insurance value of an option without charging as much premium. Common forms of barrier options are up-and-in, down-and-out, down-and-in, double-in, and double-out options.

For instance, a barrier up-and-out option with trigger level H and rebate R corresponds to the following payoff process (considering the case of a rebate R paid at T):

$$\psi(X_T, M_T) = \phi(X_T) \mathbb{1}_{\{M_T < H\}} + R \mathbb{1}_{\{M_T \geq H\}}.$$

An approximation for the price is given by

$$e^{-rT} \mathbb{E}\psi(\bar{X}_T, \bar{M}_T),$$

with

$$\mathcal{L}(\bar{M}_T | \hat{X}) = \mathcal{L}\left(\max_{0 \leq i \leq n-1} \hat{M}_i\right)$$

as above. We have

$$\begin{aligned} \mathbb{E}(\phi(\bar{X}_T) \mathbb{1}_{\{\bar{M}_T \leq H\}} | \hat{X}) &= \phi(\hat{X}_T) \prod_{i=0}^{n-1} \mathbb{P}\left(\max_{t_i \leq t \leq t_{i+1}} \bar{X}_t \leq H | \hat{X}\right) \\ &= \phi(\hat{X}_T) \prod_{i=0}^{n-1} F_{h_i}\left(\frac{H - \hat{X}_{t_i}}{\sigma(t_i, \hat{X}_{t_i})} \mid \frac{\hat{X}_{t_{i+1}} - \hat{X}_{t_i}}{\sigma(t_i, \hat{X}_{t_i})}\right). \end{aligned}$$

Likewise

$$\mathbb{E}(R \mathbb{1}_{\{\bar{M}_T > H\}} | \hat{X}) = R \left(1 - \prod_{i=0}^{n-1} F_{h_i}\left(\frac{H - \hat{X}_{t_i}}{\sigma(t_i, \hat{X}_{t_i})} \mid \frac{\hat{X}_{t_{i+1}} - \hat{X}_{t_i}}{\sigma(t_i, \hat{X}_{t_i})}\right)\right),$$

and finally

$$\mathbb{E}\psi(\bar{X}_T, \bar{M}_T) = R + \mathbb{E}\left[\left(\phi(\hat{X}_T) - R\right) \prod_{i=0}^{n-1} F_{h_i}\left(\frac{H - \hat{X}_{t_i}}{\sigma(t_i, \hat{X}_{t_i})} \mid \frac{\hat{X}_{t_{i+1}} - \hat{X}_{t_i}}{\sigma(t_i, \hat{X}_{t_i})}\right)\right].$$

In this case no random draws are needed beyond those used for simulating \hat{X}_T , i.e. n random draws per simulation run, where n can be taken equal to one in the special case of the Black–Scholes model.

6.9.3 Asian Options

We next consider an Asian option with a payoff of the form $\phi(S_T, I_T)$, with $I_t = \int_0^T S_u du$. For instance, $\phi(x, y) = (\frac{I}{T} - K)^+$ in the case of a fixed strike Asian call. We assume a Black–Scholes underlying S .

A first possible approximation for I_T is the Riemann sum $\widehat{I}_T^1 = \sum_{i=0}^{n-1} h_i \widehat{S}_{t_i}$, but this discretization works poorly in practice. A better discretization is given by the trapezoid rule $\widehat{I}_T^2 = \sum_{i=0}^{n-1} h_i \frac{\widehat{S}_{t_i} + \widehat{S}_{t_{i+1}}}{2}$. However, one can show by Taylor expansion that this is tantamount to approximating

$$\mathbb{E}\phi(\bar{S}_T, \bar{I}_T) = \mathbb{E}[\mathbb{E}[\phi(\bar{S}_T, \bar{I}_T) | \bar{S}]] \quad (6.22)$$

by

$$\mathbb{E}[\phi(\bar{S}_T, \mathbb{E}[\bar{I}_T | \bar{S}])] \quad \text{with } \bar{I}_t = \int_0^T \bar{S}_u du.$$

But this approximation involves a nonlinearity bias. An even better way is to simulate the right-hand side of (6.22) directly (see [177]), approximating \bar{I}_T conditionally on \bar{S} by

$$\begin{aligned} & \sum_{i=0}^{n-1} \widehat{S}_{t_i} \int_{t_i}^{t_{i+1}} (1 + \kappa(t - t_i) + \sigma B_t^i dt) \\ & \approx \sum_{i=0}^{n-1} h_i \widehat{S}_{t_i} \left(1 + \frac{\kappa h_i}{2} + \frac{\sigma}{h_i} \int_{t_i}^{t_{i+1}} B_t^i dt \right) =: \widehat{I}_T^3. \end{aligned}$$

Here B^i is a Brownian Bridge between (t_i, W_{t_i}) and $(t_{i+1}, W_{t_{i+1}})$, so that $\int_{t_i}^{t_{i+1}} B_t^i dt =: \varepsilon_i$ is a Gaussian random variable with (given $W_{t_i}, W_{t_{i+1}}$):

$$\mathbb{E}(\varepsilon_i) = \int_{t_i}^{t_{i+1}} \left(W_{t_i} + \frac{(t - t_i)}{h_i} (W_{t_{i+1}} - W_{t_i}) \right) dt = \frac{h_i}{2} (W_{t_i} + W_{t_{i+1}})$$

$$\text{Var}(\varepsilon_i) = 2 \int_{u=t_i}^{t_{i+1}} \int_{t=t_i}^u \text{Cov}(B_t^i, B_u^i) dt du = 2 \int_{v=0}^{h_i} \left(1 - \frac{v}{h_i} \right) \frac{v^2}{2} dv = \frac{h_i^3}{12},$$

where the identity $\text{Cov}(B_t^i, B_u^i) = (t - t_i)(1 - \frac{u - t_i}{h_i})$ resulting from (6.10) was used in the second line.

The previous discretization schemes can be used in conjunction with variance reduction [165, 177]. The arithmetic average $A_T = \frac{I_T}{T}$ is “close” to the geometric average $J_T = \exp(\frac{1}{T} \int_0^T \ln(S_t) dt)$ for r and σ “small”. This suggests the use of $\phi(S_T, T J_T)$ as a control variable for the payoff $\phi(S_T, I_T)$.

Example 6.9.3 In the case of a fixed strike Asian call, we have

$$\phi(S_T, T J_T) = (T J_T - K)^+,$$

where J_T is lognormally distributed, so that $\mathbb{E}\phi(S_T, T J_T)$ is known explicitly. Thus

$$\begin{aligned} J_T &= \exp\left(\frac{1}{T} \int_0^T \ln(S_t) dt\right) = S_0 \exp\left(\frac{1}{T} \int_0^T (\sigma W_t + bt) dt\right) \\ &= S_0 \exp\left(\frac{\sigma}{T} \int_0^T W_t dt + \frac{bT}{2}\right), \end{aligned}$$

with $\mathbb{V}\text{ar}(\int_0^T W_t dt) = 2 \int_0^T \int_0^u t dt du = \int_0^T u^2 du = \frac{T^3}{3}$. Therefore

$$T J_T = \tilde{S}_0 \exp\left(\tilde{\sigma} \sqrt{T} \varepsilon - \frac{\tilde{\sigma}^2 T}{2}\right),$$

with $\tilde{\sigma} = \frac{\sigma}{\sqrt{3}}$, $\tilde{S}_0 = T S_0 \exp(\frac{bT}{2} + \frac{\tilde{\sigma}^2 T}{2})$. Thus

$$\mathbb{E}(T J_T - K)^+ = \pi^{bl}(0, \tilde{S}_0, T, K; \tilde{\sigma}).$$

The above Brownian Bridge techniques can be extended to arbitrary jump-diffusions. As will be illustrated in Chaps. 10 and 11, the nonlinear simulation pricing schemes of the next section also work in any Markovian model X .

6.10 American Monte Carlo Pricing Schemes

We saw in Sect. 3.5.4 (see also Chaps. 12 and 13) that the prices of American options are the Snell envelopes of the related payoff processes, as opposed to straight expectations (or conditional expectations, at future times) in the case of European options. The supremum in the pricing formula (3.92) (at time 0; see Remark 3.5.2) bears on a huge set of stopping times that cannot be discretized numerically. As a consequence, American options cannot be priced by the standard Monte Carlo loops of previous sections. Simulation pricing schemes for American options do exist however, in the form of hybrid “forward (as in Monte Carlo loops) / backward (as deterministic finite difference and tree) schemes”, in which a dynamic programming equation is run backward in time at the points of a stochastic grid simulated in a forward manner, according to the underlying factor dynamics. With respect to deterministic schemes, these simulation pricing schemes present the advantage of being less severely impacted by “the curse of dimensionality”.

Remark 6.10.1 Such nonlinear simulation pricing schemes can also be viewed as special cases of simulation schemes for reflected BSDEs or for the equivalent PDE obstacle problems; see Sect. 3.5.2, Chaps. 12–13 and [96].

Remark 6.10.2 Alternative nonlinear simulation pricing schemes not treated in this book are purely forward branching particle simulation schemes (see e.g. [135]).

For pricing an American option by Monte Carlo, we can thus write a dynamic programming equation on a stochastically generated (hence nonrecombining) mesh $(X_i^j)_{0 \leq i \leq n}^{1 \leq j \leq m}$, so that: $\Pi_n^j = \phi(X_n^j)$ for $j = 1 \dots m$ and, for $i = n - 1, \dots, 0$, $j = 1 \dots m$:

$$\Pi_i^j = \max(\phi(X_i^j), e^{-rh} \mathbb{E}_i^j \Pi_{i+1}), \quad (6.23)$$

where $\mathbb{E}_i^j \Pi_{i+1}$ is the conditional expectation of Π_{i+1} given $X_i = X_i^j$. The issue of computation of the conditional expectations on the nonrecombining mesh $(X_i^j)_{0 \leq i \leq n}^{1 \leq j \leq m}$ is dealt with in Sect. 6.10.2. This is only required for $i \geq 1$, since for $i = 0$ the conditional expectation reduces to an expectation, and all the Π_0^j in (6.23) reduce to $\widehat{\Pi}_0 := \Pi_0^0$.

As American Monte Carlo estimates are not based on the law of large numbers and the central limit theorem of Sect. 6.3.1, they do not provide a confidence interval. It is possible, however, to derive an upper bound on the true price by resorting to a dual Monte Carlo approach of Rogers [235] (see also Sect. 6.10.1). Of course, one can always derive a confidence interval for the estimated (if not the true) price by running the simulation loop for various (say 50) seeds of the generator and computing a standard deviation of the estimates.

One can also recover, from the pricing function estimated by (6.23), the following estimates of the exercise region and of the related optimal stopping policy (starting from time 0 for the latter):

$$\begin{aligned} \mathcal{E} &= \{(i, X_i^j); \widehat{\Pi}_i^j = \phi(X_i^j)\}, \\ v^j &= \inf\{i \in \mathbb{N}_n; X_i^j \in \mathcal{E}\} \wedge T. \end{aligned} \quad (6.24)$$

Finally it is also possible, by simulation, to compute the option delta (see Chap. 10).

6.10.1 Time-0 Price

The above approach, which in the language of Markov decision theory corresponds to iteration on the values, was first developed by Tsitsiklis and VanRoy [251]. We now present an approach by iteration on the policies due to Longstaff and Schwartz [195] for pricing the option at time 0. The idea is to use the estimated optimal stopping policy v in (6.24) for computing an alternative estimate of the option price at time 0 as

$$\widetilde{\Pi}_0 = \mathbb{E} e^{-\sum_0^{v^j-1} rh} \phi(X_{v^j}^j). \quad (6.25)$$

Whereas $\widehat{\Pi}_0$, obtained from (6.23), typically overestimates the exact price Π_0 , on the contrary $\widetilde{\Pi}_0$ typically underestimates it. Computing both estimates gives a way to end up with an interval. If this interval is too large it typically means that the basis of functions which is used for computing the conditional expectations is not well chosen (see below).

6.10.2 Computing Conditional Expectations by Simulation

The nonlinear Monte Carlo pricing schemes of this section, and those of Chaps. 10 and 11, ultimately reduce to the numerical computation of conditional expectations. As we will now see, this can be done by a combination of simulation and regression tools. Let ξ and X denote real and \mathbb{R}^d -valued square integrable random variables. Under suitable conditions the conditional expectation $\mathbb{E}(\xi|X)$ is equal to the L^2 -projection of ξ over the vector space of random variables spanned by Borel functions of X . So, using a basis $(\varphi^l)_{l \in \mathbb{N}}$ of the set of functions from \mathbb{R}^d to \mathbb{R} ,

$$\mathbb{E}(\xi|X) = \mathbb{L}(\xi | (\varphi^l(X))_{l \in \mathbb{N}}),$$

where \mathbb{L} represents the L^2 -projection operator. Given pairs $(X^j, \xi^j)_{1 \leq j \leq m}$ simulated independently according to the law of (X, ξ) , the conditional expectation $\mathbb{E}(\xi|X)$ may thus be simulated by linear regression of $(\xi^j)_{1 \leq j \leq m}$ against $(\varphi^l(X^j))_{1 \leq l \leq \rho}^{1 \leq j \leq m}$, where the truncation order ρ is a parameter of the method. The computational cost of this regression is $O(m\rho^2)$ for forming the regression matrix, plus the time needed for solving a (typically numerically ill-conditioned) linear system of dimension ρ .

We refer the interested reader to the monograph by Györfi et al. [136] for details about these simulation/regression approaches for computing a regression function

$$x \mapsto \varrho(x) = \mathbb{E}(\xi | X = x).$$

Succinctly, the (truncated) regression basis may be:

- either parametric, i.e. formed of functions parameterized by a few parameters, or nonparametric, meaning in practice that it is formed of a very large set of functions, like one function per point of a discretization of the state space;
- either global, that is, formed of functions supported by the whole state space or with “large” support, or local, formed of functions with “small” support.

One typically uses either a parametric and global regression basis, such as a regression basis formed of a few monomials parameterized by their coefficients, or a nonparametric and local basis, such as a regression basis formed of the indicator functions of the cells of a grid of hypercubes partitioning the state space. The latter method is referred to as the method of cells in Chaps. 10 and 11. Theory tells that a global basis is preferred in the case of a “regular” regression function $\varrho(x)$, especially when a good guess is available for the shape of ϱ ; this guess can then be used to define the regression basis. Otherwise a local basis is preferred, being simpler and often more robust in terms of implementation. Note that, in the simplest case of the method of cells, the “regression” doesn’t involve the solution of a linear system, the “regression matrix” being diagonal in this case.

Note that there exist alternatives to nonlinear regression for computing conditional expectations by simulation, in particular Malliavin calculus based methods

[51, 189] and quantization methods [18, 19]. However, the former are typically difficult to implement and the latter suffer significantly from the curse of dimensionality (see [53, 133]).

Chapter 7

Tree Methods

Tree pricing schemes are natural in finance because of their Markov chain interpretation as discrete time pricing models. From a practical point of view, trees are often rather obsolete as compared with more sophisticated finite difference or finite element technologies. However, in a number of situations, they remain an adequate and simple alternative. Moreover, from the theoretical point of view, the Markov chain interpretation underlies interesting probabilistic convergence proofs of the related (deterministic) pricing schemes.

7.1 Markov Chain Approximation of Jump-Diffusions

Let X^h represent a continuous-time Markov chain approximation for the \mathbb{R}^d -valued jump-diffusion X in (4.30). By the identity in the first line of (3.67), for every regular test-function $\varphi = \varphi(x)$ we have

$$\lim_{h \rightarrow 0} h^{-1} \mathbb{E}_t (\varphi(X_{t+h}) - \varphi(X_t)) = \mathcal{A}\varphi(X_t),$$

in which \mathcal{A} represents the infinitesimal generator (3.66) of X . Convergence in law of X^h to X is essentially equivalent to having, similarly, for every test-function φ ,

$$\lim_{h \rightarrow 0} h^{-1} (\mathbb{E}_t \varphi(X_{t+h}^h) - \varphi(X_t^h)) = \mathcal{A}\varphi(x), \quad (7.1)$$

on every random set $\{\lim_{h \rightarrow 0} X_t^h = x\}$. See Kushner and Dupuis [170], Ethier and Kurtz [119] and Jacod and Shiryaev [153] for the related mathematics.

7.1.1 Kushner's Theorem

Convergence in law of processes implies convergence of prices of European and American options in the approximating Markov chains to their counterparts in the

limiting jump-diffusion model X . However, for establishing convergence in law, checking (7.1) for every test-function φ is not practical. Kushner's theorem reduces much of the burden to verification that in the limit the first two conditional moments of X^h match those of X , so that

$$\begin{aligned}\lim_{h \rightarrow 0} h^{-1} \mathbb{E}_t(X_{t+h}^h - X_t^h) &= \lim_{h \rightarrow 0} h^{-1} \mathbb{E}_t(X_{t+h} - X_t) \\ &= b(t, x) + \lambda(t, x) \int_{\mathbb{R}^d} \delta(t, x, y) w(t, x, dy) \\ \lim_{h \rightarrow 0} h^{-1} \mathbb{Cov}_t(X_{t+h}^h - X_t^h) &= \lim_{h \rightarrow 0} h^{-1} \mathbb{Cov}_t(X_{t+h} - X_t) \\ &= a(t, x) + \lambda(t, x) \int_{\mathbb{R}^d} \delta \delta^\top(t, x, y) w(t, x, dy)\end{aligned}$$

(see Corollary 3.2.13). One part of Kushner's theorem is easy, i.e.:

Proposition 7.1.1 *Convergence in law of X^h to X implies the so-called local consistency conditions, i.e. on every random set $\{\lim_{h \rightarrow 0} X_t^h = x\}$*

$$\begin{aligned}\lim_{h \rightarrow 0} h^{-1} \mathbb{E}_t(X_{t+h}^h - X_t^h) &= b(t, x) + \lambda(t, x) \int_{\mathbb{R}^d} \delta(t, x, y) w(t, x, dy) \\ \lim_{h \rightarrow 0} h^{-1} \mathbb{Cov}_t(X_{t+h}^h - X_t^h) &= a(t, x) + \lambda(t, x) \int_{\mathbb{R}^d} \delta \delta^\top(t, x, y) w(t, x, dy).\end{aligned}\tag{7.2}$$

Proof For every component i and j of X , setting $\varphi = \varphi(x) = \pi^i(x) := x_i$ and $\varphi = \pi^i \pi^j$ in (7.1) yields respectively that, on $\{\lim_{h \rightarrow 0} X_t^h = x\}$,

$$\begin{aligned}\lim_{h \rightarrow 0} h^{-1} \mathbb{E}_t(X_{t+h}^{h,i} - X_t^{h,i}) &= \mathcal{A}\pi^i(x) = b_i(t, x) + \lambda(t, x) \int_{\mathbb{R}^d} \delta_i(t, x, y) w(t, x, dy) \\ \lim_{h \rightarrow 0} h^{-1} \mathbb{E}_t(X_{t+h}^{h,i} X_{t+h}^{h,j} - X_t^{h,i} X_t^{h,j}) &= \mathcal{A}(\pi^i \pi^j)(x).\end{aligned}\tag{7.3}$$

Moreover, for fixed h , we have that

$$\begin{aligned}\mathbb{Cov}_t(X_{t+h}^{h,i} - X_t^{h,i}, X_{t+h}^{h,j} - X_t^{h,j}) &+ \mathbb{E}_t(X_{t+h}^{h,i} - X_t^{h,i}) \mathbb{E}_t(X_{t+h}^{h,j} - X_t^{h,j}) \\ &= \mathbb{E}_t(X_{t+h}^{h,i} X_{t+h}^{h,j} - X_t^{h,i} X_t^{h,j}) - X_t^{h,i} \mathbb{E}_t(X_{t+h}^{h,j} - X_t^{h,j}) - X_t^{h,j} \mathbb{E}_t(X_{t+h}^{h,i} - X_t^{h,i}).\end{aligned}\tag{7.4}$$

So on $\{\lim_{h \rightarrow 0} X_t^h = x\}$:

$$\begin{aligned}\lim_{h \rightarrow 0} h^{-1} \mathbb{Cov}_t(X_{t+h}^{h,i} - X_t^{h,i}, X_{t+h}^{h,j} - X_t^{h,j}) & \\ &= \mathcal{A}(\pi^i \pi^j)(x) - x_i \mathcal{A}\pi^j(x) - x_j \mathcal{A}\pi^i(x),\end{aligned}$$

where, by application of (3.68), the right-hand side coincides with

$$a_{i,j}(t, x) + \lambda(t, x) \int_{\mathbb{R}^d} \delta_i \delta_j(t, x, y) w(t, x, dy). \quad \square$$

Conversely, in the case of a diffusion X , Kushner's theorem [169, 170] states that the local consistency conditions (7.2) (with $\lambda = 0$ there) assure the convergence in law of X^h to X . See also p. 129 of [170] for an extension of this result to jump-diffusions.

In the next sections we discuss basic tree pricing schemes in the setup of the Black–Scholes model, i.e.

$$S_t = S_0 e^{bt + \sigma W_t},$$

with $\kappa = r - q$ and $b = \kappa - \frac{1}{2}\sigma^2$, for a constant risk-free rate r and a constant dividend yield q on S .

7.2 Trees for Vanilla Options

7.2.1 Cox–Ross–Rubinstein Binomial Tree

The Cox–Ross–Rubinstein tree [74] is the following Markov chain approximation to Black–Scholes, parameterized by two positive constants $0 < d < u$: $S_0^h = S_0$ and, for $i = 0, \dots, n - 1$,

$$S_{(i+1)h}^h = u S_{ih}^h \quad (\text{resp. } d S_{ih}^h) \quad \text{with probability } p \text{ (resp. } 1 - p).$$

Here $h = \frac{T}{n}$ where T is the maturity of an option with payoff function ϕ , n is the number of time steps in the tree and

$$u = e^{\sigma\sqrt{h}}, \quad d = e^{-\sigma\sqrt{h}}, \quad p = \frac{e^{\kappa h} - d}{u - d}. \quad (7.5)$$

We find it convenient to denote the time in the tree by i rather than ih . In particular, $S_i^h \equiv S_{ih}^h$, \mathbb{E}_i refers to the conditional expectation with respect to the σ -field \mathcal{F}_i^h generated by (S_0^h, \dots, S_i^h) and \mathcal{T}_i^h , also with $\mathcal{T}_0^h = \mathcal{T}^h$, represents the set of \mathbb{F}^h -stopping times ν with values in $\{i, \dots, n\}$. The following Proposition shows that the Cox–Ross–Rubinstein tree model shares the main hedging properties of the Black–Scholes model.

Proposition 7.2.1

(i) *In the European case, the process given for $i = 0, \dots, n$ by*

$$\Pi_i^h := e^{-r(T-i)h} \mathbb{E}_i \phi(S_n^h) = u_i(S_i^h) \quad (7.6)$$

is the unique replication price process for the option, with an associated replication strategy given by

$$\delta_i^h = \frac{u_{i+1}(uS_i^h) - u_{i+1}(dS_i^h)}{(u - d)S_i^h}; \quad (7.7)$$

here the European pricing function u is defined by

- $u_n(S) = \phi(S)$ for every S in the tree at time n ,
- for $i = n - 1, \dots, 0$,

$$u_i(S) = e^{-rh} [pu_{i+1}(uS) + (1-p)u_{i+1}(dS)] \quad (7.8)$$

for every S in the tree at time i .

(ii) In the American case, the minimal superhedging price of the option is given for $i = 0, \dots, n$ by:

$$\tilde{H}_i^h := \max_{v \in \mathcal{T}_i^h} \mathbb{E}_i(e^{-r(v-i)h} \phi(S_v^h)) = v_i(S_i^h), \quad (7.9)$$

with a related superhedging strategy defined as

$$\tilde{\delta}_i^h = \frac{v_{i+1}(uS_i^h) - v_{i+1}(dS_i^h)}{(u - d)S_i^h}; \quad (7.10)$$

here the American pricing function v is defined by

- $v_n(S) = \phi(S)$ for every S in the tree at time n ,
- for $i = n - 1, \dots, 0$,

$$v_i(S) = \max(\phi(S), e^{-rh} [pv_{i+1}(uS) + (1-p)v_{i+1}(dS)]) \quad (7.11)$$

for every S in the tree at time i .

Proof We assume $r = q = 0$ for notational simplicity.

Case n=1 First considering a European option with payoff function $\phi(S_1^h)$, let (Y_0, Z_0) denote the solution to the following elementary one time-step BSDE:

$$Y_0 = \phi(S_1^h) - Z_0(S_1^h - S_0), \text{ a.s.} \quad (7.12)$$

Note that (7.12) is equivalent to the following algebraic system of two equations in the two unknown numbers Y_0, Z_0 :

$$\begin{cases} Y_0 = \phi(uS_0) - Z_0(u - 1)S_0 \\ Y_0 = \phi(dS_0) - Z_0(d - 1)S_0, \end{cases}$$

which is well-posed,¹ since $d < u$. By construction, the price-and-hedge (Y_0, Z_0) replicates the payoff $\phi(S_1^h)$ and we have that

$$Z_0 = \delta_0^h = \frac{\phi(uS_0) - \phi(dS_0)}{(u - d)S_0}.$$

Moreover, the definition of p is such that $\mathbb{E}(S_1^h - S_0) = 0$ and therefore $Y_0 = \mathbb{E}\phi(S_1^h)$.

If the option is American, let (Y_0, Z_0, α_0) denote the solution to the following one time-step reflected BSDE:

$$\begin{cases} Y_0 = \phi(S_1^h) + \alpha_0 - Z_0(S_1^h - S_0) \\ Y_0 \geq \phi(S_0), \quad \alpha_0 \geq 0, \quad (Y_0 - \phi(S_0))\alpha_0 = 0. \end{cases}$$

By inspection we see, distinguishing the two cases $\phi(S_0) \leq$ and $\geq \phi(S_1^h)$, that the unique solution is

$$(Y_0, Z_0, \alpha_0) = (\max(\phi(S_0), \mathbb{E}\phi(S_1^h)), \delta_0^h, Y_0 - \mathbb{E}\phi(S_1^h)).$$

By construction, the price-and-hedge (Y_0, Z_0) superhedges the American payoff $\phi(S^h)$ for every holder stopping policy $v \in \mathcal{T}^h = \{\{0\}, \{1\}\}$ (for $n = 1$). Moreover, for every superhedging strategy $(\tilde{Y}_0, \tilde{Z}_0)$, we have that $\tilde{Y}_0 \geq \phi(S_0)$ and $\tilde{Y}_0 + \tilde{Z}_0(S_1^h - S_0) \geq \phi(S_1^h)$ (to superhedge the payoffs respectively due in the $v = \{0\}$ and $v = \{1\}$ cases). Hence the inequality

$$\tilde{Y}_0 \geq \max(\phi(S_0), \mathbb{E}\phi(S_1^h)) = Y_0$$

results. Therefore $Y_0 = \max(\phi(S_0), \mathbb{E}\phi(S_1^h))$ is the minimal initial wealth of an issuer superhedging strategy.

General Case Applying the above results step-by-step backwards in the tree, we deduce that the European price process defined by $u_i(S_i^h)$, along with the hedging strategy (7.7) for the pricing function u^h defined by (7.8), replicates the option payoff $\phi(S_n^h)$ at time T . This also yields the probabilistic representation of $u_i(S_i^h)$ stated in (7.6).

Likewise, the American price process defined by $v_i(S_i^h)$ along with the hedge (7.10) for the pricing function v^h defined by (7.11), is a minimal superhedging strategy for the American option with payoff function ϕ . In view of (7.11), $(v_i(S_i^h))_{0 \leq i \leq n}$ is a supermartingale dominating $(\phi(S_i^h))_{0 \leq i \leq n}$, which implies the \leq inequality in the probabilistic representation of $v_i(S_i^h)$ stated in (7.9). Moreover, let v^i in \mathcal{T}_i^h be defined as

$$v^i = \inf\{j \geq i; v_j(S_j^h) = \phi(S_j^h)\}. \quad (7.13)$$

¹Unless $\frac{\phi(uS_0)}{\phi(dS_0)} = \frac{u-1}{d-1}$, in which case the system admits an infinity of solutions.

Process $(v_i(S_i^h))$ is a martingale on the random time interval $\{i, \dots, v^i\}$ in the sense that if $i \leq j < v^i$, then

$$v_j(S_j^h) = \mathbb{E}_j v_{j+1}(S_{j+1}^h),$$

by the definition (7.11) of v and the definition (7.13) of v^i . Therefore, by this martingale property of $(v_i(S_i^h))_{i \leq j < v^i}$, we have

$$v_i(S_i^h) = \mathbb{E}_i v_{v^i}(S_{v^i}^h) = \mathbb{E}_i \phi(S_{v^i}^h),$$

where the second equality holds by the definition (7.13) of v^i . Therefore v^i achieves the maximum in (7.9). \square

Observe that not only these results, but also their proofs, are essentially the same as in the Black–Scholes model. One can also check that Proposition 7.2.1 holds independently of the exact definition of u and d , provided $0 < d < u$.²

Cox–Ross–Rubinstein Algorithm The Cox–Ross–Rubinstein algorithm consists in a backward computation of the option price based on one of the dynamic programming equations (7.8) and (7.11) subsequent to a forward computation of the $n + 1$ possible values of S_n^h . Note that, since $u = 1/d$, the corresponding tree is recombining in the sense that $S = udS = duS$. In particular, there are only $2n + 1$ possible values of the underlying in the tree between time 0 and time n .

7.2.1.1 Convergence in Law of Processes

For u and d defined by (7.5), the Cox–Ross–Rubinstein tree converges in various senses to the related Black–Scholes model as $h \rightarrow 0$. Convergence in law of processes thus follows, by an application of Kushner’s theorem, for X^h of Sect. 7.1 taken as a suitable continuous-time Markov chain interpolation³ of the Cox–Ross–Rubinstein log-returns process $\ln(S^h)$.

Convergence in law of the one-dimensional marginal S_n^h to S_T can also be established by characteristic function arguments. Assume $S_0 = 1$ for notational simplicity. Letting $i^2 = -1$, for every real λ , we have:

$$\begin{aligned} \mathbb{E}[\exp(i\lambda \ln S_n^h)] &= \mathbb{E}\left[\exp\left(i\lambda \ln \prod_{j=0}^{n-1} \frac{S_{j+1}^h}{S_j^h}\right)\right] \\ &= (\mathbb{E}[\exp(i\lambda \ln S_1^h)])^n \\ &= (p \exp(i\lambda \sigma \sqrt{h}) + (1-p) \exp(-i\lambda \sigma \sqrt{h}))^n. \end{aligned}$$

²With $e^{kh} \in [d, u]$, in order to obtain that $p = \frac{e^{kh}-d}{u-d} \in [0, 1]$; otherwise p only defines a signed probability measure.

³See Sect. 4.3 of Kushner and Dupuis [170].

Since $p = \frac{e^{\kappa h} - d}{u - d} = \frac{1}{2} + \frac{b}{2\sigma} \sqrt{h} + O(h)$, letting $h \rightarrow 0$ we get:

$$\begin{aligned}\mathbb{E}[\exp(i\lambda \ln S_n^h)] &\sim \left(1 + \left[i\lambda b - \lambda^2 \frac{\sigma^2}{2}\right] \frac{T}{n}\right)^n \\ &\rightarrow \exp\left(\left[i\lambda b - \lambda^2 \frac{\sigma^2}{2}\right] T\right) \\ &= \mathbb{E}[\exp(i\lambda \ln(S_T))],\end{aligned}$$

where the last identity was established in Proposition 5.1.4.

Time-0 Prices and Deltas Convergence in law of processes grants the convergence of prices of European and American vanilla options with integrable payoffs, e.g. call and put options. As for the deltas, in the European case, (7.7) yields

$$\begin{aligned}S_0 \delta_0^h &= \frac{u_1(uS_0) - u_1(dS_0)}{(u - d)} \\ &= e^{-r(n-1)h} \frac{\mathbb{E}[\phi(uS_0 \xi^h)] - \mathbb{E}[\phi(dS_0 \xi^h)]}{u - d}\end{aligned}$$

for some random variable ξ^h . Assuming that ϕ is of class C^1 , then

$$\phi(uS_0 \xi^h) - \phi(dS_0 \xi^h) = \int_d^u S_0 \xi^h \phi'(xS_0 \xi^h) dx,$$

so that

$$\begin{aligned}S_0 \delta_0^h &= \frac{e^{-r(n-1)h}}{(u - d)} \int_d^u \mathbb{E}[S_0 \xi^h \phi'(xS_0 \xi^h)] dx \\ &= e^{-r(n-1)h} \mathbb{E}[S_0 \xi^h \phi'(x^h S_0 \xi^h)]\end{aligned}$$

for some $x^h \in [d, u]$, by the mean value property. Assuming further that the limit $y \mapsto \psi(y) = y\phi'(y)$ is Lipschitz and bounded, then

$$\lim_{h \rightarrow 0} \mathbb{E}[S_0 \xi^h \phi'(x^h S_0 \xi^h)] = \lim_{h \rightarrow 0} \mathbb{E}[\psi(S_0 \xi^h)]. \quad (7.14)$$

Moreover, by convergence in law of S_n^h to S_T , we have

$$\lim_{h \rightarrow 0} \mathbb{E}[\psi(S_0 \xi^h)] = \mathbb{E}[\psi(S_T)] = e^{rT} S_0 \Delta_0^{bs},$$

where the last identity, in which Δ_0^{bs} denotes the Black–Scholes delta of the option at time 0, was established in (6.8).

This shows that, for bounded and regular payoff functions, the Cox–Ross–Rubinstein delta δ_0^h of a European option converges towards the Black–Scholes delta as $h \rightarrow 0$. This result can be extended to a vanilla put payoff by density, and then to a vanilla call payoff by call-put parity.

Convergence of deltas also holds for American options, although this can't be proved by elementary computations as in the European case.

7.2.2 Other Binomial Trees

To achieve convergence in law, many other choices of u , d and p can be done. The limiting law of S_n^h only depends on $pe^{i\lambda \ln(u)} + (1-p)e^{i\lambda \ln(d)}$ through its Taylor expansion up to $o(h)$. Thus u , d or/and p can be altered as long as the first order terms of the development are not modified. Also note that a binomial tree is recombining as soon as u and d remain constant within the tree.

7.2.2.1 Random Walk Scheme

A natural idea consists in approximating the Brownian motion W in Black–Scholes by the random walk approximation of Sect. 2.3.2, leading to

$$u = e^{bh+\sigma\sqrt{h}}, \quad d = e^{bh-\sigma\sqrt{h}}, \quad p = \frac{1}{2}.$$

7.2.2.2 Matching Three Moments Scheme

Here the idea is to match the first three conditional moments of the approximating Markov chain with those of the Black–Scholes model. We thereby obtain the following equations in u , d , p :

$$\begin{aligned} pu + (1-p)d &= e^{\kappa h} \\ pu^2 + (1-p)d^2 - e^{2\kappa h} &= e^{2\kappa h}(e^{\sigma^2 h} - 1) \\ pu^3 + (1-p)d^3 &= e^{3\kappa h}e^{3\sigma^2 h}, \end{aligned}$$

so that

$$\begin{aligned} u &= \frac{e^{\kappa h}\rho}{2}[1 + \rho + \sqrt{\rho^2 + 2\rho - 3}] \\ d &= \frac{e^{\kappa h}\rho}{2}[1 + \rho - \sqrt{\rho^2 + 2\rho - 3}] \\ p &= \frac{e^{\kappa h} - d}{u - d}, \end{aligned}$$

with $\rho = e^{\sigma^2 h}$.

7.2.3 Kamrad–Ritchken Trinomial Tree

The Kamrad–Ritchken tree [162] is a trinomial tree with $2n + 1$ possible values of the underlying S throughout the option life. It consists of a symmetric 3-point

approximation with space step k to $X = \ln(S)$, with up and down probabilities p_+ and p_- such that

$$\begin{aligned} k(p_+ - p_-) &= bh \\ k^2(p_+ + p_-) &= \sigma^2 h \end{aligned}$$

and with forward probability $p = 1 - p_+ - p_-$. Equivalently, in terms of the so-called stretch parameter λ defined by $k = \lambda\sigma\sqrt{h}$,

$$p_- = \frac{1}{2\lambda^2} - \frac{b\sqrt{h}}{2\lambda\sigma}, \quad p = 1 - \frac{1}{\lambda^2}, \quad p_+ = \frac{1}{2\lambda^2} + \frac{b\sqrt{h}}{2\lambda\sigma}.$$

The stretch parameter λ is a free parameter of the geometry of the tree. It must satisfy $\lambda \geq 1$ to ensure “nonnegativity of the probabilities”. The value $\lambda = 1.22474$, which corresponds to $p = \frac{1}{3}$, is reported to be a good choice for pricing an at-the-money call or put (see Fig. 7.1).

Note that the Kamrad–Ritchken tree essentially coincides with the explicit finite difference scheme of Sect. 8.3.2.1, except for a slightly different treatment of the discount factor (the ru term in the Black–Scholes equation).

7.2.4 Multinomial Trees

In a generic multinomial tree with geometry and dynamics defined by “up” factors u_j and corresponding probabilities p_j , the algorithms for pricing European and American options are the backward schemes defined by:

- $u_n(S) = v_n(S) = \phi(S)$ for every S in the tree at time n ,
- for $i = n - 1, \dots, 0$ for every S in the tree at time i :

$$\begin{aligned} u_i(S) &= e^{-rh} \sum p_j v_{i+1}(u_j S) \\ v_i(S) &= \max\left(\phi(S), e^{-rh} \sum p_j v_{i+1}(u_j S)\right). \end{aligned}$$

The equation ensuring convergence in law of the characteristic functions, and therefore of the one-dimensional marginals of S , is written: for every real λ ,

$$\sum p_j \exp(i\lambda \ln u_j) = 1 + \left[i\lambda b - \lambda^2 \frac{\sigma^2}{2} \right] h + o(h).$$

Typically this condition is equivalent to Kushner’s local consistency conditions, so that convergence in law also holds at the level of processes.

Note that, in order to get a recombining tree, u_{j+1}/u_j must not depend on j . Otherwise the complexity of the scheme is combinatorially exploding with n .

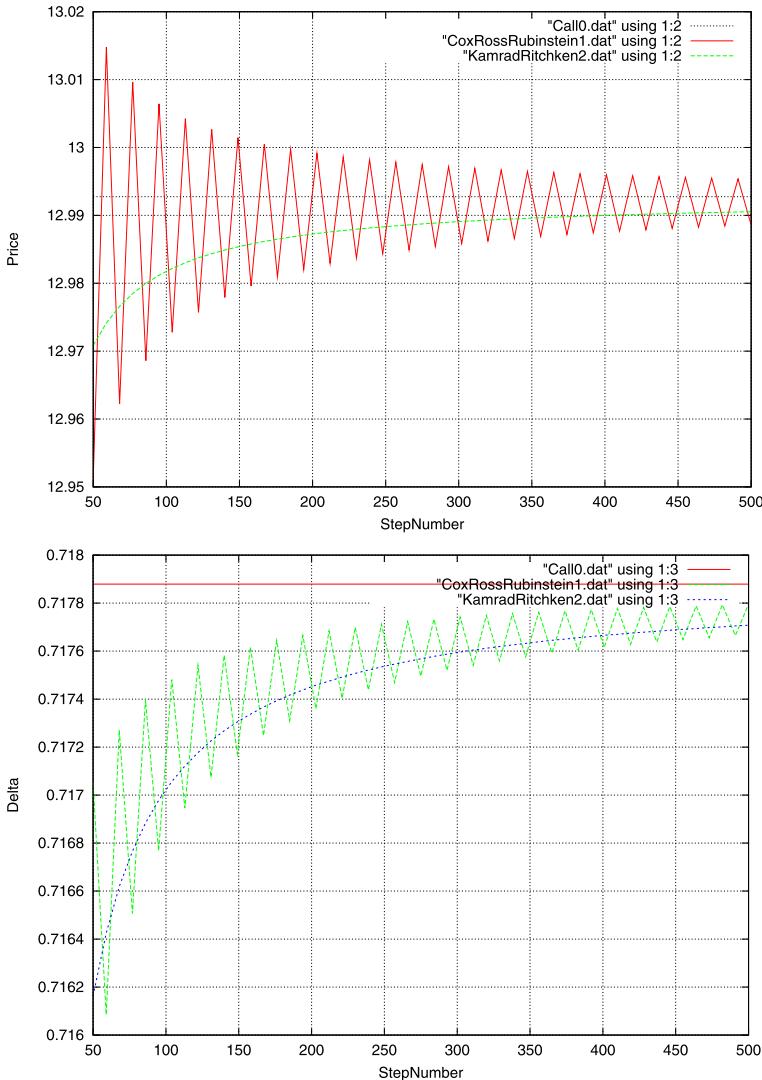


Fig. 7.1 European vanilla call in the Black–Scholes model, priced in a Cox–Ross–Rubinstein binomial tree and in a Kamrad–Ritchken trinomial tree

7.3 Trees for Exotic Options

7.3.1 Barrier Options

We now consider a down-and-out call option with a constant rebate R at the barrier level L (see Sect. 6.9.2). The first idea for pricing this option within a Cox–Ross–Rubinstein tree is to apply the usual backward induction scheme, with an option

price at or above the barrier constrained to R . It is possible to show that the resulting price converges to the right Black–Scholes limit. However, the convergence is slower than for vanilla options. To understand why, let us denote by l the tree node index such that

$$S_0 d^l \geq L > S_0 d^{l+1}.$$

For fixed n , the algorithm yields the same result for every value of the barrier between $S_0 d^l$ and $S_0 d^{l+1}$. Therefore the convergence cannot be faster than

$$\partial_L u^{bs}(d^l - d^{l+1}) = O(h^{\frac{1}{2}}),$$

where u^{bs} denotes the Black–Scholes pricing function of the barrier option. For comparison, the convergence rate in case of a European vanilla option is $O(h)$ (see Sect. 8.3.2.1).

An alternative method is to determine the stretch parameter λ of a trinomial Kamrad–Ritchken tree (see Sect. 7.2.3) so that the barrier is hit exactly [234]. Recall that λ must be greater than one in order to ensure stability of the scheme. One thus sets $\lambda = \frac{1}{m} \frac{\ln(\frac{S_0}{L})}{\sigma \sqrt{h}}$, with $m = \lceil \frac{\ln(\frac{S_0}{L})}{\sigma \sqrt{h}} \rceil$. For this choice of λ convergence is reported to be as fast as for vanilla options.

7.3.2 Bermudan Options

Bermudan options can only be exercised at specific times. Let us consider the case of an option putable on $[T_1, T]$ for some fixed T_1 in $(0, T)$. A convergent Cox–Ross–Rubinstein algorithm for pricing this option consists of the “American” backward induction formula (7.11) between step $n - 1$ and n_1 , followed by the “European” backward induction formula (7.8) before n_1 , where $(n_1 - 1)h < T_1 \leq n_1 h$. But this algorithm is very crude since it gives the same price, n being fixed, for every value of T_1 between $(n_1 - 1)h$ and $n_1 h$.

A better algorithm consists of two Kamrad–Ritchken trees pasted together at time T_1 , i.e.:

- a first tree with stretch parameter λ_1 and number of time steps n_1 between times 0 and T_1 , and
- another tree with stretch parameter λ_2 and number of time steps n_2 between times T_1 and T .

In order to get a recombining tree, we impose the following pasting condition at T_1 :

$$\lambda_1 \sqrt{\frac{T_1}{n_1}} = \lambda_2 \sqrt{\frac{T - T_1}{n_2}}.$$

For instance, we can first fix $\lambda_1 \geq 1$ (e.g., $\lambda_1 = 1.2274$) and n_1 , and then set

$$n_2 = \left\lceil \frac{n_1(T - T_1)}{T_1} \right\rceil + 1.$$

Thus $n_2 T_1 \geq n_1(T - T_1)$ and $\lambda_2^2 = \lambda_1^2 \frac{T_1}{n_1(T - T_1)} n_2 \geq \lambda_1^2 \geq 1$.

7.4 Bidimensional Trees

Until now we have only considered univariate trees with complexity $O(n^2)$. For the applications of this section one needs to consider bivariate trees of complexity $O(n^3)$.

7.4.1 Cox–Ross–Rubinstein Tree for Lookback Options

We consider a lookback option with payoff $\phi(S_T, M_T)$, where $M_t = \sup_{0 \leq s \leq t} S_s$. We can price this option in a bivariate tree $(S_i^h, M_i^h)_{0 \leq i \leq n}$, where M^h corresponds to the running maximum of a Cox–Ross–Rubinstein stock S^h . The related dynamic programming equation is written as follows:

- $u_n(S, M) = \phi(S, M)$ for every S in the Cox–Ross–Rubinstein tree at time n and M in the related lattice of all possible values of M_n^h ,
- for $i = n - 1, \dots, 0$, for every S in the Cox–Ross–Rubinstein tree at time i and M in the related lattice of all possible values of M_i^h we have

$$u_i(S, M) = e^{-rh} [pu_{i+1}(uS, \max(uS, M)) + (1-p)u_{i+1}(dS, M)], \quad (7.15)$$

having used that $dS_{i+1}^h \leq M_i^h$ since $d = e^{-\sigma\sqrt{h}} < 1$.

7.4.2 Kamrad–Ritchken Tree for Options on Two Assets

Assume a bivariate Black–Scholes model. So, for $l = 1, 2$,

$$dS_t^l = \kappa_l S_t^l dt + \sigma_l S_t^l dW_t^l,$$

where $\kappa_l = r - q_l$ and (W^1, W^2) is a bivariate Brownian motion with correlation ρ . Or, in log-returns for $l = 1, 2$,

$$dX_t^l = b_l dt + \sigma_l dW_t^l$$

with $b_l = \kappa_l - \frac{\sigma_l^2}{2}$. To approximate X we can use the bivariate Markov chain (X_i^h) such that $X_0^h = X_0$. Then for every $i = n-1, \dots, 0$ and for $l = 1, 2$,

$$X_{i+1}^{h,l} = X_i^{h,l} + \left(\kappa_l - \frac{\sigma_l^2}{2} \right) h + \sigma_l \sqrt{h} \varepsilon_i^l,$$

for i.i.d. $(\varepsilon_i^1, \varepsilon_i^2)$ such that

$$\mathbb{P}(\varepsilon_0^1 = 1, \varepsilon_0^2 = 1) = \mathbb{P}(\varepsilon_0^1 = -1, \varepsilon_0^2 = -1) = \frac{1+\rho}{4}$$

$$\mathbb{P}(\varepsilon_0^1 = 1, \varepsilon_0^2 = -1) = \mathbb{P}(\varepsilon_0^1 = -1, \varepsilon_0^2 = 1) = \frac{1-\rho}{4}.$$

We can easily check that

$$\lim_{h \rightarrow 0} h^{-1} \mathbb{E}_i(X_{i+1}^h - X_i^h) = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

$$\lim_{h \rightarrow 0} h^{-1} \mathbb{C}\text{ov}_i(X_{i+1}^h - X_i^h) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Thus Kushner's local consistency conditions are satisfied and (a suitable continuous-time Markov chain interpolation⁴ of) X^h converges in law to X .

⁴See Sect. 4.3 of Kushner and Dupuis [170].

Chapter 8

Finite Differences

Like tree methods and as opposed to simulation methods, finite difference methods can easily cope with early exercise features, and in low dimension they can give an accurate and robust computation of an option price and Greeks (delta, gamma, and theta at time 0). However, they are not practical for dimension greater than three or four, for then too many grid points are required for achieving satisfactory accuracy.

8.1 Generic Pricing PIDE

In a Markov setup, derivative prices and Greeks can be expressed in terms of a related pricing function. We thus consider a financial claim with cash flows given as functions of the jump-diffusion factor process X in (4.30). We saw in Sects. 3.5 and 4.2 (see also Chaps. 12 and 13) that the price Π_t of the claim can be represented as $\Pi_t = u(t, X_t)$ for some function u . All the randomness of Π is encoded in that of X . The pricing problem is reduced to the computation of the pricing function u . This is the basic mechanism through which deterministic methods can be used to compute derivative prices, though these are in essence stochastic processes.

Moreover (see Sect. 3.5 and Chap. 13), the pricing function u solves a parabolic partial integro-differential equation of the following form on the time-space domain $E = [0, T] \times \mathbb{R}^d$, where T is the maturity of the claim:

$$\begin{cases} F(t, x, u(t, x), \partial_t u(t, x), \partial u(t, x), \partial^2 u(t, x), \mathcal{I}u(t, x)) = 0 & \text{on } [0, T) \times \mathcal{O} \\ u(t, x) = \phi(t, x) & \text{on } E \setminus ([0, T) \times \mathcal{O}). \end{cases} \quad (8.1)$$

In this equation:

- \mathcal{O} is an open subset of \mathbb{R}^d . In particular, the terminal condition at T , which is embedded in the boundary condition ϕ in (8.1), is given by the payoff of the claim at maturity.
- ∂u and $\partial^2 u$ denote, respectively, the row-gradient and the Hessian matrix of u with respect to x .

- $\mathcal{I}u$ is the nonlocal (integral) term

$$\mathcal{I}u(t, x) = \lambda(t, x) \int_{\mathbb{R}^d} (u(t, x + \delta(t, x, y)) - u(t, x)) \eta(t, x, y) w(t, x, dy) \quad (8.2)$$

for some \mathbb{R}_+^q -valued integration kernel η (where λ, w and δ are part of the specification of X ; see Remark 3.2.8).

The precise definitions of the domain \mathcal{O} and of the operator F depend on the specifications of the derivative at hand. In the case of a European vanilla option in space dimension one, F is a linear operator and $\mathcal{O} = (-\infty, +\infty)$ or $(0, +\infty)$. In the case of a barrier option, \mathcal{O} is limited by the barriers. In the case of an American or game option, there are further obstacles in F . In any case, one deals with a monotone operator F in the sense that one has, for all $(t, x) \in [0, T] \times \mathcal{O}, u, v, a \in \mathbb{R}, p \in \mathbb{R}^d, A, B \in \mathbb{R}^{d \times d}$ and $I, J \in \mathbb{R}^q$:

$$F(t, x, u, a, p, A, I) \leq F(t, x, v, a, p, B, J) \quad \text{whenever } u \leq v, A \geq B \text{ and } I \geq J. \quad (8.3)$$

Here the inequality $I \geq J$ is meant componentwise, and the inequality $A \geq B$ in the sense of the usual order on the space of the real-valued symmetric nonnegative $d \times d$ -matrices (nonnegative eigenvalues of $A - B$).

8.1.1 Maximum Principle

A classical solution to the deterministic pricing equation (8.1) is a function $u = u(t, x)$ in $\mathcal{C}([0, T] \times \mathbb{R}^d) \cap \mathcal{C}^{1,2}([0, T] \times \mathcal{O})$ satisfying (8.1) pointwise over $[0, T] \times \mathbb{R}^d$. A classical subsolution (respectively classical supersolution) to (8.1) satisfies (8.1) pointwise over $[0, T] \times \mathbb{R}^d$, with $=$ replaced by \leq (respectively \geq).

Let us assume for simplicity that \mathcal{O} is bounded and that the monotonicity of F is strict in its third argument u . We then have the following comparison principle:

Proposition 8.1.1 *We have $u \leq v$ for every classical subsolution u and any classical supersolution v of (8.1).*

Proof Assume that $u \leq v$ doesn't hold. Then $w = u - v$ admits a positive maximum at a point $(t, x) \in [0, T] \times \mathcal{O}$. By a second-order optimality condition and, by definition of $\mathcal{I}u$, we have

$$\begin{aligned} \partial_t u(t, x) &= \partial_t v(t, x), & \mathcal{I}u(t, x) &\leq \mathcal{I}v(t, x) \\ \partial u(t, x) &= \partial v(t, x), & \partial^2 u(t, x) &\leq \partial^2 v(t, x). \end{aligned}$$

Hence the fact that

$$\begin{aligned} F(t, x, u(t, x), \partial_t u(t, x), \partial u(t, x), \partial^2 u(t, x), \mathcal{I}u(t, x)) &\leq 0 \\ \leq F(t, x, v(t, x), \partial_t v(t, x), \partial v(t, x), \partial^2 v(t, x), \mathcal{I}v(t, x)) \end{aligned}$$

implies, by the monotonicity of F in its arguments A and I :

$$\begin{aligned} F(t, x, u(t, x), \partial_t u(t, x), \partial u(t, x), \partial^2 u(t, x), \mathcal{I}u(t, x)) \\ \leq F(t, x, v(t, x), \partial_t v(t, x), \partial v(t, x), \partial^2 v(t, x), \mathcal{I}v(t, x)). \end{aligned}$$

But $>$ should hold, since $u(t, x) > v(t, x)$, taking into account the assumed strict monotonicity of F in its argument u . \square

Proposition 8.1.1 implies uniqueness for a classical solution to the pricing equation (8.1). In case F is linear, there exists a classical solution u to (8.1) (see [129] or Morton and Mayers [129]). However, if F is nonlinear, a classical solution doesn't generally exist and one must resort to suitable notions of weak solutions to (8.1).

8.1.2 Weak Solutions

8.1.2.1 Viscosity Solutions

The theory of viscosity solutions [3–5, 20, 72, 75, 122, 155, 156] defines suitable notions of weak solutions, subsolutions and supersolutions to (8.1) such that maximum and comparison principles hold for any nonlinear monotone operator F . This grants the uniqueness for a viscosity solution to (8.1). Existence is in turn established by various means, such as Perron's method (which incidentally, is itself based on comparison principles). The related viscosity solution technicalities are dealt with in Chap. 13. For practical use in the next sections, it will be enough for us to keep in mind that, under mild assumptions, the pricing equation (8.1) is well-posed in a suitable space of viscosity solutions.

8.1.2.2 Sobolev Solutions

As an alternative to viscosity solutions, it is possible to derive weak variational formulations of the deterministic pricing equation (8.1) in Sobolev functional spaces \mathcal{H} . In this approach, the boundary condition ϕ is typically accounted for by a judicious choice of the space \mathcal{H} . Existence and uniqueness for a solution to the variational formulation of (8.1) is then obtained by application of a Lax–Milgram theorem. Various choices for \mathcal{H} are possible; see Bally et al. [16, 17], Barles–Lesigne [22], Achdou–Pironneau [1], Matache et al. [200], Jaitlet et al. [154] or Ern et al. [118]. The precise variational formulation of the pricing equations is outside the scope of this book. We will simply keep in mind that, under mild technical assumptions, a variational formulation of the deterministic pricing equation (8.1) is well-posed in a suitable Sobolev space \mathcal{H} . This underlies the theory of related finite element approximation schemes.

8.2 Numerical Approximation

In the case of European vanilla options in simple models, the deterministic pricing equation (8.1) can be solved analytically. But, in general, (8.1) must be solved numerically. In order to approximate (8.1), one can either use the finite difference methods of this chapter (see also e.g. [248, 260]) or resort to more general finite element methods [1]. Note that there is no hermetic frontier between these methods. One thus has, schematically:

$$\text{Tree Methods} \subset \text{Finite Differences Methods} \subset \text{Finite Elements Methods}$$

and also, in a sense, as discussed in the preface of the book:

$$\text{MC Methods} \subset \text{Tree Methods}.$$

8.2.1 Finite Difference Methods

Finite difference methods are naturally connected with viscosity solutions of monotone equations satisfying related maximum principles. The additional complexity of using potentially more powerful finite element methods is mainly justified when the geometry of the domain makes it necessary to use an unstructured, adaptive discretization mesh. But pricing problems in finance are typically posed on rectangular domains, for which a simple finite difference grid is good enough.

8.2.1.1 Localization and Discretization

The numerical solution of the pricing equation (8.1) by finite differences is a four step process:

i. ***Transformations*** of the problem, such as:

- ***changes of variables*** (e.g. $x = \ln(S)$),
- ***changes of unknowns*** (e.g., solving the equation for $e^{-rt}u$ rather than u),
- ***changes of probability measure***.¹

ii. ***Localization*** of the problem, that is:

- replacing \mathcal{O} by a ***bounded domain*** of (8.1) (in case \mathcal{O} is not bounded from the beginning), still denoted by \mathcal{O} henceforth, and introducing a suitable boundary condition φ outside $[0, T) \times \mathcal{O}$ such that, in particular, $\varphi = \phi$ at T ,

¹See [124] for a PDE version of the Girsanov transformation.

- replacing the integration domain \mathbb{R}^q in the integral term $\mathcal{I}u$, by a **bounded integration domain** $\mathcal{O}(t, x)$ such that

$$\int_{\mathcal{O}(t,x)} w(t, x, dy) \approx 1.$$

- Discretizing** the localized problem by a finite difference numerical scheme defined on a suitable mesh over \mathcal{O} .
- Solving** the resulting linear algebra problem numerically in the values of an approximate solution defined at mesh nodes.

To exploit the parabolic structure of the pricing equations, the time dimension is typically treated separately at step (iii). The problem is then solved iteratively in time at step (iv), which saves one dimension of storage cost.

8.2.1.2 Convergence Analysis

Given a mesh discretizing $[0, T] \times \mathbb{R}^q$, with time step h and space steps $k = (k_1, \dots, k_d)$, let

$$\begin{cases} F_h^k(u_h^k) = 0 & \text{on } [0, T] \times \mathcal{O} \\ u_h^k = \varphi & \text{on } E \setminus ([0, T] \times \mathcal{O}) \end{cases} \quad (8.4)$$

denote a fully discrete finite differences approximation scheme for (a localized version of) (8.1). We assume that the discretized problem (8.4) admits a unique solution u_h^k defined on the mesh, satisfying at every mesh node in $[0, T] \times E \setminus ([0, T] \times \mathcal{O})$ the related equation which appears in functional form in (8.4).

Under mild technical assumptions, the pricing equation (8.1) has a unique solution u in a suitable space of viscosity solutions. For simplicity we assume that u is bounded. In the purely differential case ($\lambda = 0$ in X) and in case F is linear, u is in fact a classical solution to (8.1). The Lax equivalence theorem then states that any stable and consistent scheme F_h^k is convergent, which loosely means that $u_h^k \rightarrow u$ at mesh points as h and k go to 0, provided:

- **(consistency)** $F_h^k(u) \rightarrow 0$ at mesh points as $h, k \rightarrow 0$,
- **(stability)** u_h^k is bounded, uniformly over h, k ,

where u in $F_h^k(u)$ represents the solution to (8.1). The essence of the Lax equivalence theorem is that any “reasonable” (consistent) scheme converges, provided it is nonexplosive (stable). We refer the reader to Morton and Meyers [207] for the details that, in particular, involve the specification of a given norm in which u_h^k is bounded, $F_h^k(u) \rightarrow 0$ and $u_h^k \rightarrow u$ as $h, k \rightarrow 0$.

With the help of viscosity solutions, the Lax equivalence theorem can be generalized to a nonlinear monotone operator F and/or to jumps in X . For a monotone, yet still purely differential, operator F , Barles and Souganidis [23] proved the convergence of any monotone, stable and consistent approximation scheme for (8.1)

(assuming a viscosity solution comparison principle holds). Monotonicity of the scheme is a discrete version of the monotonicity condition (8.3) on F ; it is satisfied by a broad family of finite difference schemes. The convergence conditions in the Barles–Souganidis theorem are thus essentially reduced to those of the Lax equivalence theorem, namely consistency and stability. Moreover, this can be extended to equations with jumps and/or to systems of equations; see Chap. 13.

The companion issue to convergence is convergence rate, which is the speed at which u_h^k converges to u as $h, k \rightarrow 0$. The key notion here is that of the order or the consistency of a scheme, a measure of the speed at which $F_h^k(u) \rightarrow 0$ as $h, k \rightarrow 0$. More precisely, a finite difference scheme is said to be consistent at order p in time and q in space if

$$F_h^k(u) = O(h^p) + O(|k|^q) \quad (8.5)$$

(see Sect. 8.3.1 for illustration). The general idea is that the order of consistency (speed at which “the operator F defining u converges to the operator F_h^k defining u_h^k ”) determines the convergence rate of a scheme (speed at which u converges to u_h^k). The order of consistency puts a bound on the convergence rate, but there can be a deterioration between order of consistency and rate of convergence in case of a lack of regularity of the domain \mathcal{O} and/or of the boundary condition φ (the residual data defining u , along with F , through (8.1)).

8.2.2 Finite Elements and Beyond

Finite element methods are based on variational formulations of the pricing equation (8.1). They give approximate solutions defined on the whole state space E , as opposed to approximate solutions at grid points by finite difference methods. They are most naturally connected with equations expressing energy conservation principles.

They are typically heavier than finite difference methods, particularly in terms of storage cost. Indeed, a prerequisite of a finite element method is the construction of a discretization mesh, typically unstructured and adaptive, which has to be handled by the computer during the computation. This also means that finite element methods are harder to implement. One may use finite element toolboxes, but this results in less flexibility in the programming.

The additional cost of finite element methods is justified in cases of a domain \mathcal{O} with a curved boundary. This can for instance be the case for pricing a curved barrier fixed-income option, where the curved barrier arises via the exponential relation between bond rates and bond prices, or for accurately computing the exercise boundary of an American option.

With regard to the curse of dimensionality, an advantage of finite element methods is that they allow us to refine the approximation mesh in a more clever way than simply taking the product of univariate adaptive meshes with finite differences.

Another interesting feature of finite elements methods is the availability of powerful a priori and a posteriori error estimates theory to deal with convergence and convergence rate issues.

Practically speaking, the numerical solution of the pricing equation (8.1) by a finite element method, is a five steps process:

- i. ***transformations*** of the problem,
- ii. ***localization***, i.e. truncating the set \mathcal{O} and the integration domain in $\mathcal{I}u$, and introducing a suitable boundary condition outside the localized domain $[0, T) \times \mathcal{O}$,
- iii. derivation of a ***weak formulation*** of the localized problem in a Sobolev space \mathcal{H} and accounting for the boundary condition,
- iv. ***projection*** of the problem onto a finite-dimensional sub-space of finite elements $\mathcal{H}_h^k \subset \mathcal{H}$,
- v. ***solving*** the resulting high-dimensional linear algebra system in the coefficients of an approximate solution on a finite element basis.

Existence and uniqueness for a solution to the weak form of the localized problem at step iii is typically obtained by application of a Lax–Milgram theorem. As with finite difference methods, the time dimension is, in general, treated separately. The resulting problem at step v may thus be solved iteratively in time, at constant storage cost. To solve the linear systems arising at every time step in the algorithm, an iterative solver is required. In the context of finite element methods, where a PDE is rephrased as a variational problem, it is natural after discretization to use an optimization-based iterative solver. A conjugate gradient descent known as the “generalized minimal residual algorithm” (GMRES [239]), is the current industry standard in this regard.

8.2.2.1 Finite Volumes

Finite volume methods can be regarded as counterparts of finite element methods in which the test-functions that are used in the variational formulation of the problem are indicator functions instead of more regular test-functions with finite elements. Finite volume methods are particularly well suited for dealing with problems with discontinuous data, such as pricing digital options.

8.2.2.2 Sparse Grid Techniques

Sparse grid methods [232] are used to represent, integrate or interpolate high-dimensional functions. These methods rely on the seminal works of the Russian mathematician Smolyak, who found a clever quadrature rule to counter the curse of dimensionality. This direction of research underlies active developments in finite element and finite difference methods; but the related algorithms are difficult to implement.

For simplicity we focus on finite difference methods in the sequel.

8.3 Finite Differences for European Vanilla Options

We saw in Sect. 5.1.1 that, in the risk-neutral Black–Scholes model

$$\frac{dS_t}{S_t} = \kappa dt + \sigma dW_t,$$

the price process of a European vanilla option with integrable payoff $\phi(S_T)$ at T is given by $\Pi_t = v(t, S_t)$, where the pricing function v solves the following Black–Scholes PDE:

$$\begin{cases} v(T, S) = \phi(S), & S \in (0, +\infty) \\ \partial_t v + \kappa S \partial_S v + \frac{1}{2} \sigma^2 S^2 \partial_{S^2}^2 v - r v = 0 & \text{in } [0, T) \times (0, +\infty). \end{cases} \quad (8.6)$$

In log-returns $X_t = \ln(S_t)$, the option price process is given as $\Pi_t = u(t, X_t)$, where u solves the following equation:

$$\begin{cases} u(T, x) = \psi(x), & x \in \mathbb{R} \\ \partial_t u + b \partial_x u + \frac{1}{2} \sigma^2 \partial_{x^2}^2 u - ru = 0 & \text{in } [0, T) \times (-\infty, +\infty), \end{cases} \quad (8.7)$$

with $b = \kappa - \frac{\sigma^2}{2}$ and $\psi(x) = \phi(e^x)$. For “nice” terminal conditions, both (8.6) and (8.7) are well-posed in terms of classical solutions (see Sect. 8.1.1).

8.3.1 Localization and Discretization in Space

Let $x = \ln(S_0)$. To solve (8.7) numerically, we localize the space-domain to $\mathcal{O} = (x - \ell, x + \ell)$, where ℓ is chosen so that

$$\mathbb{P}(|X_t - x| \leq \ell, t \in [0, T]) \geq 1 - \alpha \quad (8.8)$$

for a “sufficiently small” $\alpha > 0$. This is achieved by setting

$$\ell = |b|T + f\sigma\sqrt{T} \quad (8.9)$$

for a “sufficiently high” quantile f of the Gaussian distribution.

Letting $k = \frac{2\ell}{m+1}$ and $x_j = x - \ell + jk$ for $0 \leq j \leq m+1$, one then approximates the differential spatial operator

$$\mathcal{A}u = \frac{1}{2} \sigma^2 \partial_{x^2}^2 u + b \partial_x u - ru$$

by a discrete operator \mathcal{A}^k acting on \mathbb{R}^m -valued vectors $u^k = (u^k(t, x_1), \dots, u^k(t, x_m))^T$.

Remark 8.3.1 For the purpose of this section, it is convenient to include the discount term $(-ru)$ in \mathcal{A} ; elsewhere in the book \mathcal{A} refers to the generator of the factor process X (see (3.66)), without the term $(-ru)$.

A common specification is

$$\mathcal{A}^k u^k(t, x_j) = \frac{1}{2} \sigma^2 \delta_{x^2}^2 u^k(t, x_j) + b \delta_x u^k(t, x_j) - ru^k(t, x_j), \quad (8.10)$$

with

$$\begin{aligned} \delta_x u^k(t, x_j) &= \frac{1}{2k} (u^k(t, x_{j+1}) - u^k(t, x_{j-1})) \\ \delta_{x^2}^2 u^k(t, x_j) &= \frac{1}{k^2} (u^k(t, x_{j+1}) - 2u^k(t, x_j) + u^k(t, x_{j-1})), \end{aligned}$$

where $u^k(t, x_0) = u^k(t, x - \ell)$ and $u^k(t, x_{m+1}) = u^k(t, x + \ell)$ are to be understood as notation for quantities to be defined below in terms of the $u^k(t, x_j)$, $j = 1, \dots, m$. By Taylor expansion, one can show that δ_x and $\delta_{x^2}^2$ are consistent approximations of order two for the spatial differential operators ∂_x and ∂_{x^2} , meaning that, for every regular test-function $\varphi(x)$, we have:

$$|\delta_x \varphi(x_j) - \partial_x \varphi(x_j)|, |\delta_{x^2}^2 \varphi(x_j) - \partial_{x^2}^2 \varphi(x_j)| = O(k^2).$$

If $|\kappa|/\sigma^2$ is “large”, a less accurate but more stable approximation for ∂_x is given by

$$\delta_x u^k(t, x_j) = \begin{cases} \frac{1}{k} (u^k(t, x_j) - u^k(t, x_{j-1})) & \text{if } b < 0 \\ \frac{1}{k} (u^k(t, x_{j+1}) - u^k(t, x_j)) & \text{if } b > 0. \end{cases}$$

This so-called upwind² discretization of $b\partial_x$ follows the characteristics of the limiting hyperbolic transport equation in which $\sigma = 0$.

One then seeks an \mathbb{R}^m -valued time functional $u^k(t)$ satisfying the following system of ODEs: for $1 \leq j \leq m$,

$$\begin{cases} u^k(T, x_j) = \psi(x_j) \\ \frac{du^k}{dt}(t, x_j) + \mathcal{A}^k u^k(t, x_j) = 0 \quad \text{for } 0 \leq t < T, \end{cases} \quad (8.11)$$

where the quantities $u^k(t, x \pm \ell)$ in (8.11) are to be understood as follows, given a function φ defined over the space-boundary $[0, T] \times \partial\mathcal{O}$ of the localized domain:

- in the case of a so-called Dirichlet boundary condition,

$$u^k(t, x \pm \ell) = \varphi(t, x \pm \ell);$$

²Recalling that the pricing equations are posed in backward time, with a terminal condition at time T .

- in the case of a so-called Neumann boundary condition,

$$u^k(t, x - \ell) = u^k(t, x_1) - k \partial_x \varphi(t, x - \ell)$$

$$u^k(t, x + \ell) = u^k(t, x_m) + k \partial_x \varphi(t, x + \ell).$$

Note that by letting

$$\alpha = \frac{\sigma^2}{2k^2} - \frac{b}{2k}, \quad \beta = -\left(\frac{\sigma^2}{k^2} + r\right), \quad \gamma = \frac{\sigma^2}{2k^2} + \frac{b}{2k}, \quad (8.12)$$

one has

$$\mathcal{A}^k u^k(t) = A^k u^k(t) + w^k(t), \quad (8.13)$$

so that in the case of a Dirichlet boundary condition:

$$A^k = \begin{bmatrix} \beta & \gamma & 0 & \cdots & 0 & 0 \\ \alpha & \beta & \gamma & 0 & \cdots & 0 \\ 0 & \alpha & \beta & \gamma & \cdots & 0 \\ 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha & \beta & \gamma \\ 0 & 0 & 0 & \cdots & \alpha & \beta \end{bmatrix}, \quad w^k(t) = \begin{bmatrix} \varphi(t, x - \ell) \alpha \\ 0 \\ \vdots \\ 0 \\ \varphi(t, x + \ell) \gamma \end{bmatrix}, \quad (8.14)$$

while in the case of a Neumann boundary condition:

$$A^k = \begin{bmatrix} \beta + \alpha & \gamma & 0 & \cdots & 0 & 0 \\ \alpha & \beta & \gamma & 0 & \cdots & 0 \\ 0 & \alpha & \beta & \gamma & \cdots & 0 \\ 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha & \beta & \gamma \\ 0 & 0 & 0 & \cdots & \alpha & \beta + \gamma \end{bmatrix}, \quad w^k(t) = \begin{bmatrix} -\alpha k \partial_x \varphi(t, x - \ell) \\ 0 \\ \vdots \\ 0 \\ \gamma k \partial_x \varphi(t, x + \ell) \end{bmatrix}. \quad (8.15)$$

8.3.2 Theta-Schemes in Time

We now discuss time-discretization of the ODE system (8.11). We will focus on the so-called θ -schemes, which in the case of the parabolic equation (8.7) may be summarized as follows (see, e.g. [174, 207]). Given a parameter $\theta \in [0, 1]$, we choose a time discretization step h such that $T = nh$ and we construct a fully discrete approximation $u_h^k(t_i, x_j) \equiv u_i^j$, where the \mathbb{R}^m -valued vectors $u_i = (u_i^j)^{1 \leq j \leq m}$ satisfy:

$$\begin{cases} u_n = \psi, & \text{and, for } i = n-1, \dots, 0 \\ h^{-1}(u_{i+1} - u_i) + \mathcal{A}^k(\theta u_i + (1-\theta)u_{i+1}) = 0. \end{cases}$$

Or equivalently in view of (8.13):

$$\begin{cases} u_n = \psi, & \text{and for } i = n-1, \dots, 0 \\ [I - h\theta A^k]u_i = [I + h(1-\theta)A^k]u_{i+1} + hw_i, \end{cases} \quad (8.16)$$

where

$$w_i = \theta w^k(t_i) + (1-\theta)w^k(t_{i+1}). \quad (8.17)$$

For $\theta = 0$ we get the so-called explicit scheme, for $\theta = 1$ the fully implicit scheme and for $\theta = \frac{1}{2}$ the Crank-Nicholson scheme.

Once we have computed the approximated prices u_i^j , we can recover the delta and the gamma,

$$\Delta = \partial_S v = e^{-x} \partial_x u, \quad \Gamma = \partial_{S^2}^2 v = e^{-2x} (\partial_{x^2}^2 u - \partial_x u),$$

by their approximations

$$\begin{aligned} \Delta_i^j &= e^{-x_j} \frac{u_i^{j+1} - u_i^{j-1}}{2k} \\ \Gamma_i^j &= e^{-2x_j} \left(\frac{u_i^{j+1} - 2u_i^j + u_i^{j-1}}{k^2} - \frac{u_i^{j+1} - u_i^{j-1}}{2k} \right). \end{aligned}$$

For m odd, $x = \ln(S_0)$ lies in the space grid. Otherwise some interpolation has to be used to recover the approximate price and Greeks at x .

8.3.2.1 The Explicit Scheme

Let us first discuss the explicit scheme $\theta = 0$. By definition of A^k in (8.14) (in the case of a Dirichlet condition), the θ -scheme (8.16), with $\theta = 0$, can be rewritten as: $u_n = \phi$ and for all $i = n-1, \dots, 0$, $j = 1, \dots, m$:

$$u_i^j = p_- u_{i+1}^{j-1} + p u_{i+1}^j + p_+ u_{i+1}^{j+1}, \quad (8.18)$$

in which

$$p_- = h\alpha, \quad p = 1 + h\beta, \quad p_+ = h\gamma$$

and $u_{i+1}(x \pm \ell) := \varphi((i+1)h, x \pm \ell)$. Using the specification of the coefficients α , β and γ in (8.12), one can show that this scheme is:³

- stable, provided $h \leq \frac{k^2}{\sigma^2 + rk^2}$ (and $\sigma^2 > |b|k$, but for $\sigma > 0$ this is always satisfied for sufficiently small k);
- consistent of order one in time and two in space.

³See Sect. 8.2.1.2 for the notions of stability and consistency.

8.3.2.2 Implicit Schemes

For $\theta > 0$, one has to solve, in (8.16), at every time step $i < n$, a linear equation

$$Mu_i = Nu_{i+1} + hw_i, \quad (8.19)$$

where w_i was defined in (8.17), and $M = I - h\theta A^k$ and $N = I + h(1 - \theta)A^k$ are tridiagonal matrices of the form

$$\begin{pmatrix} p & p_+ & 0 & \cdots & 0 & 0 \\ p_- & p & p_+ & 0 & \cdots & 0 \\ 0 & p_- & p & p_+ & \cdots & 0 \\ 0 & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & p_- & p & p_+ \\ 0 & 0 & 0 & \cdots & p_- & p \end{pmatrix}. \quad (8.20)$$

Example 8.3.2 In the case (8.14) of a Dirichlet boundary condition, one has for M :

$$p_- = -\theta h\alpha, \quad p = 1 - \theta h\beta, \quad p_+ = -\theta h\gamma, \quad (8.21)$$

and for N :

$$p_- = (1 - \theta)h\alpha, \quad p = 1 + (1 - \theta)h\beta, \quad p_+ = (1 - \theta)h\gamma,$$

(consistent with (8.18) for the explicit scheme $\theta = 0$).

One can show that implicit schemes for the Black–Scholes equation (8.7) are:

- unconditionally stable for $\theta \geq \frac{1}{2}$; otherwise they are like the explicit scheme subject to a suitable stability condition;
- consistent of order one in time and two in space, except for the Crank–Nicholson scheme ($\theta = \frac{1}{2}$), which is consistent of order two in time and space.

In Fig. 8.1 we have plotted the relative error at time 0 as a function of the spot price S_0 , when pricing a European vanilla call option with the explicit, the fully implicit and the Crank–Nicholson schemes. As expected the Crank–Nicholson is more accurate, at least around the money.

An implicit θ -scheme (with $\theta > 0$) requires, at each time step, the solution of a linear system of the form $Mu = v$, where u and v are m -dimensional vectors and the matrix M is tridiagonal. We now describe two algorithms for solving such linear systems.

Solution by Gauss Factorization Let $M = LU$ denote the Gauss factorization of a (regular) matrix M , where L is lower triangular and U is upper triangular, with only ones on the diagonals. The linear system $MUu = v$ can be decomposed into $Lz = v$, $Uu = z$. It is easy to see that if, as in our case, the matrix M is tridiagonal, then so are also L and U , so that only the upper diagonal of U and the two diagonals

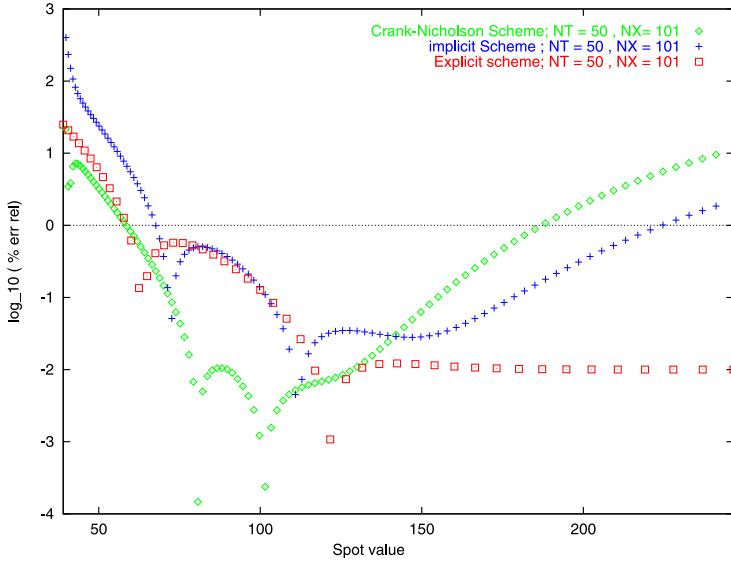


Fig. 8.1 Pricing of a European call option by θ -schemes, with $\theta = 0, \frac{1}{2}$ and 1 ($r = 10\%$, $\sigma = 20\%$, $T = 1$ y, $K = 100$)

of L need be found. We thereby obtain the following procedure for solving $Mu = v$, known as the Thomas algorithm [174, 207]:

- $b'_m = b_m$, $z_m = v_m$ and, for j decreasing from $m - 1$ to 1,

$$\begin{aligned} b'_j &= b_j - c_j a_{j+1}/b'_{j+1}, \\ z_j &= v_j - c_j z_{j+1}/b'_{j+1}; \end{aligned}$$

- $u_1 = z_1/b'_1$, and, for j increasing from 2 to m ,

$$u_j = (z_j - a_j u_{j-1})/b'_j,$$

(assuming non zero pivots b'_j).

Iterative Solution An alternative, which in the case of a tridiagonal system is only justified by its programming simplicity, is to use an iterative scheme. The matrix M resulting from the discretization of a parabolic PDE is typically diagonal-dominant, meaning that it has a “large” diagonal part D as compared with $R = M - D$ (see e.g. (8.21)). With the so-called successive over-relaxation scheme, the linear system $Mu = v$ is rewritten as $Du = v - Ru$. Starting from an initial condition u^0 , a solution is then computed as the limit of the “contracting Picard iteration” (since D is “large” and D^{-1} is therefore “small”):

$$u^{k+1} = D^{-1}(v - Ru^k); \quad (8.22)$$

Or, more generally, given an over-relaxation parameter $1 \leq \omega < 2$,

$$u^{k+1} = u^k + \omega(\tilde{u}^{k+1} - u^k) \quad \text{where } \tilde{u}^{k+1} = D^{-1}(v - Ru^k).$$

The detailed algorithm appears as follows: set $k = 0$ and until $|u^{k+1} - u^k|$ is less than some specified tolerance:

- (**Jacobi iteration**) Form an auxiliary vector $\tilde{u}^{k+1} = (\tilde{u}_j^{k+1})_{1 \leq j \leq m}$, for $1 \leq j \leq m$, by

$$\tilde{u}_j^{k+1} = \frac{1}{M_{jj}} \left(v_j - \sum_{l < j} M_{j,l} u_l^k - \sum_{l > j} M_{j,l} u_l^k \right). \quad (8.23)$$

Here a possible refinement known as Gauss–Seidel iteration consists in using \tilde{u}_l^{k+1} instead of u_l^k in the first sum.

- (**over-relaxation**) Let $u^{k+1} = u^k + \omega(\tilde{u}^{k+1} - u^k)$ and set $k = k + 1$.

8.3.3 Adding Jumps

We now add jumps in S , considering the Merton model

$$\frac{dS_t}{S_{t-}} = (\kappa - \lambda \bar{J}) dt + \sigma dW_t + J_{(t)} dN_t \quad (8.24)$$

of Sect. 5.1.3. As reviewed in Sect. 3.5, the price process of the option can be represented as $\Pi_t = u(t, X_t)$, with $X_t = \ln(S_t)$, where the pricing function u solves the partial integro-differential equation

$$\begin{cases} u(T, x) = \psi(x), & x \in \mathbb{R} \\ \partial_t u + Cu + \mathcal{I}u = 0 & \text{in } [0, T) \times \mathbb{R}, \end{cases} \quad (8.25)$$

with

$$\begin{aligned} Cu(t, x) &= \left(\frac{1}{2}\sigma^2 \partial_{x^2}^2 u + a \partial_x u - ru \right)(t, x), \mathcal{I}u(t, x) \\ &= \lambda \int_{\mathbb{R}} (u(t, x+y) - u(t, x)) w(y) dy, \end{aligned}$$

in which $a = b - \lambda \bar{J} = \kappa - \frac{1}{2}\sigma^2 - \lambda \bar{J}$ and where w is the $\mathcal{N}(\alpha, \beta)$ -density.

8.3.3.1 Localization

Localization is essentially performed as in Sect. 8.3.1, except that:

- the coefficient b is replaced by a in (8.9);

- the boundary $\partial\mathcal{O} = \{x - \ell\} \cup \{x + \ell\}$ is replaced by the “thick” boundary layer $\partial\mathcal{O}_j = [x - \ell - \underline{z}^-, x - \ell] \cup [x + \ell, x + \ell + \bar{z}^+]$, where \underline{z} and \bar{z} are such that $\int_{\underline{z}}^{\bar{z}} w(z) dz \approx 1$ (cf. step (ii) in Sect. 8.2.1.1);
- we solve the following localized problem on $\overline{\mathcal{O}}_j = \mathcal{O} \cup \partial\mathcal{O}_j$:

$$\begin{cases} u(T, x) = \psi(x), & x \in \mathbb{R} \\ u(t, x) = \varphi(t, x), & (t, x) \in [0, T) \times \partial\mathcal{O}_j \\ (\partial_t u + \mathcal{C}u + \mathcal{J}u + \Phi)(t, x) = 0, & (t, x) \in [0, T) \times \mathcal{O}, \end{cases} \quad (8.26)$$

where φ is a Dirichlet condition such that $\varphi(T, \cdot) = \psi$ and, for $(t, x) \in [0, T) \times \mathcal{O}$, with

$$\begin{aligned} \mathcal{J}u(t, x) &= \lambda \int_{\mathcal{O}} u(t, y) w(y - x) dy - \lambda u(t, x) \\ \Phi(t, x) &= \lambda \int_{\partial\mathcal{O}_j} \varphi(t, y) w(y - x) dy. \end{aligned} \quad (8.27)$$

8.3.3.2 Discretization

Let

$$k = \frac{2\ell}{m+1}, \quad \iota = \left\lceil \frac{\underline{z}^-}{k} \right\rceil + 1, \quad J = \left\lceil \frac{\bar{z}^+}{k} \right\rceil + 1;$$

let $x_j = x - \underline{z}^- + jk$ for $j = 1 - \iota, \dots, m + J$; we write $w_l = w(lk)$ for any integer l .

Finite Differences in Space To approximate the differential operator \mathcal{C} , one can use the finite difference

$$\mathcal{C}^k u(t, x_j) = \left[a\delta_x^\eta + \frac{\sigma^2}{2}\delta_{x^2}^2 - r \right] u(t, x_j),$$

with

$$\begin{aligned} \delta_{x^2}^2 u(t, x_j) &= \frac{u(t, x_{j+1}) - 2u(t, x_j) + u(t, x_{j-1})}{k^2} \\ \delta_x^\eta u(t, x_j) &= \frac{u(t, x_j) - u(t, x_{j-1})}{k} + \eta \frac{u(t, x_{j+1}) - 2u(t, x_j) + u(t, x_{j-1})}{k}, \end{aligned}$$

where, for stability reasons, η is chosen so that

$$\eta = \begin{cases} \frac{1}{2} & \text{if } ka \leq \frac{\sigma^2}{2} \\ 0 & \text{if } ka > \frac{\sigma^2}{2} \text{ and } a > 0 \\ 1 & \text{if } ka > \frac{\sigma^2}{2} \text{ and } a < 0. \end{cases}$$

Jumps Approximation We approximate

$$\begin{aligned}\mathcal{J}u(t, x_j) &\approx \mathcal{J}^k u(t, x_j) = \lambda k \sum_{l; x_l \in \mathcal{O}} w_{l-j} u(t, x_l) - \lambda u(t, x_j) \\ \Phi(t, x_j) &\approx \Phi^k(t, x_j) = \lambda k \sum_{l; x_l \in \partial \mathcal{O}_j} w_{l-j} \varphi(t, x_l),\end{aligned}\tag{8.28}$$

in which the sums can be quickly computed for every j by discrete FFT as follows (see also [101] and the end of Sect. 5.5.3). The discrete correlation f of two real sequences g and h of period m is defined by

$$f_j = \sum_{l=1}^m g_{j+l} h_l, \quad 1 \leq j \leq m.$$

The discrete correlation theorem says that $Ff = (Fg)(\overline{Fh})$, where Ff, Fg and Fh are the discrete Fourier transforms⁴ of f, g and h , and \overline{Fh} denotes the complex conjugate of Fh . So, to compute f (e.g., either sum in (8.28)), we FFT the two data sets g and h , we multiply one Fourier transform by the complex conjugate of the other and we inverse transform the product. Formally, the result is a complex vector of length m . However, all its imaginary parts are equal to zero since the original data sets are both real.

Theta-Schemes in Time Given a time grid $(t_i)_{0 \leq i \leq n}$ with uniform time-step h , one can solve (8.26) by the following θ -scheme, in which $u_i^j \approx u(t_i, x_j)$, $u_i = (u_i^j)_{1 \leq j \leq m}$ and where C^k, J^k and Φ_i^k are the matrix-vector forms of \mathcal{C}^k , and \mathcal{J}^k and $\Phi^k(t, \cdot)$:

$$\begin{aligned}h^{-1}(u_{i+1} - u_i) + C^k[\theta_c u_i + (1 - \theta_c)u_{i+1}] + J^k[\theta_j u_i + (1 - \theta_j)u_{i+1}] \\ + [\theta_j \Phi_i^k + (1 - \theta_j)\Phi_{i+1}^k] = 0,\end{aligned}\tag{8.29}$$

for some constant parameters $\theta_c, \theta_j \in [0, 1]$. We give the developed j by j form of (8.29) in two specific cases. In both cases we impose a Dirichlet condition φ on $[0, T] \times \partial \mathcal{O}_j$, with u_l^l understood as φ_l^l for every $l \in \{1 - i, \dots, 0\} \cup \{m + 1, \dots, m + j\}$ in (8.30) and (8.31).

Explicit Scheme $\theta_c = \theta_j = 0$ For every $i = n - 1, \dots, 0$, for every $j = 1, \dots, m$:

$$\begin{aligned}u_i^j = h \left(\frac{\sigma^2}{2k^2} - \frac{a}{k} \eta \right) u_{i+1}^{j-1} + \left(1 - h \left(\frac{\sigma^2}{k^2} + \frac{a}{k} (1 - 2\eta) + r \right) \right) u_{i+1}^j \\ + h \left(\frac{\sigma^2}{2k^2} + \frac{a}{k} (1 - \eta) \right) u_{i+1}^{j+1} + h \lambda \left(k \sum_{l \in \{1-i, \dots, m+j\}} w_{l-j} u_{i+1}^l - u_{i+1}^j \right).\end{aligned}\tag{8.30}$$

⁴See the equation (5.71).

This scheme is computationally feasible but potentially unstable, and it only has order one of time-consistency.

Asymmetric Scheme $\theta_c = \frac{1}{2}$, $\theta_j = 0$ For $i = n-1, \dots, 0$, one must solve the linear system $Mu_i = v_{i+1}$, where M is the m -dimensional tridiagonal matrix

$$M = \begin{pmatrix} p & p_+ & 0 & \cdots & \cdots & 0 \\ p_- & p & p_+ & 0 & \ddots & 0 \\ 0 & p_- & p & p_+ & 0 & \vdots \\ \vdots & 0 & & & 0 & \\ \vdots & \ddots & 0 & p_- & p & p_+ \\ 0 & \cdots & \cdots & 0 & p_- & p \end{pmatrix} \quad \text{with } \begin{cases} p_- = -\frac{h}{2}\left(\frac{\sigma^2}{2k^2} - \frac{a}{k}\eta\right) \\ p = 1 + \frac{h}{2}\left(\frac{\sigma^2}{k^2} + \frac{a}{k}(1-2\eta) + r\right) \\ p_+ = -\frac{h}{2}\left(\frac{\sigma^2}{2k^2} + \frac{a}{k}(1-\eta)\right) \end{cases}$$

and where, for $j = 1, \dots, m$,

$$\begin{aligned} v_{i+1}^j &= -p_- u_{i+1}^{j-1} + (2-p)u_{i+1}^j - p_+ u_{i+1}^{j+1} \\ &\quad + h\lambda \left(k \sum_{l \in \{1-i, \dots, m+j\}} w_{l-j} u_{i+1}^l - u_{i+1}^j \right) \\ &\quad + \mathbb{1}_{j=1} p_- u_i^{j-1} + \mathbb{1}_{j=m} p_+ u_i^{j+1}. \end{aligned} \quad (8.31)$$

This scheme is stable and efficient but some accuracy is lost due to the asymmetric treatment of the continuous and jump parts.

8.4 Finite Differences for American Vanilla Options

By the arguments of Sect. 3.5.4 (see also Chaps. 12 and 13), the price of an American vanilla option in the Black–Scholes model at time 0 is given by

$$\sup_{\tau \in \mathcal{T}} \mathbb{E} e^{-r\tau} \psi(X_\tau) = v(0, x),$$

with $X_t = \ln(S_t)$ and $x = X_0$, where v is the unique viscosity solution to the following obstacle problem:

$$\begin{cases} v(T, \cdot) = \psi & \text{on } \mathbb{R} \\ \max(\partial_t v + \mathcal{A}v, \psi - v) = 0 & \text{on } [0, T) \times \mathbb{R}. \end{cases} \quad (8.32)$$

8.4.1 Splitting Scheme

The so-called splitting θ -scheme for the obstacle problem (8.32) is written in vector-form as follows, in terms of M , N and w_i as in (8.19): $v_n = \psi$ and, for i decreasing

from $n - 1$ to 0,

$$M\tilde{v}_i = Nv_{i+1} + hw_i, \quad v_i = \max(\psi, \tilde{v}_i). \quad (8.33)$$

By the results of Barles et al. [23], reviewed in Sect. 13.2.3, this scheme converges to the unique viscosity solution v of (8.32).

8.5 Finite Differences for Bidimensional Vanilla Options

Alternate Direction Implicit (ADI) schemes are the industry finite difference standard to cope with multivariate pricing problems. We now describe such a scheme in a bivariate Black–Scholes setting, where two underlying stock prices satisfy the following stochastic differential equations:

$$\begin{cases} dS_t^1 = S_t^1 (\kappa_1 dt + \sigma_{11} dW_t^1 + \sigma_{12} dW_t^2) \\ dS_t^2 = S_t^2 (\kappa_2 dt + \sigma_{21} dW_t^1 + \sigma_{22} dW_t^2), \end{cases}$$

for independent Brownian motions W^1 and W^2 , with

$$\begin{pmatrix} \kappa_1 \\ \kappa_2 \end{pmatrix} = \begin{pmatrix} r - q_1 \\ r - q_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{pmatrix}$$

(see Sect. 6.7.2). In order to apply the ADI method, it is better to work directly with the independent Brownian motions W^1 and W^2 . Let us introduce

$$\varphi(t, x, y) = \phi(t, S_0^1 e^{b_1 t + \sigma_1 x}, S_0^2 e^{b_2 t + \sigma_2 (\rho x + \sqrt{1-\rho^2} y)}),$$

with $(b_1, b_2) = (\kappa_1 - \frac{1}{2}\sigma_1^2, \kappa_2 - \frac{1}{2}\sigma_2^2)$. The payoff process $\phi(S_t^1, S_t^2)$ is thus rewritten as $\varphi(t, W_t^1, W_t^2)$.

The time- t price of a European option with payoff $\phi(S_T^1, S_T^2)$ at T is then given by

$$\mathbb{E}_t e^{-r(T-t)} \phi(S_T^1, S_T^2) = w(t, S_t^1, S_t^2),$$

or equally well

$$\mathbb{E} e^{-r(T-t)} \varphi(T, W_T^1, W_T^2) = u(t, W_t^1, W_t^2),$$

where w satisfies a bivariate Black–Scholes equation in the S -variables and $u = u(t, x, y)$ solves the following bivariate heat equation:

$$\begin{cases} u(T, x, y) = \varphi(T, x, y) & \text{on } \mathbb{R}^2 \\ \partial_t u(t, x, y) + \frac{1}{2} \partial_{x^2}^2 u(t, x, y) + \frac{1}{2} \partial_{y^2}^2 u(t, x, y) - ru(t, x, y) = 0 & \text{on } [0, T] \times \mathbb{R}^2. \end{cases} \quad (8.34)$$

For the numerical solution of (8.34) by finite differences:

- **localize** the domain in space to a set $\mathcal{O} = (-\ell, \ell)^2$, introducing a suitable condition at the spatial boundary of $[0, T] \times \mathcal{O}$;
- introduce a time-space mesh $(t, x, y) = (ih, j_1 k_1, j_2 k_2)$ on $[0, T] \times \mathcal{O}$, with mesh steps h, k_1, k_2 , and **discretize** the localized problem on the mesh by a suitable finite difference scheme, such as the one described in the next subsection.

Regarding the deltas, note that we have $w(t, S_1, S_2) = u(t, x, y)$, where

$$\begin{pmatrix} \ln S_1 \\ \ln S_2 \end{pmatrix} = \begin{pmatrix} \ln S_0^1 + b_1 T \\ \ln S_0^2 + b_2 T \end{pmatrix} + \Sigma \begin{pmatrix} x \\ y \end{pmatrix}.$$

Therefore

$$\begin{aligned} \begin{pmatrix} S_1 \partial_{S_1} \\ S_2 \partial_{S_2} \end{pmatrix} w(t, S_1, S_2) &= \Sigma^{-1} \begin{pmatrix} \partial_x \\ \partial_y \end{pmatrix} u(t, x, y) \\ &= \frac{1}{\sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \begin{pmatrix} \sigma_2 \sqrt{1 - \rho^2} & 0 \\ -\sigma_2 \rho & \sigma_1 \end{pmatrix} \begin{pmatrix} \partial_x \\ \partial_y \end{pmatrix} u(t, x, y). \end{aligned}$$

The time-0 deltas $\Delta_0^1 = \partial_{S_1} w(0, S_0^1, S_0^2)$ and $\Delta_0^2 = \partial_{S_2} w(0, S_0^1, S_0^2)$ are then given in terms of u as

$$\Delta_0^1 = \frac{e^{-x}}{\sigma_1} (\partial_x u)(0, 0, 0), \quad \Delta_0^2 = \frac{e^{-y}}{\sqrt{1 - \rho^2}} \left(\frac{-\rho \partial_x u}{\sigma_1} + \frac{\partial_y u}{\sigma_2} \right)(0, 0, 0). \quad (8.35)$$

Let $u_i^{j_1, j_2}$ denote the solution to the discretized pricing equation. Approximate deltas are then retrieved from (8.35) by substituting finite differences $\delta_x u$ and $\delta_y u$ for $\partial_x u$ and $\partial_y u$, e.g.

$$(\delta_x u)_i^{j_1, j_2} = \frac{u_i^{j_1+1, j_2} - u_i^{j_1-1, j_2}}{2k_1}, \quad (\delta_y u)_i^{j_1, j_2} = \frac{u_i^{j_1, j_2+1} - u_i^{j_1, j_2-1}}{2k_2}.$$

8.5.1 ADI Scheme

ADI schemes [207, 220] consist in decomposing each time step into two parts, the first implicit in x and the second implicit in y , resulting in the following approximation scheme for (8.34): $u_n = \varphi$ and, for $i = n - 1, \dots, 0$,

$$\begin{cases} \frac{2}{h}(u_{i+1} - u_{i+\frac{1}{2}}) + \frac{1}{2}\delta_{x^2}^2 u_{i+\frac{1}{2}} + \frac{1}{2}\delta_{y^2}^2 u_{i+1} - \frac{1}{2}r u_{i+\frac{1}{2}} - \frac{1}{2}r u_{i+1} = 0 \\ \frac{2}{h}(u_{i+\frac{1}{2}} - u_i) + \frac{1}{2}\delta_{x^2}^2 u_{i+\frac{1}{2}} + \frac{1}{2}\delta_{y^2}^2 u_i - \frac{1}{2}r u_{i+\frac{1}{2}} - \frac{1}{2}r u_i = 0 \end{cases}$$

or, equivalently,

$$\begin{cases} \left[\left(1 + \frac{hr}{4} \right) I - \frac{h}{4} \delta_{x^2}^2 \right] u_{i+\frac{1}{2}} = \left[\left(1 - \frac{hr}{4} \right) I + \frac{h}{4} \delta_{y^2}^2 \right] u_{i+1} \\ \left[\left(1 + \frac{hr}{4} \right) I - \frac{h}{4} \delta_{y^2}^2 \right] u_i = \left[\left(1 - \frac{hr}{4} \right) I + \frac{h}{4} \delta_{x^2}^2 \right] u_{i+\frac{1}{2}}, \end{cases} \quad (8.36)$$

in which

$$\begin{aligned} (\delta_{x^2}^2 u)_i^{j_1, j_2} &= \frac{u_i^{j_1+1, j_2} - 2u_i^{j_1, j_2} + u_i^{j_1-1, j_2}}{k_1^2}, \\ (\delta_{y^2}^2 u)_i^{j_1, j_2} &= \frac{u_i^{j_1, j_2+1} - 2u_i^{j_1, j_2} + u_i^{j_1, j_2-1}}{k_2^2}. \end{aligned}$$

Each time step i takes the form

$$\begin{cases} M^{j_2} u_{i+\frac{1}{2}}^{j_1, j_2} = N^{j_2} u_{i+1}^{j_1, j_2}, & \text{for every } j_2 \\ P^{j_1} u_i^{j_1, \cdot} = Q^{j_1} u_{i+\frac{1}{2}}^{j_1, \cdot}, & \text{for every } j_1, \end{cases} \quad (8.37)$$

for suitable “one-dimensional tridiagonal” matrices M^{j_2} , N^{j_2} , P^{j_1} , Q^{j_1} . So each time step reduces to $(m_1 + m_2)$ implicit tridiagonal one-dimensional problems, each solvable by the Thomas algorithm of paragraph “Solution by Gauss Factorization” on p. 224. This is in general a far better alternative than having to solve the $(m_1 m_2)$ -dimensional linear system that would arise from a bivariate implicit discretization.⁵

Unless simple transformations as in the above case allow elimination of the correlation from a pricing problem, additional cross-derivatives show up in the equations. These can be dealt with explicitly (i.e. put on the right-hand side in (8.36)) and the ADI scheme is still applicable, but subject to stability conditions which become stringent in multivariate settings. See Hout and Welfert [145] or Duffy [104] for alternative schemes.

8.5.1.1 American Options

The time- t price of an American vanilla option is given as

$$\underset{\tau \in \mathcal{T}_t}{\text{esssup}} \mathbb{E} e^{-r(\tau-t)} \varphi(\tau, W_\tau^1, W_\tau^2) = v(t, W_t^1, W_t^2),$$

⁵However, the sparseness of the corresponding matrix may be exploited in an iterative solution.

where v is the solution to the following obstacle problem

$$\begin{cases} v(T, x, y) = \varphi(T, x, y) & \text{on } \mathbb{R}^2 \\ \max\left(\partial_t v + \frac{1}{2}\partial_{x^2}^2 v + \frac{1}{2}\partial_{y^2}^2 v - rv, \varphi - v\right) = 0 & \text{on } [0, T) \times \mathbb{R}^2. \end{cases} \quad (8.38)$$

To solve (8.38):

- we **localize** the equation to a set $\mathcal{O} = (-\ell, \ell)^2$, introducing a suitable condition at the spatial boundary of $[0, T] \times \mathcal{O}$;
- we **discretize and solve** the localized problem on a grid.

A convergent finite difference approximation scheme is obtained by combining the previous ADI finite difference method with the splitting method of Sect. 8.4.1 (see [253]).

8.6 Finite Differences for Exotic Options

In this section we deal with finite difference methods for path-dependent options. Note that a unified perspective on the various situations that are studied below is possible in the setup of the recently developed theory of the functional Itô calculus (see Cont and Fournié [69]). See also Lipton [190] for various connections between different categories of path-dependent options.

8.6.1 Lookback Options

The payoff of a lookback option is of the form $\phi(S_T, M_T)$, where $M_t = \sup_{0 \leq s \leq t} S_s$. In the Black–Scholes model, the pair (S_t, M_t) is a Markov process so that, as explained in Sect. 3.5 (see also Chaps. 12 and 13), the price process Π_t of a lookback option can be represented as $u(t, S_t, M_t)$ for a pricing function $u = u(t, S, M)$. The (random) measure defined by the nondecreasing process $M_t = \sup_{0 \leq s \leq t} S_s$ can be shown to be singular with respect to Lebesgue measure. Process M_t is therefore not an Itô process in the sense of (3.24), therefore the process $u(t, S_t, M_t)$ is, in principle, outside the scope of the most general Itô formula (3.65) in this book. But since M_t is nondecreasing, it doesn't contribute to any bracket (see Sect. 3.2.4) and therefore the expected Itô formula is written as

$$\begin{aligned} du(t, S_t, M_t) &= \left(\partial_t u + \frac{1}{2}\sigma^2 S^2 \partial_{S^2}^2 u + \kappa S \partial_S u - ru \right)(t, S_t, M_t) dt \\ &\quad + \partial_M u(t, S_t, M_t) dM_t + \sigma S_t \partial_S u(t, S_t, M_t) dW_t. \end{aligned}$$

This can indeed be established by an application of the semimartingale Itô formula mentioned at the end of Sect. 3.2.4. From the local martingale property of the discounted price ($e^{-rt} \Pi_t$) we deduce that, for $t < T$,

$$\begin{aligned} \partial_t u + \frac{1}{2}\sigma^2 S^2 \partial_{S^2}^2 u + \kappa S \partial_S u - ru &= 0 \quad \text{on } \{S < M\} \\ \partial_M u &= 0 \quad \text{on } \{S = M\}, \end{aligned}$$

along with the terminal condition $u(T, S, M) = \phi(S, M)$. The pricing function u of a lookback option thus satisfies the Black–Scholes PDE on the subset $\{S \leq M\}$ of a bivariate state-space (S, M) . On the boundary $\{S = M\}$, u satisfies an oblique homogeneous Neumann condition $\partial_M u = 0$.

8.6.2 Barrier Options

Under certain covenants, the rebate (if triggered) of a barrier option (see Sects. 6.9.2 and 7.3.1) is delivered in arrears at the maturity time T of the option. Barrier options are then special cases of lookback options, which may be handled by finite differences as in the previous subsection. But under the most common covenant the rebate is paid at time

$$\tau = \inf\{t \in [0, T]; S_t \geq H\}.$$

We thus consider an up-and-out barrier option with trigger level H , rebate R and “vanilla component” (payoff were it not for the barrier) $\phi(S_T)$. Under the latter covenant, the effective payoff of the option is

$$\mathbb{1}_{\{\tau > T\}}\phi(S_T) + \mathbb{1}_{\{\tau \leq T\}}R =: \bar{\phi}(\bar{\tau}, S_{\bar{\tau}}), \quad (8.39)$$

paid at the random time $\bar{\tau} = \tau \wedge T$, with related price process $\Pi_t = \beta_t^{-1} \mathbb{E}_t[\beta_{\bar{\tau}} \bar{\phi}(\bar{\tau}, S_{\bar{\tau}})]$ (for $t \leq \bar{\tau}$). In the Black–Scholes model, it follows from the usual arguments (see Sect. 3.5 and Chaps. 12–13) that $\Pi_t = u(t, X_t)$, with $X_t = \ln(S_t)$, where the pricing function u satisfies with $h = \ln(H)$ and $\psi(x) = \phi(e^x)$:

$$\begin{cases} u(T, x) = \psi(x) & \text{on } \mathcal{O} \\ u(t, x) = R & \text{on } [0, T] \times \partial\mathcal{O} \\ \partial_t u + \frac{1}{2}\sigma^2 \partial_x^2 u + b\partial_x u - ru = 0 & \text{on } [0, T) \times \mathcal{O}. \end{cases} \quad (8.40)$$

More generally, denoting by $v(t, x)$ the pricing function of the “vanilla component” of a barrier option (option with payoff $\phi(S_T)$ at time T), the pricing function $u = u(t, x)$ of a barrier option satisfies

$$\begin{cases} u(T, x) = \varphi(x) & \text{on } \mathcal{O} \\ u(t, x) = R(t, x) & \text{on } [0, T] \times \partial\mathcal{O} \\ \partial_t u + \frac{1}{2}\sigma^2 \partial_x^2 u + b\partial_x u - ru = 0 & \text{on } [0, T) \times \mathcal{O}, \end{cases} \quad (8.41)$$

where, letting also $l = \ln(L)$ below:

- in the already seen case of an up-and-out barrier at the level H : $\mathcal{O} = (x - \ell, h)$, $\varphi(x) = \psi(x)$, $R(t, h) = R$;
- in the case of a down-and-out barrier at the level L : $\mathcal{O} = (l, x + \ell)$, $\varphi(x) = \psi(x)$, $R(t, l) = R$;
- in the case of double-out barriers at levels L and H : $\mathcal{O} = (l, h)$, $\varphi(x) = \psi(x)$, $R(t, l) = R(t, h) = R$;
- in the case of an up-and-in barrier at the level H : $\mathcal{O} = (x - \ell, h)$, $\varphi(x) = R$, $R(t, h) = v(t, h)$;
- in the case of a down-and-in barrier at the level L : $\mathcal{O} = (l, x + \ell)$, $\varphi(x) = R$, $R(t, l) = v(t, l)$;
- in the case of double-in barriers at levels L and H : $\mathcal{O} = (l, h)$, $\varphi(x) = R$, $R(t, l) = v(t, l)$, $R(t, h) = v(t, h)$.

The pricing equation (8.41) can be solved by finite differences.

8.6.3 Asian Options

Asian options correspond to payoff processes of the form $\phi(t, S_t, I_t)$, where $I_t = \int_0^t S_s ds$. In the Black–Scholes model the pair (S_t, I_t) is Markov with generator

$$\mathcal{A}_{S,I} = \frac{1}{2}\sigma^2 S^2 \partial_{S^2}^2 + \kappa S \partial_S + S \partial_I. \quad (8.42)$$

This generator is degenerate (hyperbolic, without diffusion term) in the I -variable, which results in PDEs “in one-and-a-half space dimension” below.

8.6.3.1 Asian Fixed Strike Put Option

This is the option with payoff

$$\xi = \left(K - \frac{I_T}{T} \right)^+ = \phi(I_T) \quad (8.43)$$

and related price process on $[0, T]$

$$\Pi_t = \beta_t^{-1} \mathbb{E}_t[\beta_T \xi] = u(t, S_t, I_t). \quad (8.44)$$

The pricing equation is written:

$$\begin{cases} u(T, S, I) = \phi(I) \\ \partial_t u + \mathcal{A}_{S,I} u = ru, \quad 0 \leq t < T. \end{cases} \quad (8.45)$$

See [261] for the numerical issues related to the degeneracy of the equation in the I -variable.

Alternatively (see Rogers and Shi [236]), one can observe that $\frac{\phi(I_T)}{S_T} = \eta_T^+$ for the process η_t such that

$$\eta_t = \frac{1}{S_t} \left(K - \frac{I_t}{T} \right), \quad d\eta_t = -\frac{dt}{T} - \eta_t (\kappa - \sigma^2) dt - \eta_t \sigma dW_t,$$

which is Markov with generator

$$\mathcal{A}_\eta = -\left[\frac{1}{T} + (\kappa - \sigma^2)\eta \right] \partial_\eta + \frac{1}{2} \eta^2 \sigma^2 \partial_{\eta^2}. \quad (8.46)$$

Now, for the measure $\tilde{\mathbb{P}}$ associated with the numéraire S , so that $\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \frac{S_T}{S_0 e^{\kappa T}}$ (see Sect. 4.3.1.1), the price process (8.44) can be represented as⁶

$$\Pi_t = e^{qt} S_t \tilde{\mathbb{E}}_t [e^{-qT} S_T^{-1} \phi(I_T)] = e^{-q(T-t)} S_t \tilde{\mathbb{E}}_t \eta_T^+ = S_t v(t, \eta_t) \quad (8.47)$$

(one can check that η_t is a Markov process under $\tilde{\mathbb{P}}$, which justifies the last equality for a pricing function $v = v(t, \eta)$). From the results stated in Sect. 4.3.1.1, a process Θ_t is a $\tilde{\mathbb{P}}$ -local martingale if and only if $(e^{-\kappa t} S_t \Theta_t)$ is a \mathbb{P} -local martingale. Since $e^{-q t} v(t, \eta_t)$ is a $\tilde{\mathbb{P}}$ -local martingale, thus $(e^{-r t} S_t v(t, \eta_t))$ is a \mathbb{P} -local martingale. But Itô calculus yields, with “ \doteq ” standing for “equality up to a \mathbb{P} -local martingale”:

$$\begin{aligned} & e^{rt} d(e^{-rt} S_t v(t, \eta_t)) \\ & \doteq (-r S_t v(t, \eta_t) + v(t, \eta_t) \kappa S_t dt + S_t (\partial_t v + \mathcal{A}_\eta v)(t, \eta_t) - \sigma^2 S_t \eta \partial_\eta v(t, \eta_t)) dt \\ & = S_t \left(\partial_t v - \left(\frac{1}{T} + \kappa \eta \right) \partial_\eta v + \frac{1}{2} \sigma^2 \eta^2 \partial_{\eta^2}^2 v - qv \right) dt. \end{aligned}$$

Therefore, via Proposition 3.1.12:

Proposition 8.6.1 *We have $\Pi_t = S_t v(t, \eta_t)$, where the pricing function $v = v(t, \eta)$ in the numéraire S solves the following one-dimensional PDE:*

$$\begin{cases} v(T, \eta) = \eta^+ \\ \partial_t v - \left(\frac{1}{T} + \kappa \eta \right) \partial_\eta v + \frac{1}{2} \sigma^2 \eta^2 \partial_{\eta^2}^2 v - qv = 0. \end{cases} \quad (8.48)$$

This equation can be solved numerically by finite differences. For η close to 0 the advection term is dominant in this equation (due to the first-order coefficient $\frac{1}{T}$); this poses specific numerical issues which are dealt with in [102].

8.6.3.2 Hawaiian Fixed Strike Put Option

This is the American counterpart of the previous option, with the following payoff process paid at a stopping time τ at the holder's convenience:

⁶Adequately adjusting the pricing formula (4.62) for the dividends of S ; see the related comment in Sect. 4.3.1.

$$\left(K - \frac{I_\tau}{\tau} \right)^+ = \phi(\tau, I_\tau). \quad (8.49)$$

By the results stated in Sect. 3.5.4 (see also Chaps. 12 and 13), the related price process is written, for $t \in [0, T]$, as

$$\Pi_t = \beta_t^{-1} \operatorname{esssup}_{\tau \in \mathcal{T}_t} \mathbb{E}_t [\beta_\tau \phi(\tau, I_\tau)] = v(t, S_t, I_t), \quad (8.50)$$

where the pricing function $v = v(t, S, I)$ solves the following obstacle problem:

$$\begin{cases} v(T, S, I) = \phi(T, I), & S, I > 0 \\ \max((\partial_t v + \mathcal{A}_{S,I} v - rv)(t, S, I), \phi(t, I) - v(t, S, I)) = 0, & t \in (0, T), S, I > 0, \end{cases} \quad (8.51)$$

in which $\mathcal{A}_{S,I}$ was defined in (8.42). The obstacle $\phi(t, I)$ is singular at time $t = 0$, so that the second line of (8.51) only holds for $t > 0$. The proof of the following result can be found in Crépey [76].

Proposition 8.6.2 *We have $\Pi_t = v(t, S_t, I_t)$, $t \in (0, T]$, where v is the unique bounded viscosity solution to (8.51). The price of the option at time 0 is given by the radial limit*

$$\Pi_0 = \lim_{t \rightarrow 0^+} v(t, S_0, S_0 t). \quad (8.52)$$

Figure 8.2 provides a numerical illustration of the radial convergence in (8.52). Numerically, the convergence also occurs along radii αt other than $S_0 t$, yet at a slower rate than for $\alpha = S_0$. As in the European case, special numerical care is required to deal with the degeneracy of $\mathcal{A}_{S,I}$ in I [261].

8.6.4 Discretely Path Dependent Options

Discretely path-dependent options correspond to payoffs of the form $\xi = \phi(S_{t_0}, S_{t_1}, \dots, S_{t_n})$ for a discrete set of monitoring times ($t_0 = 0, \dots, t_n = T$). This includes important classes of assets, such as cliquet options, volatility and variance swaps and discretely sampled Asian options [97, 254, 256, 258]. The pricing of such products by Monte Carlo is obvious in the Black–Scholes model, but it becomes tricky in a nonlinear extension of Black–Scholes known as the Uncertain Volatility Model [13], where the volatility process is only known to remain in a range $[\underline{\sigma}, \bar{\sigma}]$. The PDE approach of this subsection, on the contrary, can easily be adapted to this setup [255, 257].

From the theoretical point of view, discrete path-dependence is dealt with in Sect. 14.1.2 (see Remark 14.1.5), to which readers are referred for all the justifications regarding the statements below.

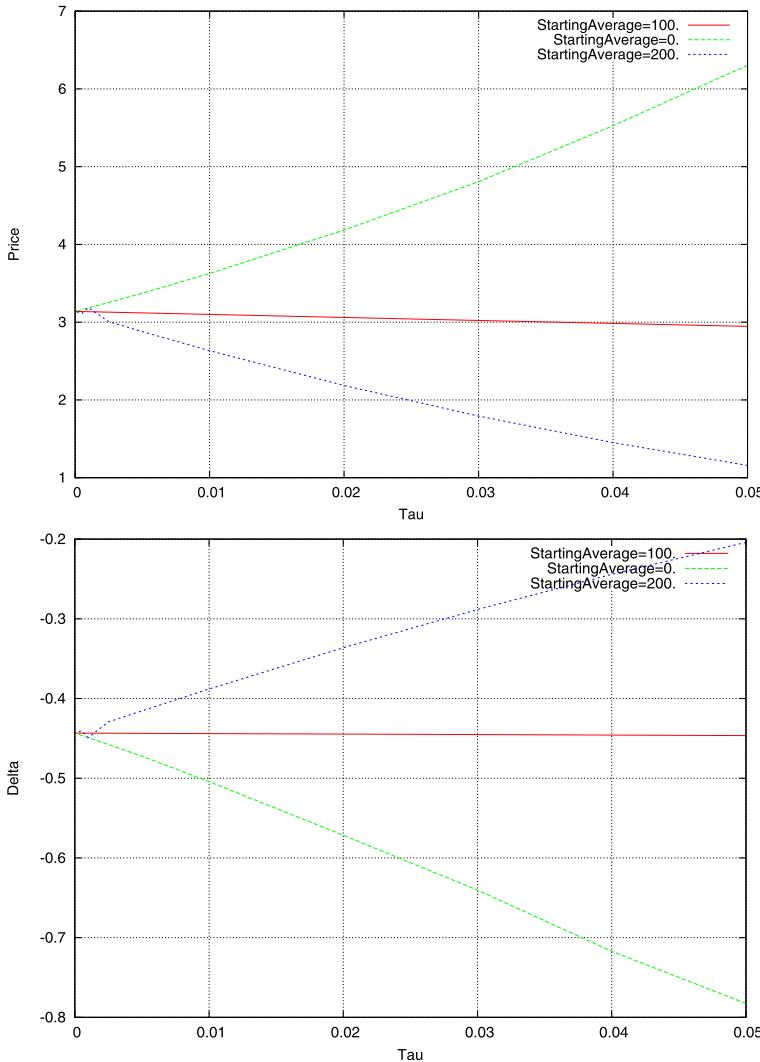


Fig. 8.2 Convergence of the price and delta of a Hawaiian put at $(t, S_0, \alpha t)$ as $t \rightarrow 0$, for $\alpha = 0$, S_0 and $2S_0$

8.6.4.1 Cliquet Options

Let $R_i = \frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}}$ denote the return of the underlying asset on the period $[t_{i-1}, t_i]$, for $i = 1, \dots, n$. The payoff of a cliquet option is defined by

$$\xi = \max \left(\min \left(\sum_{i=1}^n \max(\min(R_i, C), F), \mathcal{C} \right), \mathcal{F} \right)$$

for given thresholds $F < C$ and $\mathcal{F} < \mathcal{C}$. We introduce two auxiliary processes P and Z such that, on every time interval $[t_i, t_{i+1})$,

$$P_t = S_{t_i}, \quad Z_t = \frac{1}{i} \sum_{l=1}^i \max(F, \min(C, R_l)). \quad (8.53)$$

In the Black–Scholes model for S , the triple (S_t, P_t, Z_t) is Markov with generator

$$\mathcal{A}_{S,P,Z}\varphi(S, P, Z) = \mathcal{A}_S^{bs}\varphi(S, P, Z) = \frac{1}{2}\sigma^2 S^2 \partial_{S^2}^2 \varphi + \kappa S \partial_S \varphi.$$

Moreover, we have that

$$\xi = \max(\min(nZ_T, \mathcal{C}), \mathcal{F}) = \phi(Z_T).$$

Following the approach of Sect. 14.1.2 (see Remark 14.1.5), one can then show:

Proposition 8.6.3 *The cliquet option Black–Scholes price process $\Pi_t = e^{-r(T-t)} \mathbb{E}_t \xi$ can be represented as $\Pi_t = u_i(t, S_t, Z_t)$ on every time interval $[t_i, t_{i+1})$, for a sequence of continuous functions $u_i = u_i(t, S, P, Z)$ on $[t_i, t_{i+1}] \times (0, \infty)^3$ such that $(u_i)_{0 \leq i \leq n}$ is the unique sequence of viscosity solutions to the following PDE cascade:*

$u_n = \phi$ on $\{T\} \times (0, \infty)^3$ and for i decreasing from $n - 1$ to 0:

$$\begin{cases} u_i(t_{i+1}, S, P, Z) = u_{i+1}(t_{i+1}, S, P_+, Z_+) & \text{on } (0, \infty)^3 \\ \partial_t u_i + \frac{1}{2}\sigma^2 S^2 \partial_{S^2}^2 u_i + \kappa S \partial_S u_i - ru_i = 0 & \text{on } [t_i, t_{i+1}] \times (0, +\infty)^3, \end{cases} \quad (8.54)$$

where in the jump condition in the first line,

$$P_+ = S, \quad Z_+ = \frac{i}{(i+1)}Z + \frac{\rho}{(i+1)}, \quad (8.55)$$

in which $\rho = \max(\min(\frac{S-P}{P}, C), F)$.

The jump condition stems from the continuity of the Brownian martingale $(e^{-rt} \Pi_t)$.

To solve (8.54) numerically, we localize the domain in space to

$$\mathcal{O} = (\underline{S}, \overline{S}) \times (\underline{S}, \overline{S}) \times (\underline{Z}, \overline{Z}),$$

with

$$\underline{S} = 0, \quad \overline{S} = S_0(1 + f\sigma\sqrt{T}), \quad \underline{Z} = F, \quad \overline{Z} = C,$$

for a suitable factor f , e.g. $f = 4$. We then discretize the state space. Given an adaptive P -mesh $(P^{j_1})^{1 \leq j_1 \leq m_1}$ concentrated around S_0 , one can use the following

(P, S) -grid, concentrated around S_0 and around the diagonal $S = P$ (see Windcliff et al. [257]):

$$(P, S)^{j_1, j_2} = \left(P^{j_1}, \frac{P^{j_1} P^{j_2}}{S_0} \right), \quad 1 \leq j_1 \leq m_1, \quad 1 \leq j_2 \leq m_2.$$

We then consider the “product” of this (P, S) -grid with a uniform grid in the Z -variable. Between monitoring times, (8.54) can then be solved by a finite difference θ -scheme on the grid (standard Black–Scholes θ -scheme in the S -variable operating in the three dimensional state space (S, P, Z)). Some kind of interpolation is required for implementing the jump condition (first line of (8.54)).

8.6.4.2 Volatility and Variance Swaps

We use the same approach as for cliquet options, except that R_i and Z are now defined by $R_i = \ln(\frac{S_{t_i}}{S_{t_{i-1}}})$ and, for every $t_i \leq t < t_{i+1}$,

$$Z_t = \frac{1}{i} \sum_{l=1}^i R_l^2. \quad (8.56)$$

Variance and volatility swaps correspond to the payoffs $\xi = V^2 - K^2$ and $V - K$, in which $V^2 = \sum_{i=1}^n \ln(\frac{S_{t_i}}{S_{t_{i-1}}})^2$ is the realized variance of S and in K is a constant.

In both cases the payoff ξ is of the form $\phi(Z_T)$.

We proceed as for cliquet options, obtaining in this case (in Black–Scholes): $\Pi_t = u_i(t, S_t, P_t, Z_t)$ on every time interval $[t_i, t_{i+1}]$, where the sequence of continuous functions $u_i = u_i(t, S, P, Z)$ on $[t_i, t_{i+1}] \times (0, +\infty)^2$ satisfies (8.54)–(8.55), but for ϕ and Z suitably redefined for volatility and variance swaps, and where ρ in the jump condition (8.55) is now meant as $\ln(\frac{S}{P})^2$.

8.6.4.3 Discretely Monitored Asian Options

Finally, we consider discretely sampled Asian options with payoffs of the form

$$\xi = \left(K - \frac{T}{n} \sum_{l=1}^n S_{t_l} \right)^+.$$

Let

$$Y_t = \frac{1}{i} \sum_{k=1}^i S_{t_k}$$

on $t_i \leq t < t_{i+1}$. The pair (S_t, Y_t) is Markov in the Black–Scholes model for S , with generator

$$\mathcal{A}_{S,Y}\varphi(S, Y) = \mathcal{A}_S^{bs}\varphi(S, Y),$$

on every time interval (t_i, t_{i+1}) . Moreover, we have that $\xi = (K - TY_T)^+ = \phi(Y_T)$. We proceed as before, obtaining in this case (in Black–Scholes):

$\Pi_t = u_i(t, S_t, Y_t)$ on every time interval $[t_i, t_{i+1}]$, for a sequence of continuous functions u_i on $[t_i, t_{i+1}] \times (0, +\infty)^2$ such that $u_n = \phi$ and, for i decreasing from $n - 1$ to 0:

$$\begin{cases} u_i(t_{i+1}, S, Y) = u_{i+1}(t_{i+1}, S, Y_+) & \text{on } (0, +\infty)^2 \\ \partial_t u_i + \frac{1}{2} \sigma^2 S^2 \partial_{S^2}^2 u_i + \kappa S \partial_S u_i - r u_i = 0 & \text{on } [t_i, t_{i+1}] \times (0, +\infty)^2, \end{cases}$$

with

$$Y_+ = \frac{i}{(i+1)} Y + \frac{S}{(i+1)}$$

in the first line.

Chapter 9

Calibration Methods

An important financial engineering issue is model calibration. Calibrating a model means finding numerical values of its parameters such that the prices of market instruments computed within the model, at a given time, coincide with their market prices. The simplest example of a calibration problem was encountered in Sect. 5.4, when we discussed the notions of the implied volatility of an option and the implied correlation of a CDO tranche. In these cases the calibration problem is easy since there is only one parameter to calibrate to only one market quote.

Calibration thus corresponds to estimation of a model. However, in finance the term “estimation” specifically refers to statistical estimation, i.e. estimation based on historical data by maximum likelihood or any other statistical procedure. Statistical estimation is thus backward looking, whereas calibration is forward looking, since derivative prices at the current time are based on the views of the market regarding the future dynamics of the underlyings. It is generally acknowledged that, whenever option data are available, it is better to use them to calibrate the model than to estimate a model statistically on past data.

9.1 The Ill-Posed Inverse Calibration Problem

The calibration problem is the inverse of the pricing problem: instead of computing prices in a model for given values of its parameters, we compute values of model parameters consistent with observed prices. It is well-known to physicists that inverse problems are ill-posed, where a problem is said to be well-posed if its solution exists, is unique and depends continuously on its input data. A problem is ill-posed for any of the following reasons:

- it admits no solution,
- it admits multiple solutions,
- it has a solution that doesn’t depend continuously on the input data.

Except for trivial situations, there exists no model achieving a perfect fit with a full calibration data set, including a zero-coupon curve, expected dividend yield

curves on the underlyings and a number of vanilla option prices. But there are typically many models that fit the data within the bid-ask spread (see e.g. Lipton [192]). Then, if one perturbs the data (e.g., if the observed prices move from some small amount between today and tomorrow), a numerical solution to the calibration problem tends to switch from one locally best fit solution to another, resulting in numerical instability of the calibrated parameters. In order to get a well-posed problem, we need to introduce some regularization.

9.1.1 Tikhonov Regularization of Nonlinear Inverse Problems

The most widely known and applicable regularization method is due to Tikhonov and Phillips [225, 249]. The monograph [116] is the general reference for this subsection.

Let there be given a closed convex nonvoid subset \mathcal{C} of a Hilbert space \mathcal{H} , a direct operator

$$\mathcal{H} \supseteq \mathcal{C} \ni \varrho \xrightarrow{\Pi} \Pi(\varrho) \in \mathbb{R}^d,$$

noisy data π^δ , and a prior $\varrho_* \in \mathcal{H}$. In the financial interpretation, Π , ϱ , π^δ and ϱ_* correspond respectively to a pricing functional at the current time (time at which the calibration is performed, say $t = 0$), a set of model parameters, current market prices known up to the bid-ask spread δ and an a priori guess for the set of model parameters. The Tikhonov regularization method for inverting Π at π^δ , or for calibrating the model parameter ϱ given the observation π^δ , consists of:

- reformulating the inverse problem as the following nonlinear least squares problem:

$$\min_{\varrho \in \mathcal{C}} \|\Pi(\varrho) - \pi^\delta\|^2 \quad (9.1)$$

- to ensure existence of a solution,
- selecting the solution to this nonlinear least squares problem that minimizes $\|\varrho - \varrho_*\|$, to grant uniqueness, and
 - introducing a trade-off between accuracy and regularity, parameterized by a level of regularization $\alpha > 0$, to ensure stability.

More precisely, let us introduce the following cost criterion:

$$J_\alpha^\delta(\varrho) \equiv \|\Pi(\varrho) - \pi^\delta\|^2 + \alpha \|\varrho - \varrho_*\|_{\mathcal{H}}^2. \quad (9.2)$$

Definition 9.1.1 Given α , δ and an additional parameter η , which represents an error tolerance on the minimization, a regularized solution to the inverse problem for Π at π^δ means any model parameter $\varrho_\alpha^{\delta, \eta} \in \mathcal{C}$ such that

$$J_\alpha^\delta(\varrho_\alpha^{\delta, \eta}) \leq J_\alpha^\delta(\varrho) + \eta, \quad \varrho \in \mathcal{C}.$$

We will now see that under suitable assumptions, the regularized inverse problem is well-posed. We first postulate the following continuity assumption on the direct operator Π .

Assumption 9.1.2 (Compactness) $\Pi(\varrho_n)$ converges to $\Pi(\varrho)$ in \mathbb{R}^d if ϱ_n converges to ϱ weakly in \mathcal{H} .

We then have the following stability result.

Proposition 9.1.3 (Stability) *Let $\pi^{\delta_n} \rightarrow \pi^\delta$, $\eta_n \rightarrow 0$ as $n \rightarrow \infty$. Then any sequence of regularized solutions $\varrho_{\alpha_n}^{\delta_n, \eta_n}$ admits a subsequence which converges on \mathcal{H} toward a regularized solution $\varrho_\alpha^{\delta, \eta=0}$ as $n \rightarrow \infty$.*

An interesting feature of Tikhonov regularization is that the calibration input data do not need to belong to the range of the direct operator for the method to be applicable. However, assuming that the data lie in the range of the model leads to convergence properties of regularized solutions to the inverse problem as $\alpha \rightarrow 0$. We thus make the following additional

Assumption 9.1.4 (Range property) $\pi \in \Pi(\mathcal{C})$.

By a ϱ_* -solution of the inverse problem for Π at π , we mean any ϱ in the set $\text{Argmin}_{\{\Pi(\varrho)=\pi\}} \varrho - \varrho_* \mathcal{H}$ (which is nonempty by Assumption 9.1.4). For a proof of the following result, see Theorem 2.3 in Engl et al. [117].

Proposition 9.1.5 (Convergence) *Let the perturbed parameters $\alpha_n, \delta_n, \eta_n$ and the perturbed data $\pi_n \in \mathbb{R}^d$ satisfy*

$$(n \in \mathbb{N}) \quad \|\pi - \pi_n\| \leq \delta_n$$

$$(n \rightarrow \infty) \quad \alpha_n, \delta_n^2/\alpha_n, \eta_n/\alpha_n \longrightarrow 0.$$

Then any sequence of regularized solutions $\varrho_{\alpha_n}^{\delta_n, \eta_n}$ admits a subsequence which converges toward a ϱ_ -solution ϱ of the inverse problem for Π at π as $n \rightarrow \infty$. In particular, in case a ϱ_* -solution ϱ is unique, then $\varrho_{\alpha_n}^{\delta_n, \eta_n}$ converges to ϱ as $n \rightarrow \infty$.*

Assuming further regularity of Π , we can get convergence rates uniform over all data $\pi \in \Pi(\mathcal{C})$ “sufficiently close” to the prior ϱ_* , in the sense of the additional condition (9.3) below. We thus make the following additional regularity assumption on Π .

Assumption 9.1.6 (Twice Gateaux differentiability) There exist linear and bilinear forms $d\Pi(\varrho)$ on \mathcal{H} and $d^2\Pi(\varrho)$ on $\mathcal{H} \times \mathcal{H}$, such that

$$\Pi(\varrho + \epsilon h) = \Pi(\varrho) + \epsilon d\Pi(\varrho) \cdot h + \frac{\epsilon^2}{2} d^2\Pi(\varrho)(h, h) + o(\epsilon^2)$$

for every $\varrho, \varrho + h \in \mathcal{C}$,

$$\|d\Pi(\varrho) \cdot h\| \leq C\|h\|, \quad \|d^2\Pi(\varrho) \cdot (h, h')\| \leq C\|h\|_{\mathcal{H}}\|h'\|_{\mathcal{H}}$$

for every $\varrho \in \mathcal{C}$ and $h, h' \in \mathcal{H}$,

where the C constant is uniform in $\varrho \in \mathcal{C}$.

In the next statement, the operator

$$\mathbb{R}^d \ni \lambda \xrightarrow{d\Pi(\varrho)^*} d\Pi(\varrho)^* \cdot \lambda \in \mathcal{H},$$

denotes the adjoint of

$$\mathcal{H} \ni h \xrightarrow{d\Pi(\varrho)} d\Pi(\varrho) \cdot h \in \mathbb{R}^d,$$

in the sense that $\langle h, d\Pi(\varrho)^* \cdot \lambda \rangle_{\mathcal{H}} = \lambda^T d\Pi(\varrho) \cdot h$ for every $(\lambda, h) \in \mathbb{R}^d \times \mathcal{H}$.

Proposition 9.1.7 (Convergence Rates; see Theorem 10.4 of Engl et al. [116]) *Assume*

$$(n \in \mathbb{N}) \quad \|\pi - \pi_n\| \leq \delta_n$$

$$(n \rightarrow \infty) \quad \alpha_n \longrightarrow 0, \quad \alpha_n \sim \delta_n, \quad \eta_n = O(\delta_n^2).$$

Then $\varrho_{\alpha_n}^{\delta_n, \eta_n} - \varrho = O(\sqrt{\delta_n})$ for every ϱ_ -solution ϱ of the inverse problem for Π at π such that*

$$\varrho - \varrho_* = d\Pi(\varrho)^* \cdot \lambda \tag{9.3}$$

for λ sufficiently small in \mathbb{R}^d . In particular, there exists at most one such ϱ_ -solution ϱ .*

An important practical issue is the choice of the regularization parameter α that determines the trade-off between accuracy and regularity in the method. To set α , the main approaches are:

- a priori methods, in which the choice of α only depends on the data noise δ (size of the bid-ask spread),
- more general a posteriori methods, in which α may depend on the data in a less specific way.

In financial applications, the most commonly used method for choosing α is an a posteriori method based on the so-called discrepancy principle, which computes iteratively the greatest α for which $\|\Pi(\varrho_{\alpha}^{\delta, \eta}) - \pi^{\delta}\|$ doesn't exceed the data noise δ (for given δ, η).

9.1.2 Calibration by Nonlinear Optimization

Regularization is used in practice for calibrating nonparametric models. In the case of parametric pricing models, i.e. models with a small number of scalar parameters such as the Heston or the Merton model, the need for regularization is less stringent and the choice of a regularization term is not obvious; the industry standard is then to solve the unregularized nonlinear least squares problem (9.1).

In any case, the calibration of a model is effectively reduced to a nonlinear least squares problem of the form (9.2), with $\alpha = 0$ if no regularization is used. When it comes to implementation, the minimization problem (9.2) is discretized, which results in a nonlinear minimization problem on (some subset of) \mathbb{R}^k , where k is the number of model parameters to be estimated.

In the case of a convex and differentiable cost criterion J in (9.2), various gradient descent algorithms are proven to be convergent to the unique minimum of J . Gradient descent algorithms consist of moving at each step by some amount in a direction defined by the gradient ∇J at the current step of the algorithm, in combination with, in conjugate gradient or quasi-Newton algorithms, the gradient(s) ∇J at the previous step(s).

In the context of calibration problems in finance, J is typically not convex in ϱ . Sometimes too, as in the American calibration problem mentioned at the end of Sect. 9.2.2.1, J is only almost everywhere differentiable. In such cases gradient descent algorithms can still be used, but they typically converge to one among many local minima of J .

When the gradient ∇J doesn't exist, or is not computable explicitly or numerically with the required accuracy, an alternative to gradient descent methods is the nonlinear simplex method.¹ As opposed to gradient descent methods, the nonlinear simplex algorithm uses only the values (not the gradient) of J .

9.2 Extracting the Effective Volatility

In this section, we consider the problem of inferring a local volatility function $\sigma(t, S)$ from observed vanilla option prices. The effective volatility function thus inferred may then be used for various purposes, such as pricing exotic options and/or Greeking consistent with the market, or calibrating more general stochastic volatility models (see Gatheral [131]).

¹Not to be confused with Danzig's simplex linear programming algorithm.

9.2.1 Dupire Formula

Local volatility models [98, 105] are the straightforward generalization of the Black–Scholes model in which the volatility parameter is no longer constant, but is given by a function $\sigma = \sigma(t, S)$. So, under the unique risk-neutral probability measure \mathbb{P}

$$dS_t = S_t(\kappa dt + \sigma(t, S_t) dW_t), \quad (9.4)$$

where κ denotes, as in Black–Scholes, the difference $r - q$ between a risk-free funding rate and a dividend yield q on S (both assumed constant or time-deterministic). The corresponding European vanilla call option price process is given, for any positive maturity T and strike K and for $t \leq T$, by

$$\Pi_t(T, K) = e^{-r(T-t)} \mathbb{E}_t(S_T - K)^+ = u(t, S_t, T, K; \sigma(\cdot, \cdot)). \quad (9.5)$$

Here u is the pricing function of the state variables t and S , parameterized by the local volatility function $\sigma(\cdot, \cdot)$ and the data (T, K) of the call. This pricing function can be characterized as usual by the related pricing equation in the state variables t and S (this is part (i) in the next result). But more interestingly, in a local volatility model and for the vanilla call (or put) payoff, a “dual” equation (part (ii) below) holds for u in the variables T and K . Thus:

Proposition 9.2.1

- (i) *For fixed (T, K) , the function $(t, S) \mapsto u(t, S, T, K; \sigma(\cdot, \cdot))$ is the unique solution to the following pricing equation:*

$$\begin{cases} u(T, S) = (S - K)^+, & S > 0 \\ \left(\partial_t u + \kappa S \partial_S u + \frac{1}{2} \sigma(t, S)^2 S^2 \partial_{S^2}^2 u - ru \right)(t, S) = 0, & t < T, S > 0. \end{cases} \quad (9.6)$$

- (ii) *For fixed (t, S) , the function $(T, K) \mapsto u(t, S, T, K; \sigma(\cdot, \cdot))$ is the unique solution to the so-called Dupire equation:*

$$\begin{cases} u(t, K) = (S - K)^+, & K > 0 \\ \left(\partial_T u + \kappa K \partial_K u - \frac{1}{2} \sigma(T, K)^2 K^2 \partial_{K^2}^2 u + qu \right)(T, K) = 0, & T > t, K > 0. \end{cases} \quad (9.7)$$

Proof (i) The pricing equation (9.6) follows from the standard martingale and Markov arguments of Sect. 3.5.1.

(ii) We have the following Breeden–Litzenberger relation [55] between European vanilla prices and the density of transition probability $\gamma = \gamma(t, S, T, K; \sigma(\cdot, \cdot))$ of S_t between (t, S) and (T, K) in the dynamics (9.4):

$$\begin{aligned}\gamma(t, S, T, K; \sigma(\cdot, \cdot)) &= \partial_K \mathbb{P}(S_T \leq K \mid S_t = S) \\ &= \partial_K \mathbb{E}(\mathbb{1}_{S_T \leq K} \mid S_t = S) = -\partial_{K^2}^2 \mathbb{E}((S_T - K)^+ \mid S_t = S) \\ &= -e^{r(T-t)} \partial_{K^2}^2 u(t, S, T, K; \sigma(\cdot, \cdot)).\end{aligned}\quad (9.8)$$

Besides, we have the following Fokker-Planck equation (see Remark 9.2.2) in the forward variables (T, K) for γ , at fixed (t, S) :

$$\begin{cases} \gamma(t, K) = \delta_S(dK), & K > 0 \\ \left(\partial_T \gamma + \partial_K (\kappa K \gamma) - \partial_{K^2}^2 \left(\frac{1}{2} \sigma(T, K)^2 K^2 \gamma \right) \right)(T, K) = 0, & T > t, K > 0, \end{cases} \quad (9.9)$$

where δ_S in the first line denotes a Dirac measure at S . Substituting $-e^{r(T-t)} \partial_{K^2}^2 u$ for γ in the second line yields that

$$\begin{aligned}0 &= \partial_T \partial_{K^2}^2 u + r \partial_{K^2}^2 u + \kappa (\partial_{K^2}^2 u + K \partial_{K^3}^3 u) - \partial_{K^2}^2 \left(\frac{1}{2} \sigma(T, K)^2 K^2 \partial_{K^2}^2 u \right) \\ &= \partial_{K^2}^2 \left(\partial_T u + ru + \kappa(u + K \partial_K u - 2u) - \frac{1}{2} \sigma(T, K)^2 K^2 \partial_{K^2}^2 u \right) \\ &= \partial_{K^2}^2 \left(\partial_T u + \kappa K \partial_K u - \frac{1}{2} \sigma(T, K)^2 K^2 \partial_{K^2}^2 u + qu \right),\end{aligned}\quad (9.10)$$

where the identity

$$K \partial_{K^3}^3 u = \partial_{K^2}^2 (K \partial_K u - 2u)$$

was used in the second equality. The Dupire equation (9.7) follows by double integration in K in the last line of (9.10). \square

The Dupire equation (9.7) is thus nothing but the Fokker-Planck equation integrated twice with respect to K ; see e.g. [77] for an alternative, probabilistic proof of the Dupire equation (based on local time).

Remark 9.2.2 The Fokker-Planck equation (9.9) can be regarded as an analytic weak form of the following integrated Itô formula, for every test-function φ :

$$\mathbb{E}(\varphi(S_T) \mid S_t = S) = \int_t^T \mathbb{E}(\mathcal{A}\varphi(S_s) \mid S_t = S) ds \quad (9.11)$$

where the generator \mathcal{A} of S_t in (9.4) appears as

$$\mathcal{A}\varphi(S) = \kappa S \partial_S \varphi(S) + \frac{1}{2} \sigma^2 S^2 \partial_{S^2}^2 \varphi(S).$$

Indeed (9.11) can be rewritten

$$\int_{K>0} \varphi(K) \gamma(t, S, T, K) dK = \int_t^T \int_{K>0} \mathcal{A}\varphi(K) \gamma(t, S, s, K) dK ds,$$

or

$$\int_{K>0} \varphi(K) \gamma(t, S, T, K) dK = \int_t^T \int_{K>0} \varphi(K) \mathcal{A}^* \gamma(t, S, s, K) dK ds, \quad (9.12)$$

with

$$\mathcal{A}^* \gamma(t, S, s, K) = -\partial_K (\kappa K \gamma(t, S, s, K)) + \partial_{K^2}^2 \left(\frac{1}{2} \sigma^2 K^2 \gamma(t, S, s, K) \right).$$

Equation (9.12), which is valid for every regular test-function φ , is the weak form of the second line of (9.9). The initial motivation of Itô [151] for developing his calculus was precisely to provide some direct probabilistic insights into the analysis of Markov transition semigroup equations such as (9.9).

The above pricing and Dupire equations can be used to compute local volatility option prices and Greeks numerically. Uniqueness for a solution to the Dupire equation also implies the following corollary, provided the right-hand side in (9.13) defines a positive function; yet this essentially follows from no-arbitrage arguments (up to the regularity issue regarding $(T, K) \mapsto \Pi_0^{ma}(T, K)$) that qualify the nature of a “reasonable” market in the statement.

Corollary 9.2.3 *In any “reasonable” market of vanilla options and at any time t , say $t = 0$ below, there exists a unique local volatility function calibrated to the market at that time, the so-called effective volatility function $\sigma_0(\cdot, \cdot)$. Moreover the function $\sigma_0(\cdot, \cdot)$ is given by the following explicit Dupire formula: for every positive T and K ,*

$$\frac{\sigma_0(T, K)^2}{2} = \frac{\partial_T \Pi_0^{ma}(T, K) + \kappa K \partial_K \Pi_0^{ma}(T, K) + q \Pi_0^{ma}(T, K)}{K^2 \partial_{K^2}^2 \Pi_0^{ma}(T, K)}, \quad (9.13)$$

where Π_0^{ma} stands for “market price at time 0”.

9.2.2 The Local Volatility Calibration Problem

To test the different calibration methods reviewed in this subsection, we use a DAX index options real data set consisting of about 300 European vanilla option prices distributed throughout 6 maturities with moneyness $K/S_0 \in [0.8, 1.2]$, corresponding to the implied volatility surface² displayed in Fig. 9.1.

The local volatility calibration problem of extracting effective volatility from market prices is an ill-posed inverse problem. A naive approach, based on numerical differentiation, using the above Dupire’s formula, gives a local volatility surface that

²See Sect. 5.4.1.1.

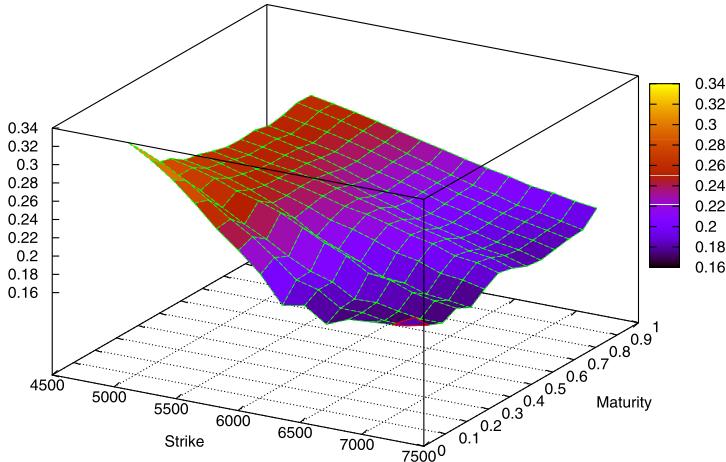


Fig. 9.1 Implied volatility surface corresponding to the DAX index options real data set that is used in the calibrations of Sect. 9.2.2

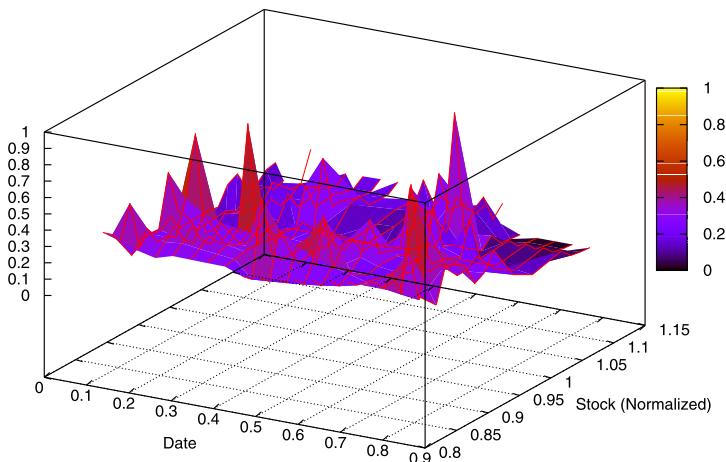


Fig. 9.2 Local volatility surface obtained by the Dupire formula using the implied volatility surface of Fig. 9.1 as input data

is both very irregular at a fixed calibration time t (see Fig. 9.2) and unstable in t . Since market prices are only available for a finite set of strikes and maturities, the calibration problem is also under-determined.

To meet these ill-posedness and under-determination issues, we might want to seek σ within a parameterized family of functions. However there is no parsimonious class of functions flexible enough to fit an arbitrary implied volatility surface

of market quotes. Of course it is always possible, as in [68], to resort to large families of splines, but then the ill-posedness of the calibration problem shows up once again.

9.2.2.1 Approach by Tikhonov Regularization

As in (9.2), we can rewrite the local volatility calibration problem in the form of the following nonlinear minimization problem:

$$\min_{\{\sigma \equiv \sigma(\cdot, \cdot); \underline{\sigma} \leq \sigma \leq \bar{\sigma}\}} J(\sigma) = \Pi(\sigma) - \pi^2 + \alpha \|\sigma - \sigma_*\|_{\mathcal{H}^1}^2, \quad (9.14)$$

where:

- the bounds $\underline{\sigma}$ and $\bar{\sigma}$ are positive constants,
- π is a vector of vanilla option prices observed in the market at the calibration time $t = 0$,
- $\Pi(\sigma)$ is the corresponding vector of prices in the local volatility model with volatility function σ ,
- σ_* is a prior on σ and
- $\|\cdot\|_{\mathcal{H}^1}^2 \equiv \int_0^\infty \int_0^\infty (u(t, S)^2 + (\partial_t u(t, S))^2 + (\partial_S u(t, S))^2) dt dS$.

This formulation of the calibration problem as a regularized minimization problem, along with a gradient descent method to solve it numerically, were first introduced in Lagnado and Osher [173]. It was shown in [77] (see also [109]) that all the assumptions of Sect. 9.1.1 are satisfied in this case, granting stability, convergence and convergence rates. A trinomial tree implementation developed in [78] draws its efficiency from an exact computation of the gradient of the (discretized) cost criterion J in (9.14). Figure 9.3 displays the local volatility surface thus calibrated, along with the accuracy of the calibration, using the implied volatility surface of Fig. 9.1 as calibration input data.

This approach can also be extended to the calibration of a local volatility function using American option quotes (see [78]).

9.2.2.2 Approach by Entropic Regularization

An alternative is to use entropic regularization, rewriting the calibration problem as the following nonlinear minimization problem (see Avellaneda et al. [11] and Samperi [240]):

$$\min_{\{\sigma \equiv \sigma(\cdot, \cdot); \underline{\sigma} \leq \sigma \leq \bar{\sigma}\}} J(\sigma) = \Pi(\sigma) - \pi^2 + \alpha \|\sigma - \sigma_*\|_{\mathcal{L}^2}^2, \quad (9.15)$$

with

$$\|\sigma - \sigma_*\|_{\mathcal{L}^2}^2 := \mathbb{E} \int_0^\infty (\sigma(t, S_t) - \sigma_*(t, S_t))^2 dt.$$

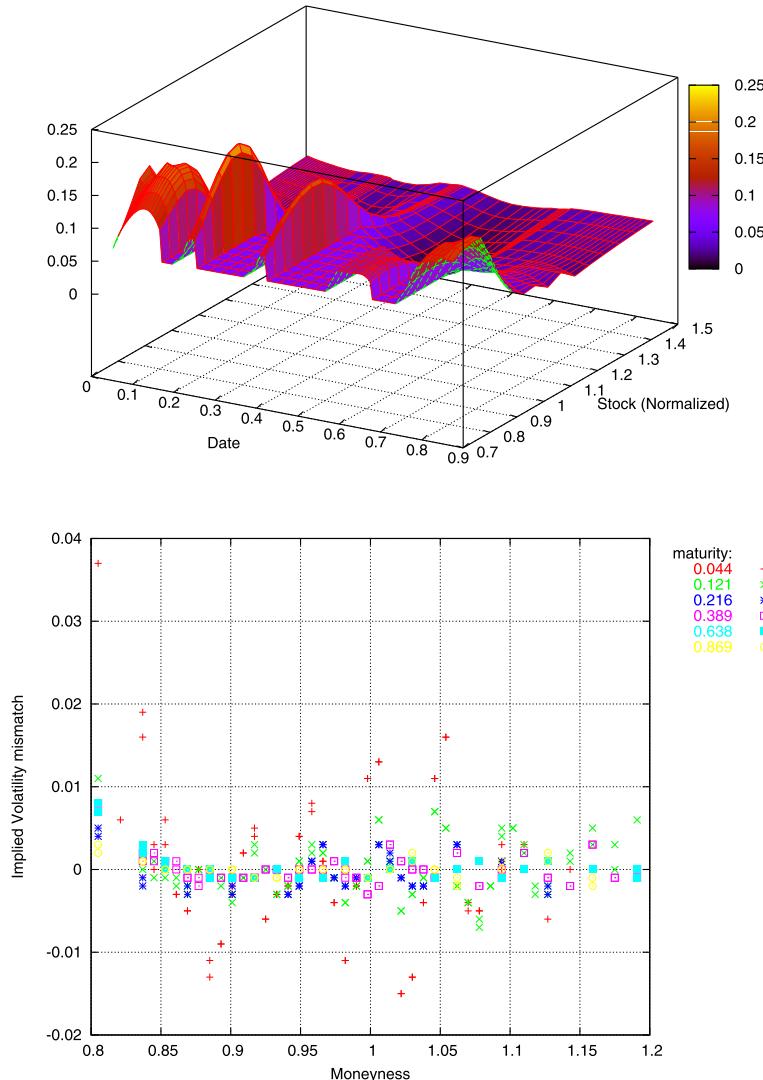


Fig. 9.3 Calibration by Tikhonov regularization to the implied volatility surface of Fig. 9.1: local volatility surface and calibration accuracy

Using a dual formulation, the minimization problem (9.15) can be solved in time $O(d)$, where d is the number of options in the calibration data set, versus $O(n^2)$ in the case of Tikhonov regularization implemented on a trinomial tree with n time steps. The numerical solution is thus typically faster than by Tikhonov regularization. However, it is also less stable, since the regularization term doesn't involve the gradient, but only the values, of $(\sigma - \sigma_*)$. See Fig. 9.4, which displays the results obtained by this method for the data set of Fig. 9.1.

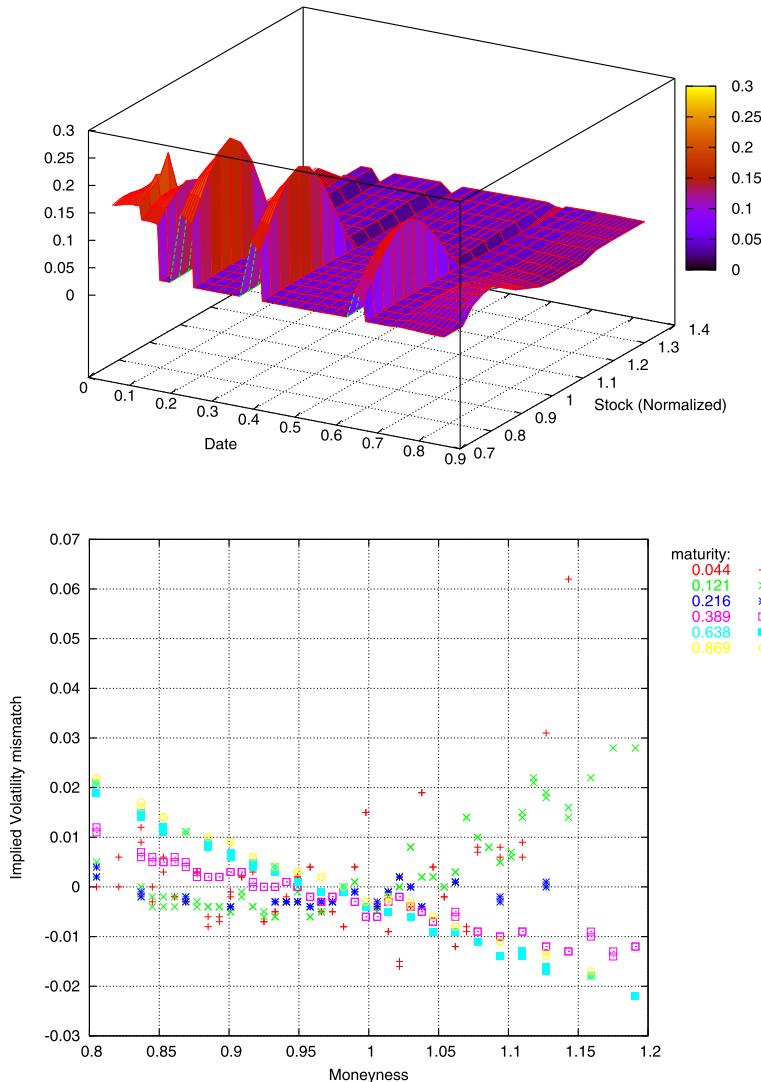


Fig. 9.4 Calibration by entropic regularization to the implied volatility surface of Fig. 9.1: local volatility surface and calibration accuracy

9.3 Weighted Monte Carlo

We finally present a method of calibration by simulation, the so-called Weighted Monte Carlo method, initially introduced by Avellaneda et al. [9, 10] to deal with equity or equity index derivatives and also used in portfolio credit risk [237]. In this approach, one first simulates m trajectories of a risk-neutral prior model. By

construction, the simulated trajectories are equiprobable in the prior model. One then weights the trajectories, looking for calibrated risk-neutral probabilities such that option prices in the weighted model match market prices. Calibration here thus takes the form of a reponderation of different scenarios, a bit in the spirit of the Girsanov transforms of Sect. 3.4.

Mathematically, this leads to the following problem, to be solved in the set \mathcal{P} of all positive probability measures p over the set of simulated trajectories: find $p \in \mathcal{P}$ such that

$$\Pi(p) = \Pi^*$$

or, equivalently (assuming zero funding rates),

$$\sum_{j=1}^m p_j \xi_l^j = \Pi_l^*, \quad l = 1, \dots, d, \quad (9.16)$$

where Π_l^* denotes the market price of the l th option in the calibration data set and ξ_l^j represents the payoff of the l th option on the j th simulated trajectory.

Since the number of simulated trajectories m is typically much larger than the number d of quoted prices (typically $m \geq 10^4$ versus $d \leq$ several hundreds even on the main index derivative markets), we further impose the constraint that the probability measure p is as close as possible to the (prior) uniform distribution on the simulated trajectories, in the sense of a suitable entropy criterion. The relative entropy of a probability measure μ with respect to another probability measure ν is defined by

$$\mathcal{E}_\nu(\mu) = \mathbb{E}^\nu \ln \left(\frac{d\mu}{d\nu} \right).$$

The convex mapping $\mu \mapsto \mathcal{E}_\nu(\mu)$ achieves a null minimum at $\mu = \nu$. In our case, we denote by $\mathcal{E}(p) = \sum_{j=1}^m p_j \ln(p_j)$ the relative entropy³ of the probability measure p with respect to the uniform distribution on the space of simulated trajectories. We then consider the following formulation of the calibration problem:

$$\inf_{\{p \in \mathcal{P}; \Pi(p) = \Pi^*\}} \mathcal{E}(p). \quad (9.17)$$

By convexity of the entropy \mathcal{E} , this problem admits a unique solution, provided the constraints are feasible, and this solution can be computed by duality as shown in Sect. 9.3.1. Feasibility of the constraints typically holds in practice since the number of unknowns m is much larger than the number of constraints d . However, it doesn't hold in general. A relaxed least squares approach not subject to this feasibility condition is presented in Sect. 9.3.2.

³Up to the constant $\ln(m)$.

9.3.1 Approach by Duality

Introducing Lagrange multipliers, we get the following equivalent form of (9.17):

$$\sup_{p \in \mathcal{P}} \inf_{\lambda \in \mathbb{R}^d} \mathcal{J}(\lambda, p) := -\mathcal{E}(p) + \sum_{l=1}^d \lambda_l \left(\sum_{j=1}^m p_j \xi_l^j - \Pi_l^* \right). \quad (9.18)$$

A minimax theorem applies, showing that there is no duality gap in this problem, so that

$$\sup_{p \in \mathcal{P}} \inf_{\lambda \in \mathbb{R}^d} \mathcal{J}(\lambda, p) = \inf_{\lambda \in \mathbb{R}^d} \sup_{p \in \mathcal{P}} \mathcal{J}(\lambda, p). \quad (9.19)$$

Assuming feasibility of the constraints, the unique solution (p^*, λ^*) of the dual is also the unique solution to the primal, and p^* is in turn the unique solution to (9.17). Now, due to the properties of the entropy \mathcal{E} , for every given $\lambda \in \mathbb{R}^d$, the probability measure p^λ which maximizes $\mathcal{J}(\lambda, p)$ is explicitly given, for $j = 1 \dots m$, as

$$p_j^\lambda = \frac{1}{Z_\lambda} \exp \left(\sum_{l=1}^d \lambda_l \xi_l^j \right), \quad (9.20)$$

where the normalization factor Z_λ is therefore given by

$$Z_\lambda = \sum_{j=1}^m \exp \left(\sum_{l=1}^d \lambda_l \xi_l^j \right).$$

So, using (9.20), for every $1 \leq k, l \leq d$ we have:

$$\begin{aligned} \mathcal{E}(p^\lambda) &= \sum_{j=1}^m p_j^\lambda \ln(p_j^\lambda) = -\ln(Z_\lambda) + \sum_{j=1}^m \sum_{l=1}^d p_j^\lambda \lambda_l \xi_l^j \\ \partial_{\lambda_l} \ln(Z_\lambda) &= \frac{\partial_{\lambda_l} Z_\lambda}{Z_\lambda} = \sum_{j=1}^m p_j^\lambda \xi_l^j = \mathbb{E}^{p^\lambda} \xi_l \\ \partial_{\lambda_k, \lambda_l}^2 \ln(Z_\lambda) &= \mathbb{C}\text{ov}^{p^\lambda}(\xi_k, \xi_l). \end{aligned} \quad (9.21)$$

In view of the first line of (9.21), the dual problem (right-hand side of (9.19)) reduces to

$$\inf_{\lambda \in \mathbb{R}^d} \mathcal{F}(\lambda) := -\mathcal{E}(p^\lambda) + \sum_{l=1}^d \lambda_l \left(\sum_{j=1}^m p_j^\lambda \xi_l^j - \Pi_l^* \right) = \ln(Z_\lambda) - \sum_{l=1}^d \lambda_l \Pi_l^*, \quad (9.22)$$

where by the second and third lines in (9.21) we have, for every $1 \leq k, l \leq d$,

$$\partial_{\lambda_l} \mathcal{F}(\lambda) = \partial_{\lambda_l} \ln(Z_\lambda) - \Pi_l^* = \mathbb{E}^{p^\lambda} \xi_l - \Pi_l^*$$

$$\partial_{\lambda_k, \lambda_l}^2 \mathcal{F}(\lambda) = \partial_{\lambda_k} \mathbb{E}^{p^\lambda} \xi_l = \text{Cov}^{p^\lambda}(\xi_k, \xi_l).$$

Thus the criterion \mathcal{F} in (9.22) is convex with respect to λ , and if λ^* minimizes \mathcal{F} , then with $p^* = p^{\lambda^*}$ we have, for every $l = 1, \dots, d$:

$$0 = \partial_{\lambda_l} \mathcal{F}(\lambda^*) = \partial_{\lambda_l} \ln(Z_\lambda) - \Pi_l^* = \mathbb{E}^{p^*} \xi_l - \Pi_l^*.$$

The calibration constraints are thus satisfied for the probability measure p^* associated with a minimizer λ^* of \mathcal{F} in (9.22).

9.3.1.1 Algorithm

The weighted Monte Carlo calibration algorithm proceeds as follows:

- i. simulate m trajectories in the prior model,
- ii. compute the payoffs ξ_l^j for $l = 1 \dots d$, $j = 1 \dots m$,
- iii. use a descent gradient routine to minimize $\mathcal{F}(\lambda)$ in (9.22),
- iv. compute the related probabilities p_j^* , for $j = 1 \dots m$, by (9.20).

9.3.2 Relaxed Least Squares Approach

To get rid of the feasibility constraints, we can relax the equality constraints (9.16) into inequality constraints. We thereby obtain the following relaxed formulation of (9.17):

$$\inf_{p \in \mathcal{P}} \mathcal{E}_\omega(p) := \mathcal{E}(p) + \frac{1}{2} \sum_{l=1}^d \frac{1}{\omega_l} (\mathbb{E}^p(\xi_l) - \Pi_l^*)^2 \quad (9.23)$$

for positive weights ω_l . Note that, in the limit $\omega \rightarrow 0$, (9.23) reduces to (9.17), whereas in the limit $\omega \rightarrow +\infty$ the solution to (9.23) converges to $\mathcal{U}_{\{1, \dots, m\}}$.

It is shown in Sect. 4 of [10] that the minimization problem (9.23) admits the following dual formulation:

$$\begin{aligned} & \inf_{\lambda \in \mathbb{R}^d} \sup_{p \in \mathcal{P}} \left(-\mathcal{E}(p) + \sum_{l=1}^d \lambda_l (\mathbb{E}^p(\xi_l) - \Pi_l^*) \right) + \frac{1}{2} \sum_{l=1}^d \omega_l \lambda_l^2 \\ &= \inf_{\lambda \in \mathbb{R}^d} \mathcal{F}(\lambda) + \frac{1}{2} \sum_{l=1}^d \omega_l \lambda_l^2 =: \inf_{\lambda \in \mathbb{R}^d} \mathcal{F}_\omega(\lambda), \end{aligned}$$

where $\mathcal{F}(\lambda)$ is defined in (9.22). The calibration algorithm is then the same as in Sect. 9.3.1.1, except that we minimize $\mathcal{F}_\omega(\lambda)$ instead of $\mathcal{F}(\lambda)$ in step (iii). This calibration procedure achieves a trade-off, controlled by the vector of weights ω , between entropy with respect to the prior and accuracy of the calibration.

Remark 9.3.1 In either approach (with exact or relaxed calibration constraints), a calibrated probability measure p^* can be used to price other options, or to compute sensitivities, consistent with the market prices of traded options. The weighted Monte Carlo method is particularly well suited for static hedging. Suppose that we want to minimize the variance⁴ of the final output of a strategy consisting of a short position in a derivative with payoff χ , statically hedged by a portfolio of constant positions, given in the form of a row-vector ξ , in the vanilla options previously calibrated through a weighted Monte Carlo procedure. Then the min-variance hedge is given by the following multilinear regression formula, which can be regarded as a static analog of the formula (4.48):

$$\zeta^{va} = \text{Cov}^{p^*}(\chi, \xi) (\text{Cov}^{p^*}(\xi))^{-1}. \quad (9.24)$$

⁴Risk-neutral variance for simplicity.

Part IV

Applications

We now present two applications.

In Chap. 10 we deal with the issue of pricing convertible bonds numerically, by simulation. A convertible bond can be regarded as a coupon-paying and callable American option. Moreover, call times are subject to constraints, known as call protections, preventing the issuer from calling the bond at certain sub-periods of time. The nature of the call protection may be very path-dependent, leading to high-dimensional nonlinear pricing problems. Deterministic pricing schemes are then ruled out by the curse of dimensionality, and simulation methods are the only viable alternative.

In Chap. 11 we devise simulation/regression numerical schemes in pure jump models. There the idea is to perform the nonlinear regressions, used for computing conditional expectations, in the time variable for a given state of the model rather than in the space variables at a given time in the diffusive setups of Chap. 10. This idea is stated in the form of a generic lemma that is valid in any continuous-time Markov chain model. This is then tested in the context of two credit risk applications, the first of which values the sensitivities of a CDO tranche in a homogeneous groups model of portfolio credit risk by Monte Carlo without resimulation. The second computes by Monte Carlo the credit valuation adjustment (CVA) on a CDO tranche in a common shock model of counterparty credit risk.

These two chapters provide a thorough illustration of the abilities of simulation/regression numerical schemes for solving high-dimensional pricing equations: very large systems of partial differential equations in Chap. 10 and Markov chain related systems of ODEs in Chap. 11.

Chapter 10

Simulation/Regression Pricing Schemes in Diffusive Setups

In this chapter we consider the issue of pricing numerically by simulation American and game options on an underlying S . Our standing example of a game option is a convertible bond written on a stock S . A convertible bond pays coupons, from time 0 onwards, until a terminal payoff

$$\mathbb{1}_{v=\tau < T} \ell(\tau, S_\tau) + \mathbb{1}_{\theta < \tau} h(\theta, S_\theta) + \mathbb{1}_{v=T} \phi(S_T) \quad (10.1)$$

occurs at the minimum $v = \tau \wedge \theta$ of two $[0, T]$ -valued stopping times τ and θ . In (10.1) the put time τ and the call time θ are $[0, T]$ -valued stopping times under the control of the holder and the issuer of the bond, respectively, and:

- $\phi(S_T)$ corresponds to a terminal payoff that is paid by the issuer to the holder at time T if the contract was not exercised before the maturity time T ;
- $\ell(\tau, S_\tau)$, respectively $h(\theta, S_\theta)$, corresponds to an early put payoff, respectively early call payoff, that is paid by the issuer to the holder of the claim in the event of early termination of the contract at the initiative of the holder, respectively issuer.

Note that “put” and “call” above are in fact to be understood as “put or conversion” and “call or conversion”, since in either a “put” or “call” event the holder has the choice between being granted a certain amount of cash or enforcing conversion of the bond against a number¹ of stocks S . The latter possibility explains the name of the product. Convertible bonds are thus products with early exercise clauses both on the part of the holder and on the part of the issuer of the claim. They represent the main practical example of a game option as formally introduced by Kifer in [166].

Remark 10.0.1 With their triple equity, bond and credit features, convertible bonds have always played a very important role in the sphere of capital structure arbitrage strategies, particularly equity-to-credit. They have been an even more topical issue subsequent to the great crisis, since regulators advocate their use in dealing with the systemic risk of banks. However, this is in the form of the so-called contingent

¹Taken equal to one for notational simplicity in the sequel.

convertible bonds (“CoCo”), in which the conversion is automatically triggered (as opposed to optional in this chapter) as certain events related to distress of the issuing bank occur; see [99] regarding the theory and practice of contingent convertible bonds.

Call times θ are typically subject to constraints, called call protections, that prevent the issuer from calling the bond on certain random time intervals. Moreover, the nature of the call protection may be very path-dependent. Thus, let a trigger level \bar{S} and nonnegative integers $l \leq d$ be given. The “ l out of the last d ” call protection clause means that call is allowed at a given time t if and only if S has been $\geq \bar{S}$ at $\geq l$ among the last d daily closing prices preceding the current time t . This can be found in actual convertible bond contracts with d of the order of thirty (for thirty days, i.e. one month). Mathematically, this clause corresponds to a path dependence based on the d Boolean parameters “ $S \geq \bar{S}$ or $< \bar{S}$ ” at each of the last d call protection monitoring times. After extension of the state space to ensure Markovianity for the set-up, this leads to a system of 2^d pricing partial differential equations. For d greater than, say, ten, deterministic pricing schemes are ruled out by the curse of dimensionality (see Table 10.10), and simulation methods are the only computationally tractable alternative.

From the mathematical point of view, the study of game options with call protection leads to doubly reflected backward stochastic differential equations with an upper barrier that is only active on random time intervals. These equations are henceforth called doubly reflected BSDEs with an intermittent upper barrier, or RIBSDE. Such RIBSDEs and, in the Markovian case, the related variational inequalities (VI for short), are studied theoretically in Sect. 14.2. A convergence rate for a time-discretization scheme for an RIBSDE is established in [65]. In the present chapter, published in article form in [88], we assess the practical value of this scheme on the benchmark problem of pricing convertible bonds with call protection by simulation.

Outline Section 10.1 lays out the setup and Sect. 10.2 presents the convertible bond pricing equations and their approximation. In Sect. 10.3 we treat the simplest cases of no call (American option) or no call protection (game option). Section 10.4 deals with stylized cases of continuously monitored call protection. More realistic forms of discretely monitored call protection are handled in Sect. 10.5.

All the viscosity solutions, subsolutions or supersolutions in this chapter are assumed to have polynomial growth in S . We refer the reader to Chap. 13 for the related technicalities.

10.1 Market Model

10.1.1 Underlying Stock

We work throughout with the following local drift and volatility pre-default model of an underlying stock price process S : $S_0 = x$ and, for $t \in [0, T]$,

$$dS_t = S_t (b(t, S_t) dt + \sigma(t, S_t) dW_t), \quad (10.2)$$

where W is a univariate \mathbb{P} -Brownian motion and $b(t, S)$ and $\sigma(t, S)$ are local drift and volatility coefficients so that the SDE (10.2) admits a unique strong solution over $[0, T]$. Thus, writing $\partial u = \frac{\partial u}{\partial S}$ and $\partial^2 u = \frac{\partial^2 u}{\partial S^2}$ for every function $u = u(t, S)$, S_t follows a one-dimensional diffusion with generator given by

$$\mathcal{A}u = bS\partial u + \frac{1}{2}\sigma^2 S^2\partial^2 u. \quad (10.3)$$

More precisely, we work with a credit risk adjusted risk-neutral drift coefficient b of the form

$$b(t, S) = r(t) - q(t) + \Lambda\lambda(t, S), \quad (10.4)$$

where:

- the riskless short interest rate $r(t)$, the equity dividend yield $q(t)$ on S , and the local default intensity $\lambda(t, S) \geq 0$ of the firm issuing the bond are bounded functions, and
- $\Lambda \in [0, 1]$ is a real constant representing the fractional loss on S in case of a default of the firm issuing the bond.

We refer the reader to Sect. 4.3.2 for the credit risk background and to [40] for all the details regarding this model. Note that this is in fact a pre-default model of a stock price process, for a default time τ_d of the firm issuing the bond defined as

$$\tau_d = \inf \left\{ t \in [0, T]; \int_0^t \lambda(u, S_u) du \geq \epsilon \right\}, \quad (10.5)$$

for some unit exponential random variable ϵ independent from the Brownian motion W . Thus the filtration of W only plays the role of a reference filtration, whereas the full model filtration is given as \mathbb{F}^W progressively enlarged by τ_d (see Sect. 4.3.2.2).

We henceforth let

$$\lambda(t, S) = \lambda_0(S_0/S)^\gamma, \quad \sigma(t, S) = \sigma, \quad (10.6)$$

for nonnegative constants λ_0 , γ and σ .

The “total default” case with $\Lambda = 1$ and $\gamma > 0$ in (10.4), (10.6) corresponds to the case where S represents the pre-default value of the equity of the firm issuing the bond, so that the value of the equity vanishes at τ_d , as reflected by the $\lambda(t, S)$ -adjustment (for $\Lambda = 1$) to the risk-neutral drift coefficient of S in (10.4).

The “partial default” case with $\Lambda = 0$ and $\gamma = 0$ corresponds to the situation of a so-called exchangeable bond, where S represents the pre-default value of the equity of a firm different from the issuer of the bond. So the value of the equity is not affected at τ_d , and the pre-default drift coefficient of S in (10.4) is a standard risk-neutral coefficient $r(t) - q(t)$.

For consistency with Sect. 4.3.2 we should denote in this chapter $\mathcal{F}_t^W = \tilde{\mathcal{F}}_t$, $\mathbb{E}(\cdot | \mathcal{F}_t^W) = \tilde{\mathbb{E}}_t$ and \tilde{S}_t instead of S_t . However, with respect to the targeted application of pricing a defaultable convertible bond (see [40] regarding hedging), a pre-default perspective is sufficient here. For simplicity, we henceforth drop any “ $\tilde{\cdot}$ ” notation and “pre-default” terminology. We likewise simply denote by β_t (instead of α_t as in Sect. 4.3.2) the risk-neutral discount factor adjusted for credit risk $\beta_t = e^{-\int_0^t \mu(s, S_s) ds}$, in which $\mu(t, S) = r(t) + \lambda(t, S)$.

10.1.1.1 First-Variation Process

The so-called first variation of S is the process ∇S such that

$$\nabla S_t = \lim_{\alpha \rightarrow 0+} (2\alpha)^{-1} (S_t^{x+\alpha} - S_t^{x-\alpha}), \quad (10.7)$$

where we write explicitly the initial condition $x = S_0$ of S as a superscript. We assume that $B(t, S) = b(t, S)S$ and $\Sigma(t, S) = \sigma(t, S)S$ are of class \mathcal{C}_b^1 with Lipschitz first derivatives in S . Then ∇S is such that $\nabla S_0 = 1$ and, for $t \geq 0$,

$$d\nabla S_t = \nabla S_t (\partial B(t, S_t) dt + \partial \Sigma(t, S_t) dW_t). \quad (10.8)$$

10.1.1.2 Time-Discretization

For numerical purposes we need to approximate (10.2) and (10.8) on a discrete time-grid $\mathbf{t} = \{0 = t_0 < t_1 < \dots < t_n = T\}$. In this context we will often find it convenient to denote the time by i rather than t_i , so that:

- X_i and $u_i(X_i)$ are shorthand for \hat{X}_{t_i} and $u(t_i, \hat{X}_{t_i})$, given a function $u = u(t, x)$ and a discrete-time process \hat{X}_{t_i} approximating a continuous-time process X ,
- conditional expectation given X_i is denoted by \mathbb{E}_i , and
- \mathcal{T}_i represents the set of stopping times² v taking their values in $\{i, \dots, n\}$.

We will consider the Euler schemes for S_t and ∇S_t defined by $S_0 = x$, $\nabla S_0 = 1$ and, for every $i = 0, \dots, n-1$:

$$\begin{aligned} S_{i+1} &= S_i (1 + b_i(S_i)(t_{i+1} - t_i) + \sigma_i(S_i)(W_{t_{i+1}} - W_{t_i})) \\ \nabla S_{i+1} &= \nabla S_i (1 + \partial B_i(S_i)(t_{i+1} - t_i) + \partial \Sigma_i(S_i)(W_{t_{i+1}} - W_{t_i})). \end{aligned} \quad (10.9)$$

10.1.2 Convertible Bond

Given a nondecreasing sequence of stopping times ϑ , which will represent times of switching of call protection in the financial interpretation (see Sect. 14.2), let \mathcal{T}_t^ϑ , or

²With respect to the filtration of the S_i .

simply \mathcal{T}^ϑ if $t = 0$, denote the set of all the $\bigcup_{l>0}[\vartheta_{2l-1}, \vartheta_{2l}) \cup \{T\}$ -valued stopping times. We consider a convertible bond continuously paying coupons $c(t, S_t) dt$ until a terminal payoff (10.1) is paid at time $\nu = \tau \wedge \theta$, where $(\tau, \theta) \in \mathcal{T} \times \mathcal{T}^\vartheta$. In (10.1) we henceforth let

$$\ell(t, S_t) = \ell(S_t) = \overline{P} \vee S_t, \quad h(t, S_t) = h(S_t) = \overline{C} \vee S_t, \quad \phi(S_T) = \xi = \overline{N} \vee S_T \quad (10.10)$$

for nonnegative constants $\overline{P} \leq \overline{N} \leq \overline{C}$. That θ has to be chosen by the bond issuer in the constrained set \mathcal{T}_t^ϑ means that we effectively deal with an intermittent call protection, forbidding issuer calls on the “even” time intervals $[\vartheta_{2l}, \vartheta_{2l+1})$. In view of this restriction on call times, in addition to the nominal call payoff process $h(S_t)$, we introduce the effective call payoff process accounting for the call protection defined, for $t \in [0, T]$, by

$$U_t = \sum_{l \geq 0} \mathbb{1}_{\{t \in [\vartheta_{2l}, \vartheta_{2l+1})\}} \infty + \sum_{l > 0} \mathbb{1}_{\{t \in [\vartheta_{2l-1}, \vartheta_{2l})\}} h(S_t). \quad (10.11)$$

We also define $L_t = \ell(S_t)$, for $t \in [0, T]$.

Moreover, accounting for defaultability of the bond with intensity $\lambda(t, S_t)$, we assume the following form of the coupon rate function c (see Sects. 10.1.1, 4.3.2 and [40]):

$$c(t, S) = \bar{c}(t) + \lambda(t, S)((1 - \Lambda)S \vee \overline{R}), \quad (10.12)$$

where \bar{c} denotes a nominal coupon rate function, and \overline{R} represents a nominal recovery on the bond upon default.

Remark 10.1.1 In practice coupons are paid discretely rather than continuously, which results in a discrete stream of nominal coupons instead of the continuously paid nominal coupon rate \bar{c} in (10.12). In the theoretical description of the model we consider continuously paid coupons for simplicity of presentation. However, following the guidelines of Sect. 14.1 and [37, 40], one can also deal with discrete coupons. In the numerical experiments we actually use discrete coupons.

10.2 Pricing Equations and Their Approximation

We showed in Sect. 3.5 and Part II (see also Part V) how, from the mathematical point of view, the task of pricing and hedging financial derivatives can be reduced to that of solving backward stochastic differential equations, called stochastic pricing equations, equivalent to the deterministic pricing equations that arise in more specific Markovian setups. In this section we discuss the convertible bond pricing equations and their approximation.

10.2.1 Stochastic Pricing Equation

By application of the results of Sect. 14.2 (see also [65]), our convertible bond pricing BSDE, denoted henceforth by (\mathcal{E}) , appears as follows, with $c_t = c(t, S_t)$, $\mu_t = \mu(t, S_t)$ and using the convention that $0 \times \infty = 0$ in the last line:

$$\begin{aligned} \Pi_T &= \xi, \quad \text{and for } t \in [0, T] : \\ \begin{cases} -d\Pi_t = (c_t - \mu_t \Pi_t) dt + dA_t - \Delta_t \sigma(t, S_t) S_t dW_t \\ L_t \leq \Pi_t \leq U_t, \quad (\Pi_t - L_t) dA_t^+ = (U_t - \Pi_t) dA_t^- = 0. \end{cases} \end{aligned}$$

This equation is to be solved in a triple of processes (Π, Δ, A) , where A is a finite variation process with square integrable Jordan components³ A^\pm , continuous except for possible jumps of A^- at the even times ϑ_{2l} in $(0, T)$. As shown in Sect. 14.2 (see also [65]), allowing for discontinuities of A^- at the ϑ_{2l} is necessary to ensure existence of a solution to (\mathcal{E}) satisfying the constraint $\ell(S_t) \leq \Pi_t \leq h(S_t)$ on the odd intervals $[\vartheta_{2l-1}, \vartheta_{2l}]$, whereas on the even intervals $[\vartheta_{2l}, \vartheta_{2l+1})$ only the lower constraint $\ell(S_t) \leq \Pi_t$ is in force. Uniqueness for a solution holds under the minimality conditions $(\Pi_t - L_t) dA_t^+ = (U_t - \Pi_t) dA_t^- = 0$.

The following verification principle (see Sect. 14.2.1.1) establishes the connection between a solution (Π, Δ, A) of (\mathcal{E}) , assumed to exist, and the problem of pricing the convertible bond under the risk-neutral measure \mathbb{P} . We write $v = \tau \wedge \theta$, for $(\tau, \theta) \in \mathcal{T}_t \times \mathcal{T}_t^\vartheta$.

Proposition 10.2.1

- (i) Π is the value process of the Dynkin game with cost criterion $\mathbb{E}_t \pi^t(\tau, \theta)$ on $\mathcal{T}_t \times \mathcal{T}_t^\vartheta$, where $\pi^t(\tau, \theta)$ is the \mathcal{F}_v -measurable random variable defined by

$$\beta_t \pi^t(\tau, \theta) = \int_t^v \beta_s c_s ds + \beta_v (\mathbb{1}_{\{v=\tau < T\}} L_\tau + \mathbb{1}_{\{v=\theta < \tau\}} U_\theta + \mathbb{1}_{\{v=T\}} \xi).$$

We thus have \mathbb{P} -almost surely, for every $t \in [0, T]$,

$$\underset{\tau \in \mathcal{T}_t}{\text{esssup}} \underset{\theta \in \mathcal{T}_t^\vartheta}{\text{essinf}} \mathbb{E}_t \pi^t(\tau, \theta) = \Pi_t = \underset{\theta \in \mathcal{T}_t^\vartheta}{\text{essinf}} \underset{\tau \in \mathcal{T}_t}{\text{esssup}} \mathbb{E}_t \pi^t(\tau, \theta). \quad (10.13)$$

More precisely, for every $t \in [0, T]$ and $\epsilon > 0$, the pair of stopping times $(\tau_\epsilon^t, \theta_\epsilon^t) \in \mathcal{T}_t \times \mathcal{T}_t^\vartheta$ defined by

$$\begin{aligned} \tau_\epsilon^t &= \inf \{s \in [t, T]; \Pi_s \leq L_s + \epsilon\} \wedge T \\ \theta_\epsilon^t &= \inf \left\{ s \in \bigcup_{l>0} [\vartheta_{2l-1} \vee t, \vartheta_{2l} \vee t); \Pi_s \geq U_s - \epsilon \right\} \wedge T \end{aligned} \quad (10.14)$$

³See Definition 4.1.12.

is an ϵ -saddle-point for this Dynkin game at time t , in the sense that for any $(\tau, \theta) \in \mathcal{T}_t \times \mathcal{T}_t^\vartheta$, we have

$$\mathbb{E}_t \pi^t(\tau, \theta_\epsilon^t) - \epsilon \leq \Pi_t \leq \mathbb{E}_t \pi^t(\tau_\epsilon^t, \theta) + \epsilon. \quad (10.15)$$

(ii) If the reflecting process A in the solution to (\mathcal{E}) is continuous, then the pair of stopping times $(\tau^t, \theta^t) \in \mathcal{T}_t \times \mathcal{T}_t^\vartheta$, obtained by setting $\epsilon = 0$ in (10.14), is a saddle-point of the game. We thus have, for every $(\tau, \theta) \in \mathcal{T}_t \times \mathcal{T}_t^\vartheta$,

$$\mathbb{E}_t \pi^t(\tau, \theta^t) \leq \Pi_t \leq \mathbb{E}_t \pi^t(\tau^t, \theta).$$

As a consequence, Π is an arbitrage price process for the bond (see Sects. 14.2 and [38, 40]). Given a suitable set of hedging instruments, the time-0 price Π_0 is also a bilateral superhedging price, i.e. there exists a self-financing super-hedging strategy for the issuer of the bond from any initial wealth greater than Π_0 , and a self-financing super-hedging strategy for the holder of the bond from any initial wealth greater than $(-\Pi_0)$. The price Π_0 is also the infimum of the initial wealth of all the issuer's self-financing super-hedging strategies, and this infimum is attained as a minimum in case the reflecting process A is continuous in the solution to (\mathcal{E}) .

10.2.2 Markovian Case

We assume henceforth that the ϑ_l are given as the successive times of exit from and entrance into a domain K , for an auxiliary finite-valued, pure jump process H . So $\vartheta_0 = 0$ and, for every $l > 0$,

$$\vartheta_{2l-1} = \inf\{t > \vartheta_{2l-2}; H_t \notin K\} \wedge T, \quad \vartheta_{2l} = \inf\{t > \vartheta_{2l-1}; H_t \in K\} \wedge T. \quad (10.16)$$

The effective call payoff process U in (10.11) can thus be written as

$$U_t = \mathbb{1}_{H_t \in K} \infty + \mathbb{1}_{H_t \notin K} h(S_t) =: \bar{h}(X_t), \quad (10.17)$$

where X_t denotes the bivariate process (S_t, H_t) . In case this process is Markov, in view of (10.17) it is expected, and will be established later, that for $t \geq 0$ we have:⁴

$$\Pi_t = u(t, X_t), \quad \Delta_t = \partial_S u(t, X_t) \quad (10.18)$$

for a suitable pricing function $u = u(t, x)$ with $x = (S, k)$. The second argument k , which corresponds to the finite-valued process H_t in the probabilistic interpretation, can be thought of as the index of a system of functions u_k of (t, S) . We will write

⁴The second identity holds provided the function u is sufficiently regular; otherwise a more general but less constructive representation for the delta can be given in terms of Malliavin calculus.

$u(t, x)$, $u(t, S, k)$ or $u_k(t, S)$, depending on what is more convenient in the context at hand. We will see below that the pricing function u can be characterized analytically as the viscosity solution to a related variational inequality,⁵ referred to as (\mathcal{VI}) henceforth.

10.2.3 Generic Simulation Pricing Schemes

Regarding simulation pricing schemes that may be used for solving the stochastic pricing equation (\mathcal{E}) , given the process $(X_t)_{t \in [0, T]} = (S, H)_{t \in [0, T]}$ and its Euler scheme $(X_i)_{0 \leq i \leq n} = (S_i, H_i)_{0 \leq i \leq n}$ with time-step h , we write further:

- $X_i^j = (S_i^j, H_i^j)$ for the value of X_i on the j th simulated trajectory,
- \mathbb{E}_i^j for the conditional expectation given $X_i = X_i^j$,

for every simulation run $j = 1, \dots, m$. On the stochastically generated mesh $(X_i^j)_{0 \leq i \leq n}^{1 \leq j \leq m}$, the generic simulation/regression estimate $(u_i^j, \delta_i^j)_{0 \leq i \leq n}^{1 \leq j \leq m}$ of [65] for $(u(t_i, X_i^j), \partial_S u(t_i, X_i^j))_{0 \leq i \leq n}^{1 \leq j \leq m}$ appears as follows, in the vein of the American Monte Carlo pricing scheme of Sect. 6.10 (see Remark 10.2.2 for more details): $u_n = \phi$ and, for $i = n - 1, \dots, 0$ and $j = 1, \dots, m$:

$$\begin{aligned} u_i^j &= \min(\bar{h}(X_i^j), \max(\ell(S_i^j), e^{-\mu_i^j h} \mathbb{E}_i^j(u_{i+1} + hc_{i+1}))) \\ \delta_i^j &= \frac{\mathbb{E}_i^j\{u_{i+1}(W_{i+1} - W_i)\}}{\sigma_i(S_i^j) S_i^j h}. \end{aligned} \tag{10.19}$$

Note that the “min” in the first line plays no role at $X_i^j = (S_i^j, H_i^j)$ if $H_i^j \in K$, since in this case $\bar{h}(X_i^j) = +\infty$.

This simulation pricing scheme thus ultimately relies on the computation of the conditional expectations residing in (10.19) (for $i \geq 1$, since for $i = 0$ the conditional expectations reduces to expectations). For computing, at time step $i \geq 1$, the conditional expectations \mathbb{E}_i^j in (10.19), which are conditional given $(t = t_i, S_i = S_i^j, H_i = H_i^j)$, we perform, for every $k \in K$, nonlinear regressions⁶ of $(u_{i+1}^j + hc_{i+1}^j)_{j \in \Omega_i^k}$ and $(u_{i+1}^j(W_{i+1}^j - W_i^j))_{j \in \Omega_i^k}$ against $(S_i^j)_{j \in \Omega_i^k}$, where Ω_i^k denotes the subset of the indices j of the trajectories such that $H_i^j = k$. Let $\tilde{u}_i(\cdot, k)$ and $\hat{\delta}_i(\cdot, k)$ denote the resulting regression functions where, for every index k such that the set Ω_i^k is empty or too small for the nonlinear regressions over Ω_i^k to be

⁵See Sect. 3.5.4 and Chap. 13.

⁶See Sect. 6.10.2.

tractable or significant, the regression functions are set to zero. We then approximate in (10.19), for every $j = 1, \dots, m$,

$$\begin{aligned} u_i^j &\approx \widehat{u}_i^j := \min(\bar{h}(X_i^j), \max(\ell(S_i^j), e^{-\mu_i^j h} \widetilde{u}_i(S_i^j, H_i^j))) \\ \delta_i^j &\approx \widehat{\delta}_i(S_i^j, H_i^j). \end{aligned}$$

Note that this procedure for computing the conditional expectations in (10.19) can be interpreted as using a method of cells (see Sect. 6.10.2) in the direction of the k variable and whatever method of choice in the direction of the S variable, a method of cells again being a simple and robust alternative for estimating the pricing function $u = u_k(t, S)$. We can also recover, from the pricing function estimated by (10.19), the following estimates of the call/put regions and of the optimal call/put policies:

$$\begin{aligned} \mathcal{E}_p &= \{(i, X_i^j); u_i^j = \ell(X_i^j)\}, \quad \tau^j = \inf\{0 \leq i \leq n; X_i^j \in \mathcal{E}_p\} \wedge n \\ \mathcal{E}_c &= \{(i, X_i^j); H_i^j \notin K \text{ and } u_i^j = h(S_i^j)\} \\ \theta^j &= \inf\{0 \leq i \leq n; H_i^j \notin K \text{ and } X_i^j \in \mathcal{E}_c\} \wedge n. \end{aligned} \tag{10.20}$$

As we saw regarding American Monte Carlo pricing schemes in Sect. 6.10, confidence intervals are not available for such time-iterative simulation/regression pricing schemes; one can only derive a confidence interval of the estimation method by running the simulation/regression scheme several times. It is also possible to derive bounds on the prices by a suitable extension to game options of the dual American Monte Carlo method of Rogers [235]; see Beveridge and Joshi [29].

10.2.3.1 Time-0 Price-and-Delta

The previous scheme suffers from an accumulation of errors through the iterated computation of the conditional expectations in (10.19). To limit this effect, an often preferred alternative regarding the prices and delta at time 0 consists in reusing the estimated optimal stopping policies τ^j and θ^j in (10.20) for computing new estimates \widetilde{u}_0 and $\widetilde{\delta}_0$ of the option price and delta at time 0. We thus have the so-called Monte Carlo forward estimates for the option price and delta at time 0, in which $v^j = \tau^j \wedge \theta^j$:

$$\begin{aligned} \widetilde{u}_0 &= \frac{1}{m} \sum_{j=1}^m \left\{ h \sum_{i=1}^{v^j} \beta_i^j c_i(S_i^j) + \beta_{v^j}^j (\mathbb{1}_{\{v^j=\tau^j < n\}} \ell(S_{\tau^j}^j) \right. \\ &\quad \left. + \mathbb{1}_{\{\theta^j < \tau^j\}} h(S_{\theta^j}^j) + \mathbb{1}_{\{v^j=n\}} \phi(S_n^j)) \right\} \end{aligned}$$

$$\begin{aligned} \tilde{\delta}_0 = & \frac{1}{m} \sum_{j=1}^m \left\{ h \sum_{i=1}^{v^j} \beta_i^j (\beta_i^j \partial c_i(S_i^j) \nabla S_i^j + \chi_i^j c_i(S_i^j)) \right. \\ & + \beta_{v^j}^j (\mathbb{1}_{\{v^j = \tau^j < n\}} \partial \ell(S_{\tau^j}^j) + \mathbb{1}_{\{\theta^j < \tau^j\}} \partial h(S_{\theta^j}^j) + \mathbb{1}_{\{\tau^j \wedge \theta^j = n\}} \partial g(S_n^j)) \nabla S_{v^j}^j \\ & \left. + \chi_{v^j}^j (\mathbb{1}_{\{v^j = \tau^j < n\}} \ell(S_{\tau^j}^j) + \mathbb{1}_{\{\theta^j < \tau^j\}} h(S_{\theta^j}^j) + \mathbb{1}_{\{\tau^j \wedge \theta^j = n\}} g(S_n^j)) \right\}, \end{aligned} \quad (10.21)$$

where $\beta_l^j = e^{-h \sum_{k=0}^{l-1} \mu_k^j}$ and where $\chi_i^j = -h \beta_i^j \sum_{l=1}^i \partial \mu_l(S_l^j) \nabla S_l^j$ is a discretization of the first-variation process

$$-\beta_t \int_0^t \partial \mu(s, S_s) \nabla S_s ds$$

of $\beta_t = e^{-\int_0^t \mu(s, S_s) ds}$.

Remark 10.2.2 The above simulation/regression price estimates are essentially the ones that were developed in the American Monte Carlo case of Sect. 6.10:

- by Tsitsiklis and VanRoy in [251] using an iteration on the values algorithm, regarding the MC backward estimate u_0 in (10.19);
- by Longstaff and Schwartz in [195] using an iteration on the policies algorithm, regarding the MC forward estimate \tilde{u}_0 in (10.21).

As for the deltas, the MC backward estimate δ_0 in (10.19) is standard in the numerical BSDE literature; the MC forward estimate $\tilde{\delta}_0$ in (10.21) is considered in Theorem 2.3 of Gobet [134].

10.2.4 Convergence Results

In this subsection we discuss the theoretical convergence of the price and delta estimates (10.19) and (10.21).

10.2.4.1 Time-Discretization

The simulation scheme (10.19) implicitly relies on a time-discretized Euler scheme $(\widehat{\Pi}_{t_i})_{0 \leq i \leq n}$ for $(\Pi_t)_{t \in [0, T]}$. We first review convergence results regarding this time-discretization. The exact statements, that can be found in the references in the proposition below, entail the introduction of a sub-grid of $t = (t_i)_{0 \leq i \leq n}$ such that the reflections in the first line of (10.19) only operate at times of the sub-grid. Since we found no significant difference in the numerical results by using the full time-grid t as reflection grid, this is omitted here for simplicity of presentation.

Denoting by $(\widehat{\Pi}_{t_i})_{0 \leq i \leq n}$ the discrete time approximation for $(\Pi_t)_{t \in [0, T]}$ that is implicit in (10.19) and setting $h = \max_{1 \leq i \leq n} (t_i - t_{i-1})$, we thus have:

Proposition 10.2.3

- (i) (Bouchard and Chassagneux [50], Chassagneux [64]) *In the American (no call) and game (no call protection) cases of Sect. 10.3 below,*

$$\max_{i \leq n-1} \sup_{t \in [t_i, t_{i+1})} \mathbb{E}[|\Pi_t - \widehat{\Pi}_{t_i}|^2] \leq C h^{\frac{1}{2}};$$

- (ii) (Chassagneux and Crépey [65]) *In the cases of discretely monitored call protection considered in Sect. 10.5 below, for every $\eta > 0$,*

$$\max_{i \leq n-1} \sup_{t \in [t_i, t_{i+1})} \mathbb{E}[|\Pi_t - \widehat{\Pi}_{t_i}|^2] \leq C_\eta h^{\frac{1}{2}-\eta}.$$

These convergence rates exploit the semi-convex regularity of the payoff functions ℓ and h in (10.10). In the case of Lipschitz payoff functions, downgraded convergence rates are available. Convergence in cases of continuously monitored call protection, considered in Sect. 10.4, will be discussed there. In the American or game cases, convergence rates are also available for the deltas; see Theorem 6.1 in [64] and Theorem 4.1 in [50].

Monte Carlo Backward Versus Forward Estimates In the American case of Sect. 10.3.1, we refer the reader to Remark 5.5 in Bouchard and Chassagneux [50] for the derivation of representations, similar to (10.21), regarding a discretely reflected BSDE associated with the continuous Euler scheme of an underlying diffusion. In the American and game cases (no call and no call protection), working as in [50], but using a discrete time Euler scheme for the underlying diffusion instead of the continuous Euler scheme in [50], would give rise to space-continuous analogs of the “forward” representations (10.21) for the quantities denoted in [65] by \tilde{Y}_0 and \tilde{Z}_0 . These quantities represent the values of discrete time approximations, convergent, with the rates stated in Proposition 10.2.3, to the option price and delta at time 0.⁷ The MC forward estimates (10.21) can thus be considered to be based on the same time-discretization of (\mathcal{E}) as the MC backward estimates (10.19), i.e. the one with the rates recalled in Proposition 10.2.3. Thus (10.19) and (10.21) only differ by space-discretization effects, to be commented upon in Sect. 10.2.4.2 below.

Remark 10.2.4 Beyond the American and game cases, the above-mentioned forward representation of the price at time 0 is still valid, but not the one for the delta. So, in all the cases of Sects. 10.4 and 10.5, the MC forward estimate $\tilde{\delta}_0$ is currently unsupported in terms of mathematical convergence, even from the mere point of view of the time-discretization which is underlying the estimates (10.21), not to mention space-discretization issues.

⁷To be precise, \tilde{Z} is the integrand of the Brownian motion in the stochastic integral representation of the discrete time approximation of Y , denoted by \overline{Y} in [65].

10.2.4.2 Space-Discretization

Building on the above time-discretization results, a complete time-space convergence analysis can be conducted by proceeding along the lines of Lemor, Gobet and Warin [185] and Lemor [184], who show how to control the cumulative regression approximation error resulting from nonlinear regressions in space that are repeatedly performed over time.

Note that in practice (see e.g. Sect. 10.3.3 below), the MC forward estimates (10.21) typically exhibit less variance than the MC backward estimates (10.19). A common interpretation is that, as mentioned in Sect. 10.2.3.1, (10.21) doesn't directly suffer from the accumulation of errors in (10.19). However, this interpretation is a bit fallacious since (10.21) actually relies on (10.19), which is used in a first stage for deriving the τ^j and θ^j in (10.20). But the pricing function u is typically not very sensitive to the optimal stopping policy, which explains why ultimately (10.21) is more accurate than (10.19).

10.3 American and Game Options

In this section we review basic results and methodologies in the cases of American and game options. Other cases of call protection will be considered in later sections.

10.3.1 No Call

We first consider the simplest no call case in which $\vartheta_1 = T$, so $\bigcup_{l>0}[\vartheta_{2l-1}, \vartheta_{2l}) \cup \{T\} = \{T\}$ and the set \mathcal{T}^ϑ of admissible call times is reduced to the singleton $\{T\}$. This corresponds to the case of a (dividend-paying) American option. Given such a claim with coupon rate c , early payoff process L and payoff at maturity ξ , we define the cumulative discounted payoff process, for $t \in [0, T]$, by

$$\int_0^t \beta_s c_s ds + \beta_t (\mathbb{1}_{t < T} L_t + \mathbb{1}_{t=T} \xi). \quad (10.22)$$

10.3.1.1 Pricing Equations

In the risk-neutral model (10.2), the discounted price process $\beta \Pi$ of the above claim is given by the Snell envelope of the cumulative discounted payoff process (10.22) (see Sect. 3.5.4 and Chap. 12). This results in the following simple form of (10.13), for $t \in [0, T]$:

$$\beta_t \Pi_t = \text{esssup}_{\tau \in \mathcal{T}_t} \mathbb{E}_t \left\{ \int_t^\tau \beta_s c_s ds + \beta_\tau (\mathbb{1}_{\{\tau < T\}} L_\tau + \mathbb{1}_{\{\tau=T\}} \xi) \right\}.$$

In this case the stochastic pricing equation (\mathcal{E}) simplifies to

$$\begin{aligned} \Pi_T &= \xi, \quad \text{and for } t \in [0, T] : \\ \begin{cases} -d\Pi_t = (c_t - \mu_t \Pi_t) dt + dA_t - \Delta_t \sigma(t, S_t) S_t dW_t \\ L_t \leq \Pi_t, \quad (\Pi_t - L_t) dA_t = 0, \end{cases} \end{aligned} \quad (10.23)$$

for a nondecreasing reflecting process $A = A^+$, continuous since there are no ϑ_{2l} in $(0, T)$ that are involved. So the price process Π is continuous as well.

By application of the results recalled in Sect. 3.5.4 (see also [112] and Chaps. 12–13), the reflected BSDE (10.23) is well-posed and the Feynman-Kac representations (10.18) holds with $X = S$ (no auxiliary process H is needed here). The pricing function $u = u(t, S)$ in this representation is the unique (continuous) viscosity solution to the deterministic pricing equation (\mathcal{VI}), which reduces to the following scalar variational inequality:

$$\begin{cases} u = \phi & \text{at } T \\ \max(\partial_t u + \mathcal{A}u + c - \mu u, \ell - u) = 0 & \text{on } [0, T) \times (0, +\infty). \end{cases} \quad (10.24)$$

10.3.1.2 Deterministic Numerical Scheme

Let us briefly recall the generic multinomial tree algorithm of Sect. 7.2.4 for solving (10.24) on a tensorized (“product”) time-space grid (t_i, S^j) , where i and j index the time and space steps in the algorithm: $u_n(j) = \phi(S^j)$ for $j = 1, \dots, m$ and, for $i = n-1, \dots, 0$ and $j = 1, \dots, m$,

$$u_i^j = \max \left(\ell(S^j), e^{-\mu_i^j h} \sum_l p_i^{j,l} (u_{i+1}^{j+l} + h c_{i+1}^{j+l}) \right). \quad (10.25)$$

The $p_i^{j,l}$ are suitable weights which can be interpreted as the conditional transition probabilities of a time-inhomogeneous Markov chain $(S_i)_{0 \leq i \leq n}$ approximating $(S_t)_{t \in [0, T]}$, so that

$$p_i^{j,l} = \mathbb{P}(S_{i+1} = S^{j+l} | S_i = S^j).$$

As we saw in Chaps. 7–8, the $p_i^{j,l}$ are typically obtained by substitution of Taylor expansions for u and its derivatives into (10.24), followed by the numerical solution of a linear system in the case of implicit schemes. The scheme is stable and converges if $p \geq 0$.

An approximation $(\delta_i^j)_{0 \leq i \leq n}^{2 \leq j \leq m-1}$ for the delta function $\Delta(t, S) = \partial u(t, S)$ at the interior points of the time-space grid can then be obtained, for all $i = 0, \dots, n$ and $j = 2, \dots, m-1$, by

$$\delta_i^j = \frac{u_i^{j+1} - u_i^{j-1}}{S^{j+1} - S^{j-1}}. \quad (10.26)$$

10.3.1.3 Simulation Pricing Schemes

For solving (10.24) by simulation, a possible procedure consists in writing a dynamic programming equation such as (10.25), but on a stochastically generated, nontensorized mesh $(S_i^j)_{0 \leq i \leq n}^{1 \leq j \leq m}$, using an appropriate discretization scheme S_i for the underlying diffusion \bar{S} (such as the Euler scheme of Sect. 10.1.1.2). We thus get the following “stochastic version” of (10.25), in which we recognize the generic simulation scheme of (10.19) particularized to the present case: $u_n^j = \phi(S_n^j)$ for $j = 1, \dots, m$ and, for $i = n - 1, \dots, 0$, and $j = 1, \dots, m$,

$$u_i^j = \max(\ell(S_i^j), e^{-\mu_i^j h} \mathbb{E}_i^j(u_{i+1} + hc_{i+1})), \quad (10.27)$$

where \mathbb{E}_i^j represents the conditional expectation given $S_i = S_i^j$. A numerical approximation $\delta = (\delta_i^j)$ of the delta function on the grid could be deduced from (u_i^j) by the formula (10.26), but convergence of this estimate has not been established, and it typically exhibits a very high variance in practice. In approximating the delta, the regression estimates in (10.19) and (10.21) provide better, convergent alternatives.

At step $i \geq 1$, the conditional expectations involved can be computed by nonlinear regression of $(u_{i+1}^j + hc_{i+1}^j)_{1 \leq j \leq m}$ and $(u_{i+1}^j(W_{i+1}^j - W_i^j))_{1 \leq j \leq m}$ against $(S_i^j)_{1 \leq j \leq m}$. For the price and delta at time 0, MC forward estimates \tilde{u}_0 and $\tilde{\delta}_0$ can be used as alternatives for u_0 and δ_0 .

10.3.2 No Protection

We now discuss the case of game options without call protection, so that $\vartheta_1 = 0$, $\vartheta_2 = T$ and $\mathcal{T}_t^\vartheta = \mathcal{T}_t$.

10.3.2.1 Pricing Equations

In this case the stochastic pricing equation (\mathcal{E}) is rewritten as

$$\begin{aligned} \Pi_T &= \xi, \quad \text{and for } t \in [0, T] : \\ \begin{cases} -d\Pi_t = (c_t - \mu_t \Pi_t) dt + dA_t - \Delta_t \sigma(t, S_t) S_t dW_t \\ L_t \leq \Pi_t \leq U_t, \quad (\Pi_t - L_t) dA_t^+ = (U_t - \Pi_t) dA_t^- = 0, \end{cases} \end{aligned} \quad (10.28)$$

to be solved for (Π, Δ, A) , where A and Π are again sought as continuous processes. Moreover, we have $\Pi_t = v(t, S_t)$, where the deterministic pricing equation (\mathcal{VI}) in the pricing function v is:

$$\begin{cases} v = \phi & \text{at } T \\ \min(\max(\partial_t v + \mathcal{A}v + c - \mu v, \ell - v), h - v) = 0 & \text{on } [0, T) \times (0, +\infty). \end{cases} \quad (10.29)$$

The American deterministic pricing algorithm (10.25) simply needs to be amended as: $v_n(j) = \phi(S^j)$ for $j = 1, \dots, m$, and, for $i = n - 1, \dots, 0$ and $j = 1, \dots, m$,

$$v_i^j = \min\left(h(S^j), \max\left(\ell(S^j), e^{-\mu_i^j h} \sum_l p_i^{j,l} (v_{i+1}^{j+l} + hc_{i+1}^{j+l})\right)\right). \quad (10.30)$$

A delta estimate $(\delta_i^j)_{0 \leq i \leq n}^{2 \leq j \leq m-1}$ can then be deduced from $(v_i^j)_{0 \leq i \leq n}^{1 \leq j \leq m}$ by the formula (10.26), with v there instead of u . Moreover, it is possible to conduct a convergence analysis of (10.30) similar to that of (10.25) in the American no-call case.

10.3.2.2 Simulation Schemes

The stochastic pricing scheme (10.27) simply needs to be amended as follows (cf. (10.19)): $v_n^j = \phi(S_n^j)$ for $j = 1, \dots, m$ and, for $i = n - 1, \dots, 0$ and $j = 1, \dots, m$,

$$v_i^j = \min\left(h_i(S_i^j), \max\left(\ell(S_i^j), e^{-\mu_i^j h} \mathbb{E}_i^j (v_{i+1} + hc_{i+1})\right)\right). \quad (10.31)$$

We also have the delta estimate δ_i^j in (10.19), with v_i^j there instead of u_i^j . For $i \geq 1$, the conditional expectations at time step i in the algorithm can be computed by nonlinear regression against $(S_i^j)_{1 \leq j \leq m}$. For the price and delta at time 0, we may use MC forward estimates as in (10.21), as an alternative to the estimates v_0 and δ_0 .

10.3.3 Numerical Experiments

The case of American options is well documented in the numerical literature (see [53, 133]). We consider here a game option. The general data of Table 10.1 are used by default in the sequel. Also, in all the numerical experiments, we use a constant time-step $t_{i+1} - t_i = h$, with:

- $h =$ six hours (four time steps per day) in the case of simulation pricing schemes for solving (\mathcal{E}) ;
- $h =$ one day in the case of deterministic schemes for solving (\mathcal{VI}) .

The superscript j referring to a generic space step index,⁸ the space-steps in the S variable are:

- $S^{j+1} - S^j = 0.5$ in the case of deterministic schemes, and

⁸Space step in the sense of a trajectory's index varying between 1 and m , in the case of simulation pricing schemes.

Table 10.1 General data

\bar{P}	\bar{N}	\bar{C}	\bar{R}	Λ	σ	r	q	λ_0	γ	m
0	100	103	0	1	0.2	0.05	0	0.02	1.2	10^4

Table 10.2 VI Values and MC backward and forward standard deviations over 50 trials, for the option prices and deltas ($S_0 = 100.55$)

	Value VI	Dev MC Bd	Dev MC Fd
Price	102.049	0.821	0.010
Delta	0.416	0.071	0.019

- cells of diameter one (segments of \mathbb{R}_+ of length one) in the case of simulation/regression methods involving a method of cells in the direction of the S variable (see Sect. 6.10.2).

Regarding the deterministic numerical schemes, fully implicit finite difference schemes are used throughout (see Sect. 8.3.2.2).

For a maturity $T = 125$ days and a nominal coupon rate $\bar{c} = 0$ in (10.12), Table 10.2 shows the standard deviation of fifty estimates of the game option price and delta obtained by changing the seed of the random generator, using the MC backward estimate on the one hand, and the MC forward estimate on the other hand. A global parametric regression basis 1, S , S^2 is used in both cases for estimating the conditional expectations (see Sect. 6.10.2). The MC forward estimates show a much lower deviation, as expected.

Table 10.3 shows the option prices and deltas computed for various S_0 by the MC forward estimates, or by the deterministic scheme (10.30) for the VI (10.29). The “%Err” in this and the following tables are the unsigned percentage relative errors of the MC prices with respect to the VI prices. So a “%Err” of “1” in the table means a relative difference of $\pm 1\%$ between the MC price and the reference VI price.

MC forward estimates are used by default henceforth. Moreover, in view of Remark 10.2.4, we will in all nontrivial cases of call protection (i.e. except in the American and game cases of Sect. 10.3) focus on prices, disregarding deltas for which convergence of the MC forward estimate $\tilde{\delta}_0$ is not established mathematically.

Computation times are an important issue in this chapter, especially with the path-dependent call protection clauses of Sect. 10.5. To make the computation time information below less contingent on the implementation, it will be stated in terms of multiples of the average time needed to solve the scalar variational inequalities of the present subsection. So a computation time of one second, henceforth in this chapter, is to be understood not in the sense of a physical second, but of the average time needed to solve a variational inequality in one space-dimension—which is less than one second on present day computers in a compiled programming language such as C++.

Table 10.4 compares, in terms of accuracy and computation time, the results for $S_0 = 100.55$ (also corresponding to the third line of Table 10.3) for various nonlinear regression schemes in S (see Sect. 6.10.2):

Table 10.3 MC versus VI prices and deltas

S_0	VI Price	%Err MC Bd	%Err MC Fd	VI delta	%Err MC Bd	%Err MC Fd
98.55	101.246	1.90	0.04	0.376	1.07	0.07
99.55	101.637	1.92	0.01	0.396	0.95	0.50
100.55	102.049	1.99	0.01	0.416	2.77	0.67
101.55	102.479	1.65	0.07	0.435	3.97	3.47

Table 10.4 MC errors and CPU times depending on the regression basis which is used for S

	%Err MC Price	%Err MC Delta	CPU Time
(1, S)	2.149	5.03	0.5 s
(1, S, S^2)	0.014	0.67	0.7 s
(1, S, S^2, S^3)	0.018	0.69	1.5 s
Cells	0.049	0.89	0.5 s

- either global regression schemes against powers of S , which require the inversion of an empirical covariance matrix at every time step in the simulation,
- or a method of cells, for which no matrix inversion is required.

Regarding the global regression schemes, the computation times increase significantly with the size of the basis, due to the matrix inversion involved. The choice of $(1, S, S^2)$ appears as the best compromise (as far as global regression schemes are concerned), for the number $m = 10^4$ of simulations and the time step $h =$ six hours for the Euler scheme which are used.

10.4 Continuously Monitored Call Protection

In this section we study the two following forms of continuously monitored call protection, given a trigger level $\bar{S} > S_0$:

- call forbidden until the first time ϑ_1 such that $S \geq \bar{S}$ and call possible for $t \geq \vartheta_1$, which corresponds to the sequence ϑ given by

$$\vartheta_1 = \inf\{t \geq 0; S_t \geq \bar{S}\} \wedge T$$

and $\vartheta_l = T$ for $l \geq 2$;

- the “intermittent analog” of the previous clause and so, essentially, call possible whenever $S \geq \bar{S}$ and forbidden otherwise. But this clause is ill-posed mathematically (see below), so that the following modified clause will be considered instead, given a second trigger level $\underline{S} < \bar{S}$: call forbidden until the first time ϑ_1 such that $S \geq \bar{S}$, then call possible until the first time $\vartheta_2 > \vartheta_1$ that $S < \underline{S}$ etc., which cor-

responds to the nondecreasing sequence ϑ of stopping times given by, for every $l > 0$:

$$\vartheta_{2l-1} = \inf\{t > \vartheta_{2l-2}; S_t \geq \bar{S}\} \wedge T, \quad \vartheta_{2l} = \inf\{t > \vartheta_{2l-1}; S_t \leq \underline{S}\} \wedge T. \quad (10.32)$$

Note that for $\underline{S} = \bar{S}$, (10.32) would typically not define an increasing sequence of stopping times, which explains why we need to set $\underline{S} < \bar{S}$.

10.4.1 Vanilla Protection

We first consider a call protection until a stopping time ϑ_1 , so that $\vartheta_2 = T$ and the effective call payoff process U_t in (\mathcal{E}) reduces to

$$U_t = \mathbb{1}_{\{t < \vartheta_1\}}\infty + \mathbb{1}_{\{t \geq \vartheta_1\}}h(S_t). \quad (10.33)$$

Since $\vartheta_2 = T$, the processes A^\pm and Π , in a solution to (\mathcal{E}) , are again continuous in this case.

The price process Π of a convertible bond with call protection before ϑ_1 coincides on $[\vartheta_1, T]$ with the no protection price process of Sect. 10.3.2 (see [40]). The only remaining issue is thus the “protection pricing problem”, in the sense of the characterization and computation of the price process on the no-call time interval $[0, \vartheta_1]$.

Remark 10.4.1 In the case of a constant lifting time of a call protection $\vartheta_1 = T_1 \in [0, T)$, the protection pricing problem reduces to the no call pricing problem of Sect. 10.3.1.

In this subsection we will consider the case of continuously-monitored call protection, dubbed vanilla protection, corresponding to

$$\vartheta_1 = \inf\{t \in \mathbb{R}_+; S_t \geq \bar{S}\} \wedge T \quad (10.34)$$

for some trigger level $\bar{S} > S_0$. This case is studied in detail in [40] (see also Chaps. 12 and 13).

10.4.1.1 Deterministic Pricing Equation

The following equation for the protection pricing function⁹ $u = u(t, S)$ on the domain $[0, T] \times (0, \bar{S}]$ is established in [40]:

$$\begin{cases} u = v & \text{on } (\{T\} \times (0, \bar{S})) \cup ([0, T] \times \{\bar{S}\}) \\ \max(\partial_t u + \mathcal{A}u + c - \mu u, \ell - u) = 0 & \text{on } [0, T] \times (0, \bar{S}). \end{cases} \quad (10.35)$$

⁹Pricing function which applies before ϑ_1 .

Here v is the no call protection pricing function of Sect. 10.3.2.1, which was characterized there as the solution to the deterministic pricing equation (10.29). Moreover, provided the functions u and v are sufficiently regular for Itô formulas to be applicable or in a weak sense otherwise, for $t \in [0, T]$ we have:

$$\Delta_t = \mathbb{1}_{\{t \leq \vartheta_1\}} \partial_S u(t, S_t) + \mathbb{1}_{\{t > \vartheta_1\}} \partial_S v(t, S_t). \quad (10.36)$$

Knowing an approximation (v_i^j) of v , computed e.g. by the deterministic scheme (10.30) of Sect. 10.3.2.1, the Cauchy-Dirichlet problem (10.35) can be solved by standard finite difference deterministic numerical schemes, such as

$$u_i^j = \max \left(\ell(S^j), e^{-\mu_i^j h} \sum_l p_i^{j,l} (u_{i+1}^{j+l} + hc_{i+1}^{j+l}) \right) \quad (10.37)$$

at the grid points interior to the domain $[0, T] \times (0, \bar{S}]$, with a Dirichlet boundary condition $u = v$ at the grid points of the time-space boundary “ $\{t = T\} \cup \{S = \bar{S}\}$ ”. Convergence results for this scheme can be found in Sect. 13.2.3. Note, however, that these can only be considered as partial results, since one only gets the convergence of the scheme for u on $[0, T] \times (0, \bar{S}]$ provided it converges on $[0, T] \times \{\bar{S}\}$, for which no explicit condition is given. Moreover, the convergence analysis of Sect. 13.2.3 is conducted under the assumption that the true value for v on $[0, T] \times \{\bar{S}\}$ is substituted into the approximation scheme for u , whereas in practice one only uses an approximation (v_i^j) of v . Finally, this analysis only yields convergence, not rates of convergence.

10.4.1.2 Simulation Scheme

Given a stochastically generated mesh $(S_i^j)_{0 \leq i \leq n}^{1 \leq j \leq m}$ and, for every $j = 1, \dots, m$, setting (cf. (10.34))

$$\vartheta_1^j = \inf \{0 \leq i \leq n; S_i^j \geq \bar{S}\} \wedge n,$$

a simulation algorithm for estimating the $u(t_i, S_i^j)$ is:

$u_n = v_n = \phi$ and, for $i = n - 1, \dots, 0$ and $j = 1, \dots, m$:

- $v_i^j = \min(h_i(S_i^j), \max(\ell(S_i^j), e^{-\mu_i^j h} \mathbb{E}_i^j (v_{i+1} + hc_{i+1})))$;
- if $i \geq \vartheta_1^j$, $u_i^j = v_i^j$, else

$$u_i^j = \max \left(\ell(S_i^j), e^{-\mu_i^j h} \mathbb{E}_i^j (u_{i+1} + hc_{i+1}) \right). \quad (10.38)$$

Thus (v_i^j) here is nothing but the no call protection price of Sect. 10.3.2.2, estimated by simulation as in (10.31). For $i \geq 1$ the conditional expectations in (10.38) can be computed by nonlinear regression of $(u_{i+1}^j + hc_{i+1}^j)_{j \in \mathcal{Q}_i}$ against $(S_i^j)_{j \in \mathcal{Q}_i}$, where \mathcal{Q}_i denotes the subset of the trajectories j for which $\{i < \vartheta_1^j\}$.

Convergence Analysis This case falls outside the scope of [65], where the more realistic setup of a discretely monitored call protection, also considered in Sect. 10.5 below, is favored in the analysis. It is likely that convergence rates for the above scheme for the solution (Π, Δ, A) of the related BSDE (\mathcal{E}) could be obtained by combining the reflected BSDE convergence techniques of Chassagneux [64] with approximation for the call time by the method of Bouchard and Menozzi [52]. The related convergence results regarding Π would then not suffer from the same limitations as those mentioned in Sect. 10.4.1.1 regarding the deterministic approximation for u . Moreover, in this way we would also get convergence and convergence rates with respect to the Δ -component of a solution to (\mathcal{E}) .

Time-0 Price We can recover, from the protection pricing function u above, the following estimates of the protection put region and of the optimal protection put policy:

$$\tilde{\mathcal{E}}_p = \{(i, S_i^j); u_i^j = \ell(S_i^j)\}, \quad \tilde{\tau}^j = \inf\{0 \leq i \leq \vartheta_1^j; S_i^j \in \tilde{\mathcal{E}}_p\} \wedge n. \quad (10.39)$$

We then have the following MC forward estimate for the protection price at time 0, with $\tilde{\nu}^j = \tilde{\tau}^j \wedge \vartheta_1^j$:

$$\tilde{u}_0 = \frac{1}{m} \sum_{j=1}^m \left\{ h \sum_{i=1}^{\tilde{\nu}^j} \beta_i^j c_i(S_i^j) + \beta_{\tilde{\nu}^j}^j (\mathbb{1}_{\{\tilde{\tau}^j < \vartheta_1^j\}} \ell_{\tilde{\tau}^j}(S_{\tilde{\tau}^j}^j) + \mathbb{1}_{\{\vartheta_1^j \leq \tilde{\tau}^j\}} v_{\vartheta_1^j}^j) \right\}. \quad (10.40)$$

It would be likewise possible to write a forward estimate $\tilde{\delta}_0$ of the time-0 protection delta $\partial_S u(0, S_0)$ (cf. (10.21)). However, as explained in Remark 10.2.4, convergence of the delta is not guaranteed in this case.

10.4.2 Intermittent Vanilla Protection

We now consider the “intermittent analog” of the “until” clause of Sect. 10.4.1. The nondecreasing sequence ϑ of stopping times is thus defined by (10.32) for activating and deactivating trigger levels \bar{S} and \underline{S} , with $\underline{S} < \bar{S}$.

10.4.2.1 Deterministic Pricing Equation

We are in the Markovian case of Sect. 10.2.2 for the call protection set $K = \{0\}$ and the Boolean-valued call protection process H_t defined by: $H_0 = 0$ and

- H jumping from 0 to 1 at the ϑ_l such that $S_{\vartheta_l} \geq \bar{S}$ and $H_{\vartheta_l-} = 0$;
- H jumping from 1 to 0 at the ϑ_l such that $S_{\vartheta_l} \leq \underline{S}$ and $H_{\vartheta_l-} = 1$.

In particular we have $H_t = 1$ and hence call is possible on $\{(t, \omega); S_t \geq \bar{S}\}$, and $H_t = 0$ and hence call is forbidden on $\{(t, \omega); S_t \leq \underline{S}\}$. The pair $X = (S, H)$ is a

Markov process. It is thus expected that

$$\Pi_t = u(t, S_t, H_t) \quad \text{for a pricing function } u = u(t, S, k) = u_k(t, S) \text{ with } k \in \{0, 1\}. \quad (10.41)$$

Note however that the call protection ϑ is monitored continuously in time, not only at the dates of a discrete time grid \mathfrak{T} . Continuously monitored intermittent call protection clauses fall outside the scope of the results of Sect. 14.2. Yet, in the present case, existence of a solution to the related pricing RIBSDE (\mathcal{E}) follows by application of the results of Peng and Xu [222]. More precisely, existence follows from an immediate extension of these results to the case of an $\mathbb{R} \cup \{+\infty\}$ -valued upper barrier U . Indeed, the results of Peng and Xu, even if stated for real-valued barriers, only use the square integrability of the random variable $\sup_{t \in [0, T]} U_t^-$, a condition which is satisfied in the present case where $U_t^- = h(S_t)^-$; cf. Remark 14.2.8.

Building upon this existence result, we can follow the methodology developed in Part V (see also [65]) to prove (10.41). The deterministic pricing equation (\mathcal{VI}) is written:

$$\begin{cases} u_0 = u_1 = \phi & \text{at } T \\ u_0(t, \bar{S}) = u_1(t, \bar{S}) \\ u_1(t, \underline{S}) = \min(u_0(t, \underline{S}), h(t, \underline{S})) \\ \max(\partial_t u_0 + \mathcal{A}u_0 + c - \mu u_0, \ell - u_0) = 0 & \text{on } [0, T] \times (0, \bar{S}) \\ \min(\max(\partial_t u_1 + \mathcal{A}u_1 + c - \mu u_1, \ell - u_1), h - u_1) = 0 & \text{on } [0, T] \times (\underline{S}, +\infty). \end{cases} \quad (10.42)$$

We thus get a system of two equations in $u = (u_0, u_1)$, with u_0 and u_1 respectively defined on $[0, T] \times (0, \bar{S})$ and $[0, T] \times (\underline{S}, +\infty)$. This system can be solved practically by finite differences of the following form on a time-space grid (see Chap. 8): $u_{0,n}^j = u_{0,n}^j = \phi(S_n^j)$ for $j = 1, \dots, m$ and, for $i = n-1, \dots, 0$ and $j = 1, \dots, m$:

$$\begin{cases} u_{0,i}^j = \max\left(\ell(S^j), e^{-\mu_i^j h} \sum_l p_i^{j,l} (u_{0,i+1}^{j+l} + hc_{i+1}^{j+l})\right) \\ u_{1,i}^j = \min\left(h(S^j), \max\left(\ell(S^j), e^{-\mu_i^j h} \sum_l p_i^{j,l} (u_{1,i+1}^{j+l} + hc_{i+1}^{j+l})\right)\right), \end{cases} \quad (10.43)$$

where on the right-hand side $u_{0,i+1}^{j+l}$ is meant as $u_{1,i+1}^{j+l}$ for $S^{j+l} \geq \bar{S}$, and $u_{1,i+1}^{j+l}$ is meant as $\min(u_{0,i+1}^{j+l}, h_{i+1}^{j+l})$ for $S^{j+l} \leq \underline{S}$.

Yet the analytic characterization of the pricing function $u = (u_0, u_1)$ as the unique solution to (10.43) in some sense (continuous viscosity solution, presumably), and the related deterministic approximation results, are not yet established. This is due to the absence of stability results so far, beyond existence and uniqueness, for the stochastic pricing equation (\mathcal{E}) in this case. More precisely, the absence of a priori error estimates (see Sect. 12.1.2.2) makes it difficult to establish the continuity of the pricing function $u = (u_0, u_1)$, which would be a prerequisite to proving that u satisfies (10.42) in the continuous viscosity sense.

Table 10.5 Vanilla protection ($\bar{S} = 103$)

S_0	100.55	101.55	102.55	103.55
MC price	103.841	103.724	103.713	103.55
VI price	103.874	103.785	103.693	103.55
%Err	0.032	0.059	0.019	0

Table 10.6 Same as Table 10.5, but for $\bar{S} = 120$

S_0	100.55	101.55	102.55	103.55
MC price	110.082	110.819	111.351	111.809
VI price	110.324	110.896	111.488	112.099
%Err	0.22	0.069	0.12	0.26

10.4.2.2 Simulation Pricing Scheme

The related simulation pricing algorithm is (10.19), particularized to the setup of Sect. 10.4.2.1. We stress again that convergence rates for this scheme fall outside the scope of [65], since we are dealing here with a continuously monitored intermittent call protection.

10.4.3 Numerical Experiments

10.4.3.1 Vanilla protection

Using the general data of Table 10.1 in Sect. 10.3.3 with a maturity $T = 180$ days and a nominal discrete coupon¹⁰ of $\bar{c} = 1.2$ every month, Tables 10.5 and 10.6 show the vanilla protection convertible bond prices computed in two ways:

- i. by the simulation pricing scheme of Sect. 10.4.1.2, where the conditional expectations are estimated by a method of cells in S ;
- ii. by the deterministic numerical scheme of Sect. 10.4.1.1.

The accuracy of the simulation pricing scheme is satisfactory. In the last column of Table 10.5 we see that both schemes compute the exact bond value, which is equal to S_0 in the case $S_0 = 103.55 > 103 = \bar{S}$ (see [40]).

10.4.3.2 Intermittent Vanilla protection

We now consider convertible bonds with intermittent call protection as in Sect. 10.4.2 that we evaluate by the deterministic scheme of Sect. 10.4.2.1 or, alternatively, by

¹⁰See Remark 10.1.1.

Table 10.7 Intermittent Vanilla protection

S_0	92	93	94	102	103	104
VI price	104.128	104.147	104.154	103.716	103	104
MC %Err	0.094	0.051	0.012	0.001	0	0

the simulation method of Sect. 10.4.2.2. The results are presented in Table 10.7 for the same data as above and $\bar{S} = 103$, $\underline{S} = 97$. In spite of the above-mentioned lack of theoretical convergence results, the numerical results are quite accurate. As already observed in Table 10.5, both schemes give the exact value, which is equal to the stock price S_0 , for $S_0 \geq \bar{S} = 103$.

10.5 Discretely Monitored Call Protection

In the previous section we studied two forms of continuously monitored call protection. In the intermittent case, mathematical difficulties arose. However, in practice call protection is monitored in discrete time, e.g. reexamined at the end of each trading day, rather than in continuous time. As we will now see, it turns out that it is not only closer to actual contracts, but also easier from the mathematical point of view, to work with a discretely monitored call protection. The path dependence of certain clauses can then be computationally challenging, but this is where the simulation pricing schemes reveal their interest, as compared with deterministic schemes that are ruled out by the curse of dimensionality.

Let $\mathfrak{T} = \{T_0 = 0 < T_1, \dots, < T_N = T\}$ denote an increasing sequence of fixed monitoring times, so that the ϑ_l are now \mathfrak{T} -valued stopping times. Well-posedness of (\mathcal{E}) and the Feynman-Kac representation (10.18) for the solution to (\mathcal{E}) then hold by application of the results of Sect. 14.2 (see also [65]). As we will see in this section, the deterministic pricing equation (\mathcal{VI}) in $u = u(t, x) = u_k(t, S)$ assumes the form of a system of variational inequalities indexed by k , to be solved in the largest viscosity subsolution sense by the upper semi-continuous envelope \bar{u} (in the sense made precise in [65]) of the pricing function u .

Discontinuities of the pricing function u will typically arise at T_l in connection with the discontinuities of the pricing process Π at the even times ϑ_{2l} in $(0, T)$. In spite of this technicality, the deterministic pricing equation (\mathcal{VI}) can, in principle, be solved by standard deterministic pricing schemes such as finite difference θ -schemes (see Chap. 8), which are shown to be convergent in [65] (under a suitable refinement of the scheme to deal with the discontinuities of the function u). But in this case such deterministic pricing schemes are a bit theoretical, since the potentially highly path-dependent nature of the call protection can lead to huge, impractical systems of variational inequalities.

The particular form of the deterministic pricing equation (\mathcal{VI}) depends on the nature of the call protection process H ; in the next subsections (\mathcal{VI}) will be specified in two different cases.

10.5.1 “ l Last” Protection

Given $\bar{S} > S_0$ and an integer l , we now consider the “ l last” call protection clause, under which call is possible at a given time t if and only if S was $\geq \bar{S}$ at the last l monitoring dates $\leq t$. In this subsection we accordingly let H_t represent the minimum of l and the number of consecutive monitoring dates T_I with $S_{T_I} \geq \bar{S}$, from time t backwards. In particular, at any given time t , we have $H_t = 0$ if S was smaller than \bar{S} at the last monitoring date $\leq t$. The “ l last” call protection clause corresponds to this protection process H and to the protection set $K = \{0, \dots, l - 1\}$.

10.5.1.1 Deterministic Pricing Equation

By application of the results of Sect. 14.2 (see also [65]), we then have that $\Pi_t = u(t, S_t, H_t)$ on $[0, T]$, for a pricing function $u = u_k(t, S)$ with $k = 0, \dots, l$; the restriction of any u_k to every set $[T_{I-1}, T_I] \times [0, +\infty)$ is continuous and the limit

$$u_k(T_I-, S) = \lim_{(t,x) \rightarrow (T_I, S) \text{ with } t < T_I} u_k(t, x) \quad (10.44)$$

exists for every $k = 0, \dots, l$, $I \geq 1$ and $S \neq \bar{S}$. The deterministic pricing equation (\mathcal{VI}) now assumes the form of the following Cauchy cascade of variational inequalities:

For I decreasing from N to 1:

- at $t = T_I$, for $k = 0, \dots, l$,

$$u_k(T_I-, S) = \begin{cases} u_{k+1}(T_I, S), & \text{or } u_k(T_I, S) \text{ if } k = l, \text{ for } \{S > \bar{S}\}, \\ u_0(T_I, S), & \text{or } \min(u_0(T_I, S), h(T_I, S)) \\ & \text{if } k = l, \text{ for } \{S < \bar{S}\}, \end{cases} \quad (10.45)$$

or, if $I = N$, $u_k(T_I-, S) = \phi(S)$ for $S > 0$;

- on the time interval $[T_{I-1}, T_I]$,

$$\begin{aligned} \max(\partial_t u_k + \mathcal{A} u_k + c - \mu u_k, \ell - u_k) &= 0, \quad k = 0, \dots, l - 1 \\ \min(\max(\partial_t u_l + \mathcal{A} u_l + c - \mu u_l, \ell - u_l), h - u_l) &= 0. \end{aligned} \quad (10.46)$$

As shown in [65], the upper semi-continuous envelope \bar{u} of the pricing function u can then be characterized as the largest viscosity subsolution to (\mathcal{VI}) ; the Cauchy cascade (\mathcal{VI}) can be solved numerically by finite difference numerical schemes that converge under a suitable refinement of the scheme so as to deal with the singularities of u .

10.5.1.2 Simulation Scheme

Let there be given a time mesh $\mathbf{t} = (t_i)_{0 \leq i \leq n}$ refining the protection monitoring grid \mathfrak{T} . To solve the above Cauchy cascade of VIs by simulation, we first generate

a stochastic grid $(S_i^j, H_i^j)_{0 \leq i \leq n}^{1 \leq j \leq m}$ by an Euler scheme for S on t , using past values of S to fill H .

Example 10.5.1 With $r = q = \Lambda = 0$, $h =$ six hours (four time-steps per day) and $T_{I+1} - T_I =$ one day (daily monitored call protection):

- we simulate, starting from S_0 given, $S_{6h} = S_0(1 + \sigma\sqrt{h}\varepsilon_1)$, $S_{12h} = S_{6h}(1 + \sigma\sqrt{h}\varepsilon_2)$, S_{18h} , S_{24h} , S_{30h}, \dots, S_T , for independent Gaussian draws ε_i ;
- whenever t_i coincides with one of the T_I (i.e. every fourth i), we update the variable H , so that: $H_0 = 0$, $H_{6h} = H_0$, $H_{12h} = H_0$, $H_{18h} = H_0$, $H_{1day} = 1$ if $S_{1day} \geq \bar{S}$, otherwise $H_{1day} = 0$, and so on (with H capped at l) until we get H_T ;

We repeat this m times independently, which yields m trajectories $(S_i^j, H_i^j)_{0 \leq i \leq n}^{1 \leq j \leq m}$.

We then run, on the stochastic grid thus generated, the generic simulation pricing scheme of Sect. 10.2.3, particularized to the context of this subsection and referred to henceforth as MC_l .

10.5.2 “ l Out of the Last d ” Protection

Given a further integer $d \geq l$, we now consider the “ l out of the last d ” call protection clause under which call is possible at a given time t whenever $S \geq \bar{S}$ on at least l among the last d monitoring dates $\leq t$. Now let $H_t = (H_t^p)_{1 \leq p \leq d}$ accordingly represent the vector of the indicator functions of the events $S_{T_p} \geq \bar{S}$ at the last d monitoring dates preceding time t . The “last d ” call protection clause corresponds to this new process H , and to the protection set $K = \{k = (k_p)_{1 \leq p \leq d} \in \{0, 1\}^d; |k| < l\}$, with $|k| := \sum_{1 \leq p \leq d} k_p$.

10.5.2.1 Deterministic Pricing Equation

By application of the results of Sect. 14.2 (see also Sect. 4.3.3), we then have $\Pi_t = u(t, S_t, H_t)$ on $[0, T]$ for a pricing function $u = u_k(t, S)$ over $[0, T] \times (0, +\infty) \times \{0, 1\}^d$; the restriction of any u_k to any of the sets $[T_{I-1}, T_I) \times [0, +\infty)$ is continuous and the limit $u_k(T_I-, S)$ exists in the sense of (10.44), for every $k \in \{0, 1\}^d$, $I \geq 1$ and $S \neq \bar{S}$. The deterministic pricing equation (VI) assumes the form of the following Cauchy cascade of variational inequalities, with $k_+ = k_+(k, S) = (k_1, \dots, k_{d-1}, \mathbb{1}_{S \geq \bar{S}})$ in the jump condition (10.47):

for I decreasing from N to 1:

- at $t = T_I$, for every $k \in \{0, 1\}^d$,

$$u_k(T_I-, S) = u_{k_+}(T_I, S) \quad \text{or} \quad \min(u_{k_+}(T_I, S), h(T_I, S)) \quad (10.47)$$

if $k \notin K$ and $k_+ \in K$, for every $S \neq \bar{S}$,

or, if $I = N$, $u_k(T_I-, S) = \phi(S)$ for $S > 0$;

- on the time interval $[T_{I-1}, T_I]$, for every $k \in \{0, 1\}^d$,

$$\max(\partial_t u_k + \mathcal{A}u_k + c - \mu u_k, \ell - u_k) = 0, \quad k \in K$$

$$\min(\max(\partial_t u_k + \mathcal{A}u_k + c - \mu u_k, \ell - u_k), h - u_k) = 0, \quad k \notin K.$$

Again (see [65]), the upper semi-continuous envelope \bar{u} of the pricing function u can be characterized as the largest viscosity subsolution to (\mathcal{VI}) . In theory, u can be computed numerically by rather standard deterministic numerical schemes (provided appropriate care is taken of the singularities of u), but (\mathcal{VI}) is here a system of 2^d equations, which precludes the use of this scheme for d greater than, say, ten or so.

10.5.2.2 Simulation Scheme

To solve this problem by simulation, given a time grid $(t_i)_{0 \leq i \leq n}$ refining the protection monitoring grid $\mathfrak{T} = (T_I)$, we generate a stochastic grid $(S_i^j, H_i^j)_{0 \leq i \leq n}^{1 \leq j \leq m}$, using past values of S to fill H .

Example 10.5.2 In the setup of Example 10.5.1 and for $d = 30$:

- we simulate the S_i as in Example 10.5.1;
- whenever t_i coincides with one of the T_I (i.e. every fourth i), we update the vector H thus: $H_0 = (0, \dots, 0)$ is the null element of $\{0, 1\}^{30}$, $H_{6 \text{ h}} = H_0$, $H_{12 \text{ h}} = H_0$, $H_{18 \text{ h}} = H_0$, $H_{1 \text{ day}} = (\overline{H}_0, \mathbb{1}_{S_1 \text{ day} \geq \overline{S}})$, where \overline{H}_0 is the vector composed of the first 29 components (zeros) of H_0 , $H_{30 \text{ h}} = H_{1 \text{ day}}$, and so on until we get H_T ;

We repeat this m times independently, which yields m trajectories $(S_i^j, H_i^j)_{0 \leq i \leq n}^{1 \leq j \leq m}$.

Then, on the stochastic grid thus generated, we run the generic simulation pricing scheme of Sect. 10.2.3, which we refer to henceforth as $\text{MC}_{l,d}$. Note that H_i^j is now given as a vector of d Boolean variables instead of an integer in $\{0, \dots, l\}$ in the MC_l scheme of Sect. 10.5.1.2. Regarding the nonlinear regressions involved, in the present setting, there are m simulated trajectories of X for 2^d cells, corresponding to the 2^d possible states of the Boolean vector k . For large¹¹ values of d , values of k for which the subset Ω_i^k of the indices j such that $H_i^j = k$ is empty or too small for the related nonlinear regressions to be tractable (see Sect. 10.2.3) will be the rule rather than the exception. However, this is in a sense the power of the simulation approach, which automatically selects the most likely states of the vector k , given an initial condition (S_0, H_0) , as opposed to the deterministic scheme, which loops over all the possible states of k .

¹¹Large but typical, e.g. $d = 30$.

Table 10.8 %Err MC no protection versus MC_d for various “ $l = 0$ out of the last d ” cases

S_0	$d = 1$	5	10	20	30
98.55	0.04	0.04	0.04	0.04	0.04
99.55	0.02	0.02	0.02	0.02	0.02
100.55	0.05	0.05	0.05	0.05	0.05
101.55	0.07	0.07	0.07	0.07	0.07

Table 10.9 %Err MC no protection versus $MC_{0,d}$ for various “ $l = 0$ out of the last d ” cases

S_0	$d = 1$	5	10	20	30
98.55	0.04	0.04	0.04	0.04	0.04
99.55	0.02	0.02	0.01	0.02	0.02
100.55	0.05	0.05	0.01	0.05	0.05
101.55	0.07	0.07	0.08	0.07	0.07

Table 10.10 $VI_{0,d}$ versus MC_d average computation times corresponding to Tables 10.8 and 10.9

d	1	5	10	20	30
$VI_{0,d}$	1.0 s	16.1 s	465.0 s	–	–
MC_d	0.6 s	0.6 s	0.7 s	0.7 s	0.7 s
$MC_{0,d}$	0.5 s	0.6 s	0.9 s	1.4 s	1.9 s

10.5.3 Numerical Experiments

10.5.3.1 No Protection

In the $l = 0$ case, both the “ l last” and the “ l out of the last d ” clauses reduce to the no protection game case of Sect. 10.3.2. For the data of Sect. 10.3.3, Tables 10.8 and 10.9 display the accuracy of the simulation/regression pricing schemes of Sects. 10.5.1.2 and 10.5.2.2 (MC_d and $MC_{0,d}$ in the tables), validated by the deterministic scheme (10.30). Indeed, in the situation of this subsubsection, the deterministic scheme (10.30) and the ones of Sects. 10.5.1.1 or 10.5.2.1 produce the same numbers. But the corresponding computation time is independent of d for (10.30), linear in d for the scheme of Sect. 10.5.1.1 and exponential in d for the one of Sect. 10.5.2.1.

Table 10.10 displays the average computation times for the MC results of Tables 10.8 and 10.9, versus those corresponding to numerical solution by the VI cascade scheme of Sect. 10.5.2.1—at least for $d \leq 10$, since for greater values of d the VI CPU times become prohibitively long. In the “ l out of the last d ” cases with $0 < l < d$, for which the only available deterministic scheme could be the one of Sect. 10.5.2.1, this scheme is in fact ruled out by the curse of dimensionality (unless d is very small), and the simulation scheme of Sect. 10.5.2.2 is the only tractable alternative.

Table 10.11 “ l last” protection ($S_0 = 100$)

l	1	5	10	20	30
VI _{l} price	103.91	105.10	106.03	107.22	108.01
MC _{l} %Err	0.04	0.16	0.47	0.88	1.34
MC _{l,l} %Err	0.04	0.15	0.03	0.04	0.24

Table 10.12 “ l last” protection ($S_0 = 90$)

l	1	5	10	20	30
VI _{l} price	104.07	104.50	104.81	105.17	105.37
MC _{l} %Err	0.04	0.20	0.32	0.40	0.60
MC _{l,l} %Err	0.098	0.087	0.066	0.007	0.037

Table 10.13 Average CPU times corresponding to the computations of Tables 10.11 and 10.12

l	1	5	10	20	30
VI _{l}	1.5 s	4.0 s	7.0 s	13.1 s	19.2 s
MC _{l}	1.3 s	1.4 s	1.4 s	1.6 s	1.7 s
MC _{l,l}	1.0 s	1.5 s	2.0 s	3.4 s	4.5 s

10.5.3.2 “ l Last” Protection

We now consider bonds with path-dependent call protection, as in Sect. 10.5.1, that we evaluate by the deterministic scheme of Sect. 10.5.1.1 or, alternatively, by one of the following simulation schemes:

- MC _{l} The “ l last” simulation scheme of Sect. 10.5.1.2, using methods of cells for the computations of the conditional expectations, or
- MC _{l,l} The “ l out of the last d ” simulation scheme of Sect. 10.5.2.2 with $d = l$ there, using methods of cells again for the computation of the conditional expectations.

The results are presented in Table 10.11 for $S_0 = 100$ and in Table 10.12 for $S_0 = 90$, using the general data of Sect. 10.4.3. Both schemes are quite accurate in practice. In these experiments the highly path-dependent, “($l + 1$)-dimensional” scheme MC _{l,l} , is in fact more accurate, even for large l , than the path-dependent, “bivariate” scheme MC _{l} . Here “($l + 1$)-dimensional” refers to the stock S plus l Boolean variables, and “bivariate” refers to the stock S plus one integer varying in $\{0, \dots, l\}$. Table 10.13 gives the average computation times corresponding to Tables 10.11 and 10.12.

10.5.3.3 “ l Out of the Last d ” Protection

We finally price bonds with highly path-dependent call protection as in Sect. 10.5.2. The bonds are evaluated by the deterministic scheme of Sect. 10.5.2.1 or, alternatively, by the “ l out of the last d ” Monte Carlo scheme of Sect. 10.5.2.2, using three alternative methods for estimating the conditional expectations involved in the simulation scheme, based on trajectories of the process $X = (S, H)$ simulated as explained in Sect. 10.5.2.2:

- $\text{MC}_{l,d}$ conditional expectations computed by a method of cells in (S, H) as described in Sect. 10.5.2.2;
- MC_d conditional expectations computed by a method of cells in $(S, |H|)$, where $|H|$ is the number of entries equal to one in H ;
- $\text{MC}_{l,d}^\sharp$ conditional expectations computed by a method of cells in $(S, |H|^\sharp)$, where $|H|^\sharp$ is defined as the number of entries equal to one in H , starting from the $(l - |H|)$ th zero in H .

Example 10.5.3 Assuming e.g. $d = 10, l = 8$:

- if $H = (1, 1, 1, 1, \mathbf{0}, 1, 1, 1, 0, 0)$, then $l - |H| = 8 - 7 = 1$ and $|H|^\sharp = 3$ (the number of ones on the right of the first zero, in bold in H),
- if $H = (1, 1, 1, 0, 1, 1, \mathbf{0}, 0, 0)$, then $l - |H| = 8 - 6 = 2$ and $|H|^\sharp = 0$ (the number of ones on the right of the second zero, in bold in H).

In the general “ l out of the last d ” case the MC_d -algorithm is of course biased. So too is the $\text{MC}_{l,d}^\sharp$ -algorithm, but the latter can be thought of as a good approximate algorithm based on the “good regressor” $|H|^\sharp$ for estimating highly path-dependent conditional expectations. The rationale for preferring $|H|^\sharp$ to $|H|$ as a regressor, and therefore $\text{MC}_{l,d}^\sharp$ to MC_d as a proxy to $\text{MC}_{l,d}$, is that in the “ l out of the last d ” case, the entries of H preceding its $(l - |H|)$ th zero are of little relevance to the price, since these entries will necessarily be erased by new ones before the bond may become callable. In the $l = d$ case, an “ l out of the last d ” call protection reduces to an “ l last” call protection, $|H|^\sharp = |H|$ and $\text{MC}_{l,d}^\sharp$ reduces to MC_d .

The interest of the $\text{MC}_{l,d}^\sharp$ -algorithm with respect to the $\text{MC}_{l,d}$ -algorithm is that it is faster—not exponentially but by some factor, since both schemes are simulation schemes, only polynomially affected by the dimension of the state space. Essentially the complexity of the $\text{MC}_{l,d}^\sharp$ -algorithm is the same as that of the MC_d -algorithm (see e.g. Table 10.19 giving the computation times of Table 10.16, or Table 10.10 above), so we do not give more comprehensive computation time reports in this regard.

Results for $S_0 = 100$ or 90 and $d = 5$ or 10 are presented in Tables 10.14–10.17. For larger values of d , Table 10.18 compares $\text{MC}_{l,d}$ and $\text{MC}_{l,d}^\sharp$ in terms of standard deviations over 50 trials corresponding to different seeds of the random generator, and of the relative difference (denoted in the table “%Err^b”) between the $\text{MC}_{l,d}$ and the $\text{MC}_{l,d}^\sharp$ 50 trials average prices. The $\text{MC}_{l,d}^\sharp$ -algorithm appears to be reasonably

Table 10.14 “ l out of the last d ” protection
($d = 5, S_0 = 100$)

l	2	3	5
VI $_{l,d}$ price	104.07	104.43	105.10
MC $_d$ %Err	0.89	1.93	2.33
MC $_{l,d}$ %Err	0.21	0.15	0.15
MC $_{l,d}^\sharp$ %Err	0.19	0.23	0.18

Table 10.15 “ l out of the last d ” protection
($d = 5, S_0 = 90$)

l	2	3	5
VI $_{l,d}$ price	104.10	104.25	104.50
MC $_d$ %Err	0.33	0.61	0.83
MC $_{l,d}$ %Err	0.01	0.01	0.10
MC $_{l,d}^\sharp$ %Err	0.155	0.108	0.034

Table 10.16 “ l out of the last d ” protection
($d = 10, S_0 = 100$)

l	2	5	10
VI $_{l,d}$ price	104.27	104.87	106.03
MC $_d$ %Err	0.91	2.43	3.01
MC $_{l,d}$ %Err	0.01	0.15	0.03
MC $_{l,d}^\sharp$ %Err	0.04	0.26	0.38

Table 10.17 “ l out of the last d ” protection
($d = 10, S_0 = 90$)

l	2	5	10
VI $_{l,d}$ price	104.24	104.41	104.82
MC $_d$ %Err	0.48	1.23	1.43
MC $_{l,d}$ %Err	0.01	0.02	0.07
MC $_{l,d}^\sharp$ %Err	0.05	0.09	0.32

Table 10.18 Standard deviations over 50 trials and %Err $^\flat$ ($d = 30, S_0 = 102.55$)

	1	5	10	20	30
Dev MC $_{l,d}$	0.056	0.061	0.086	0.152	
Dev MC $_{l,d}^\sharp$	0.060	0.069	0.092	0.175	
%Err $^\flat$	0.09	0.24	0.72	1.06	

Table 10.19 MC computation times corresponding to Table 10.16

	1	2	5	10
MC $_{l,d}$	2.0 s	2.1 s	2.0 s	
MC $_d$	1.4 s	1.4 s	1.4 s	
MC $_{l,d}^\sharp$	1.5 s	1.6 s	1.5 s	

accurate, and significantly better than the MC_d -algorithm, which is too severely biased in the general “ l out of the last d ” case.

10.5.4 Conclusions

The numerical results of this chapter illustrate the good performances of the simulation regression scheme for pricing convertible bonds with highly path-dependent call protection. Given the theoretical convergence results of [65] recalled in Sect. 10.2.4, we end up with an approach to this problem that is both practical and mathematically justified.

More generally, this chapter is an illustration of the power of simulation approaches, which automatically select and loop over *the most likely states* X_i of a potentially high-dimensional, but also often very degenerate, factor process X , *given an initial condition* $X_0 = x$. By contrast, deterministic schemes loop over all the possible states of X_i regardless of their likelihood (except for the choice of the computational domain, typically centered around x). Simulation schemes thus compute “where there is light”. By contrast, deterministic schemes compute *everywhere*, also in “obscure”, useless regions of the state space. In the context of path-dependent payoffs with a high-dimensional but very degenerate factor process, a simulation scheme thus exploits the degeneracy of a factor process, whereas this path-dependence makes the deterministic scheme impractical.

Section 10.5.3.3 illustrates another interesting feature of simulation/regression schemes, i.e. the possibility of working with an approximate, low-dimensional regressor.

Chapter 11

Simulation/Regression Pricing Schemes in Pure Jump Setups

In this chapter we consider simulation/regression pricing schemes in the context of pure jump models. Jumps were already present in the previous chapter via the indicator processes of the call protection triggering events, but the (pre-default) model filtration was Brownian, whereas we will now deal with the filtration generated by the integer-valued random measure of a pure jump process (see Sect. 12.1). The problem is knowing whether simulation/regression pricing schemes can be fruitfully used to face the curse of dimensionality that arises in the context of combinatorially exploding continuous-time Markov chain models.

Our reference applications in this regard will be the pricing, in a Markovian setup generically described in Sect. 11.1, of CDO tranches and of counterparty risk on CDO tranches. As opposed to the situation of Chap. 10, there will be no optimal stopping features involved here so that, for pricing at time 0, a standard Monte Carlo loop, even though quite heavy due to the nature of the application at hand, would be adequate for the task. Our motivation for resorting to simulation/regression pricing schemes is to avoid multi- or nested Monte-Carlo schemes in situations where we need not only the price at time 0, but various prices in order to derive sensitivities (Sects. 11.2–11.3) or where a significant part of the pricing function is needed for estimating the credit valuation adjustment (CVA) on a CDO tranche (Sects. 11.4–11.5), a topical issue since the great crisis.

In the two applications of this chapter (Greeking by Monte Carlo without resimulation and Monte Carlo CVA computations), we will work, as in Chap. 10 for convertible bonds, in dynamic models of high nominal dimension. Deterministic schemes are thus ruled out by the curse of dimensionality. However, in specific sub-cases the models can be reformulated in a much lower effective dimension, so that the accuracy of the simulation results can be assessed by means of alternative exact or high-precision schemes.

As a prerequisite let us recall the portfolio credit risk problem of pricing a CDO tranche (see Sect. 5.3). Let N_t represent the cumulative default process, i.e. the number of firms defaulted by time t within a reference pool of n credit names, with n “large” (of the order of 100 or more in applications; see Sect. 5.3.2). Also, let $L_t = (1 - R)N_t$ represent the cumulative portfolio loss process, assuming a homogeneous

fractional recovery R of the underlying names (typically taken as $R = 40\%$ before the crisis, generally considered as less since then). For simplicity we assume nil risk-free interest rates and we consider protection legs of CDO tranches, with cumulative payoff process, for $t \in [0, T]$, of the form

$$\pi(N_t) = \left(\frac{L_t}{n} - a \right)^+ \wedge (b - a); \quad (11.1)$$

here the attachment and detachment points a and b are such that $0 \leq a \leq b \leq 100\%$. In particular we will consider equity tranches (attachment $a = 0$) and super-senior tranches (detachment $b = 100\%$), with respective payoff processes

$$\pi^+(N_t) = \frac{L_t}{n} \wedge k \quad \text{and} \quad \pi^-(N_t) = \left(\frac{L_t}{n} - k \right)^+, \quad (11.2)$$

where the “strike” (respectively detachment and attachment point) k belongs to $[0, 1]$. The “credit index” corresponds to the equity tranche with $k = 100\%$ (or the senior tranche with $k = 0$), with payoff process

$$p(N_t) = \frac{L_t}{n} = (\pi^+ + \pi^-)(N_t). \quad (11.3)$$

We will refer to our stylized loss derivatives as index and equity/senior tranches, thus ignoring the fees legs of the actual products.

Again, as opposed to the situation of Chap. 10, there are no optimal stopping features involved here. Our stylized loss derivatives are European options with payoff $\xi = \pi(N_T)$ and price process

$$\Pi_t = \mathbb{E}_t \xi. \quad (11.4)$$

Note that Π thus defined is in fact the cumulative price of the tranche (including past dividends), whereas the ex-dividend price $\check{\Pi}_t$ would be defined in the manner of (5.36) as $\mathbb{E}_t(\xi - \pi(N_t))$ (with the additional difference that, for simplicity, we only consider protection legs of CDO tranches in this chapter).

11.1 Generic Markovian Setup

Let there be given a continuous-time d -variate Markov chain $\mathcal{N} = (N^1, \dots, N^d)$ living in a finite set \mathcal{I} . The cumulative default process N_t on a credit portfolio can then be modeled as $N_t = \Lambda(\mathcal{N}_t)$ for some integer-valued function Λ , in the filtration $\mathcal{F}_t = \mathcal{F}_t^{\mathcal{N}}$. Many dynamic models of credit risk can be cast in this framework, directly or after discretization in space. Two concrete specifications will be investigated in Sects. 11.2–11.3 and 11.4–11.5.

Given a payoff $\xi = \pi(N_T) = \phi(\mathcal{N}_T)$ (with $\phi = \pi \circ \Lambda$), by the Markov property of \mathcal{N} the corresponding price process satisfies, for $t \in [0, T]$:

$$\Pi_t = \mathbb{E}_t \xi = \mathbb{E}_t \phi(\mathcal{N}_T) = u(t, \mathcal{N}_t). \quad (11.5)$$

Here $u(t, \iota)$ or $u_\iota(t)$, for $t \in [0, T]$ and $\iota = (i_1, \dots, i_d) \in \mathcal{I}$, defines the pricing function, i.e. the solution to a related system of Kolmogorov equations (see Sect. 2.2). In low-dimensional cases with small d , it is possible to devise efficient deterministic computational schemes for the pricing function u (see e.g. [34]):

- ODE schemes for the related Kolmogorov equations,
- or, in the time-homogeneous case, numerical matrix exponentiation schemes.

However, due to the curse of dimensionality, these schemes cease to be practical as soon as d exceeds a few units. Our aim in this chapter is to compute the pricing function u (or part of it) in high dimension, numerically by simulation/regression. The following lemma will be key for this purpose.

Lemma 11.1.1 *Given a fixed state $\iota \in \mathcal{I}$, let τ denote the first hitting time of ι by \mathcal{N} , with the convention that $\tau = +\infty$ if $\tau > T$ (so that τ is a $[0, T] \cup \{+\infty\}$ -valued stopping time). Then, almost surely:*

$$\mathbb{1}_{\tau < T} u_\iota(\tau) = \mathbb{E}(\mathbb{1}_{\tau < T} \xi | \tau). \quad (11.6)$$

Proof An application of Doob's optional sampling theorem to the martingale¹ $u(t, \mathcal{N}_t)$ with bounded stopping time $\bar{\tau} = \tau \wedge T$ yields that

$$\mathbb{E}_{\bar{\tau}}(\xi) = \mathbb{E}_{\bar{\tau}}(\phi(\mathcal{N}_T)) = u(\bar{\tau}, \mathcal{N}_{\bar{\tau}}), \quad (11.7)$$

where $\mathcal{N}_{\bar{\tau}} = \iota$ on $\{\tau < T\}$. Multiplying (11.7) by $\mathbb{1}_{\tau < T}$ we obtain

$$\mathbb{E}_{\bar{\tau}}(\mathbb{1}_{\tau < T} \xi) = \mathbb{1}_{\tau < T} u_\iota(\bar{\tau}). \quad (11.8)$$

Since $\sigma(\bar{\tau}) \subseteq \mathcal{F}_{\bar{\tau}}^{\mathcal{N}}$, taking conditional expectations given $\bar{\tau}$ in (11.8) yields (11.6). \square

11.1.1 Generic Simulation Pricing Scheme

The practical output of Lemma 11.1.1 is the following algorithm, based on (11.6), for estimating the function $t \mapsto u_\iota(t)$ over $(0, T)$:

- simulate m ($=10^4$ to 10^6 , say) trajectories of \mathcal{N} over $[0, T]$ and related values of the payoff $\xi = \phi(\mathcal{N}_T)$, storing the pairs $(\tau^j, \xi^j)^{1 \leq j \leq m}$, where the superscript j refers to the j^{th} trajectory. Note that this can typically be done exactly, by standard simulation methods, without any discretization error (see the following sections for concrete examples);
- estimate the function $u_\iota(\cdot)$ over $(0, T)$ by nonlinear regression (see Sect. 6.10.2) of the ξ^j with respect to the τ^j , where the nonlinear regression is performed over the subset of the indices j such that $\tau^j < T$ (for which $u(\tau, \mathcal{N}_\tau) = u_\iota(\tau)$).

The nonlinear regression provides an estimator of the following regression function (cf. Sect. 6.10.2)

¹In a càdlàg version, as always by default throughout this book.

$$(0, T) \ni t \mapsto \mathbb{E}(\xi | \tau = t) = u_t(t) \in \mathbb{R}_+, \quad (11.9)$$

in which the equality holds by (11.6). This results in an estimator of the function $u_t(\cdot)$ over $(0, T)$. As for $\Pi_0 = u_{\mathbf{0}}(0)$, it is estimated by the sample average $\widehat{\Pi}_0$ of the ξ^j (standard Monte Carlo estimate). We thus have a method for estimating the function $u_t(\cdot)$ that is exact up to the error involved in the estimation of the conditional expectations in (11.9). By exact we mean here that this estimator doesn't involve any time-discretization error, but only Monte Carlo simulation/regression errors, whereas in a deterministic ODE scheme, e.g., default times need to be “discretized” on a fixed time-grid.

The simulation step (first bullet item above) is only performed once, and can be used in the second bullet item for estimating the functions $u_t(\cdot)$ corresponding to different states t . Given a state $t \in \mathcal{I}$, we denote by:

- τ_i^j in $[0, T] \cup \{+\infty\}$, the first hitting time of the state $i \in \mathcal{I}$ on the j th trajectory of \mathcal{N} , with the convention that $\tau_i^j = +\infty$ if $\tau_i^j > T$,
- Ω_t , the subset of indices j of the simulated trajectories such that $\tau_i^j < T$, so that $u(\tau_i^j, \mathcal{N}_{\tau_i^j}^j) = u_t(\tau_i^j)$ for $j \in \Omega_t$.

As in Chap. 10, for most states $t \in \mathcal{I}$ the set Ω_t will typically be very small, even actually empty, so the nonlinear regression related to the state t will not be tractable. Now the key point of the method (the main advantage of a simulation/regression scheme over a deterministic one) is that instead of looping over all states $t \in \mathcal{I}$ and tentatively performing the regression over each set Ω_t , we can loop instead over the trajectories $j = 1, \dots, m$, making a record of the states $t \in \mathcal{I}$ for which the set Ω_t has cardinality greater than a given threshold μ , and then only compute the regressions corresponding to those t .

The rationale for this procedure is that the states t which are not hit (or not hit a sufficient number of times) in the simulation are low probability states t , starting from the initial condition $\mathcal{N}_0 = \mathbf{0}$. So the values of the pricing functions at these states are essentially irrelevant and we do not need, nor would we be able, to compute them.

11.2 Homogeneous Groups Model of Portfolio Credit Risk

As a first application of the generic Markov chain simulation/regression pricing scheme of Sect. 11.1, we will consider the problem of computing by simulation (but without resimulation) the prices and sensitivities of a CDO tranche in a homogeneous groups model of credit risk considered in [35, 61, 128, 143]. We will describe the theoretical aspects of the model in this section, whereas the numerical results will be presented in the next one.

In the homogeneous groups model, the n names of a credit pool are divided into d groups of $v - 1 = \frac{n}{d}$ homogeneous obligors (taking n as a multiple of d). The cumulative default processes N^l , $l = 1, \dots, d$, in the different groups (so that $N =$

$\sum N^l$) are jointly modeled as a d -variate Markov point process \mathcal{N} with intensity of N^l given as

$$\lambda^l(t, \mathcal{N}_t) = (\nu - 1 - N_t^l) \tilde{\lambda}^l(t, \mathcal{N}_t), \quad (11.10)$$

for some pre-default individual intensity functions $\tilde{\lambda}^l = \tilde{\lambda}_i^l(t)$, where $i = (i_1, \dots, i_d) \in \mathcal{I} = \{0, 1, \dots, \nu - 1\}^d$. Moreover, we exclude simultaneous jumps of the N^l , so that the generator of \mathcal{N} is given in the form of a very sparse ν^d -dimensional matrix. Precluding simultaneous jumps of the N^l allows us to reduce the simulation of a d -dimensional model with a ν^d -dimensional matrix generator to the coupled simulation of d one-dimensional models with ν -dimensional matrix generators. More precisely, for the purpose of simulating the next jump time and the state of \mathcal{N} conditional on \mathcal{N}_t , rather than dealing with a ν^d -dimensional intensity measure, it is sufficient to deal with d intensity measures (“which component of \mathcal{N} jumps”), each of dimension ν (“where it jumps”); see [34].

For $d = 1$ we recover the so-called local intensity model (pure birth process stopped at level n) of Laurent, Cousin and Fermanian [178] or Herbertsson [143] for the portfolio cumulative default process N . This model is, in a sense, the analog for credit derivatives of the local volatility model of Sect. 9.2.1. At the other extreme, for $d = n$, we are modeling the vector of all default indicator processes of the pool names. As d varies between 1 and n , we thus get a variety of models of credit risk, ranging from pure top models for $d = 1$ to pure bottom-up models for $d = n$; see [35]. In any case, we are in the generic setup of Sect. 11.1 and so, for $t \in [0, T]$:

$$\Pi_t = \mathbb{E}_t \xi = \mathbb{E}_t \phi(\mathcal{N}_T) = u(t, \mathcal{N}_t), \quad (11.11)$$

where $u(t, i)$ or $u_i(t)$ for $(t, i) \in [0, T] \times \mathcal{I}$ defines the pricing function to be determined numerically by simulation. For small d , deterministic numerical schemes are efficient and typically preferred alternatives. These will be used in Sect. 11.3 for assessing the accuracy of the simulation schemes.

11.2.1 Hedging in the Homogeneous Groups Model

The following representation property is valid in the homogeneous groups model for every bounded martingale $\Pi_t = u(t, \mathcal{N}_t)$: for $t \in [0, T]$ we have

$$\Pi_t = \Pi_0 + \sum_{l=1}^d \int_0^t \delta_l u(s, \mathcal{N}_s) dM_s^l, \quad (11.12)$$

where

$$M^l = N^l - \int_0^\cdot \lambda^l(t, \mathcal{N}_t) dt$$

is the compensated martingale of N^l and where

$$\delta_l u_i(t) = u_{i^l}(t) - u_i(t), \quad (11.13)$$

in which i^l represents i with its l th component i_l replaced by $i_l + 1$.

Let there thus be given a loss derivative with price process Π , to be hedged using a set of $q \leq d$ other loss derivatives with \mathbb{R}^q -valued price process denoted by Θ . We use the riskless constant asset for funding the strategy. Using the representation (11.12) for Π and the analogous representation for Θ (component by component), by applying the results of Sect. 4.2.3, we get explicit formulas for various hedging ratios such as the min-variance deltas of Π against Θ . By min-variance hedging we mean minimization of the hedging error under the prevailing pricing probability measure \mathbb{P} . In the $q = d$ case and under a suitable nonredundancy condition on the hedging instruments, min-variance hedging coincides with hedging by replication. See also Frey and Backhaus [127] and Bielecki, Vidozzi and Vidozzi [45].

11.2.1.1 Single-Instrument Hedge

We now consider in more detail the $q = 1$ case of min-variance hedging an equity or senior tranche by the credit index. Let Π and P (respectively u and v) denote the model price processes (respectively pricing functions) of the tranche and of the index. By application of the formula (4.48), we can min-variance hedge the tranche by the index and the riskless constant asset, using the strategy ζ in the index defined, for $t \in [0, T]$, by:

$$\begin{aligned} \zeta_t &= \frac{\sum_{l=1}^d \lambda^l (\delta_l u)(\delta_l v)}{\sum_{l=1}^d \lambda^l (\delta_l v)^2} (t, \mathcal{N}_{t-}) = \sum_{l=1}^d w^l \left(\frac{\delta_l u}{\delta_l v} \right) (t, \mathcal{N}_{t-}) \\ \text{with } w^l &= \frac{\lambda^l (\delta_l v)^2}{\sum_{j=1}^d \lambda^j (\delta_j v)^2}. \end{aligned} \quad (11.14)$$

Case of a Local Intensity Model If $d = 1$, in view of the martingale representation property (11.12) for Π and of the analogous representation for P , we have that

$$\Pi_t = \Pi_0 + \int_0^t \delta_s dP_s, \quad (11.15)$$

where

$$\delta_i = \delta(t, N_{t-}), \quad \text{with } \delta_i(t) = \frac{\delta u_i(t)}{\delta v_i(t)} = \frac{u_{i+1}(t) - u_i(t)}{v_{i+1}(t) - v_i(t)}. \quad (11.16)$$

In the $q = d = 1$ case, it is thus possible to replicate the tranche by the index using the strategy δ_t in the index defined by (11.16) (see also Laurent, Cousin and Fermanian [178]). In this case, the min-variance hedging strategy ζ_t of (11.14) coincides with the perfect replication strategy δ_t of (11.16).

For $d > 1$, it is not possible to replicate a tranche by the index in this model. We refer the reader to [35] for a numerical illustration of the fact that this potential

lack of replicability is not purely speculative, but can be very significant in practice.

11.2.2 Simulation Scheme

Having simulated m trajectories of \mathcal{N} over $[0, T]$ we can estimate, independently for every state i of interest in \mathcal{I} , the function $u_i(\cdot)$ over $(0, T)$ by the algorithm of Sect. 11.1.1. Moreover $\Pi_0 = u_0(0)$ is estimated by the arithmetic average $\widehat{\Pi}_0$ of the ξ^j .

Given the pricing function u of a loss derivative and the \mathbb{R}^q -valued pricing function v of a set of $q \leq d$ hedging instruments, where v is estimated like u by nonlinear regression based on the same set of simulated trajectories of \mathcal{N} , formula (11.13) yields, at no further computational cost, an estimator of the delta functions $\delta_l u_i(t)$ and $\delta v_i^l(t)$. Of course this only holds for states i such that both sets Ω_i and $\Omega_{i'}$ have a sufficient cardinality ($\geq \mu$ as explained in Sect. 11.1.1), but these are precisely the states that matter in practice.

Note, however, that the error on the delta functions is typically an order of magnitude larger than the error on the prices, since the estimator proposed above for the function $\delta_l u_i$ (say) is given as the difference between the estimators of the functions $u_{i'}(\cdot)$ and $u_i(\cdot)$.

Remark 11.2.1 In a simulation environment, an alternate way of estimating the deltas should be by regression. E.g., we have in the context of Sect. 11.2.1.1, by applications of formulas (11.16) and (3.68):

$$\delta_t = \delta(t, N_{t-}) = \frac{d\langle \Pi, P \rangle_t}{d\langle P \rangle_t} = \frac{\lim_{h \rightarrow 0} \mathbb{C}\text{ov}_t(\Pi_{t+h} - \Pi_t, P_{t+h} - P_t)}{\lim_{h \rightarrow 0} \mathbb{V}\text{ar}_t(P_{t+h} - P_t)}. \quad (11.17)$$

See Remark 10.2.2 for a discussion on the linear regression estimate δ_i^j in (10.19) for the delta of a convertible bond, called the delta MC backward estimate there, as opposed to the simulation/regression finite difference delta estimate introduced in Sect. 10.3.1.3. The linear regression, or MC backward delta estimate, is the standard delta estimate in the numerical BSDE literature. However, in the present jump setup with simulation/regression in time, as opposed to simulation/regression in space on a discrete time-grid in diffusive setups, it is not clear how (11.17) can be exploited numerically.

11.3 Pricing and Greeking Results in the Homogeneous Groups Model

We now report on numerical experiments made in the above homogeneous groups model. In addition to the τ_i of Sect. 11.1.1, we introduce the t_i denoting the consecutive jump times of the cumulative portfolio loss process N , with the convention that $t_i = +\infty$ in case the number of jumps on $[0, T]$ is less than i ; we also set $t_0 = 0$.

We consider an equity tranche of strike k and the related senior tranche as well as the credit index, with payoff functionals π^\pm and p in (11.2)–(11.3), on an underlying pool of $n = 8$ obligors. We set $T = 5$ years and $R = 40\%$.

Remark 11.3.1 The computation times for the exact matrix exponentiations, used for benchmarking our simulation/regression results, grow as v^{2d} . This explains the small value of $n = 8$ used, whereas the simulation/regression pricing scheme can be run for the much larger values of n , such as $n \sim 100$ that are needed for real-life applications.

Let Π^\pm and P , and u^\pm and v denote, respectively, the price processes and pricing functions of the equity/senior tranche and of the index.

11.3.1 Fully Homogeneous Case

We first consider the case of a fully homogeneous model, or local intensity model (see Sect. 11.2) for the cumulative default process N , so that $d = 1$ and $v = n + 1$. We thus deal here with a “one-dimensional” model with generator given as a v^d -dimensional matrix, where $v^{2d} = 9^2 = 81$. The state t reduces to a single component i and the τ_t reduce to the t_i . The individual pre-default intensity function $\tilde{\lambda}$ is taken as

$$\tilde{\lambda}_i(t) = \frac{1+i}{n}. \quad (11.18)$$

11.3.1.1 Pricing

We simulated by standard means $m = 10^4$ trajectories of N over $[0, T]$, 9925 out of which were such that $t_1 < T = 5$ yr (with at least one default in the pool before the expiry of the CDO tranche). Figure 11.1 displays the results obtained by application of the simulation pricing scheme for estimating $\Pi_{t_1}^\pm$ (“red +”/“blue ×”) and P_{t_1} (“green o”) on $\{t_1 \in (0, T)\}$.

The left panel shows the simulated points (t_1, ξ) , where $\xi = \pi^\pm(N_T)$ or $p(N_T)$, restricting attention to the “useful” points (regarding estimation of $u_1^\pm(\cdot)$ and $v_1(\cdot)$) for which $t_1 < T$. To every simulated trajectory of N over $[0, T]$, with $t_1 < T$, we thus associate on the left panel three points with common abscissa t_1 and respective ordinates $\xi = \pi^\pm(N_T)$ and $p(N_T)$. For example:

- if, on a specific trajectory, $t_1 < T$ but $t_2 > T$, meaning that exactly one default occurred before time T on this trajectory, then $N_T = 1$ and $\xi = \pi^\pm(1)$ or $p(1)$, and the corresponding points (t_1, ξ) are displayed on the panel (with the above color code);
- if $t_1 > T$, the trajectory is dropped and the corresponding points (t_1, ξ) , not used in the nonlinear regressions for estimating $u_1^\pm(\cdot)$ and $v_1(\cdot)$, are not shown on the panel.

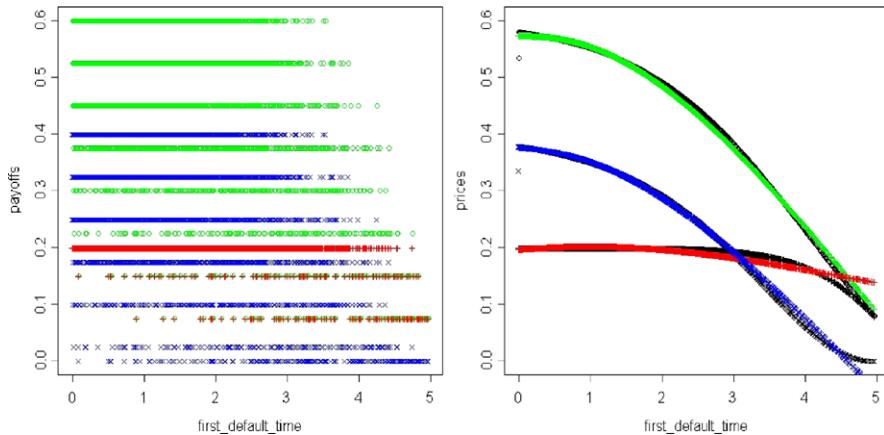


Fig. 11.1 (Left) index and tranches payoffs: simulated points (t_1, ξ) ; (Right) index and tranches prices at t_1 (trajectories j with $t_1^j < T = 5$ yr; equity tranche “red +”, senior tranche “blue \times ” and index “green o”)

Points on the same level of the “layers” that can be seen on the panel (green, blue and red layers, one for each of the three products under consideration) correspond to trajectories sharing a common value of $\xi = \pi^\pm(N_T)$ or $p(N_T)$ that arise due to the granular nature of the payoffs ξ .

The right panel presents the pricing results: points $(t_1, \Pi_{t_1}^\pm)$ and (t_1, P_{t_1}) with $t_1 < T$, where $\Pi_{t_1}^\pm$ and P_{t_1} are computed in two different ways:

- first (color curves with the above color code) by the simulation pricing scheme of Sects. 11.1.1/11.2.2, using parabolic regressions of the payoffs ξ^j with respect to the t_1^j over the subset $\tilde{\Omega}$ of the trajectories such that $t_1^j < T$ (global parametric regressions in the sense of Sect. 6.10.2);
- second (black curves) by matrix exponentiation.

We also put, on the right panel of the figure, the points $(0, \Pi_0^\pm)$ for the tranches and $(0, P_0)$ for the index, computed by matrix exponentiation.

The relative error of the simulation method varies from less than one percent to a few percent (visually the black and color curves are almost superposed for each of the three products), depending on the location of t_1 in the time interval $(0, T = 5$ yr)—with a greater error for short time to maturity, when t_1 gets closer to T .

11.3.1.2 Greeking

We then computed, as explained in Sect. 11.2.2, the quantities $\frac{\delta u^\pm}{\delta v}(t, i)$ at the origin $(t, i) = (0, 0)$, so that

$$\frac{u_1^\pm - u_0^\pm}{v_1 - v_0}(0), \quad (11.19)$$

using a common set of $m = 4 \times 10^4$ simulated trajectories of N over $[0, T]$ for deriving all the simulation estimates. The exact values in column 1 of Table 11.2 were computed by matrix exponentiation. Moreover, we considered six different estimates for the prices $u_1^\pm(0)$ and $v_1(0)$ in (11.19) ($u_0^\pm(0)$ and $v_0(0)$ in (11.19) being estimated by standard Monte Carlo over the m simulated trajectories of \mathcal{N}) corresponding to (see Sect. 6.10.2):

- methods of cells consisting in estimating the price $u_1^\pm(0)$ or $v_1(0)$ in (11.19) by the empirical average of the related payoffs over the set of trajectories j for which $t_1^j < T_1$, with $T_1 = 0.01$ or 0.1 , whence two first estimates for $u_1^\pm(0)$ and $v_1(0)$;
- affine or parabolic regression of the related payoffs ξ^j with respect to the t_1^j , where a nonlinear regression is performed over the subset Ω_1 of the indices j such that $t_1^j < T_1$, with $T_1 = 1$ or 0.1 , whence four further estimates of $u_1^\pm(0)$ and $v_1(0)$.

Out of $m = 4 \times 10^4$ simulated trajectories that were used for this purpose, we obtained 25369 trajectories with $t_1 < 1$, among which 3870 were such that $t_1 < 0.1$. The relative error percentages corresponding to the six estimates in the case of equity and senior tranches with strike $k = 45\%$ are displayed Table 11.1 in columns 2 and 3 (methods of cells with $T_1 = 1$ or 0.1), 4 and 5 (affine and parabolic regressions over $(0, 1)$) and 6 and 7 (affine and parabolic regressions over $(0, 0.1)$), respectively. Table 11.2 displays the analogous results for the equity and senior tranches with strike $k = 20\%$.

Note that in these tables the results for the senior tranche can be deduced from those for the equity tranche by parity (the related exact values as simulated deltas sum up to one, by construction). However a given absolute error on a tranche has, of course, a very different impact, in terms of relative error, on a far-from-the-index equity or senior tranche with delta less than 0.5 than on a close-to-the-index senior or equity tranche with delta greater than 0.5.

The best results are obtained with the estimator $\widehat{\delta}_1^1$ (affine regression with $T = 1$), which yields a relative error smaller than 1 % in all cases.

11.3.2 Semi-Homogeneous Case

We now set $d = 2$ and $v = 5$. We thus still consider a pool of $n = 2 \times 4 = 8$ obligors, yet the model is now bivariate, with generator given as a v^d -dimensional matrix with $v^{2d} = 5^4 = 625$. Moreover, we take $\tilde{\lambda}_l^l$ given, for $l = 1, 2$, by:

$$\tilde{\lambda}_l^l(t) = l \left(\frac{1 + i_l}{n} \right). \quad (11.20)$$

The pre-default individual intensities of the obligors of group 1 and 2 are thus given as $\frac{1+i_1}{8}$ and $\frac{1+i_2}{4}$ where, in the probabilistic interpretation, i_1 and i_2 represent the number of defaulted obligors in groups 1 and 2.

Table 11.1 Exact values (column 1) and relative error percentages (columns 2 to 7) for various estimators of $\delta = \frac{\delta v_0(0)}{\delta v_0(0)}$ in the fully homogeneous model with $d = 1$ (u^\pm = equity or senior pricing function for the strike $k = 45\%$, v = index pricing function; $m = 4 \times 10^4$ trajectories)

$k = 45\%$	Val δ	Err $\hat{\delta}_{0,01}^0$	Err $\hat{\delta}_{0,1}^0$	Err $\hat{\delta}_1^1$	Err $\hat{\delta}_1^2$	Err $\hat{\delta}_{0,1}^1$	Err $\hat{\delta}_{0,1}^2$
Eq	0.4151304	1.323378	2.364615	0.2919607	-0.5240502	2.072199	-1.088964
Sen	0.5848696	-0.9393109	-1.678363	-0.2072287	0.3719618	-1.470811	0.7729282

Table 11.2 Like Table 11.1, but for $k = 20\%$

$k = 20\%$	Val δ	Err $\hat{\delta}_{0,01}^0$	Err $\hat{\delta}_{0,1}^0$	Err $\hat{\delta}_1^1$	Err $\hat{\delta}_1^2$	Err $\hat{\delta}_{0,1}^1$	Err $\hat{\delta}_{0,1}^2$
Eq	0.06509018	2.232646	4.302896	0.5870334	2.273163	3.948764	0.6395232
Sen	0.9349098	-0.1554410	-0.2995757	-0.04087037	-0.1582619	-0.2749204	-0.04452481

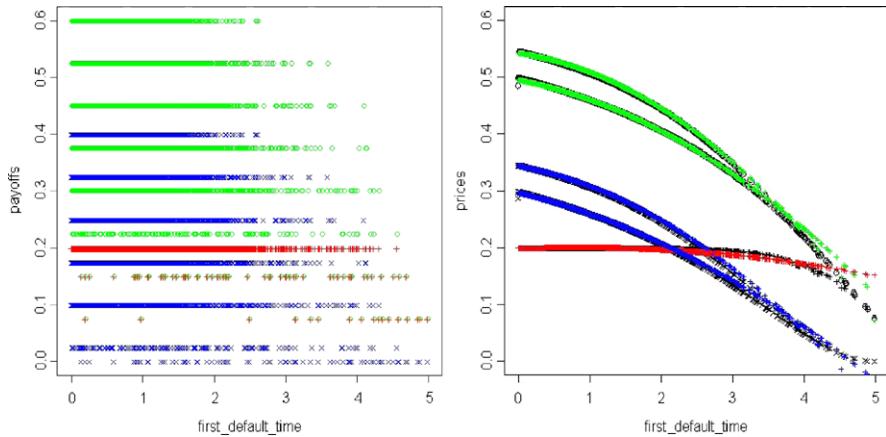


Fig. 11.2 Like Fig. 11.1, but in the semi-homogeneous model

11.3.2.1 Pricing

Simulating $m = 10^4$ trajectories, we got 9994 draws with $t_1 < 5$, corresponding to 3368 defaults in the first group of obligors and 6626 defaults in the second one. Figure 11.2 is the analog of Fig. 11.1 in the present semi-homogeneous model, where the parabolic regressions for estimating u_t^\pm and v_t , for $t = (1, 0)$ or $(0, 1)$, were performed over the subset Ω_i of the trajectories j such that $t_1 < T$ and $\mathcal{N}_{t_1} = i$. Observe in the right panel of Fig. 11.2 that:

- In the case of the equity tranche (“red +”), conditionally on t_1 , the probability that the tranche will be wiped out before maturity is essentially one unless t_1 is close to T . Therefore the equity tranche price at t_1 is essentially equal to $k = 20\%$ for t_1 not to close to T . Moreover, for t_1 close to T the intrinsic value of the tranche at t_1 is the main contributor to the value of the equity tranche at t_1 (since the tranche has low time-value close to maturity). In conclusion, the state of \mathcal{N} at t_1 (depending on the group of the defaulted obligor) has a low impact on $\Pi_{t_1}^+$, whatever the value of t_1 in $[0, T]$.
- In the case of the senior tranche (“blue x”) or of the index (“green o”), the state of \mathcal{N} at t_1 has a high impact on the corresponding price, unless t_1 is close to T (in which case the intrinsic value is dominant in the tranche price). This explains the two-track profiles which emerge in the cases of the senior tranche and of the index on the right panel of Fig. 11.2, whereas the two-tracks are essentially superimposed in the case of the equity tranche.

The relative error of the simulation method varies from less than one percent to a few percent (visually, the black and color curves are almost superposed for each of the three products and for both groups of obligors).

Table 11.3 Number of simulated trajectories with $t_1 < T_1$ and $\mathcal{N}_{t_1} = (1, 0)$ or $\mathcal{N}_{t_1} = (0, 1)$, for $T_1 = 1$ and 0.1 (over a total of $m = 4 \times 10^4$ simulated trajectories)

$ t_1 < T_1 $	$\mathcal{N}_{t_1} = (1, 0)$	$\mathcal{N}_{t_1} = (0, 1)$	Total
$T_1 = 1$	10434	20629	31063
$T_1 = 0.1$	1879	3717	5596

11.3.2.2 Greeking

We then computed, as explained in Sect. 11.2.2, the quantities $\frac{\delta_1 u^\pm}{\delta_1 v}(t, \iota)$ and $\frac{\delta_2 u^\pm}{\delta_2 v}(t, \iota)$ at the origin (for $t = 0$ and $\iota = (0, 0)$), so that

$$\frac{\delta_1 u_{0,0}^\pm}{\delta_1 v_{0,0}}(0) = \frac{u_{1,0}^\pm - u_{0,0}^\pm}{v_{1,0} - v_{0,0}}(0), \quad \frac{\delta_2 u_{0,0}^\pm}{\delta_2 v_{0,0}}(0) = \frac{u_{0,1}^\pm - u_{0,0}^\pm}{v_{0,1} - v_{0,0}}(0). \quad (11.21)$$

In view of the formula (11.14), the time-0 min-variance hedging delta of the tranche \pm by the index is given by the following weighted average of these two quantities (valued at $t = 0$):

$$w_{0,0}^1 \left(\frac{\delta_1 u_{0,0}^\pm}{\delta_1 v_{0,0}} \right) + w_{0,0}^2 \left(\frac{\delta_2 u_{0,0}^\pm}{\delta_2 v_{0,0}} \right) \quad \text{with } w_{0,0}^{1/2} = \frac{\lambda_{0,0}^{1/2} (\delta_1 v_{0,0})^2}{\lambda_{0,0}^1 (\delta_1 v_{0,0})^2 + \lambda_{0,0}^2 (\delta_2 v_{0,0})^2}.$$

All our Monte Carlo estimates were obtained by using a common set of $m = 4 \times 10^4$ simulated trajectories of $\mathcal{N} = (N^1, N^2)$ over $[0, T]$ (see Table 11.3). Exact values were computed by matrix exponentiation (see columns 1 of Tables 11.4 and 11.5). Moreover we considered the six estimates introduced in Sect. 11.3.1.2 for the prices $u_\iota^\pm(0)$ and $v_\iota(0)$ involved in (11.21) for $\iota \neq (0, 0)$, whereas $u_{0,0}^\pm(0)$ and $v_{0,0}(0)$ in (11.21) were computed by standard Monte Carlo over the m simulated trajectories of \mathcal{N} .

Tables 11.4 and 11.5 are the respective analogs, in the present semi-homogeneous model, of Tables 11.1 and 11.2 in the fully homogeneous model of Sect. 11.3.1.2. As in the homogeneous case, the best results were obtained with the estimator $\hat{\delta}_1^1$ (affine regression with $T = 1$), but we observe a significant deterioration of the accuracy with respect to the homogeneous case, with a relative error attaining e.g. 7.18 % for the second delta on the senior tranche with strike $k = 45$ % in Table 11.4.

11.4 Common Shocks Model of Portfolio Credit Risk

In the remaining two sections of this chapter, we deal with counterparty risk on CDO tranches. Counterparty risk, on credit derivatives in particular, has been a topical issue since the great crisis; see e.g. [8] and other references in the same volume, as well as [30, 59]. We consider a risk-free bank holding a buy-protection CDO tranche from a risky counterparty.

Table 11.4 Exact values (column 1) and percentage relative errors (columns 2 to 7) for various estimators of $\delta = \frac{\delta_1 u^\pm}{\delta_1 v}(0, 0, 0)$ or $\frac{\delta_2 u^\pm}{\delta_2 v}(0, 0, 0)$ in the semi-homogeneous model (u^\pm = equity or senior pricing function of strike $k = 45\%$, v = index pricing function; $m = 4 \times 10^4$)

$k = 45\%$	Val δ	Err $\hat{\delta}_{0,1}^0$	Err $\hat{\delta}_1^0$	Err $\hat{\delta}_1^1$	Err $\hat{\delta}_1^2$	Err $\hat{\delta}_{0,1}^1$	Err $\hat{\delta}_{0,1}^2$
Eq1	0.3945317	-6.32058	-0.5255639	0.05174216	-2.534111	-5.689355	-5.018331
Eq2	0.5317208	-33.99359	-2.569012	-6.330396	-12.57564	-7.807325	-24.44841
Sen1	0.6054683	4.11858	0.3424649	-0.03371593	1.651263	3.707265	3.270016
Sen2	0.4682792	38.5998	0.4682792	7.188027	14.27937	8.865048	27.76064

Table 11.5 Like Table 11.4, but for $k = 20\%$

$k = 20\%$	Val δ	Err $\hat{\delta}_{0,1}^0$	Err $\hat{\delta}_{0,1}^1$	Err $\hat{\delta}_1^0$	Err $\hat{\delta}_1^1$	Err $\hat{\delta}_{0,1}^1$	Err $\hat{\delta}_{0,1}^2$
Eq1	0.008829481	10.82540	0.06888924	4.901498	-4.931537	-18.25982	-13.61765
Eq2	0.04291727	19.03478	48.66493	1.877609	18.71009	56.02024	62.72903
Sen1	0.9911705	-0.0964341	-0.0006136747	-0.04366321	0.0439308	0.1626609	0.1213079
Sen2	0.9570827	-0.853553	-2.182221	-0.08419527	-0.8389934	-2.512046	-2.81288

Remark 11.4.1 What follows also covers the case of a bank, possibly risky but disregarding, in a unilateral CVA perspective, its own default risk when assessing the counterparty risk of its position with another party. The more general situation of bilateral counterparty risk, including also a related funding issue [30, 80, 81, 216], is very involved in the case of counterparty risk on portfolio credit derivatives.

We denote by $\tau_0 = \tau$ the default time of the counterparty, and by τ_i , for $i \in \mathbb{N}_n^* = \{1, \dots, n\}$, the default times of n credit names underlying the CDO tranche. For $i \in \mathbb{N}_n = \{0, \dots, n\}$, we let N^i represent the default indicator process of τ_i , so that $N_t^i = \mathbb{1}_{\tau_i \leq t}$, and we write $\mathcal{N} = (N^i)_{i \in \mathbb{N}_n}$, $N_t = \sum_{i \in \mathbb{N}_n^*} N_t^i$. For the targeted application of CVA computations on CDO tranches, we need a model of counterparty credit risk with the following key features:

- it should be a dynamic Markovian model, where the dynamic requirement stems from the fact that, as will be apparent from the CVA formulas (11.28)–(11.29), the CVA is an option on the future values of the so-called clean price (price not accounting for the counterparty risk) of the underlying, in our case a CDO tranche; we thus need a model in which clean prices of a CDO tranche at future time points can be computed numerically, at least approximately via a Markov structure;
- the model should be calibratable jointly to individual CDS and CDO tranches data; it should therefore be a bottom-up [35] model of portfolio credit risk so that individual names are represented in the model; but for tractability of the calibration the model should also enjoy a copula-like separation property between the marginal and the dependence parameters, as well as efficient explicit pricing formulas for vanillas (CDSs and CDOs).

This motivates the use of the Markovian copula common shock model of [31, 33] (to our knowledge, this is the only model with all the required features in the currently existing literature). First, we define a certain number v (typically small: a few units) of groups $I_l \subseteq \mathbb{N}_n$ of obligors who are likely to default simultaneously, for $l = 0, \dots, v$. The idea is that, at every time t , there will be a positive probability that the survivors of the group of obligors I_l (obligors of group I_l still alive at time t) default simultaneously. Write $\mathcal{I} = \{I_0, \dots, I_v\}$, $\mathcal{Y} = \{\{0\}, \dots, \{n\}, I_0, \dots, I_v\}$. Given the so-called default-triggering event intensities λ_Y , $Y \in \mathcal{Y}$ (assumed constant for simplicity) we model $\mathcal{N} = (N^i)_{i \in \mathbb{N}_n}$ as a Markov chain with respect to its natural filtration \mathbb{F} , with infinitesimal generator \mathcal{A} of \mathcal{N} given in functional form, for $u = u_\iota(t)$ with $t \in [0, T]$ and $\iota = (i_j)_{j \in \mathbb{N}_n}$ as:

$$\mathcal{A}u(t, \iota) = \sum_{Y \in \mathcal{Y}} \lambda_Y \delta_Y u(t, \iota), \quad (11.22)$$

where $\delta_Y u(t, \iota) = u(t, \iota^Y) - u(t, \iota)$, in which ι^Y is the vector obtained from $\iota = (i_0, \dots, i_n)$ by replacing the components i_j , $j \in Y$ by ones. As shown in [31], this

structure of the generator implies the following expression for the intensity of a jump of $\mathcal{N} = (N^i)_{i \in \mathbb{N}_n}$ from $\mathcal{N}_{t-} = i$ to $\mathcal{N}_t = j$, with $j \neq i$ in $\{0, 1\}^{n+1}$:

$$\sum_{\{Y \in \mathcal{Y}; i^Y = j\}} \lambda_Y. \quad (11.23)$$

The intensity of a jump of \mathcal{N} from i to j at time t is thus equal to the sum of the intensities of the sets $Y \in \mathcal{Y}$ such that, if the default of the survivors in set Y occurred at time t , the state of \mathcal{N} would move from i to j . Therefore, the pre-default intensity of a given name i is given as

$$\lambda_i := \sum_{\mathcal{Y} \ni Y \ni i} \lambda_Y = \lambda_{\{i\}} + \sum_{\mathcal{I} \ni I \ni i} \lambda_I,$$

which does not depend on the default status of other obligors. As shown in [31], this implies the so-called Markov copula property [41, 42, 45] that, for every $i \in \mathbb{N}_n$, N^i is a $\{0, 1\}$ -valued Markov chain admitting the following generator, for $u_i = u_i(t, j)$ with $(t, j) \in [0, T] \times \{0, 1\}$:

$$\mathcal{A}_i u_i(t, j) = \lambda_i (u_i(t, 1) - u_i(t, j))$$

with

$$\lambda_i := \sum_{\mathcal{Y} \ni Y \ni i} \lambda_Y = \lambda_{\{i\}} + \sum_{\mathcal{I} \ni I \ni i} \lambda_I,$$

or in an equivalent matrix-form:

$$\mathcal{A}_i \equiv \begin{bmatrix} -\lambda_i & \lambda_i \\ 0 & 0 \end{bmatrix}. \quad (11.24)$$

11.4.1 Example

The following illustrative example is drawn from [31]. Take $n = 2$ (case of three credit names labeled 0 to 2), so that the state space of \mathcal{N} contains the following 8 elements:

$$\{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1)\}.$$

Let \mathcal{Y} be given as $\mathcal{Y} = \{\{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 1, 2\}\}$. There are five different shocks, corresponding to the elements of \mathcal{Y} . In particular, obligors 0 and 1 can default simultaneously if either the shock corresponding to $\{0, 1\}$ arrives, or the shock corresponding to $\{0, 1, 2\}$ arrives. This is an example of a nested structure of \mathcal{I} : $I_0 = \{0, 1\} \subset I_1 = \{0, 1, 2\}$. Here the generator of the chain \mathcal{N} takes the following matrix form:

$$\mathcal{A} \equiv \begin{bmatrix} \cdot & \lambda_{\{0\}} & \lambda_{\{1\}} & \lambda_{\{2\}} & \lambda_{\{0,1\}} & 0 & 0 & \lambda_{\{0,1,2\}} \\ 0 & \cdot & 0 & 0 & \lambda_{\{1\}} + \lambda_{\{0,1\}} & \lambda_{\{2\}} & 0 & \lambda_{\{0,1,2\}} \\ 0 & 0 & \cdot & 0 & \lambda_{\{0\}} + \lambda_{\{0,1\}} & 0 & \lambda_{\{2\}} & \lambda_{\{0,1,2\}} \\ 0 & 0 & 0 & \cdot & 0 & \lambda_{\{0\}} & \lambda_{\{1\}} & \lambda_{\{0,1,2\}} + \lambda_{\{0,1\}} \\ 0 & 0 & 0 & 0 & \cdot & 0 & 0 & \lambda_{\{2\}} + \lambda_{\{0,1,2\}} \\ 0 & 0 & 0 & 0 & 0 & \cdot & 0 & \lambda_{\{1\}} + \lambda_{\{0,1,2\}} + \lambda_{\{0,1\}} \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdot & \lambda_{\{0\}} + \lambda_{\{0,1,2\}} + \lambda_{\{0,1\}} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

where a “.” represents the negative of the sum of all the other elements in the same row. Now consider group $\{0, 1, 2\}$. Suppose that at some point of time obligor 1 is defaulted, but obligors 0 and 2 are still alive, so that process \mathcal{N} is in state $(0, 1, 0)$. In this case the two survivors in the group $\{0, 1, 2\}$ may default simultaneously with intensity $\lambda_{\{0,1,2\}}$. Note that here $\lambda_{\{0,1,2\}}$ cannot be interpreted as the intensity of all three defaulting simultaneously, as obligor 1 has already defaulted. The only state of the model in which $\lambda_{\{0,1,2\}}$ can be interpreted as the intensity of all three defaulting is the origin $(0, 0, 0)$. Regarding the Markov copula property, observe that as long as obligor 0 survives, the corresponding default intensity is $\lambda_{\{0\}} + \lambda_{\{0,1,2\}} + \lambda_{\{0,1\}}$, regardless of the state of the pool. Similar remarks apply to obligors 1 and 2.

11.4.2 Marshall-Olkin Representation

In [31], a connection is established between the above Markov chain model \mathcal{N} with default times τ_i (actually, an extension of this model involving diffusive factor processes) and a static Marshall-Olkin model [199].

Let $\text{supp}^c(\mathcal{N}_t)$ denote the set of the surviving names in model \mathcal{N} at time t , so that

$$\text{supp}^c(\mathcal{N}_t) = \{i \in \mathbb{N}_n; N_t^i = 0\}.$$

We can define a Marshall-Olkin model of default times $(\tilde{\tau}_i)_{i \in \text{supp}^c(\mathcal{N}_t)}$ of the surviving names in model \mathcal{N} at time t such that, given \mathcal{N}_t , $\tau = (\tau_i)_{i \in \text{supp}^c(\mathcal{N}_t)}$ and $\tilde{\tau} = (\tilde{\tau}_i)_{i \in \text{supp}^c(\mathcal{N}_t)}$ are equal in law. Given $\mathcal{Y}_t = \{Y \in \mathcal{Y}; Y \not\subseteq \text{supp}(\mathcal{N}_t)\}$, let us thus define, for every $Y \in \mathcal{Y}_t$,

$$\tau_Y = t + \frac{\epsilon_Y}{\lambda_Y}, \tag{11.25}$$

for i.i.d. exponential random variables ϵ_Y . Let us set, for every $i \in \text{supp}^c(\mathcal{N}_t)$,

$$\tilde{\tau}_i = \bigwedge_{Y \in \mathcal{Y}_t; i \in Y} \tau_Y. \tag{11.26}$$

The τ_i denoting, as before, the default times of the credit names in the Markovian model \mathcal{N} , then for every $t_1, \dots, t_n \geq t$:

$$\mathbb{P}(\tau_i > t_i, i \in \text{supp}^c(\mathcal{N}_t) | \mathcal{N}_t) = \mathbb{P}(\tilde{\tau}_i > t_i, i \in \text{supp}^c(\mathcal{N}_t) | \mathcal{N}_t). \quad (11.27)$$

Crucially for the targeted application of CVA computations on CDO tranches, this interpretation can be used for deriving fast exact recursive procedures for pricing a CDO tranche conditionally on \mathcal{N}_t (see [31]). The Marshall-Olkin representation can also be used for simulation purposes. So, given \mathcal{N}_t , the simulation of a set of random times $(\tau_i)_{i \in \text{supp}^c(\mathcal{N}_t)}$ or, equally in law by (11.27), $(\tilde{\tau}_i)_{i \in \text{supp}^c(\mathcal{N}_t)}$, is very fast: all we need do is simulate the i.i.d. exponential random variables ϵ_Y in (11.25) for Y in \mathcal{Y}_t . Finally, thanks to the Markov copula property of the model, the marginal survival probabilities can essentially be considered as inputs for the model (they can be bootstrapped independently in a preprocessing step from individual CDS quotes); the only parameters we need calibrate are then the few model dependence parameters λ_I .

11.5 CVA Computations in the Common Shocks Model

The numerical results of [31, 32] show that the model is indeed jointly calibratable to CDS and CDO real data sets. Moreover, as this is a Markovian model of portfolio credit risk, it is possible to deal with intrinsic dynamic issues such as counterparty risk (see [8, 33]).

The Credit Valuation Adjustment (CVA) refers to the correction in value accounting for the credit risk of a counterparty, in the context of a bilateral OTC derivative transaction. We consider the risk-free bank buying credit protection from her defaultable counterparty (modeled as name 0) through a buy-protection CDO tranche with maturity T on names 1 to n . By application of the results of [8, 33], one can represent the CVA of the bank by the following process Θ_t , for $0 \leq t \leq \tau \wedge T$:

$$\Theta_t = \mathbb{E}_t(\mathbb{1}_{\tau < T} \chi) \quad (11.28)$$

for an \mathcal{F}_τ -measurable counterparty exposure χ of the bank given as

$$\chi = (1 - R_0)(\Pi_\tau - \pi(N_{\tau-}) - \Gamma_\tau)^+, \quad (11.29)$$

in which:

- R_0 is the recovery rate of the counterparty,
- $\Pi_\tau = \mathbb{E}_\tau \xi = u(\tau, \mathcal{N}_\tau)$ is the counterparty clean (not accounting for counterparty risk) price of the CDO tranche at the time of default τ of the counterparty,
- π is the tranche payoff function and
- Γ represents a nonnegative margin amount, value of the accumulated collateral posted by the counterparty to the bank in order to mitigate counterparty risk.

Remark 11.5.1 As noted after (11.4), Π defined by (11.4) is in fact the cumulative price of the (protection leg of a) CDO tranche. This explains the expression (11.29) for χ , consistent with the corresponding expression of [33] in terms of the ex-dividend price process $\check{\Pi}_t = \Pi_t - \pi(N_t)$:

$$\chi = (1 - R_0)(\check{\Pi}_\tau + \pi(N_\tau) - \pi(N_{\tau-}) - \Gamma_\tau)^+. \quad (11.30)$$

We thus have, under the so-called naked (no collateralization) scheme $\Gamma = 0$ and the (collateralized) scheme $\Gamma_t = \Pi_{t-} - \pi(N_{t-})$, respectively:

$$\begin{aligned}\Theta_t &= \Theta_t^0 = \mathbb{E}_t(\mathbb{1}_{\tau < T} \chi^0) \quad \text{with } \chi^0 = (1 - R_0)(u(\tau, \mathcal{N}_\tau) - \pi(N_{\tau-}))^+ \\ \Theta_t &= \Theta_t^1 = \mathbb{E}_t(\mathbb{1}_{\tau < T} \chi^1) \quad \text{with } \chi^1 = (1 - R_0)(u(\tau, \mathcal{N}_\tau) - u(\tau, \mathcal{N}_{\tau-}))^+.\end{aligned}\quad (11.31)$$

Remark 11.5.2 Accounting for various frictions and delays (notably the so-called cure period [33]) regarding formation and delivery of the collateral, the above “collateralized” scheme can be considered the most extreme case of collateralization, and is sometimes referred to as continuous collateralization (with an “infinitesimal” cure period reflected by the left-limit in “ $\Gamma = \Pi_- - \pi_-$ ”, whereas more realistic models of collateral account for a positive cure period; see [82]).

Using the Marshall-Olkin representation of Sect. 11.4.2, the simulation of the counterparty’s default time $\tau = \tau_0$ and the computation of $\Pi_\tau = \mathbb{E}_\tau \xi = u(\tau, \mathcal{N}_\tau)$ in χ^0 , and also of $\Pi_{\tau-} = u(\tau, \mathcal{N}_{\tau-})$ in χ^1 , can be done in a very fast and exact way. We thus have the following exact Monte Carlo estimates $\tilde{\Theta}_0^0$ and $\tilde{\Theta}_0^1$ for the naked and collateralized CVA at time 0 of a CDO tranche in this model, based on m simulated trajectories of \mathcal{N} :

$$\begin{aligned}\tilde{\Theta}_0^0 &= (1 - R_0) \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{\tau^j < T} (u(\tau^j, \mathcal{N}_{\tau^j}) - \pi(N_{\tau^j-}))^+ \\ \tilde{\Theta}_0^1 &= (1 - R_0) \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{\tau^j < T} (u(\tau^j, \mathcal{N}_{\tau^j}) - u(\tau^j, \mathcal{N}_{\tau^j-}))^+.\end{aligned}\quad (11.32)$$

By exact Monte Carlo estimates we mean that these involve no time-discretization error, nor any approximation error regarding valuation of the tranche at time τ . The only error is the statistical $O(m^{-\frac{1}{2}})$ -Monte Carlo error (see Sect. 6.3.2).

As alternatives to the above exact Monte Carlo estimates, the approximate Monte Carlo estimates $\widehat{\Theta}_0^0$ and $\widehat{\Theta}_0^1$ are defined as

$$\begin{aligned}\widehat{\Theta}_0^0 &= (1 - R_0) \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{\tau^j < T} (\widehat{u}(\tau^j, \mathcal{N}_{\tau^j}) - \pi(N_{\tau^j-}))^+ \\ \widehat{\Theta}_0^1 &= (1 - R_0) \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{\tau^j < T} (\widehat{u}(\tau^j, \mathcal{N}_{\tau^j}) - \widehat{u}(\tau^j, \mathcal{N}_{\tau^j-}))^+,\end{aligned}\quad (11.33)$$

in which $\widehat{u}(\tau^j, \mathcal{N}_{\tau^j})$ and $\widehat{u}(\tau^j, \mathcal{N}_{\tau^j-})$ are numerical approximations of $\Pi_{\tau^j} = u(\tau^j, \mathcal{N}_{\tau^j})$ and $\Pi_{\tau^j-} = u(\tau^j, \mathcal{N}_{\tau^j-})$, computed by the simulation/regression approach of Sect. 11.1.1.

Note that in a real life CVA setup, we deal not with a single CDO tranche, but with a netted portfolio of thousands of contracts, collateralized in a path-dependent way (see [33, 82]). The interest of the estimates (11.33) is that they may still be practical in more realistic situations where the exact Monte Carlo estimates (11.32) will no longer be available (see Cesari et al. [63]). The estimates (11.33) are also typically faster to compute (see the “CPU+” rows of the numerical tables in the next section).

11.5.1 Numerical Results

We implemented in Matlab the exact and approximate Monte Carlo CVA estimates in the naked and collateralized cases, $\widetilde{\Theta}_0^0$, $\widetilde{\Theta}_0^1$, $\widehat{\Theta}_0^0$ and $\widehat{\Theta}_0^1$ thus, on a stylized credit portfolio of 100 obligors, for CDO tranches with maturity $T = 2$ years and for recovery rates $R_0 = R = 40\%$. The individual default intensities were taken as $\lambda_{\{i\}} = 10^{-4} \times (200 - i)$, which decreases from about 200 bps to 100 bps as i increases from 0 to 99. Moreover, we considered four nested groups of joint defaults (so $v = 3$), respectively consisting of the riskiest 3, 9, 21 and 100 (so all) names, with corresponding joint default intensities $\lambda_I = 10^{-3} \times \frac{2}{1+I}$, which decreases from about 20 bps to 5 bps as v increases from 0 to 3. The counterparty was taken as the fiftieth riskiest name (name with median risk) in the portfolio. Also, regarding the approximate Monte Carlo estimates $\widehat{\Theta}_0^0$ and $\widehat{\Theta}_0^1$, various values were tried for the degree ρ of the nonlinear regression in the time variable, which is used for computing the conditional expectations (\widehat{u} -terms in (11.33)), and for the threshold μ of Sect. 11.1.1.

The results are presented in Tables 11.6 (naked CVA) and 11.7 (collateralized CVA). With the approximate Monte Carlo estimate $\widehat{\Theta}_0^1$ in the collateralized case, we needed to estimate not only the $u(\tau^j, \mathcal{N}_{\tau^j})$, but also the $u(\tau^j, \mathcal{N}_{\tau^j-})$. This required a greater number of simulated trajectories to achieve a given level of accuracy. All our naked, respectively collateralized, Monte Carlo estimates were thus based on a common set of $m = 1.5 \times 10^5$, respectively $m = 3 \times 10^5$, simulated trajectories. The times needed to simulate these m trajectories and construct the tables of payoffs used in all the corresponding Monte Carlo estimates were 362 s and 2700 s, respectively. In Tables 11.6 and 11.7:

- the CVA numbers correspond to a nominal of one hundred per obligor e.g. the maximal loss on the equity tranche in this scale is $\text{Nom} \times n \times 5\% = 500$;
- σ is the standard error associated with a Monte Carlo CVA estimate;
- $\% \sigma$ is the corresponding percentage error in the sense of $10^2 \times \frac{\sigma}{\text{CVA}}$;
- $\% \text{Err}$ is the percentage relative error of the approximate Monte Carlo estimate with respect to the exact Monte Carlo estimate;

Table 11.6 Naked CVA ($n = 100$ obligors, $m = 1.5 \times 10^5$ trajectories)

$\mu = 1$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	4.75	2.93	2.42	4.75	2.93	2.42	4.76	2.93	2.42	4.78	2.96	2.44
σ	0.08	0.24	0.20	0.08	0.24	0.20	0.08	0.24	0.20	0.08	0.24	0.20
$\% \sigma$	1.6	8.2	8.3	1.6	8.2	8.3	1.6	8.2	8.3	1.6	8.1	8.2
%Err	–0.6	–1.1	–0.8	–0.6	–1.0	–0.8	–0.6	–1.0	–0.7			
CPU+	190 s			200 s			210 s			308 s		
$\mu = 3$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	3.42	2.91	2.42	3.42	2.91	2.42	3.42	2.91	2.42	4.78	2.96	2.44
σ	0.07	0.24	0.20	0.07	0.24	0.20	0.07	0.24	0.20	0.08	0.24	0.20
$\% \sigma$	2.0	8.2	8.3	2.0	8.2	8.3	2.0	8.2	8.3	1.6	8.1	8.2
%Err	–28.5	–1.6	–0.8	–28.5	–1.6	–0.8	–28.4	–1.6	–0.7			
CPU+	190 s			200 s			210 s			308 s		
$\mu = 5$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	3.36	2.91	2.42	3.36	2.91	2.42	3.36	2.91	2.42	4.78	2.96	2.44
σ	0.07	0.24	0.20	0.07	0.24	0.20	0.07	0.24	0.20	0.08	0.24	0.20
$\% \sigma$	2.0	8.2	8.3	2.0	8.2	8.3	2.0	8.2	8.3	1.6	8.1	8.2
%Err	–29.8	–1.6	–0.8	–29.9	–1.6	–0.8	–29.8	–1.6	–0.7			
CPU+	3.2 s			3.2 s			3.2 s			308 s		
$\mu = 10$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	3.35	2.91	2.42	3.35	2.91	2.42	3.35	2.91	2.42	4.78	2.96	2.44
σ	0.07	0.24	0.20	0.07	0.24	0.20	0.07	0.24	0.20	0.08	0.24	0.20
$\% \sigma$	2.0	8.2	8.3	2.0	8.2	8.3	2.1	8.2	8.3	1.6	8.1	8.2
%Err	–29.9	–1.6	–0.8	–29.9	–1.6	–0.8	–29.9	–1.6	–0.7			
CPU+	3.2 s			3.2 s			3.2 s			308 s		
$\mu = 20$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	3.19	2.91	2.42	3.19	2.91	2.42	3.19	2.91	2.42	4.78	2.96	2.44
σ	0.07	0.24	0.20	0.07	0.24	0.20	0.07	0.24	0.20	0.08	0.24	0.20
$\% \sigma$	2.1	8.2	8.3	2.1	8.2	8.3	2.1	8.2	8.3	1.6	8.1	8.2
%Err	–33.4	–1.7	–0.8	–33.4	–1.6	–0.8	–33.4	–1.6	–0.7			
CPU+	3.2 s			3.2 s			3.2 s			308 s		

Table 11.7 Collateralized CVA ($n = 100$ obligors, $m = 3 \times 10^5$ trajectories)

$\mu = 1$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	2.76	2.73	2.27	2.91	2.76	2.27	3.37	2.77	2.27	3.41	2.73	2.26
σ	0.09	0.16	0.14	0.14	0.17	0.14	0.23	0.17	0.14	0.05	0.16	0.14
$\% \sigma$	3.2	6.0	6.0	4.7	6.0	6.0	6.8	6.0	6.0	1.4	6.0	6.0
%Err	–19.1	–0.2	0.4	–14.6	1.2	0.4	–1.0	1.3	0.4			
CPU+	168 s			168 s			168 s			1033 s		

$\mu = 2$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	2.76	2.73	2.27	2.91	2.76	2.27	3.37	2.77	2.27	3.41	2.73	2.26
σ	0.09	0.16	0.14	0.14	0.17	0.14	0.23	0.17	0.14	0.05	0.16	0.14
$\% \sigma$	3.2	6.0	6.0	4.7	6.0	6.0	6.8	6.0	6.0	1.4	6.0	6.0
%Err	–19.1	–0.2	0.4	–14.6	1.2	0.4	–1.0	1.3	0.4			
CPU+	168 s			168 s			168 s			1033 s		

$\mu = 3$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	2.53	2.73	2.27	2.58	2.76	2.27	2.87	2.77	2.27	3.41	2.73	2.26
σ	0.04	0.16	0.14	0.05	0.17	0.14	0.07	0.17	0.14	0.05	0.16	0.14
$\% \sigma$	1.7	6.0	6.0	1.9	6.0	6.0	2.4	5.8	6.0	1.4	6.0	6.0
%Err	–25.7	–0.2	0.4	–24.3	1.2	0.4	–15.8	5.0	0.4			
CPU+	168 s			168 s			168 s			1033 s		

$\mu = 5$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	2.53	2.73	2.27	2.53	2.73	2.27	2.81	2.73	2.27	3.41	2.73	2.26
σ	0.04	0.16	0.13	0.04	0.16	0.13	0.06	0.16	0.13	0.05	0.16	0.14
$\% \sigma$	1.8	6.0	5.7	1.8	6.0	5.7	2.3	6.0	5.7	1.4	6.0	6.0
%Err	–25.7	–0.2	0.4	–25.7	–0.2	0.4	–17.4	0.0	0.4			
CPU+	5.3 s			5.3 s			5.3 s			1033 s		

$\mu = 10$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	2.53	2.73	2.27	2.53	2.73	2.27	2.81	2.73	2.27	3.41	2.73	2.26
σ	0.04	0.16	0.13	0.04	0.16	0.13	0.06	0.16	0.13	0.05	0.16	0.14
$\% \sigma$	1.8	6.0	5.7	1.8	6.0	5.7	2.3	6.0	5.7	1.4	6.0	6.0
%Err	–25.7	–0.2	0.4	–25.7	–0.2	0.4	–17.4	0.0	0.4			
CPU+	5.3 s			5.3 s			5.3 s			1033 s		

- CPU+ is the additional time (in seconds) needed to compute a specific CVA estimate, on top of the common times of 362 s or 2700 s which were required for constructing the tables of payoffs corresponding to the $m = 1.5 \times 10^5$ or 3×10^5 trajectories;
- μ and ρ are the values of the threshold and of the order of polynomials which are used in the simulation/regression estimates of the conditional expectations (see Sect. 11.1.1).

Observe that the impact of the collateralization, as assessed by the difference between the exact Monte Carlo CVA results of Tables 11.6 and 11.7, is mainly significant for the equity tranche. So, for the equity tranche 0–5, the exact Monte Carlo value of the naked CVA is $\tilde{\Theta}_0^0 = 4.78$, whereas the exact Monte Carlo estimate of the collateralized CVA is $\tilde{\Theta}_0^1 = 3.41$. For the 35+ tranche these numbers become, respectively, 2.44 and 2.26.

This reflects the fact that the equity tranche CVA is more due to the clean price term \check{I}_{τ} in (11.30) and less to the joint defaults term $\pi(N_{\tau}) - \pi(N_{\tau-})$, where the latter is scarcely collateralizable (see [8]). Conversely, joint defaults are the main explanatory factor of the more senior tranches CVA. Collateralization having little impact on the senior tranches conveys the important message that, due to wrong-way risk, counterparty risk on credit derivatives may be scarcely collateralizable. Wrong-way risk refers to a positive dependence between the exposure of the bank and the default of the counterparty. As explained in [86], this is represented in our model by the possibility of joint defaults.

In order to keep the construction of the payoff tables under reasonable CPU budget, we first simulate the default times following the approach of Sect. 11.4.2, and then only construct the corresponding trajectories of \mathcal{N} with default of the counterparty at a time $\tau^j < T$ (therefore in particular, ignoring in the regressions the “rejected” paths for which $\tau^j > T$). Indeed simulating the default times is very fast, whereas reconstructing the sequence of configurations successively reached at the subsequent default times on the corresponding trajectory of \mathcal{N} is much more time-consuming. Moreover even in case $\tau^j < T$ we do not store the full trajectory, but only the last two states \mathcal{N}_{τ^j} and \mathcal{N}_{τ^j-} . These reductions drastically diminish the CPU budget for constructing the tables, which still took 362 s and 2700 s above for $m = 1.5 \times 10^5$ and 3×10^5 , but would have been intractable otherwise for $n = 100$ obligors and comparable values of m (see however Sect. 11.5.1.2 for possible improvements in this regard).

Note that this rejection of paths for which $\tau^j > T$ induces a specific bias into the $\hat{u}(\tau^j, \mathcal{N}_{\tau^j-})$ which intervene in $\hat{\Theta}_0^1$ in (11.33). Indeed, these are estimates of \hat{u} at states $i = \mathcal{N}_{\tau^j-}$ also possibly attained on some of the rejected paths, which should then have been taken into account in the time-regression estimate of such $\hat{u}_i(\cdot)$. The computation of the $\hat{u}(\tau^j, \mathcal{N}_{\tau^j-})$ is therefore slightly biased with respect to that of the $\hat{u}(\tau^j, \mathcal{N}_{\tau^j})$, hence a specific bias in $\hat{\Theta}_0^1$ follows.

However, in both the naked case of Table 11.6 and the collateralized case of Table 11.7, the accuracy of the approximate Monte Carlo estimates $\widehat{\Theta}_0^0$ and $\widehat{\Theta}_0^1$ is satisfactory with regard to both $\% \sigma$ and $\% \text{Err}$, provided we appropriately choose the threshold μ and the degree ρ of the regression. The accuracy of the approximate Monte Carlo estimates is best and of the order of 1 % with respect to both $\% \sigma$ and $\% \text{Err}$, for $\mu = 1$ and any ρ in the naked case, respectively for $\mu = 1$ or 2 and $\rho = 3$ in the collateralized case. The additional times CPU+ needed to compute the approximate Monte Carlo estimates are always less than for the exact estimates.

Even in the collateralized case of Table 11.7, the accuracy of the approximate Monte Carlo estimate² is quite satisfactory, since for well-chosen μ and ρ , the error $\% \text{Err}$ is typically within the confidence interval error $\% \sigma$; this holds even though for the latter we chose a relatively narrow one-standard-error confidence interval.

Regarding the threshold μ , the value $\mu = 1$ (or equally $\mu = 2$ in the collateralized case), gives the best results in both cases. It thus seems preferable to retain as many states as possible in the estimates, including states only reached a very small number of times (in fact, even only once) on the simulated trajectories.

One may finally wonder why, in the naked case, the approximate Monte Carlo estimate $\widehat{\Theta}_0^0$ essentially doesn't depend on the degree ρ of the nonlinear regression which is used for computing the conditional expectations. This is because

$$\begin{aligned}\widehat{\Theta}_0^0 &= (1 - R_0) \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{\tau^j < T} (\widehat{u}(\tau^j, \mathcal{N}_{\tau^j}) - \pi(N_{\tau^j-}))^+ \\ &\approx (1 - R_0) \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{\tau^j < T} (u(\tau^j, \mathcal{N}_{\tau^j}) - \pi(N_{\tau^j-}))^+ \\ &= (1 - R_0) \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{\tau^j < T} u(\tau^j, \mathcal{N}_{\tau^j}) - (1 - R_0) \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{\tau^j < T} \pi(N_{\tau^j-}) \\ &\approx (1 - R_0) \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{\tau^j < T} \widehat{u}(\tau^j, \mathcal{N}_{\tau^j}) - (1 - R_0) \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{\tau^j < T} \pi(N_{\tau^j-}),\end{aligned}$$

where the next-to-last identity holds because we deal with nondecreasing payoff functions π and cumulative default process N_t . Therefore, in $\widehat{\Theta}_0^0$, the $\widehat{u}(\tau^j, \mathcal{N}_{\tau^j})$ only matter through their empirical average on $\tau^j < T$. Now recall that the $\widehat{u}(\tau^j, \mathcal{N}_{\tau^j})$ are obtained by multilinear regression of the payoffs $\xi^j = \pi(N_T^j)$ against the first ρ powers of the τ^j for which $\tau^j < T$, for every fixed state i reached by the \mathcal{N}_{τ^j} . We thus have, by a standard property of multilinear regression estimates:

²Though endowed with a specific bias as explained above.

Table 11.8 Collateralized CVA with reduced payoffs table ($n = 60$ obligors, $m = 2 \times 10^4$ trajectories)

$\mu = 1$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	1.70	1.60	1.32	2.40	1.63	1.35	2.90	1.63	1.35	2.19	1.66	1.36
σ	0.11	0.37	0.31	0.11	0.36	0.30	1.49	0.38	0.32	0.11	0.38	0.32
% σ	6.7	23.2	23.4	4.4	22.1	22.2	51.4	23.5	23.6	5.2	23.0	23.4
%Err	–22.2	–3.9	–2.8	10.0	–2.1	–0.8	32.7	–2.1	–0.8			
CPU+	68 s			68 s			68 s			33 s		

$\mu = 3$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	1.60	1.60	1.32	2.35	1.63	1.35	2.50	1.63	1.35	2.19	1.66	1.36
σ	0.11	0.37	0.31	0.23	0.36	0.30	0.25	0.26	0.20	0.11	0.38	0.32
% σ	7.1	23.2	23.4	9.8	22.1	22.2	9.9	16.1	15.0	5.2	23.0	23.4
%Err	–26.8	–3.9	–2.8	7.6	–2.1	–0.8	14.4	–2.1	–0.8			
CPU+	68 s			68 s			68 s			33 s		

$$\sum_{j=1}^m \mathbb{1}_{\tau^j < T} \widehat{u}(\tau^j, \mathcal{N}_{\tau^j}) = \sum_{j=1}^m \mathbb{1}_{\tau^j < T} \xi^j = \sum_{j=1}^m \mathbb{1}_{\tau^j < T} \pi(N_T^j),$$

which is independent of the order ρ of the regression.

11.5.1.1 Reduced Versus Full Payoff Tables

To further investigate the impact, in the collateralized case, of working with reduced tables of payoffs, we finally present comparative results obtained on a portfolio of $n = 60$ obligors, with reduced payoff tables as above, or with the full payoff tables. For $m = 2 \times 10^4$ simulated trajectories, the construction of the reduced and full payoff tables took 3 s and 3800 s. We have found that, for a given number of trajectories, the time of construction of the full payoffs table increases very fast (“exponentially”) with the number of obligors, whereas the time of construction of the reduced payoff tables increases only “linearly” with the number of obligors. With a portfolio of $n = 100$ obligors as before, the construction of the full payoff tables was thus far out of reach.

It is apparent in Tables 11.8 (reduced payoffs table) and 11.9 (full payoffs table) that the accuracy of the CVA results is less robust with the reduced payoffs table than with the full payoffs table. Indeed, in the first case accurate results are only obtained for $\mu = 3$ and $\rho = 2$, whereas the accuracy of the results is much more uniform in μ and ρ in the second case. To get better results by the reduced payoffs table, we need to use a greater number m of simulated trajectories. But for a given level of accuracy this is still much less costly numerically than using a full payoffs table with fewer trajectories.

Table 11.9 Collateralized CVA with full payoffs table ($n = 60$ obligors, $m = 2 \times 10^4$ trajectories)

$\mu = 1$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	2.05	1.60	1.32	2.04	1.53	1.26	2.06	1.68	1.39	2.19	1.66	1.36
σ	0.11	0.37	0.31	0.11	0.36	0.30	0.11	0.41	0.34	0.11	0.38	0.32
$\% \sigma$	5.2	23.2	23.4	5.2	23.5	23.8	5.5	24.5	24.4	5.2	23.0	23.4
%Err	–6.1	–3.9	–2.8	–6.6	–7.7	–7.4	–5.6	1.0	2.3			
CPU+	137 s			137 s			137 s			55 s		

$\mu = 3$	MC $\rho = 1$			MC $\rho = 2$			MC $\rho = 3$			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+	0–5	5–35	35+
CVA	2.05	1.60	1.32	2.05	1.53	1.26	2.09	1.68	1.39	2.19	1.66	1.36
σ	0.11	0.37	0.31	0.11	0.36	0.30	0.11	0.41	0.34	0.11	0.38	0.32
$\% \sigma$	5.2	23.2	23.4	5.2	23.5	23.8	5.4	19.7	24.4	5.2	23.0	23.4
%Err	–6.1	–3.9	–2.8	–6.1	–7.7	–7.4	–4.2	1.0	2.3			
CPU+	137 s			137 s			137 s			55 s		

Table 11.10 Collateralized CVA with reduced payoffs table and probit regression ($n = 120$ obligors, $m = 5 \times 10^4$ trajectories)

$\mu = 1$	MC Probit			Exact		
Tranche	0–5	5–35	35+	0–5	5–35	35+
CVA	3.70	3.25	2.70	3.66	3.27	2.70
σ	0.23	0.48	0.40	0.13	0.48	0.4
$\% \sigma$	6.2	14.8	14.8	3.6	14.7	14.8
%Err	1.1	–0.6	0.0			
CPU+	46 s			241 s		

11.5.1.2 Further Improvements

One can improve the above results in several ways. First, more elaborate regressions can be used instead of the basic polynomial regressions of degrees one to three above. Table 11.10 thus shows the results obtained for computing the collateralized CVA with reduced payoffs table on CDO tranches bearing on an underlying portfolio of $n = 120$ obligors, with a probit regression based on $m = 5 \times 10^4$ model trajectories and a threshold $\mu = 1$. By this, we mean that for every state t reached by the simulations, we regress the corresponding payoffs of the (a, b) -tranche by $(b - a)\mathcal{N}^{-1}(\alpha_t + \beta_t t)$, where α_t and β_t are the regression parameters and \mathcal{N} represents the standard Gaussian cumulative distribution function (Matlab function “`gmlfit`” with the arguments “`binomial, link and probit`”). The accuracy of the results is similar to those of Table 11.7, although six times fewer trajectories are used and more obligors are considered. Moreover, the table construction took only 20 s (on top of the additional times CPU+ that appear in the last row of Table 11.10) instead of 2700 s as in the case of the experiment of Table 11.7.

Table 11.11 Collateralized CVA computed by Latin Hypercube Monte Carlo with reduced payoffs table and probit regression ($n = 120$ obligors, $m = 3 \times 10^4$ trajectories)

$\mu = 1$	LH Probit			Exact		
	Tranche	0–5	5–35	35+	0–5	5–35
CVA	3.72	3.48	2.90	3.70	3.50	2.90
σ	0.25	0.60	0.50	0.16	0.6	0.5
% σ	6.7	17.2	17.2	4.3	17.1	17.2
%Err	0.5	–0.6	0.0			
CPU+	40 s			159 s		

Moreover, one can use various variance reduction techniques or simulation add-ons. Table 11.11 thus shows the results obtained with the same specifications as in Table 11.10, except that a Latin Hypercube Monte Carlo was used instead of a standard Monte Carlo (Matlab function “lhsdesign” used instead of “rand” for generating the uniform numbers underlying the default times). The accuracy of the results is similar as in Table 11.10 (at least in terms of %Err), although a lower number of trajectories was used ($m = 3 \times 10^4$ versus 5×10^4 in Table 11.10). This is, however, at the cost of an increased table construction time of 220 s (on top of the additional times CPU+ in the last row of Table 11.11), as opposed to 20 s in Table 11.10.

11.5.2 Conclusions

There is numerical evidence that nonlinear regressions in time can fruitfully be used in the context of high-dimensional Markov chain models of portfolio credit risk or, more generally, for solving any Markov chain related system of Kolmogorov ODEs. This could be pushed further in several directions.

In the context of the Greeking application of Sects. 11.2–11.3, how to choose the relative values of the method parameters (e.g., how to let T_1 and m tend to 0 and $+\infty$ in the method of cells) in order to ensure convergence to the exact values and avoid excessive deterioration of the accuracy as the dimension increases (cf. the observation made after Tables 11.4 and 11.5) is an interesting question for future investigation. Another issue is the possible use of regression delta estimates instead of the finite difference Monte Carlo deltas which are used at the present stage (see Remark 11.2.1).

Also, in the context of the CVA application of Sects. 11.4–11.5, having in mind that real life CVA applications bear on portfolios of thousands of contracts, computation times are still quite high. So the question arises of suitable variance reduction techniques such as importance sampling or particle methods [61] that could be used to efficiently address the issue of rare default events simulation. Another challenging question is how one might possibly extend the computations to the more general CVA setup of bilateral counterparty risk under funding constraints [80, 81]. One then has to deal with implicit, nonlinear CVA equations, no longer amenable to ex-

plicit Monte Carlo loops. Perhaps particle simulation schemes could also be useful here [82].

In both cases the choice and number of basis functions that should theoretically be used³ in the nonlinear regressions is a pending issue.

³However, as demonstrated in this chapter, the practical side of this can ultimately only be, to a large extent, experimental.

Part V

Jump-Diffusion Setup with Regime Switching (**)

We saw in Sect. 3.5 and Chap. 4 that the problem of pricing and hedging financial derivatives can be modeled in terms of (possibly reflected) backward stochastic differential equations (BSDEs) or, equivalently in the Markovian setup, by partial integro-differential equations or variational inequalities (PDEs for short). This part deals with the mathematics of these BSDEs and PDEs. The most demanding proofs are deferred to Chap. 15.

Chapter 12 establishes the well-posedness of a Markovian reflected BSDE in a rather generic jump-diffusion model with regime switching, which covers all the models considered in this book. In Chap. 13 we study the associated partial integro-differential systems of obstacle problems; as an aside we establish the convergence of stable, monotone and consistent approximation schemes. The main results of Chaps. 12 and 13 are synthesized in Propositions 12.5.2 and 13.2.5. Chapter 14 provides various extensions of these results needed for applications to pricing problems involving discrete dividends on a financial derivative or on an underlying asset, as well as various forms of discrete path dependence such as those of Chap. 10.

Chapter 12

Backward Stochastic Differential Equations

In this chapter we construct a jump-diffusion with regime switching denoted by \mathcal{X} . We then establish the well-posedness of related Markovian reflected and doubly reflected BSDEs.

The reader may want to think of process \mathcal{X} as either an illustrative specification of the abstract factor process X of Chap. 4, informally defined by (4.30), or as a generalization with regime switching of this factor process X .

Outline In Sect. 12.1 we present different kinds of SDEs and (possibly reflected) BSDEs. In Sect. 12.2 we introduce a versatile factor process $\mathcal{X} = (X, N)$ composed of a jump-diffusion-like component X interacting with a pure jump process N . The interaction between X and N is characterized by the dependence of the coefficients of the dynamics of X on N , and by mutual dependence of the jump intensities for the two processes. This framework encompasses a vast majority of the factor models that are used in pricing applications and covers, in particular, all the models previously considered in this book. The model is constructed by change of measure in Sect. 12.3. In Sect. 12.4 we study reflected and doubly reflected Markovian BSDEs in this setup, showing that they are well-posed in the sense that they have unique solutions, which depend continuously on their input data. In Sect. 12.5 we derive the associated Markov and flow properties.

12.1 General Setup

We consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$ ¹ for a finite time horizon $T > 0$. Let there also be given a measured mark space¹ $(E, \mathcal{B}_E, \varrho)$, where ϱ is a nonnegative finite measure on (E, \mathcal{B}_E) . By default, we

¹Typically finite or Euclidean.

consider the right-continuous and completed versions of all filtrations;² a random variable has to be \mathcal{F} -measurable; a process is defined on the time interval $[0, T]$ and is \mathbb{F} -adapted; all semimartingales are taken in a càdlàg version; all inequalities between random quantities are to be understood $d\mathbb{P}$ -almost surely, $dt \otimes d\mathbb{P}$ -almost everywhere or $dt \otimes d\mathbb{P} \otimes \varrho(de)$ -almost everywhere, as appropriate.

Let $B = (B_t)_{t \in [0, T]}$ be a standard d -dimensional Brownian motion; let

$$\mu = (\mu(dt, de))_{t \in [0, T], e \in E}$$

be an integer-valued random measure (or counting measure) on $([0, T] \times E, \mathcal{B}([0, T]) \otimes \mathcal{B}_E)$, i.e. an $\mathbb{N} \cup \{+\infty\}$ -valued random measure on $[0, T] \times E$ such that $\mu(\{t\} \times E, \omega) \leq 1$ for every (t, ω) (see Definition II.1.13 on p. 68 of Jacod and Shiryaev [153]). We denote by \mathcal{P} the predictable σ -field on $[0, T] \times \Omega$, which is the σ -field generated by the class of all left-continuous (and adapted) processes. By a predictable integrand $V = V_t(e)$ we mean a $\mathcal{P} \otimes \mathcal{B}_E$ -measurable real-valued random function $V_t(\omega, e)$. We assume that the compensator of the counting measure μ is given as $\varrho_t(de) dt$, for some (possibly random) jump intensity measure ϱ_t of the form $\varrho_t(de) = \zeta_t(e)\varrho(de)$, where ζ is a nonnegative and bounded predictable integrand. Since $\mu(dt, de)$ and $\varrho_t(de) dt$ are nonnegative measures, we can define the integrals $V_t \cdot d\mu_t$ and $V_t \cdot \varrho_t dt$ of a predictable integrand V with respect to $\mu(dt, de)$ and $\varrho_t(de) dt$ in a pathwise Lebesgue-Stieltjes sense as, respectively,

$$\begin{aligned} \int_0^\cdot V_t \cdot d\mu_t &= \int_0^\cdot \int_E V_t^+(e) \mu(dt, de) - \int_0^\cdot \int_E V_t^-(e) \mu(dt, de), \\ \int_0^\cdot V_t \cdot \varrho_t dt &= \int_0^\cdot \int_E V_t^+(e) \varrho_t(de) dt - \int_0^\cdot \int_E V_t^-(e) \varrho_t(de) dt, \end{aligned} \tag{12.1}$$

where the term on the right-hand side is well defined whenever

$$\int_0^\cdot \int_E |V_t(e)| \mu(dt, de) < +\infty, \quad \text{respectively} \quad \int_0^\cdot \int_E |V_t(e)| \varrho_t(de) dt < +\infty. \tag{12.2}$$

Moreover, one can show that the second condition in (12.2) implies the first (both conditions being understood almost surely). When the second condition holds, we define the integral of V against the compensated measure $\tilde{\mu}(dt, de) = \mu(dt, de) - \varrho_t(de) dt$ by

$$\int_0^\cdot V_t \cdot d\tilde{\mu}_t = \int_0^\cdot V_t \cdot d\mu_t - \int_0^\cdot V_t \cdot \varrho_t dt. \tag{12.3}$$

Remark 12.1.1 In case $V_t \cdot \varrho_t dt$ and $V_t \cdot d\mu_t$ fail to exist separately, it is sometimes possible to define a stochastic integral $V_t \cdot d\tilde{\mu}_t$ of V with respect to the compensated measure $\tilde{\mu}$. We refer the reader to [149, 153, 167] for the construction of

²See Sect. 4.1.

such stochastic integrals, for which in particular (12.3) holds for every $\varrho_t(de) dt$ -integrable random function V . This is important for extending the methods and results of this book to σ -finite jump measures under suitable integrability conditions. However, in the context of this book where all intensity measures are finite for simplicity, this is immaterial since we can consider the ordered jump times T_n of μ , so that the integral in the first line of (12.1) is always well defined, as

$$\int_0^\cdot V_t \cdot d\mu_t = \sum_{T_n \leq \cdot} V_{T_n}(e_n),$$

where e_n is the mark of the n th jump of μ .

The above “.” notation is extended componentwise to the case of real-matrix valued integrands. We likewise write, for every measurable function v on $(E, \mathcal{B}_E, \varrho)$,

$$v \cdot \varrho_t \equiv \int_E v(e) \cdot \varrho_t(de). \quad (12.4)$$

We also use the following notations:

- $|X|$, the Euclidean norm of a vector or row-vector X in \mathbb{R}^d ;
- $|M|$, the supremum of $|MX|$ as X ranges over the unit ball of \mathbb{R}^d , for M in $\mathbb{R}^{d \times d}$;
- $\mathcal{M}_\varrho = \mathcal{M}(E, \mathcal{B}_E, \varrho; \mathbb{R})$, the set of measurable functions from $(E, \mathcal{B}_E, \varrho)$ to \mathbb{R} endowed with the topology of convergence in measure and, for $v \in \mathcal{M}_\varrho$ and $t \in [0, T]$:

$$|v|_t = [v^2 \cdot \varrho_t]^{\frac{1}{2}} = \left[\int_E v(e)^2 \varrho_t(de) \right]^{\frac{1}{2}} \in \mathbb{R}_+ \cup \{+\infty\}; \quad (12.5)$$

- $\mathcal{B}(\mathcal{O})$, the Borel σ -field on \mathcal{O} , for any topological space \mathcal{O} .

We now introduce some Banach spaces³ of random variables or processes, where p denotes a real number in $[2, \infty)$:

- L^p , the space of real random variables ξ such that

$$\|\xi\|_{L^p} = (\mathbb{E}[\xi^p])^{\frac{1}{p}} < +\infty;$$

- \mathcal{S}_d^p , or \mathcal{S}^p in case $d = 1$, the space of \mathbb{R}^d -valued càdlàg processes X such that

$$\|X\|_{\mathcal{S}_d^p} = \left(\mathbb{E} \left[\sup_{t \in [0, T]} |X_t|^p \right] \right)^{\frac{1}{p}} < +\infty;$$

³Or Hilbert spaces, in the cases of L^2 , \mathcal{H}_d^2 or \mathcal{H}_μ^2 .

- \mathcal{H}_d^p , or \mathcal{H}^p in case $d = 1$, the space of \mathbb{R}^d -valued predictable row-vector processes Z such that

$$\|Z\|_{\mathcal{H}_d^p} = \left(\mathbb{E} \left[\int_0^T |Z_t|^2 dt \right] \right)^{\frac{1}{p}} < +\infty;$$

- \mathcal{H}_μ^p , the space of $\mathcal{P} \otimes \mathcal{B}_E$ -measurable functions $V : [0, T] \times \Omega \times E \rightarrow \mathbb{R}$ such that

$$\|V\|_{\mathcal{H}_\mu^p} = \left(\mathbb{E} \left[\int_0^T (|V_t|^p \cdot \varrho_t) dt \right] \right)^{\frac{1}{p}} < +\infty,$$

so that, in particular, (cf. (12.5))

$$\|V\|_{\mathcal{H}_\mu^2} = \left(\mathbb{E} \left[\int_0^T |V_t|^2 dt \right] \right)^{\frac{1}{2}};$$

- \mathcal{A}^2 , the space of finite variation continuous processes A with continuous Jordan components⁴ $A^\pm \in \mathcal{S}^2$;
- \mathcal{A}_+^2 , the space of nondecreasing processes in \mathcal{A}^2 .

By abuse of notation, we also write $\|X\|_{\mathcal{H}^p}$ for $(\mathbb{E}[\int_0^T X_t^2 dt]^{\frac{p}{2}})^{\frac{1}{p}}$ in the case of a progressively measurable, not necessarily predictable, real process X . The following facts are well known.

Proposition 12.1.2

- The processes $\int_0^\cdot Z_t dB_t$ and $\int_0^\cdot V_t \cdot d\tilde{\mu}_t$ are true martingales, for every $Z \in \mathcal{H}_d^p$ and $V \in \mathcal{H}_\mu^p$.*
- There exist positive constants c_p and C_p , depending only on p , $\varrho(E)$, T and on a bound on ζ , such that*

$$c_p \|V\|_{\mathcal{H}_\mu^p} \leq \left\| \int_0^\cdot V_t \cdot d\tilde{\mu}_t \right\|_{\mathcal{S}_d^p} \leq C_p \|V\|_{\mathcal{H}_\mu^p} \quad (12.6)$$

for every $V \in \mathcal{H}_\mu^p$.

12.1.1 Semimartingale Forward SDE

We first consider the following forward stochastic differential equation: X_0 written in the form of an \mathcal{F}_0 -measurable random variable and, for $t \in [0, T]$,

$$dX_t = b_t(X_t) dt + \sigma_t(X_t) dB_t + \delta_t(X_{t-}) \cdot d\mu_t, \quad (12.7)$$

⁴See Definition 4.1.12.

where $b_t(x)$, $\sigma_t(x)$ and $\delta_t(x, e)$ are, respectively, d -dimensional drift vector, dispersion matrix and jump size vector random coefficients, such that:

- $b_t(x)$, $\sigma_t(x)$ and $\delta_t(x, e)$ are Lipschitz in x , uniformly in $t \geq 0$ and $e \in E$;
- $b_t(0)$, $\sigma_t(0)$ and $\delta_t(0, e)$ are bounded in $t \geq 0$ and $e \in E$.

The following a priori bound and error estimates for a solution X_t of (12.7), assumed to exist, are classical. They can be shown by repeated application of the Burkholder-Davis-Gundy inequality and the Gronwall lemma (see Protter [228]), and also by using (12.6) and the fact that, for $p \geq 2$,

$$\|Z\|_{\mathcal{H}_d^p} = \mathbb{E} \left[\int_0^T |Z_t|^2 dt \right]^{\frac{p}{2}} \leq C \mathbb{E} \left[\int_0^T |Z_t|^p dt \right]$$

(note that, conversely, $\mathbb{E}[\int_0^T |Z_t|^p dt] \leq \mathbb{E}[\int_0^T |Z_t|^2 dt]^{\frac{p}{2}}$). In particular, in the explicit case where $b_t(x)$, $\sigma_t(x)$ and $\delta_t(x)$ do not depend on x in (12.7), the bound estimate (12.8) readily follows from a Burkholder-Davis-Gundy inequality, needed for controlling the martingale terms in

$$dX_t = b_t dt + \sigma_t dB_t + \delta_t \cdot d\mu_t = (b_t + \delta_t \cdot \varrho_t) dt + \sigma_t dB_t + \delta_t \cdot d\tilde{\mu}_t,$$

and a Jensen inequality, needed for controlling the drift term $(b_t + \delta_t \cdot \varrho_t) dt$.

Proposition 12.1.3 *For every solution X of the stochastic differential equation (12.7) with initial condition $X_0 \in \mathcal{F}_0 \cap L^p$, the following bound and error estimates are available:*

$$\begin{aligned} \|X\|_{\mathcal{S}_d^p}^p &\leq C_p \mathbb{E} \left[|X_0|^p + \int_0^T |b_t(0)|^p dt \right. \\ &\quad \left. + \int_0^T |\sigma_t(0)|^p dt + \int_0^T |\delta_t(0)|^p \cdot \varrho_t dt \right] \end{aligned} \tag{12.8}$$

$$\begin{aligned} \|X - X'\|_{\mathcal{S}_d^p}^p &\leq C_p \mathbb{E} \left[|X_0 - X'_0|^p + \int_0^T |b_t(X_t) - b'_t(X_t)|^p dt \right. \\ &\quad \left. + \int_0^T |\sigma_t(X_t) - \sigma'_t(X_t)|^p dt + \int_0^T |\delta_t(X_t) - \delta'_t(X_t)|^p \cdot \varrho_t dt \right] \end{aligned} \tag{12.9}$$

where, in the second line, X' represents a solution to a stochastic differential equation of the form (12.7), with coefficients b' , σ' , δ' and initial condition $X'_0 \in \mathcal{F}_0 \cap L^p$.

In particular, uniqueness holds for (12.7).

12.1.2 Semimartingale Reflected and Doubly Reflected BSDEs

Let there be given a terminal condition ξ and a driver coefficient $g : [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^d \times \mathcal{M}_\varrho \rightarrow \mathbb{R}$ such that

$$(H.0) \quad \xi \in L^2,$$

(H.1.i) $g_\cdot(y, z, v)$ is a progressively measurable process such that $\|g_\cdot(y, z, v)\|_{\mathcal{H}^2} < +\infty$, for $y \in \mathbb{R}$, $z \in \mathbb{R}^d$, $v \in \mathcal{M}_\varrho$,

(H.1.ii) g is uniformly Λ -Lipschitz continuous with respect to (y, z, v) , i.e. Λ is a constant such that for $t \in [0, T]$, $y, y' \in \mathbb{R}$, $z, z' \in \mathbb{R}^d$, $v, v' \in \mathcal{M}_\varrho$, we have:

$$|g_t(y, z, v) - g_t(y', z', v')| \leq \Lambda(|y - y'| + |z - z'| + |v - v'|_t).$$

Note that, given the Lipschitz property (H.1.ii) of g , the requirement that

$$\|g_\cdot(y, z, v)\|_{\mathcal{H}^2} < +\infty \quad \text{for every } y \in \mathbb{R}, z \in \mathbb{R}^d, v \in \mathcal{M}_\varrho$$

in (H.1.i) reduces to $\|g_\cdot(0, 0, 0)\|_{\mathcal{H}^2} < +\infty$.

For reflected BSDEs below we also introduce barriers L and U such that

(H.2.i) L and U are càdlàg processes in \mathcal{S}^2 ,

(H.2.ii) we have \mathbb{P} -almost surely that $L_T \leq \xi \leq U_T$ and

$$L_t \leq U_t, \quad t \in [0, T]. \tag{12.10}$$

Definition 12.1.4

- (a) An (\mathbb{F}, \mathbb{P}) , (B, μ) -solution \mathcal{Y} to the BSDE with the data (g, ξ) is a triple $\mathcal{Y} = (Y, Z, V)$ such that:
 - (i) $Y \in \mathcal{S}^2$, $Z \in \mathcal{H}_d^2$, $V \in \mathcal{H}_\mu^2$
 - (ii) $Y_t = \xi + \int_t^T g_s(Y_s, Z_s, V_s) ds - \int_t^T Z_s dB_s - \int_t^T V_s \cdot d\tilde{\mu}_s$ for every $t \in [0, T]$, \mathbb{P} -a.s.
- (b) An (\mathbb{F}, \mathbb{P}) , (B, μ) -solution \mathcal{Y} to the reflected BSDE (RBSDE for short) with the data (g, ξ, L) is a quadruple $\mathcal{Y} = (Y, Z, V, A)$ such that:
 - (i) $Y \in \mathcal{S}^2$, $Z \in \mathcal{H}_d^2$, $V \in \mathcal{H}_\mu^2$, $A \in \mathcal{A}_+^2$
 - (ii) $Y_t = \xi + \int_t^T g_s(Y_s, Z_s, V_s) ds + A_T - A_t - \int_t^T Z_s dB_s - \int_t^T V_s \cdot d\tilde{\mu}_s$ for every $t \in [0, T]$, \mathbb{P} -a.s.
 - (iii) $L_t \leq Y_t$ for every $t \in [0, T]$, \mathbb{P} -a.s. and $\int_0^T (Y_t - L_t) dA_t = 0$, \mathbb{P} -a.s.
- (c) An (\mathbb{F}, \mathbb{P}) , (B, μ) -solution \mathcal{Y} to the doubly reflected backward stochastic differential equation (R2BSDE for short)⁵ with the data (g, ξ, L, U) is a quadruple $\mathcal{Y} = (Y, Z, V, A)$ such that:

⁵The notation R2BSDE refers to the fact that there are two barriers involved; it has nothing to do with any notion of second-order BSDEs such as in [247].

- (i) $Y \in \mathcal{S}^2, Z \in \mathcal{H}_d^2, V \in \mathcal{H}_\mu^2, A \in \mathcal{A}^2$
- (ii) $Y_t = \xi + \int_t^T g_s(Y_s, Z_s, V_s) ds + A_T - A_t - \int_t^T Z_s dB_s - \int_t^T V_s \cdot d\tilde{\mu}_s$ for every $t \in [0, T]$, \mathbb{P} -a.s.
- (iii) $L_t \leq Y_t \leq U_t$ for every $t \in [0, T]$, \mathbb{P} -a.s. and $\int_0^T (Y_t - L_t) dA_t^+ = \int_0^T (U_t - Y_t) dA_t^- = 0$, \mathbb{P} -a.s.

The components of a solution \mathcal{Y} are respectively called the value process Y , the control processes Z, V and the reflecting processes A^\pm .

We now consider two variants of the above problems involving a $[0, T]$ -valued stopping time ϑ . Note that $(\mathbb{1}_{\cdot \leq \vartheta} g, \xi, L_{\cdot \wedge \vartheta}, , U_{\cdot \wedge \vartheta})$ satisfies (H.0), (H.1) and (H.2) like (g, ξ, L, U) .

Definition 12.1.5 Assume that ξ is \mathcal{F}_ϑ -measurable.

- (a) The RDBSDE⁶ with the data $(g, \xi, L, U, \vartheta)$ is the generalization of an R2BSDE in which the upper barrier U is inactive before ϑ . Formally, we replace U by

$$\overline{U}_t = \mathbb{1}_{\{t < \vartheta\}} \infty + \mathbb{1}_{\{t \geq \vartheta\}} U_t \quad (12.11)$$

in Definition 12.1.4(c)(iii), with the convention that $0 \times \infty = 0$.

- (b) (i) A solution to the stopped BSDE with the data (g, ξ, ϑ) is a triple (Y, Z, V) which satisfies the BSDE with the data $(\mathbb{1}_{\cdot \leq \vartheta} g, \xi)$ so that $Y = Y_\vartheta$ and $Z = V = 0$ on $[\vartheta, T]$.
- (ii) A solution to the stopped RBSDE with the data (g, ξ, L, ϑ) is a quadruple (Y, Z, V, A) which satisfies the RBSDE with the data $(\mathbb{1}_{\cdot \leq \vartheta} g, \xi, L_{\cdot \wedge \vartheta})$ so that $Y = Y_\vartheta, A = A_\vartheta$ and $Z = V = 0$ on $[\vartheta, T]$.
- (iii) A solution to the stopped R2BSDE with the data $(g, \xi, L, U, \vartheta)$, is a quadruple (Y, Z, V, A) which satisfies the R2BSDE with the data $(\mathbb{1}_{\cdot \leq \vartheta} g, \xi, L_{\cdot \wedge \vartheta}, U_{\cdot \wedge \vartheta})$ so that $Y = Y_\vartheta, A = A_\vartheta$ and $Z = V = 0$ on $[\vartheta, T]$.

Remark 12.1.6

- (i) In the special case $\vartheta = 0$ (respectively $\vartheta = T$), the RDBSDE with the data $(g, \xi, L, U, \vartheta)$ reduces to the R2BSDE with the data (g, ξ, L, U) (respectively to the RBSDE with the data (g, ξ, L)).
- (ii) If (Y, Z, V, A) is a solution to the RDBSDE with the data $(g, \xi, L, U, \vartheta)$, then the process

$$(Y_{\cdot \wedge \vartheta}, \mathbb{1}_{\cdot \leq \vartheta} Z, \mathbb{1}_{\cdot \leq \vartheta} V, A_{\cdot \wedge \vartheta})$$

is a solution to the stopped RBSDE with the data $(g, Y_{\cdot \wedge \vartheta}, L, \vartheta)$.

In Crépey and Matoussi [87], BSDEs stopped at a random time are presented as BSDEs with random terminal time in the manner of Darling and Pardoux [91],

⁶The “D” in RDBSDE stands for “delayed”.

only defined over the time interval $[0, \vartheta]$. Such BSDEs stopped at a random time are equivalent to the stopped BSDEs of Definition 12.1.5(b).

It will result from Proposition 12.4.4 that the solution to an RDBSDE is essentially given as the solution to a stopped RBSDE before ϑ , appropriately pasted at ϑ with the solution to an R2BSDE after ϑ . In Chap. 13 we will not deal explicitly with RDBSDEs. Yet, given the above observation, the results of Chap. 13 are applicable to RDBSDEs, giving a way to compute their solutions in two pieces, before and after ϑ .

In Sect. 14.2 we will consider a doubly reflected BSDE with an intermittent upper barrier, or RIBSDE, generalizing an RDBSDE to an effective upper barrier \overline{U} of the form

$$\overline{U}_t = \Omega_t^c \infty + \Omega_t U_t, \quad (12.12)$$

for càdlàg event-processes⁷ Ω_t , $\Omega_t^c = 1 - \Omega_t$ more general than $\Omega_t = \mathbb{1}_{\{t \geq \vartheta\}}$ in (12.11).

Remark 12.1.7 All these definitions admit obvious extensions to problems in which the driver contains a further finite variation process D , not necessarily absolutely continuous.

12.1.2.1 Verification Principle

Originally R2BSDEs were developed in connection with Dynkin games, or optimal stopping game problems (see Lepeltier and Maingueneau [186], Cvitanić and Karatzas [90]). We thus have the following verification principle, which justifies the name of the value process for the component Y in a solution. We state it for an RDBSDE, which, in view of Remark 12.1.6(i), covers RBSDEs and R2BSDEs as special cases. Note that, in the case of an RBSDE (special case $\vartheta = T$), the related Dynkin game reduces to an optimal stopping problem.

Given a $[0, T]$ -valued stopping time θ , let \mathcal{T}_θ , or simply \mathcal{T} for $\theta = 0$, denote the set of $[\theta, T]$ -valued stopping times.

Proposition 12.1.8 *If $\mathcal{Y} = (Y, Z, V, A)$ satisfies the RDBSDE with the data $(g, \xi, L, U, \vartheta)$, then the process Y is the value process of the Dynkin game with payoff functional given, for $t \in [0, T]$ and $(\tau, \theta) \in \mathcal{T}_t \times \mathcal{T}_\vartheta$, by:*

$$J^t(\tau, \theta) = \int_t^{\tau \wedge \theta} g_s(Y_s, Z_s, V_s) ds + \mathbb{1}_{\{\theta \wedge \tau = \tau < T\}} L_\tau + \mathbb{1}_{\{\theta < \tau\}} U_\theta + \mathbb{1}_{\{\theta \wedge \tau = T\}} \xi.$$

More precisely, a saddle-point of that game at time t is given by

$$\tau^t = \inf \{s \in [t, T]; Y_s = L_s\} \wedge T, \quad \theta^t = \inf \{s \in [t \vee \vartheta, T]; Y_s = U_s\} \wedge T.$$

⁷Boolean-valued processes.

So that, for $t \in [0, T]$ and $(\tau, \theta) \in \mathcal{T}_t \times \mathcal{T}_{\vartheta}$,

$$\mathbb{E}[J^t(\tau, \theta^t) | \mathcal{F}_t] \leq Y_t = \mathbb{E}[J^t(\tau^t, \theta^t) | \mathcal{F}_t] \leq \mathbb{E}[J^t(\tau^t, \theta) | \mathcal{F}_t]. \quad (12.13)$$

Proof We prove the right-hand side inequality in (12.13). The left-hand side inequality is shown similarly. Let τ denote $\tau^t \wedge \theta$. By definition of τ^t , we have that A^+ equals 0 on $[t, \tau]$. Since A^- is nondecreasing, taking conditional expectations in the RDBSDE and also using the facts that⁸ $Y_{\tau^t} \leq L_{\tau^t}$ if $\tau^t < T$, $Y_\theta \leq U_\theta$ if $\theta < T$, and $Y_T = \xi$, we obtain:

$$\begin{aligned} Y_t &\leq \mathbb{E}\left(\int_t^\tau g_s(Y_s, Z_s, V_s) ds + Y_\tau | \mathcal{F}_t\right) \\ &\leq \mathbb{E}\left(\int_t^\tau g_s(Y_s, Z_s, V_s) ds + (\mathbb{1}_{\{\tau=\tau^t < T\}} L_{\tau^t} + \mathbb{1}_{\{\theta < \tau^t\}} U_\theta + \mathbb{1}_{\{\tau=T\}} \xi) | \mathcal{F}_t\right). \end{aligned}$$

We conclude that $Y_t \leq \mathbb{E}(J^t(\tau^t, \theta) | \mathcal{F}_t)$ for every $\theta \in \mathcal{T}_{\vartheta}$. \square

12.1.2.2 A Priori Estimates

Like Proposition 12.1.3 regarding a forward SDE, the following a priori bound and error estimates for solutions (assumed to exist) of the various kinds of BSDEs above can be established by repeated applications of the Burkholder-Davis-Gundy inequality and the Gronwall lemma (for proofs, see e.g. [87]).

For doubly reflected problems in particular,⁹ some quasimartingale conditions will be needed. Quasimartingales are dealt with in Protter [228], and in Section VI.40 and Appendix 2.4 of Dellacherie and Meyer [95]. A quasimartingale can be defined as a difference between two nonnegative supermartingales. Among the decompositions of a quasimartingale X as a difference of two nonnegative supermartingales X^1 and X^2 , there exists a unique decomposition $X = \bar{X}^1 - \bar{X}^2$, called the Rao decomposition of X , which is minimal in the sense that $X^1 \geq \bar{X}^1$, $X^2 \geq \bar{X}^2$ for every such decomposition $X = X^1 - X^2$. Any quasimartingale X belonging to \mathcal{S}^2 is a special semimartingale with canonical Doob-Meyer decomposition

$$X_t = X_0 + D_t + M_t, \quad t \in [0, T] \quad (12.14)$$

for a uniformly integrable martingale M and a predictable process of integrable variation D with Jordan components denoted by D^\pm .

Proposition 12.1.9 *Under assumptions (H.0) and (H.1), and also (H.2) for any barrier involved, we consider a sequence of BSDEs, RBSDEs or RDBSDEs, with*

⁸Recalling $\theta \in \mathcal{T}_{\vartheta}$, so that $\theta \geq \vartheta$ and $\bar{U}_\theta = U_\theta$.

⁹Simply reflected BSDEs can be treated without quasimartingale conditions; see [112].

the data and solutions indexed by n , with lower barriers L^n (in the case of reflected problems) given as \mathcal{S}^2 -quasimartingales with the predictable finite variation component of L^n denoted by D^n , and for a common stopping time ϑ in the case of RDBSDEs. The data are assumed to be bounded in the sense that the driver coefficients $g^n = g_t^n(y, z, v)$ are uniformly Λ -Lipschitz continuous in (y, z, v) and that there is a constant c such that:

$$\|\xi^n\|_{L^2}^2 + \|g^n(0, 0, 0)\|_{\mathcal{H}^2}^2 + \|L^n\|_{\mathcal{S}^2}^2 + \|U^n\|_{\mathcal{S}^2}^2 + \|D^{n,-}\|_{\mathcal{S}^2}^2 \leq c, \quad (12.15)$$

where L^n and $D^{n,-}$ (resp. as well as U^n) here, and $A^{n,\pm}$ in (12.16), are to be understood as zero in the case of problems without lower (resp. and upper) barrier. We then have, for some constant $c(\Lambda)$:

$$\|Y^n\|_{\mathcal{S}^2}^2 + \|Z^n\|_{\mathcal{H}_d^2}^2 + \|V^n\|_{\mathcal{H}_\mu^2}^2 + \|A^{n,+}\|_{\mathcal{S}^2}^2 + \|A^{n,-}\|_{\mathcal{S}^2}^2 \leq c(\Lambda)c. \quad (12.16)$$

We also have, indexing by n,p the differences $^{n,-} - {}^{p,+}$ and taking $L^{n,p}$ and/or $U^{n,p}$ and the associated reflecting processes to be zero in the case of problems without lower and/or upper barrier:

$$\begin{aligned} & \|Y^{n,p}\|_{\mathcal{S}^2}^2 + \|Z^{n,p}\|_{\mathcal{H}_d^2}^2 + \|V^{n,p}\|_{\mathcal{H}_\mu^2}^2 + \|A^{n,p}\|_{\mathcal{S}^2}^2 \\ & \leq c(\Lambda)c(\|\xi^{n,p}\|_{L^2}^2 + \|g^{n,p}(Y^n, Z^n, V^n)\|_{\mathcal{H}^2}^2 + \|L^{n,p}\|_{\mathcal{S}^2}^2 + \|U^{n,p}\|_{\mathcal{S}^2}^2). \end{aligned} \quad (12.17)$$

Assume further that $dD^{n,-} \leq c_t^n dt$ for some progressively measurable processes c^n with $\|c^n\|_{\mathcal{H}^2}$ finite for every n . We can then replace $\|L^n\|_{\mathcal{S}^2}^2$ and $\|L^{n,p}\|_{\mathcal{S}^2}^2$ by, respectively, $\|L^n\|_{\mathcal{H}^2}^2$ and $\|L^{n,p}\|_{\mathcal{H}^2}$ in (12.15) and (12.17).

Suppose, additionally, that $\|c^n\|_{\mathcal{H}^2}$ is bounded in n and that, when $n \rightarrow \infty$:

- $g^n(Y, Z, V)$ converges on \mathcal{H}^2 to $g(Y, Z, V)$ locally uniformly with respect to $(Y, Z, V) \in \mathcal{S}^2 \times \mathcal{H}_d^2 \times \mathcal{H}_\mu^2$, and
- (ξ^n, L^n, U^n) converges on $\mathcal{L}^2 \times \mathcal{H}^2 \times \mathcal{S}^2$ to (ξ, L, U) .

Then (Y^n, Z^n, V^n, A^n) converges on $\mathcal{S}^2 \times \mathcal{H}_d^2 \times \mathcal{H}_\mu^2 \times \mathcal{S}^2$ to a solution (Y, Z, V, A) of the BSDE with the data $(g, \xi, L, U, \vartheta)$ (taking L and/or U and the associated reflecting processes to be zero in the case of problems without lower and/or upper barrier). Moreover, (Y, Z, V, A) satisfies (12.16)–(12.17) with $n = \infty$ there.

Finally, in the special case where $L^{n,p}$ and/or $U^{n,p} = 0$, the estimate (12.17) holds with $L^{n,p}$ and/or $U^{n,p} = 0$ there, irrespective of the above specific assumptions on the L_n (beyond the standing assumptions contained in (H.2)).

12.1.2.3 Comparison

In this subsubsection we specialize the assumption (H.1) to the case (H.1)' where, for some nonnegative integer q ,

$$g_t(y, z, v) = \tilde{g}_t(y, z, (v\eta_t) \cdot \varrho_t) \quad (12.18)$$

for an \mathbb{R}_+^q -valued predictable integrand $\eta_t(e)$ with $|\eta_t|_t$ bounded, and for a $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^q)$ -measurable function $\tilde{g} : [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^q \rightarrow \mathbb{R}$ such that:

- (H.1.i)' $\tilde{g}_t(y, z, r)$ is a progressively measurable process, for every $y \in \mathbb{R}$, $z \in \mathbb{R}^d$, $r \in \mathbb{R}^q$;
- (H.1.ii)' $\|\tilde{g}_t(0, 0, 0)\|_{\mathcal{H}^2} < +\infty$;
- (H.1.iii)' $|\tilde{g}_t(y, z, r) - \tilde{g}_t(y', z', r')| \leq A(|y - y'| + |z - z'| + |r - r'|)$, for every $t \in [0, T]$, $y, y' \in \mathbb{R}$, $z, z' \in \mathbb{R}^d$ and $r, r' \in \mathbb{R}^q$;
- (H.1.iv)' $\mathbb{R}^q \ni r \mapsto \tilde{g}_t(y, z, r) \in \mathbb{R}$ is nondecreasing componentwise in r , for every $(t, y, z) \in [0, T] \times \mathbb{R} \times \mathbb{R}^d$.

In particular, using the fact that

$$|((v - v')\eta_t) \cdot \varrho_t| \leq ((v - v')^2 \cdot \varrho_t)^{\frac{1}{2}} (\eta_t^2 \cdot \varrho_t)^{\frac{1}{2}} = |v - v'|_t |\eta_t|$$

with $|\eta_t|_t$ bounded, we see that g defined by (12.18) satisfies (H.1).

We write the following comparison result for RDBSDEs (see [87] for a proof), leaving to readers the expected amended statements in the cases of more specific BSDEs, RBSDEs or R2BSDEs (see also Barles et al. [20] and Royer [238] for related statements and proofs).

Proposition 12.1.10 *Let (Y, Z, V, A) and (Y', Z', V', A') be solutions of the RDBSDEs with the respective data $(g, \xi, L, U, \vartheta)$ and $(g', \xi', L', U', \vartheta')$ that satisfy the assumptions (H.0)–(H.2). We assume further that g satisfies (H.1)'. Then $Y \leq Y'$, $dt \otimes d\mathbb{P}$ -almost everywhere, whenever:*

- (i) $\xi \leq \xi'$, \mathbb{P} -almost surely,
- (ii) $g(Y'_-, Z'_-, V'_-) \leq g'(Y'_-, Z'_-, V'_-)$, $dt \otimes d\mathbb{P}$ -almost everywhere,
- (iii) $L \leq L'$ and $\bar{U} \leq \bar{U}'$, $dt \otimes d\mathbb{P}$ -almost everywhere, where \bar{U} is defined by (12.11) and \bar{U}' is the analogous process with respect to ϑ' .

Remark 12.1.11 In [87], it is postulated that the coefficients only depend on their argument v through one average of v , rather than through q such averages here (the last argument r of \tilde{g} lives in \mathbb{R}^q). However, by inspection of the proof in [87], we can see that the comparison principle which is established there can be elevated from the scalar case to any finite number of averages. Of course this comparison is, as already in the scalar case, subject to a monotonicity condition of \tilde{g} with respect to r , i.e. (H.1.iv)'. See also [238] for such technicalities.

12.1.2.4 Existence and Uniqueness

As a consequence of Proposition 12.1.9 (using, in particular, the last statement there; see [87] for more detail), the assumptions (H.0)–(H.2) are enough to provide uniqueness for a solution to any of the above BSDEs. For existence, we need to postulate the following martingale representation property:

Every square integrable martingale M admits a representation

$$M_t = M_0 + \int_0^t Z_s dB_s + \int_0^t V_s \cdot d\tilde{\mu}s, \quad t \in [0, T] \quad (12.19)$$

for some $Z \in \mathcal{H}_d^2$ and $V \in \mathcal{H}_\mu^2$.

For a càdlàg process, quasi-left continuity is equivalent to continuity at predictable stopping times (see for instance Nikeghbali [211]). Regarding the barriers (whenever involved), we also need to strengthen the assumption (H.2.i) so that it takes the form:

(H.2.i)' L (for a problem with a lower barrier) and U (for a problem with an upper barrier) are càdlàg quasi-left continuous processes in \mathcal{S}^2 .

In the case of problems where two barriers are involved, we also need to strengthen (12.10) in the assumption (H.2.ii) into the following Mokobodski condition:

There exists a quasimartingale X with Rao components in \mathcal{S}^2 such that $L \leq X \leq U$ over $[0, T]$.

The Mokobodski condition (see e.g. [87, 139, 140]) is tantamount to the existence of two nonnegative square integrable supermartingales X^1 and X^2 such that $L \leq X^1 - X^2 \leq U$ over $[0, T]$. In particular, it holds when L is a quasimartingale with Rao components in \mathcal{S}^2 .

Let (H.2)' denote (H.2) with (H.2.i) strengthened into (H.2.i)' (in case at least one lower barrier is involved), and (in case two barriers are involved) with (12.10) in (H.2.ii) strengthened in the form of the Mokobodski condition. Let (H) denote the collection of the assumptions (H.0), (H.1) and (H.2)' along with the above martingale representations assumption. In summary, we then have (see [87] for more detail),

Proposition 12.1.12 *Existence (and uniqueness) of a solution to any of the above BSDEs holds under the assumption (H).*

12.2 Markovian Setup

We now introduce a Markovian specification of the above semimartingale setup. Let the state space \mathcal{E} be given as $[0, T] \times \mathbb{R}^d \times I$, with $I = \{1, \dots, k\}$, for some integers d and k . A function u on $[0, T] \times \mathbb{R}^d \times I$ is equivalently referred to as a system $u = (u^i)_{i \in I}$ of functions $u^i = u^i(t, x)$ on $[0, T] \times \mathbb{R}^d$. Given a domain D of $\mathbb{R}^d \times I$ and integers m and n , a function u is said to be of class $\mathcal{C}^{m,n}(\mathcal{D})$, where $\mathcal{D} = [0, T] \times D$, if u^i is of class $\mathcal{C}^{m,n}$ on \mathcal{D} for every $i \in I$. We denote by \mathcal{G} the following linear operator acting on functions u in $\mathcal{C}^{1,2}(\mathcal{E})$:

$$\begin{aligned}
\mathcal{G}u^i(t, x) &= \partial_t u^i(t, x) + \sum_{1 \leq n \leq d} \partial_{x_n} u^i(t, x) b_n^i(t, x) \\
&\quad + \frac{1}{2} \sum_{1 \leq l, m, n \leq d} \sigma_{m,l}^i(t, x) \sigma_{n,l}^i(t, x) \partial^2_{x_m, x_n} u^i(t, x) \\
&\quad + \int_{\mathbb{R}^d} (u^i(t, x + \delta^i(t, x, y)) - u^i(t, x)) m^i(t, x, dy) \\
&\quad + \sum_{j \in I} (u^j(t, x) - u^i(t, x)) n^{i,j}(t, x).
\end{aligned} \tag{12.20}$$

In (12.20), $m^i(t, x, dy) = f^i(t, x, y) \hat{m}(dy)$, where $\hat{m}(dy)$ is a probability measure on \mathbb{R}^d and all the function coefficients are such that, for every $(t, x, i) \in \mathcal{E}$:

- $\sigma^i(t, x)$ is a d -dimensional diffusion matrix,
- $b^i(t, x)$ is a d -dimensional drift vector coefficient,
- $y \mapsto f^i(t, x, y)$ is a bounded jump intensity function,
- $y \mapsto \delta^i(t, x, y)$ is a bounded jump size function of y , bounded in y locally uniformly in (t, x) ,
- $j \mapsto n^{i,j}(t, x)$ is a nonnegative and bounded regime switching intensity function, from regime i to $j \neq i$, with $n^{i,i}(t, x) = 0$ by convention.

We now introduce a more compact matrix notation. We first define “.” notations in the manner of (12.4) for jump space integrals so that, for every function $\varphi = \varphi^i(t, x, y)$:

$$\begin{aligned}
(\varphi \cdot \hat{m})^i(t, x) &= \varphi^i(t, x) \cdot \hat{m} = \int_{\mathbb{R}^d} \varphi^i(t, x, y) \hat{m}(dy) \\
(\varphi \cdot m)^i(t, x) &= \varphi^i(t, x) \cdot m^i(t, x) = \int_{\mathbb{R}^d} \varphi^i(t, x, y) m^i(t, x, dy) = ((\varphi f) \cdot \hat{m})^i(t, x)
\end{aligned}$$

and, for every function $\psi = \psi^{i,j}(t, x)$:

$$\begin{aligned}
(\psi \cdot n)^i(t, x) &= \psi^i(t, x) \cdot n^i(t, x) = \sum_{j \in I} \psi^{i,j}(t, x) n^{i,j}(t, x) \\
&= \sum_{j \in I \setminus \{i\}} \psi^{i,j}(t, x) n^{i,j}(t, x).
\end{aligned}$$

For every $u = u^i(t, x)$ in $\mathcal{C}^{1,2}(\mathcal{E})$, we let ∂u (respectively $\partial^2 u$) denote the row-gradient (respectively the Hessian matrix) of u with respect to x . Let finally

$$\delta u^i(t, x, y) = u^i(t, x + \delta^i(t, x, y)) - u^i(t, x), \quad \Delta u^{i,j}(t, x) = u^j(t, x) - u^i(t, x).$$

Using the matrix notation, (12.20) takes the form

$$\begin{aligned}
\mathcal{G}u^i(t, x) &= \partial_t u^i(t, x) + \partial u^i(t, x) b^i(t, x) + \frac{1}{2} \partial^2 u^i(t, x) : a^i(t, x) \\
&\quad + \delta u^i(t, x) \cdot m^i(t, x) + \Delta u^{i,j}(t, x) \cdot n^i(t, x),
\end{aligned} \tag{12.21}$$

where $\partial^2 u^i(t, x) : a^i(t, x)$ represents the trace (sum of the diagonal elements) of the product of the Hessian matrix $\partial^2 u^i(t, x)$ by the covariance matrix $a^i(t, x) = \sigma^i(t, x)\sigma^i(t, x)^\top$. We also write $\Delta^{i,j} = j - i$ and $u(t, x) = (u^i(t, x))_{i \in I}$. Whenever convenient we will write $v(t, x, i, \dots)$ rather than $v^i(t, x, \dots)$ for a function v of (t, x, i, \dots) , and $n^j(t, x, i)$ rather than $n^{i,j}(t, x)$. In this spirit, with $\mathcal{X}_t = (X_t, N_t)$ below, the notation $\Delta^j(N_{t-})$, $m(t, \mathcal{X}_t, dy)$ or $n^j(t, \mathcal{X}_t)$ will typically be used rather than $\Delta^{N_{t-}, j}$, $m^{N_t}(t, X_t, dy)$ or $n^{N_t, j}(t, X_t)$.

12.2.1 Dynamics

Definition 12.2.1 A model with generator \mathcal{G} and initial condition (t, x, i) is a triple

$$(\mathbb{F}^t, \mathbb{P}^t), \quad (B^t, \chi^t, v^t), \quad \mathcal{X}^t = (X^t, N^t),$$

where the superscript t refers to an initial condition $(t, x, i) \in \mathcal{E}$, such that $(\Omega, \mathbb{F}^t, \mathbb{P}^t)$ is a stochastic basis on $[t, T]$, with respect to which the following processes and random measures are defined:

- a standard d -dimensional Brownian motion B^t starting at t , along with integer-valued random measures χ^t on $[t, T] \times \mathbb{R}^d$ and v^t on $[t, T] \times I$ such that χ^t and v^t cannot jump simultaneously at stopping times;
- an $\mathbb{R}^d \times I$ -valued process $\mathcal{X}^t = (X^t, N^t)$ on $[t, T]$, with initial condition (x, i) at t , such that for $s \in [t, T]$:

$$\begin{cases} dN_s^t = \Delta(N_{s-}^t) \cdot dv_s^t \\ dX_s^t = b(s, \mathcal{X}_s^t) ds + \sigma(s, \mathcal{X}_s^t) dB_s^t + \delta(s, \mathcal{X}_{s-}^t) \cdot d\chi_s^t, \end{cases} \quad (12.22)$$

where the \mathbb{P}^t -compensated measures \tilde{v}^t of v^t and $\tilde{\chi}^t$ of χ^t are given as

$$\begin{cases} d\tilde{v}_s^t(j) = dv_s^t(j) - n^j(s, \mathcal{X}_s^t) ds \\ \tilde{\chi}^t(ds, dy) = \chi^t(ds, dy) - m(s, \mathcal{X}_s^t, dy) ds \end{cases} \quad (12.23)$$

and where in (12.22) we write

$$\begin{aligned} \delta(s, \mathcal{X}_{s-}^t) \cdot d\chi_s^t &= \int_{\mathbb{R}^d} \delta(s, \mathcal{X}_{s-}^t, y) \chi^t(ds, dy) \\ \Delta(N_{s-}^t) \cdot dv_s^t &= \sum_{j \in I} (j - N_{s-}^t) dv_s^t(j). \end{aligned} \quad (12.24)$$

Thus $v_s^t(j)$ counts the number of transitions of N^t to state j between times t and s , and $\chi^t((t, s] \times O)$ counts the number of jumps of X^t with mark $y \in O$ between times t and s , for every $s \in [t, T]$ and $O \in \mathcal{B}(\mathbb{R}^d)$. We likewise write:

$$\begin{aligned}\delta(s, \mathcal{X}_{s-}^t) \cdot d\tilde{\chi}_s^t &= \int_{\mathbb{R}^d} \delta(s, \mathcal{X}_{s-}^t, y) \chi^t(ds, dy) - \int_{\mathbb{R}^d} \delta(s, \mathcal{X}_s^t, y) m(s, \mathcal{X}_s^t, dy) ds \\ \Delta(N_{s-}^t) \cdot d\tilde{v}_s^t &= \sum_{j \in I} (j - N_{s-}^t) dv_s^t(j) - \sum_{j \in I} (j - N_s^t) n^j(s, \mathcal{X}_s^t) ds.\end{aligned}\quad (12.25)$$

If we suppose that the coefficients b , σ , δ and f do not depend on i , then X is a standard jump-diffusion. If n doesn't depend on x , then N is a continuous-time Markov chain with finite state space I . In general, the above framework defines a rather generic Markovian factor process $\mathcal{X} = (X, N)$ in the form of an N -modulated jump-diffusion-like component X and an X -modulated Markov-chain-like component N . The pure jump process N may be interpreted as defining the so-called regime of X , whence the name jump-diffusion with regimes for \mathcal{X} ; see Pardoux et al. [219] or Becherer and Schweizer [25] for related setups. Constructing a model with mutual dependence between X and N is a nontrivial issue that will be addressed by change of measure in Sect. 12.3.

For simplicity of presentation we do not consider the case of an infinite intensity measure m . The model could be extended to Lévy measures without major changes, but this would entail some technicalities regarding, in particular, the PDE viscosity solutions arguments of Chap. 13.

By application of a “continuous” Itô formula between the jump times of \mathcal{X} and of the relevant correction at jump times, we get the following Itô formula. Note that the correction for the jumps is elementary since jump times can be ordered (we only deal with finite jump intensity measures); see also (3.65) or [149].

Proposition 12.2.2 *Given a model $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, χ^t, v^t) , $\mathcal{X}^t = (X^t, N^t)$ with generator \mathcal{G} , for every function u in $\mathcal{C}^{1,2}(\mathcal{E})$ we have that, for $s \in [t, T]$:*

$$\begin{aligned}du(s, \mathcal{X}_s^t) &= \mathcal{G}u(s, \mathcal{X}_s^t) ds + (\partial u \sigma)(s, \mathcal{X}_s^t) dB_s^t \\ &\quad + \delta u(s, X_{s-}^t) \cdot d\tilde{\chi}_s^t + \Delta u(s, \mathcal{X}_{s-}^t) \cdot d\tilde{v}_s^t.\end{aligned}\quad (12.26)$$

In particular, in the terminology of Sect. 7 of Ethier and Kurtz [119], $(\Omega, \mathbb{F}^t, \mathbb{P}^t)$, \mathcal{X}^t is a solution to the time-dependent local martingale problem with generator¹⁰ \mathcal{G} and initial condition (t, x, i) .

Remark 12.2.3 Once the related semigroup and Markov properties have been established in Propositions 12.4.3, 12.4.10 and Theorem 12.5.1, it is then possible to restrict attention to a single process \mathcal{X} corresponding to an initial condition (t^*, x^*, i^*) of interest. In the context of pricing problems in finance, this initial condition is given as the current values of the underlyings and other model parameters calibrated to the current market data (see Chap. 9). Yet, for deriving these results in this part, it is necessary to consider Markov families of processes \mathcal{X}^t parameterized by an initial condition $(t, x, i) \in \mathcal{E}$.

¹⁰Strictly speaking, the operator \mathcal{G} is the generator of the time-extended process (t, \mathcal{X}_t) .

12.2.1.1 Financial Motivation

The above jump-diffusion with regimes $\mathcal{X} = (X, N)$ can be adapted to most pricing and hedging purposes; see Sect. 4.2.1 for a review of applications where this model is represented, denoted by X , in the stylized formalism of the abstract jump-diffusion (4.30).

In standard applications, the main component of the model, in which the pay-offs of a derivative are expressed, is X . An additional model component N can be used to represent a pricing regime, which may also be viewed as a degenerate form of stochastic volatility (see the corrected problems of Sects. 17.2 and 17.3). More standard diffusive forms of stochastic volatility may also be accounted for in X . The presence of jumps in X is motivated by the empirical evidence of the short-term volatility smile in the market. See e.g. Albanese et al. [2], Lipton and McGhee [194] or Lipton [193] for flexible and tractable stochastic volatility and/or jumps specifications.

In credit and counterparty risk modeling, the main model component (the one which drives the cash flows) is the Markov-chain-like-component N , representing a vector of default status and/or credit ratings of reference obligors; a jump-diffusion-like-component X can be used to represent the evolution of economic variables modulating the dynamics of N . Frailty and default contagion are accounted for by the coupled interaction between N and X (see e.g. Bielecki et al. [36]).

12.2.2 Mapping with the General Set-Up

The model $\mathcal{X}^t = (X^t, N^t)$ can be viewed as a Markovian specification of the semi-martingale setup of Sect. 12.1, where:¹¹

- E is the set $(\mathbb{R}^d \times \{0\}) \cup (\{0_d\} \times I)$, where 0_d represents the origin in \mathbb{R}^d ;
- \mathcal{B}_E is the σ -field generated by $\mathcal{B}(\mathbb{R}^d) \times \{0\}$ and $\{0_d\} \times \mathcal{B}_I$ on E , where $\mathcal{B}(\mathbb{R}^d)$ and \mathcal{B}_I , respectively, represent the Borel σ -field on \mathbb{R}^d and the power set (σ -field of all parts) of I ;
- for every $s \in [t, T]$ and $e = (y, j) \in E$:

$$\varrho(de) = \begin{cases} \widehat{m}(dy) & \text{if } j = 0 \\ 1 & \text{if } y = 0_d, \end{cases} \quad \xi_s^t(e) = \begin{cases} f(t, \mathcal{X}_s^t, y) & \text{if } j = 0 \\ n^j(t, \mathcal{X}_s^t) & \text{if } y = 0_d \end{cases}$$

and $\varrho_s^t = \xi_s^t \varrho$;

- μ^t is the integer-valued random measure on $([t, T] \times E, \mathcal{B}([t, T]) \otimes \mathcal{B}_E)$ counting the jumps of X with mark $y \in O$ and the jumps of N to state j between t and s , for $s \geq t$, $O \in \mathcal{B}(\mathbb{R}^d)$ and $j \in I$.

¹¹Note that the initial time is t here instead of 0 in Sect. 12.1. Superscripts t are thus added, where needed, to the notation of Sect. 12.1.

We summarize this correspondence by the following “ \oplus ” notation:

$$(E, \mathcal{B}_E, \varrho) = (\mathbb{R}^d \oplus I, \mathcal{B}(\mathbb{R}^d) \oplus \mathcal{B}_I, \widehat{m}(dy) \oplus \mathbb{1}),$$

and we also write $\mu^t = \chi^t \oplus \nu^t$ on $([t, T] \times E, \mathcal{B}([t, T]) \otimes \mathcal{B}_E)$. The compensator of the random measure μ^t is thus given, in the “ \oplus ” notation, for $s \geq t$, $O \in \mathcal{B}(\mathbb{R}^d)$ and $j \in I$, by

$$\int_t^s \int_{O \oplus \{j\}} \zeta_r^t(e) \varrho(de) dr = \int_t^s \int_O m(r, \mathcal{X}_r^t, dy) dr + \int_t^s n^j(r, \mathcal{X}_r^t) dr,$$

with $O \oplus \{j\} := (O \times \{0\}) \cup (\{0_d\} \times \{j\})$. Note that $\mathcal{H}_{\mu^t}^2$ can be identified with the product space $\mathcal{H}_{\chi^t}^2 \times \mathcal{H}_{\nu^t}^2$ and that $\mathcal{M}_\varrho = \mathcal{M}(E, \mathcal{B}_E, \varrho; \mathbb{R})$ can be identified with the product space $\mathcal{M}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \widehat{m}(dy); \mathbb{R}) \times \mathbb{R}^k$. These identifications are used freely in the sequel. Thus let \tilde{v} denote a generic pair (v, w) in $\mathcal{M}_\varrho \equiv \mathcal{M}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \widehat{m}(dy); \mathbb{R}) \times \mathbb{R}^k$. Consistent with (12.5), we write for $s \geq t$:

$$|\tilde{v}|_s^t = [(v^2 \cdot m)(s, \mathcal{X}_s^t) + (w^2 \cdot n)(s, \mathcal{X}_s^t)]^{\frac{1}{2}}. \quad (12.27)$$

12.2.3 Cost Functionals

We denote by \mathcal{P} the class of Borel functions u of $x \in \mathbb{R}^d$ and possibly other variables, with polynomial growth in x , uniformly with respect to the other variables. Let there be given a system \mathcal{C} of real continuous cost functions, i.e. a running cost function $g^i(t, x, u, z, r)$ where $(u, z, r) \in \mathbb{R}^k \times \mathbb{R}^d \times \mathbb{R}^q$, a terminal cost function $\Phi^i(x)$, and lower and upper cost functions $\ell^i(t, x)$ and $h^i(t, x)$, such that:

- (M.0) Φ is in \mathcal{P} ;
- (M.1.i) g is in \mathcal{P} , for every $(u, z, r) \in \mathbb{R}^k \times \mathbb{R}^d \times \mathbb{R}^q$;
- (M.1.ii) g is uniformly Λ -Lipschitz continuous with respect to (u, z, r) in the sense that Λ is a constant such that for every $(t, x, i) \in \mathcal{E}$ and $(u, z, r), (u', z', r') \in \mathbb{R}^k \times \mathbb{R}^d \times \mathbb{R}^q$:

$$|g^i(t, x, u, z, r) - g^i(t, x, u', z', r')| \leq \Lambda(|u - u'| + |z - z'| + |r - r'|);$$

- (M.1.iii) g is nondecreasing with respect to r , componentwise in r ;

- (M.2.i) ℓ and h are in \mathcal{P} ;

- (M.2.ii) $\ell \leq h$, $\ell(T, \cdot) \leq \Phi \leq h(T, \cdot)$.

Fixing an initial condition $(t, x, i) \in \mathcal{E}$ for $\mathcal{X} = (X, N)$, we define for every $(s, y, z, \tilde{v}) \in [t, T] \times \mathbb{R} \times \mathbb{R}^d \times \mathcal{M}_\varrho$, with $\tilde{v} = (v, w)$:¹²

$$\tilde{g}(s, \mathcal{X}_s^t, y, z, \tilde{v}) = g(s, \mathcal{X}_s^t, \tilde{u}_s^t, z, \tilde{r}_s^t) - (w \cdot n)(s, \mathcal{X}_s^t), \quad (12.28)$$

¹²Recall our notational conventions such as “ $g^i(t, x, \dots) \equiv g(t, x, i, \dots)$ ”.

where $\tilde{r}_s^t = \tilde{r}_s^t(v)$ and $\tilde{u}_s^t = \tilde{u}_s^t(y, w)$ are defined by

$$\tilde{r}_s^t = ((v\eta) \cdot m)(s, \mathcal{X}_s^t), \quad (\tilde{u}_s^t)^j = \begin{cases} y, & j = N_s^t \\ y + w_j, & j \neq N_s^t \end{cases} \quad (12.29)$$

for an \mathbb{R}_+^q -valued function $\eta = \eta^i(t, x, y)$, bounded in y locally uniformly in (t, x) . The driver coefficient $\tilde{g} = \tilde{g}(s, \mathcal{X}_s^t, y, z, \tilde{v})$ thus only depends on the functions v and w in $\tilde{v} = (v, w)$ through their respective y - and j -integrals¹³ with respect to $\eta(s, \mathcal{X}_s^t, y)m(s, \mathcal{X}_s^t, dy)$ and $n^j(s, \mathcal{X}_s^t)$. Under this structural assumption on the driver and with g nondecreasing in r (componentwise), we will be able to derive a comparison theorem for related BSDEs with jumps below, using Proposition 12.1.10. As will be apparent in the proof of Theorem 13.2.1, a comparison theorem is key to establishing the connection between a BSDE and a related PDE.

In the financial interpretation, in view of identities (12.32)–(12.33) below, we can think of the variables y and w in (12.28)–(12.29) as representing the price and the regime switching deltas. As can be seen by an application of an Itô formula in regular cases and through Malliavin calculus [96, 214] in greater generality, the variable z represents the delta with respect to the space-continuous variable x .

12.2.4 Markovian Decoupled Forward Backward SDE

Given the previous ingredients and an \mathbb{F}^t -stopping time ϑ^t , where the parameter t refers to an initial condition $(t, x, i) \in \mathcal{E}$ of \mathcal{X} , we now define our main decoupled forward backward stochastic differential equation (decoupled FBSDE), encapsulating all the SDEs and BSDEs of interest in this book.¹⁴ Recall that \tilde{v} denotes a generic pair (v, w) in \mathcal{M}_ϱ .

Definition 12.2.4

- (a) A solution to the Markovian decoupled forward backward stochastic differential equation with the data $(\mathcal{G}, \mathcal{C}, \vartheta)$ is a parameterized family of triples

$$\mathcal{Z}^t = (\mathbb{F}^t, \mathbb{P}^t), \quad (B^t, \chi^t, v^t), \quad (\mathcal{X}^t, \mathcal{Y}^t, \bar{\mathcal{Y}}^t),$$

where the superscript t refers to the initial condition $(t, x, i) \in \mathcal{E}$, such that:

- (i) $(\mathbb{F}^t, \mathbb{P}^t), (B^t, \chi^t, v^t), \mathcal{X}^t = (X^t, N^t)$ is a model with generator \mathcal{G} and initial condition (t, x, i) ;
- (ii) $\mathcal{Y}^t = (Y^t, Z^t, \mathcal{V}^t, A^t)$, with $\mathcal{V}^t = (V^t, W^t)$ in $\mathcal{H}_{\mu^t}^2 = \mathcal{H}_{\chi^t}^2 \times \mathcal{H}_{v^t}^2$, is an $(\mathbb{F}^t, \mathbb{P}^t), (B^t, \mu^t)$ -solution to the R2BSDE on $[t, T]$ with the data

$$\tilde{g}(s, \mathcal{X}_s^t, y, z, \tilde{v}), \quad \Phi(\mathcal{X}_T^t), \quad \ell(s, \mathcal{X}_s^t), \quad h(s, \mathcal{X}_s^t); \quad (12.30)$$

¹³A discrete sum in the case of w .

¹⁴Except for slightly more general RIBSDEs to be introduced of Sect. 14.2.

- (iii) $\bar{\mathcal{Y}}^t = (\bar{Y}^t, \bar{Z}^t, \bar{\mathcal{V}}^t, \bar{A}^t)$, with $\bar{\mathcal{V}}^t = (\bar{V}^t, \bar{W}^t)$ in $\mathcal{H}_{\mu^t}^2 = \mathcal{H}_{\chi^t}^2 \times \mathcal{H}_{\nu^t}^2$, is an $(\mathbb{F}^t, \mathbb{P}^t), (B^t, \mu^t)$ -solution to the stopped RBSDE on $[t, T]$ with the data

$$\tilde{g}(s, \mathcal{X}_s^t, y, z, \tilde{v}), \quad Y_{\vartheta^t}^t, \quad \ell(s, \mathcal{X}_s^t), \quad \vartheta^t, \quad (12.31)$$

where Y^t is the value process of \mathcal{Y}^t in (ii).

- (b) The solution is said to be Markovian if:

- (i) $Y_s^t =: u^i(t, x)$ defines, as (t, x, i) varies in \mathcal{E} , a continuous value function of class \mathcal{P} on \mathcal{E} , and we have \mathbb{P}^t -almost surely:

$$Y_s^t = u(s, \mathcal{X}_s^t), \quad s \in [t, T] \quad (12.32)$$

$$\text{for every } j \in I : \quad W_s^t(j) = u^j(s, X_{s-}^t) - u(s, \mathcal{X}_{s-}^t), \quad s \in [t, T] \quad (12.33)$$

$$\begin{aligned} & \int_t^s \tilde{g}(\zeta, \mathcal{X}_\zeta^t, Y_\zeta^t, Z_\zeta^t, \mathcal{V}_\zeta^t) d\zeta \\ &= \int_t^s [g(\zeta, \mathcal{X}_\zeta^t, u(\zeta, X_\zeta^t), Z_\zeta^t, \tilde{r}_\zeta^t) - (\Delta u \cdot n)(\zeta, \mathcal{X}_\zeta^t)] d\zeta, \quad s \in [t, T], \end{aligned} \quad (12.34)$$

with $\tilde{r}_\zeta^t := (V_\zeta^t \eta(\zeta, \mathcal{X}_\zeta^t)) \cdot m(\zeta, \mathcal{X}_\zeta^t)$ in (12.34).

- (ii) $\bar{Y}_t^t =: v^i(t, x)$ defines as (t, x, i) varies in \mathcal{E} , a continuous value function of class \mathcal{P} on \mathcal{E} , and we have \mathbb{P}^t -almost surely:

$$v(\vartheta^t, \mathcal{X}_{\vartheta^t}^t) = u(\vartheta^t, \mathcal{X}_{\vartheta^t}^t) \quad (12.35)$$

$$\bar{Y}_s^t = v(s, \mathcal{X}_s^t), \quad s \in [t, \vartheta^t] \quad (12.36)$$

$$\text{for every } j \in I : \quad \bar{W}_s^t(j) = v^j(s, X_{s-}^t) - v(s, \mathcal{X}_{s-}^t), \quad s \in [t, \vartheta^t] \quad (12.37)$$

$$\begin{aligned} & \int_t^s \tilde{g}(\zeta, \mathcal{X}_\zeta^t, \bar{Y}_\zeta^t, \bar{Z}_\zeta^t, \bar{\mathcal{V}}_\zeta^t) d\zeta \\ &= \int_t^s [g(\zeta, \mathcal{X}_\zeta^t, v(\zeta, X_\zeta^t), \bar{Z}_\zeta^t, \bar{r}_\zeta^t) - (\Delta v \cdot n)(\zeta, \mathcal{X}_\zeta^t)], \quad s \in [t, \vartheta^t], \end{aligned} \quad (12.38)$$

with $\bar{r}_\zeta^t := (\bar{V}_\zeta^t \eta(\zeta, \mathcal{X}_\zeta^t)) \cdot m(\zeta, \mathcal{X}_\zeta^t)$ in (12.38).

The “Markovian consistency conditions” (12.32)–(12.34) or (12.35)–(12.38) in part (b) of these definitions are key to establishing the connection between the BSDE and the PDE approaches in Chap. 13.

12.2.4.1 Markovian Verification Principle

The following proposition is a Markovian counterpart to the semimartingale verification principle of Proposition 12.1.8.

Proposition 12.2.5 *Let $\mathcal{Z}^t = (\mathbb{F}^t, \mathbb{P}^t), (B^t, \chi^t, v^t), (\mathcal{X}^t, \mathcal{Y}^t, \bar{\mathcal{Y}}^t)$ be a Markovian solution to the Markovian decoupled forward backward stochastic differential equation with the data $(\mathcal{G}, \mathcal{C}, \vartheta)$ and with related value functions u and v . Then:*

- (i) *a saddle-point (τ^t, θ^t) of the Dynkin game related to \mathcal{Y}^t is given by*

$$\begin{aligned}\tau^t &= \inf\{s \in [t, T]; (s, \mathcal{X}_s^t) \in \mathcal{E}_+\} \wedge T \\ \theta^t &= \inf\{s \in [t, T]; (s, \mathcal{X}_s^t) \in \mathcal{E}_-\} \wedge T,\end{aligned}$$

with

$$\begin{aligned}\mathcal{E}_+ &= \{(t, x, i) \in [0, T] \times \mathbb{R}^d \times I; u^i(t, x) = \ell^i(t, x)\} \\ \mathcal{E}_- &= \{(t, x, i) \in [0, T] \times \mathbb{R}^d \times I; u^i(t, x) = h^i(t, x)\};\end{aligned}$$

- (ii) *an optimal stopping time τ^t of the optimal stopping problem related to $\bar{\mathcal{Y}}^t$ is given by:*

$$\bar{\tau}^t = \inf\{s \in [t, \vartheta^t]; (s, \mathcal{X}_s^t) \in \mathcal{E}^+\} \wedge T, \quad (12.39)$$

with

$$\bar{\mathcal{E}}_+ = \{(t, x, i) \in [0, T] \times \mathbb{R}^d \times I; v^i(t, x) = \ell^i(t, x)\}.$$

Proof (i) Given the verification principle of Proposition 12.1.8, this follows immediately from identity (12.32) and from the definition of the barriers in (12.30).

- (ii) By the fact that $\bar{\mathcal{Y}}^t$ is stopped at ϑ^t , (12.36) becomes

$$\bar{Y}_s^t = v(s \wedge \vartheta^t, \mathcal{X}_{s \wedge \vartheta^t}^t), \quad s \in [t, T].$$

Using the definition of the barrier in (12.31), the stopping time $\bar{\tau}^t$ defined by (12.39) is hence optimal in the related optimal stopping problem, by application of the verification principle of Proposition 12.1.8 (special case $\vartheta = T$ there). \square

12.2.5 Financial Interpretation

In the risk-neutral pricing problems of Chap. 4, process \mathcal{X} (alias X in Chap. 4) corresponds to a vector of observable factors explaining all the dynamics and cash flows. Moreover:

- $\Phi(\mathcal{X}_T^t)$ corresponds to a terminal payoff that is paid by the issuer to the holder at time T if the contract has not been exercised before T ;
- $\ell(\mathcal{X}_s^t)$, respectively $h(\mathcal{X}_s^t)$, corresponds to a lower (resp. upper) payoff that is paid by the issuer to the holder of the claim in the event of early termination of the contract at the initiative of the holder (resp. issuer).

The stopping time ϑ^t , corresponding to ϑ in Chap. 4, is interpreted as the time of lifting of a call protection. This call protection prevents the issuer of the claim from calling before time ϑ^t . For instance, we have $\vartheta^t = T$ in the case of an American contingent claim, which may only be exercised at the convenience of the holder of the claim.

The driver coefficient \tilde{g} is most commonly given as $c^i(t, x) - \mu^i(t, x)y$, for dividend and interest rate related functions c and μ . Observe that, in order for a Markovian solution \mathcal{Z}^t of the Markovian FBSDE of Definition 12.2.4 to satisfy

$$\int_t^s \tilde{g}(\zeta, \mathcal{X}_\zeta^t, Y_\zeta^t, Z_\zeta^t, \mathcal{V}_\zeta^t) d\zeta = \int_t^s (c(\zeta, \mathcal{X}_\zeta^t) - \mu(\zeta, \mathcal{X}_\zeta^t)Y_\zeta^t) d\zeta, \quad s \in [t, T]$$

$$\int_t^s \tilde{g}(\zeta, \mathcal{X}_\zeta^t, \bar{Y}_\zeta^t, \bar{Z}_\zeta^t, \bar{\mathcal{V}}_\zeta^t) d\zeta = \int_t^s (c(\zeta, \mathcal{X}_\zeta^t) - \mu(\zeta, \mathcal{X}_\zeta^t)\bar{Y}_\zeta^t) d\zeta, \quad s \in [t, \vartheta^t]$$

for given functions c and μ on \mathcal{E} , in view of identities (12.32), (12.34) and (12.36), (12.38) it suffices to set

$$g^i(t, x, u, z, r) = c^i(t, x) - \mu^i(t, x)u^i + (\Delta u \cdot n)^i(t, x). \quad (12.40)$$

This running cost function g doesn't depend on z or r . Modeling the pricing problem under the historical probability (as opposed to a risk-neutral probability as throughout this book) would lead to a driver coefficient function g depending linearly on (z, r) . Also, we assume everywhere a perfect, frictionless financial market. Market imperfections would lead to a nonlinear coefficient g [30, 80, 81, 114].

12.3 Study of the Markovian Forward SDE

Sections 12.3 to 12.5, which conclude with Proposition 12.5.2, are devoted to finding technical conditions on the data \mathcal{G}, \mathcal{C} and ϑ that ensure the existence of a Markovian solution

$$\mathcal{Z}^t = (\mathbb{F}^t, \mathbb{P}^t), \quad (B^t, \chi^t, v^t), \quad (\mathcal{X}^t, \mathcal{Y}^t, \bar{\mathcal{Y}}^t)$$

to our main Markovian decoupled FBSDE. The first step is to construct a solution $\mathcal{X} = (X, N)$ to the Markovian forward SDE with the desired mutual dependence between X and N . Our approach is to start in Sect. 12.3.1 from a model with independent components. In Sect. 12.3.2 we then apply a Markovian change of probability measure which yields a model with mutual dependence under the changed measure.

12.3.1 Homogeneous Case

We thus first consider a data set with coefficients $n, f, b = \hat{n}, \hat{f}, \hat{b}$ and the related generator $\hat{\mathcal{G}}$ such that:

Assumption 12.3.1

- (i) $\hat{f}^i(t, x, y) = 1, \hat{n}^{i,j}(t, x) = \hat{n}^{i,j};$
- (ii) $\hat{b}^i(t, x), \sigma^i(t, x), \delta^i(t, x, y)$ and $\eta^i(t, x, y)$ are Lipschitz in x , uniformly in t, y ;
- (iii) $\hat{b}^i(t, 0), \sigma^i(t, 0), \delta^i(t, 0, y)$ and $\eta^i(t, 0, y)$ are bounded in t, y .

Let there be given a stochastic basis $(\Omega, \mathbb{F}, \mathbb{P})$, with $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$, assumed to support the following processes, independent of each other:

- a standard d -dimensional Brownian motion B ;
- a compound Poisson process P with jump intensity measure $\hat{m}(dy) ds$;
- a continuous-time Markov chain Q on the product state space $I^2 = I \times I$, with jump intensity from (l, j) to (l', j') given by $\hat{n}^{l',j'}$, for every $(l, j) \neq (l', j')$.

Note that, since P and Q are independent of each other and the jump times of P and Q are totally inaccessible, it follows that P and Q cannot jump simultaneously (see e.g. [159]). We denote by χ the random measure $\chi(ds, dy)$ on $[0, T] \times \mathbb{R}^d$ counting the jumps of P with mark y , and by v the random measure $dv_s(l, j)$ on $[0, T] \times I^2$ counting the jumps of Q to the state (l, j) .

Lemma 12.3.2 *The \mathbb{P} -compensated measures $\tilde{\chi}$ of χ and \tilde{v} of v , are given by*

$$\tilde{\chi}(ds, dy) = \chi(ds, dy) - \hat{m}(dy) ds, \quad d\tilde{v}_s(l, j) = dv_s(l, j) - \hat{n}^{l,j} ds. \quad (12.41)$$

Moreover, for every $l \neq j$ in I , $v.(l, j)$ is a \mathbb{P} -Poisson process with intensity $\hat{n}^{l,j}$ on $[0, T]$.

Proof The probability measure is fixed to \mathbb{P} . That $\hat{m}(dy) ds$ compensates χ directly follows from the definition of χ . We prove the results regarding v . Let $\lambda_s(\iota, j)$ denote the measure that counts the number of jumps of the bivariate Markov chain Q from state ι to state $j \neq \iota$ on the time interval $(0, s]$. By the characterization of Markov chains in Lemma 5.1 of Bielecki et al. [42], the compensator ℓ of the measure λ is given as

$$d\ell_s(\iota, j) = \mathbb{1}_{Q_s=\iota} \hat{n}^j ds.$$

Thus, the compensator of the measure $v_s(j)$ counting the jumps of Q to the state j on the time interval $(0, s]$, is given as

$$\sum_l d\ell_s(\iota, j) = \hat{n}^j \sum_l \mathbb{1}_{Q_s=\iota} ds = \hat{n}^j ds.$$

Consequently $\tilde{v}_\cdot(j)$ is a martingale. In view of the Watanabe characterization of a Poisson process,¹⁵ $v_\cdot(j)$ is thus a Poisson process. \square

We now consider the following stochastic differential equation: for $s \in [t, T]$,

$$\begin{cases} dN_s^t = \Delta(N_{s-}^t) \cdot dv_s(N_{s-}^t) := \sum_{j \in I} (j - N_{s-}^t) d\nu_s(N_{s-}^t, j) \\ dX_s^t = \hat{b}(s, \mathcal{X}_s^t) ds + \sigma(s, \mathcal{X}_s^t) dB_s + \delta(s, \mathcal{X}_{s-}^t) \cdot d\chi_s. \end{cases} \quad (12.42)$$

Proposition 12.3.3 *The stochastic differential equation (12.42) on $[t, T]$ with initial condition (x, i) on $[0, t]$ has a unique (\mathbb{F}, \mathbb{P}) -solution $\mathcal{X}^t = (X^t, N^t)$ on $[0, T]$. For any $p \in [2, +\infty)$, we have:*

$$\|X^t\|_{S_d^p}^p \leq C_p(1 + |x|^p) \quad (12.43)$$

$$\|\mathbb{1}_{(s,r)}(X^t - X_s^t)\|_{S_d^p}^p \leq C_p(1 + |x|^p)(r - s), \quad \text{for } s < r. \quad (12.44)$$

Moreover, t' referring to a perturbed initial condition (t', x', i) , we have:

$$\mathbb{P}(N^t \neq N^{t'}) \leq C|t - t'| \quad (12.45)$$

$$\|X^t - X^{t'}\|_{S_d^p}^p \leq C_p(|x - x'|^p + (1 + \bar{x}^p)|t - t'|^\frac{1}{2}), \quad (12.46)$$

where $\bar{x} = |x| \vee |x'|$.

Proof The stochastic differential equation (12.42) with initial condition (x, i) on $[0, t]$ is an (\mathbb{F}, \mathbb{P}) -diffusion SDE with Lipschitz and bounded coefficients between the ordered jump times of v and χ , and with jumps explicitly specified by (12.42) at the jump times of v and χ . This SDE thus has an (\mathbb{F}, \mathbb{P}) -solution $\mathcal{X}^t = (X^t, N^t)$ on $[0, T]$ (see for instance Ikeda and Watanabe [149]).

Let us now show the a priori error estimates (12.43)–(12.44) for solutions to (12.42), where (12.44) also implies the uniqueness for a solution to (12.42) with initial condition (x, i) on $[0, t]$. The estimates (12.8)–(12.9) yield, under Assumption 12.3.1:

$$\|X^t\|_{S_d^p}^p \leq C_p C_p^t \quad (12.47)$$

$$\|\mathbb{1}_{(s,r)}(X^t - X_s^t)\|_{S_d^p}^p \leq C_p C_p^t (r - s) \quad (12.48)$$

$$\|X^t - X^{t'}\|_{S_d^p}^p \leq C_p(C_p^t|t - t'| + C_p^{t,t'}), \quad (12.49)$$

¹⁵See Theorem 5 on p. 25 of Brémaud [56].

with

$$\begin{aligned} C_p^t &= |x|^p + \mathbb{E} \left[\sup_{[t, T]} |\widehat{b}(\cdot, 0, N_s^t)|^p + \sup_{[t, T]} |\sigma(\cdot, 0, N_s^t)|^p + \sup_{[t, T]} (|\delta(\cdot, 0, N_s^t)|^p \cdot \widehat{m}) \right] \\ C_p^{t,t'} &= |x - x'|^p + \mathbb{E} \left[\int_{t \wedge t'}^T |\widehat{b}(s, X_s^t, N_s^t) - \widehat{b}(s, X_s^{t'}, N_s^{t'})|^p ds \right. \\ &\quad + \int_{t \wedge t'}^T |\sigma(s, X_s^t, N_s^t) - \sigma(s, X_s^{t'}, N_s^{t'})|^p ds \\ &\quad \left. + \int_{t \wedge t'}^T (|\delta(s, X_s^t, N_s^t) - \delta(s, X_s^{t'}, N_s^{t'})|^p \cdot \widehat{m}) ds \right]. \end{aligned}$$

The bound estimates (12.43)–(12.44) result from (12.47)–(12.48) by Assumption 12.3.1(iii) on the coefficients. As for the error estimates (12.45)–(12.46), assuming without loss of generality that $t \leq t'$, we have by construction of N via Q in (12.42) that N^t and $N^{t'}$ coincide unless at least one jump of ν occurs on $(t, t']$. Thus:

$$N^t \not\equiv N^{t'} \implies \sum_{j \in I \setminus \{i\}} \nu_{t'}(i, j) > \sum_{j \in I \setminus \{i\}} \nu_t(i, j). \quad (12.50)$$

Now, in view of Lemma 12.3.2, the probability of at least one jump of $\nu(i, j)$ on $(t, t']$ is $1 - e^{-\widehat{n}^{i,j}|t-t'|}$, and therefore

$$\mathbb{P}(N^t \not\equiv N^{t'}) \leq \sum_{j \in I \setminus \{i\}} (1 - e^{-\widehat{n}^{i,j}|t-t'|}) \leq \left(\sum_{j \in I \setminus \{i\}} \widehat{n}^{i,j} \right) |t - t'|,$$

which proves (12.45). Thus

$$\begin{aligned} &\mathbb{E} \int_t^T |\widehat{b}(s, X_s^t, N_s^t) - \widehat{b}(s, X_s^{t'}, N_s^{t'})|^p ds \\ &\leq C |t - t'|^{\frac{1}{2}} \left(\mathbb{E} \int_t^T (|\widehat{b}(s, X_s^t, N_s^t)|^{2p} + |\widehat{b}(s, X_s^{t'}, N_s^{t'})|^{2p}) ds \right)^{\frac{1}{2}}, \end{aligned}$$

where by (12.43) and the properties of \widehat{b} :

$$\mathbb{E} \int_t^T |\widehat{b}(s, X_s^t, N_s^t)|^{2p} ds \leq C \mathbb{E} \int_t^T (|\widehat{b}(s, 0, N_s^t)|^{2p} + |X_s^t|^{2p}) ds \leq C_{2p} (1 + \bar{x}^{2p}),$$

and likewise for $\mathbb{E} \int_t^T |\widehat{b}(s, X_s^{t'}, N_s^{t'})|^{2p} ds$. Thus

$$\mathbb{E} \int_t^T |\widehat{b}(s, X_s^t, N_s^t) - \widehat{b}(s, X_s^{t'}, N_s^{t'})|^p ds \leq C_p (1 + \bar{x}^p) |t - t'|^{\frac{1}{2}}$$

and, by similar estimates regarding the terms in σ and δ of $C_p^{t,t'}$:

$$C_p^{t,t'} \leq |x - x'|^p + C_p(1 + \bar{x}^p)|t - t'|^{\frac{1}{2}}.$$

In view of (12.49), (12.46) results. \square

In the case $k = 1$ (no regimes), we see by inspection of the above proof that $|t - t'|^{\frac{1}{2}}$ is improved to $|t - t'|$ in (12.46).

Given the definition of N^t in the first line of (12.42), an application of Lemma 5.1 in Bielecki et al. [42] yields that N^t is an \mathbb{F} -Markov chain. This Markov property of N^t can also be recovered independently as a by-product of Theorem 12.5.1(i). Note, however, that one of the messages of this part of the book is that Markov properties are not really needed in SDE setups. As we will see, SDE uniqueness results with respect to different filtrations are enough for most related purposes.

Let us further define, for $s \in [t, T]$:

$$B_s^t = B_s - B_t, \quad \chi_s^t = \chi_s - \chi_t, \quad \nu_s^t(j) = \int_t^s d\nu_r(N_{r-}^t, j). \quad (12.51)$$

Note that B^t , χ^t and ν^t are defined over $[t, T]$, whereas B , χ and ν live over $[t, T]$. The random measure ν^t counts the number of jumps of N^t to regime j on $[t, T]$, for every $j \in I$. Of course $dB_s^t = dB_s$ and $\chi^t(ds, dy) = \chi(ds, dy)$; the reason why we introduce B^t and χ^t is mainly for notational consistency with ν^t .

Let \mathbb{F}_{B^t} , \mathbb{F}_{χ^t} , \mathbb{F}_{ν^t} and \mathbb{F}^t respectively represent the filtrations on $[t, T]$ generated by B^t , by χ^t , by ν^t , and by the three of them together. Given an initial condition at time t given in the form of a random variable \tilde{M}_t , with generated σ -field denoted by $\sigma(\tilde{M}_t)$, let $\widehat{\mathbb{F}}_{B^t}$, $\widehat{\mathbb{F}}_{\chi^t}$, $\widehat{\mathbb{F}}_{\nu^t}$ and $\widehat{\mathbb{F}}^t$ represent the filtrations on $[t, T]$ generated by $\sigma(\tilde{M}_t)$ and, respectively, \mathbb{F}_{B^t} , \mathbb{F}_{χ^t} , \mathbb{F}_{ν^t} and \mathbb{F}^t . In (12.52) and later we write

$$W_r \cdot d\tilde{\nu}_r(N_{r-}^t) \quad \text{for} \quad \sum_{j \in I} W_r^j d\nu_r(N_{r-}^t, j) - \sum_{j \in I} \widehat{n}^j(N_r^t) dr.$$

Proposition 12.3.4

- (i) Let \mathcal{X}^t be defined as in Proposition 12.3.3. The stochastic differential equation (12.42) on $[t, T]$ with initial condition (x, i) at t admits a unique $(\mathbb{F}^t, \mathbb{P})$ -solution, given by the restriction of \mathcal{X}^t to $[t, T]$. In particular,

$$(\Omega, \mathbb{F}^t, \mathbb{P}), \quad (B^t, \chi^t, \nu^t), \quad (\mathcal{X}_s^t)_{t \leq s \leq T}$$

is a solution to the time-dependent local martingale problem with generator $\widehat{\mathcal{G}}$ and initial condition (t, x, i) .

- (ii) $(\mathbb{F}^t, \mathbb{P}), (B^t, \chi^t, \nu^t)$ has the local martingale predictable representation property, in the sense that any $(\widehat{\mathbb{F}}^t, \mathbb{P})$ -local martingale M with initial condition \tilde{M}_t at time t admits a representation

$$M_s = M_t + \int_t^s Z_r dB_r + \int_t^s V_r \cdot d\tilde{\chi}_r + \int_t^s W_r \cdot d\tilde{v}_r(N_{r-}^t), \quad s \in [t, T] \quad (12.52)$$

for Z, V, W in the related spaces of predictable integrands.

Proof (i) This is straightforward, given Proposition 12.3.3 and the fact that the restriction of \mathcal{X}^t to $[t, T]$ is \mathbb{F}^t -adapted. That $(\mathbb{F}^t, \mathbb{P}), (B^t, \chi^t, v^t), \mathcal{X}^t$ is a model with generator $\widehat{\mathcal{G}}$ immediately follows from the Itô formula (12.26).

(ii) By Theorem 4.34(a) on p. 189 of Jacod and Shiryaev [153] regarding (12.53)–(12.54), and by the results of Boel et al. [48, 49] regarding (12.55), we have the following predictable representation properties, for Z, V, W in the related spaces of predictable integrands:

- every $(\widetilde{\mathbb{F}}_{B^t}, \mathbb{P})$ -local martingale M with initial condition \tilde{M}_t at time t admits a representation

$$M_s = M_t + \int_t^s Z_r dB_r, \quad s \in [t, T], \quad (12.53)$$

- every $(\widetilde{\mathbb{F}}_{\chi^t}, \mathbb{P})$ -local martingale M with initial condition \tilde{M}_t at time t admits a representation

$$M_s = M_t + \int_t^s V_r \cdot d\tilde{\chi}_r, \quad s \in [t, T], \quad (12.54)$$

- every $(\widetilde{\mathbb{F}}_{v^t}, \mathbb{P})$ -local martingale M with initial condition \tilde{M}_t at time t admits a representation

$$M_s = M_t + \int_t^s W_r \cdot d\tilde{v}_r(N_{r-}^t), \quad s \in [t, T]. \quad (12.55)$$

By independence of B, P and Q , we classically deduce¹⁶ the predictable representation property (12.52) for $(\mathbb{F}^t, \mathbb{P}), (B^t, \chi^t, v^t)$. \square

12.3.2 Inhomogeneous Case

Our next goal is to construct a model \mathcal{X} with state-dependent intensities. Towards this end, we apply to the model of Sect. 12.3.1 a Markovian change of probability measure as in Kunita and Watanabe [168] or Palmowski and Rolski [217]. Let a change of measure function γ be defined as the exponential of a function of class $C^{1,2}$ with compact support on \mathcal{E} . Starting from $\widehat{\mathcal{G}}$, we define an operator \mathcal{G} of the form (12.21) with the data n, f and b as follows, and other data as in $\widehat{\mathcal{G}}$ (where, in particular, $\widehat{f} \equiv 1$), for $(t, x, i) \in \mathcal{E}$:

¹⁶See e.g. Jeanblanc et al. [159].

$$\begin{cases} n^{i,j}(t, x) = \frac{\gamma^j(t, x)}{\gamma^i(t, x)} \hat{n}^{i,j} \\ f^i(t, x, y) = \frac{\gamma^i(t, x + \delta^i(t, x, y))}{\gamma^i(t, x)} \hat{f}^i(t, x, y) \\ b^i(t, x) = \hat{b}^i(t, x) + \delta^i(t, x) \cdot (m^i(t, x) - \hat{m}), \end{cases} \quad (12.56)$$

where here and henceforth we write $m^i(t, x, dy) = f^i(t, x, y) \hat{m}(dy)$.

Lemma 12.3.5

- (i) *The function n is bounded; the function f is positively bounded and Lipschitz in x , uniformly in t, y .*
- (ii) *Let the $(\mathbb{F}^t, \mathbb{P})$ -local martingale Γ^t be defined by $\Gamma_t^t = 1$ and, for $s \in [t, T]$,*

$$\frac{d\Gamma_s^t}{\Gamma_{s-}^t} = \left(\frac{f(s, \mathcal{X}_{s-}^t)}{\hat{f}(s, \mathcal{X}_{s-}^t)} - 1 \right) \cdot d\tilde{\chi}_s + \left(\frac{n(s, \mathcal{X}_{s-}^t)}{\hat{n}(N_{s-}^t)} - 1 \right) \cdot d\tilde{v}_s(N_{s-}^t). \quad (12.57)$$

This yields a positive $(\mathbb{F}^t, \mathbb{P})$ -martingale Γ^t such that $\mathbb{E}\Gamma_s^t = 1$ and, with Γ^t extended by one on $[0, t]$:

$$\|\Gamma^t\|_{\mathcal{S}_d^p}^p \leq C_p. \quad (12.58)$$

Proof (i) is straightforward, given Assumption 12.3.1(ii) and our assumptions on γ .

(ii) By standard Lipschitz SDE arguments (see for instance Ikeda and Watanabe [149]), the stochastic differential equation (12.57) with initial condition 1 on $[0, t]$ has a unique (\mathbb{F}, \mathbb{P}) -solution Γ^t . The bound (12.58) follows by application of the a priori estimate (12.8) to Γ^t . In particular the local martingale Γ^t is a true martingale. \square

Thus, for every $s \in [t, T]$, we can define an equivalent probability measure \mathbb{P}_s^t on $(\Omega, \mathcal{F}_s^t)$ by

$$\frac{d\mathbb{P}_s^t}{d\mathbb{P}} = \Gamma_s^t, \quad \mathbb{P}\text{-a.s.},$$

and we write $\mathbb{P}^t = \mathbb{P}_T^t$. Hence, Γ_s^t is the \mathcal{F}_s^t -measurable version of the Radon-Nikodym density of \mathbb{P}^t with respect to \mathbb{P} on \mathcal{F}_s^t , for every $s \in [t, T]$. We define

$$\begin{cases} \tilde{\chi}^t(ds, dy) = \chi^t(ds, dy) - m(s, \mathcal{X}_s^t, dy) ds \\ d\tilde{v}_s^t(j) = dv_s^t(j) - n^j(s, \mathcal{X}_s^t) ds. \end{cases} \quad (12.59)$$

The proof of the following lemma is provided in Chap. 15.1.1.

Lemma 12.3.6 *B^t is an $(\mathbb{F}^t, \mathbb{P}^t)$ -Brownian motion starting at time t , and $\tilde{\chi}^t$ and \tilde{v}^t are the \mathbb{P}^t -compensated measures of χ^t and v^t .*

We then have:

Proposition 12.3.7

- (i) The restriction to $[t, T]$ of $\mathcal{X}^t = (X^t, N^t)$ in Propositions 12.3.3 and 12.3.4(i) is the unique $(\mathbb{F}^t, \mathbb{P}^t)$ -solution to the following SDE on $[t, T]$: $\mathcal{X}_t^t = (x, i)$ and, for $s \in [t, T]$,

$$\begin{cases} dN_s^t = \Delta(N_{s-}^t) \cdot d\nu_s^t \\ dX_s^t = b(s, \mathcal{X}_s^t) ds + \sigma(s, \mathcal{X}_s^t) dB_s^t + \delta(s, \mathcal{X}_{s-}^t) \cdot d\chi_s^t. \end{cases} \quad (12.60)$$

In particular $(\Omega, \mathbb{F}^t, \mathbb{P}^t), (B^t, \chi^t, \nu^t), \mathcal{X}^t$ is a solution to the time-dependent local martingale problem with generator \mathcal{G} and initial condition (t, x, i) .

- (ii) $(\mathbb{F}^t, \mathbb{P}^t), (B^t, \chi^t, \nu^t)$ has the local martingale predictable representation property, in the sense that for every random variable \tilde{M}_t , any $(\tilde{\mathbb{F}}^t, \mathbb{P}^t)$ -local martingale M with initial condition \tilde{M}_t at time t , where $\tilde{\mathbb{F}}^t$ denotes the filtration on $[t, T]$ generated by \mathbb{F}^t and $\sigma(\tilde{M}_t)$, admits a representation

$$M_s = M_t + \int_t^s Z_r^t dB_r^t + \int_t^s V_r^t \cdot d\tilde{\chi}_r^t + \int_t^s W_r^t \cdot d\tilde{\nu}_r^t, \quad s \in [t, T] \quad (12.61)$$

for Z^t, V^t, W^t in the related spaces of predictable integrands.

Proof (i) In view of (12.56) and (12.59), \mathcal{X}^t is an $(\mathbb{F}^t, \mathbb{P}^t)$ -solution to the stochastic differential equation (12.60) with initial condition (x, i) at time t if and only if it is an $(\mathbb{F}^t, \mathbb{P})$ -solution to the stochastic differential equation (12.42) with initial condition (x, i) at time t . The result hence follows from Proposition 12.3.4(i).

(ii) By Theorem 5.24 on p. 196 of Jacod and Shiryaev [153], a local martingale predictable representation property is preserved under equivalent changes of probability measures. The result thus follows from Proposition 12.3.4(ii). \square

Remark 12.3.8 Instead of Γ_s^t in (12.57), one could work with $\tilde{\Gamma}_s^t$ such that

$$\frac{d\tilde{\Gamma}_s^t}{\tilde{\Gamma}_{s-}^t} = \frac{\partial \gamma \sigma}{\gamma}(s, \mathcal{X}_s^t) dB_s + \left(\frac{f(s, \mathcal{X}_{s-}^t)}{\widehat{f}(s, \mathcal{X}_{s-}^t)} - 1 \right) \cdot d\tilde{\chi}_s + \left(\frac{n(s, \mathcal{X}_{s-}^t)}{\widehat{n}(N_{s-}^t)} - 1 \right) \cdot d\tilde{\nu}_s(N_{s-}^t).$$

As compared with (12.57), this would have the additional effect of further changing the Brownian motion into

$$d\tilde{B}_s^t = dB_s^t - \frac{(\partial \gamma \sigma)^T}{\gamma}(s, \mathcal{X}_s^t) ds \quad (12.62)$$

in (12.59) and to modify accordingly the first-order coefficient of the generator \mathcal{G} of \mathcal{X} .

In Sects. 12.4 and 12.5, we will work with the model

$$(\mathbb{F}^t, \mathbb{P}^t), \quad (B^t, \chi^t, \nu^t), \quad \mathcal{X}^t = (X^t, N^t),$$

with generator $\widehat{\mathcal{G}}$ deduced from $\widehat{\mathcal{G}}$ by the formulas (12.56), with initial condition $t \equiv (t, x, i) \in \mathcal{E}$. We denote \mathbb{P}^t -expectation and \mathbb{P} -expectation by \mathbb{E}^t and \mathbb{E} . The original stochastic basis $(\Omega, \mathbb{F}, \mathbb{P})$ and generator $\widehat{\mathcal{G}}$ will be used in Sects. 12.4 and 12.5 for deriving error estimates with respect to the differences between $(\mathbb{F}^t, \mathbb{P}^t)$ -solutions corresponding to different initial conditions $(t, x, i) \in \mathcal{E}$. Towards this view, in addition to the notation already introduced in Sect. 12.2.2 regarding this process \mathcal{X}^t considered with respect to the stochastic basis $(\Omega, \mathbb{F}^t, \mathbb{P}^t)$, we likewise define for \mathcal{X}^t , considered with respect to $(\Omega, \mathbb{F}, \mathbb{P})$:

- \widehat{E} , the set $(\mathbb{R}^d \times \{0_2\}) \cup (\{0_d\} \times I^2)$;
- $\mathcal{B}_{\widehat{E}}$, the σ -field generated by $\mathcal{B}(\mathbb{R}^d) \times \{0_2\}$ and $\{0_d\} \times \mathcal{B}_{I^2}$ on \widehat{E} , where $\mathcal{B}(\mathbb{R}^d)$ and \mathcal{B}_{I^2} respectively represent the Borel σ -field on \mathbb{R}^d and the powerset of I^2 ;
- for every $t \in [0, T]$ and $e = (y, (l, j)) \in \widehat{E}$:

$$\widehat{\varrho}(de) = \begin{cases} \widehat{m}(dy) & \text{if } (l, j) = 0_2 \\ 1 & \text{if } y = 0_d, \end{cases} \quad \zeta_t(e) = \begin{cases} 1 & \text{if } (l, j) = 0_2 \\ \widehat{n}^{l,j} & \text{if } y = 0_d; \end{cases}$$

- μ , the (\mathbb{F}, \mathbb{P}) -integer-valued random measure on $([0, T] \times \widehat{E}, \mathcal{B}([0, T]) \otimes \mathcal{B}_{\widehat{E}})$ counting the jumps of χ with mark $y \in O$ and the jumps of v to (l, j) between 0 and t , for every $t \geq 0$, $O \in \mathcal{B}(\mathbb{R}^d)$ and $(l, j) \in I^2$.

We also abbreviate, as in Sect. 12.2.2,

$$(\widehat{E}, \mathcal{B}_{\widehat{E}}, \widehat{\varrho}) = (\mathbb{R}^d \oplus I^2, \mathcal{B}(\mathbb{R}^d) \oplus \mathcal{B}_{I^2}, \widehat{m}(dy) \oplus \mathbb{1})$$

and $\mu = \chi \oplus v$. The (\mathbb{F}, \mathbb{P}) -compensator of μ is thus given, for every $t \geq 0$, $O \in \mathcal{B}(\mathbb{R}^d)$ and $(l, j) \in I^2$, with $O \oplus \{(l, j)\} := (O \times \{0_2\}) \cup (\{0_d\} \times \{(l, j)\})$, by:

$$\int_0^t \int_{O \oplus \{(l, j)\}} \zeta_s(e) \widehat{\varrho}(de) ds = \int_0^t \int_O \widehat{m}(dy) ds + \int_0^t \widehat{n}^{l,j} ds.$$

Note that \mathcal{H}_μ^2 can be identified with the product space $\mathcal{H}_\chi^2 \times \mathcal{H}_v^2$ and that $\mathcal{M}_{\widehat{\varrho}} = \mathcal{M}(\widehat{E}, \mathcal{B}_{\widehat{E}}, \widehat{\varrho}; \mathbb{R})$ is identified with the product space $\mathcal{M}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \widehat{m}(dy); \mathbb{R}) \times \mathbb{R}^{k^2}$. For $\widehat{v} = (v, \widehat{w}) \in \mathcal{M}_{\widehat{\varrho}}$, we accordingly write (cf. (12.27)):

$$|\widehat{v}| = \left[v^2 \cdot \widehat{m} + \sum_{l,j \in I} \widehat{w}_{l,j}^2 \widehat{n}^{l,j} \right]^{\frac{1}{2}}. \quad (12.63)$$

12.4 Study of the Markovian BSDEs

We assume that the cost functions \mathcal{C} satisfy the Markovian BSDE assumptions (M.0)–(M.2) of Sect. 12.2.3. Moreover, for problems that involve a lower barrier, we postulate further:

(M.3) $\ell = \phi \vee c$, where c is a constant in $\mathbb{R} \cup \{-\infty\}$ and ϕ in $\mathcal{C}^{1,2}(\mathcal{E})$ is such that

$$\phi, \mathcal{G}\phi, \partial\phi\sigma, (t, x, i) \mapsto |\phi^i(t, x + \delta^i(t, x))| \cdot \hat{m} \in \mathcal{P}, \quad (12.64)$$

and we denote by ρ a bound on the related polynomial orders as well as on those for the data Φ , g , ℓ and h in \mathcal{C} .

By computations detailed in [87], the estimates of Proposition 12.3.3 on \mathcal{X} and (12.64) imply that $\ell(s, \mathcal{X}_s^t)$ is an \mathcal{S}^2 -quasimartingale, implying the Mokobodski condition in (H.2)' for any of our reflected BSDEs (see Sect. 12.1.2.4).

Our standing example for ϕ in (M.3) is $\phi = x_1$, the first component of $x \in \mathbb{R}^d$ (assuming $d \geq 1$), whence $\mathcal{G}\phi = b_1$. In this case (12.64) reduces to

$$b_1, \sigma_1, (t, x, i) \mapsto |\delta_1^i(t, x)| \cdot \hat{m} \in \mathcal{P}.$$

Through this example, the assumption (M.3) is met in most related financial applications.

In part (i) of the following result, building upon the $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, χ^t, ν^t) -martingale representation property of Proposition 12.3.7(ii), we establish the existence and uniqueness for an $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, μ^t) -solution \mathcal{Y}^t to the R2BSDE on $[t, T]$ with the data (12.30). This is then translated in part (ii) in terms of an (\mathbb{F}, \mathbb{P}) , (B, μ) -solution $\tilde{\mathcal{Y}}^t$ to another, auxiliary R2BSDE. The interest in the auxiliary R2BSDE arises because the solutions $\tilde{\mathcal{Y}}^t$ as (t, x, i) varies in \mathcal{E} are defined with respect to a common stochastic basis $(\Omega, \mathbb{F}, \mathbb{P})$. We can then use the a priori estimates of Proposition 12.1.9 for deriving in Proposition 12.4.2 Markovian stability estimates regarding the $\tilde{\mathcal{Y}}^t$. The latter are used in Chap. 13 to obtain an analytic representation of the Y^t , the value processes in the $\tilde{\mathcal{Y}}^t$ and \mathcal{Y}^t .

Proposition 12.4.1

(i) *The R2BSDE on $[t, T]$ with the data*

$$\tilde{g}(s, \mathcal{X}_s^t, y, z, \tilde{v}), \quad \Phi(\mathcal{X}_T^t), \quad \ell(s, \mathcal{X}_s^t), \quad h(s, \mathcal{X}_s^t) \quad (12.65)$$

*has a unique $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, μ^t) -solution $\mathcal{Y}^t = (Y^t, Z^t, \mathcal{V}^t, A^t)$.
(ii) Writing $\mathcal{V}^t = (V^t, W^t)$ with $V^t \in \mathcal{H}_{\chi^t}^2$, $W^t \in \mathcal{H}_{\nu^t}^2$, we extend Y^t by Y_t^t and A^t , Z^t and \mathcal{V}^t by 0 on $[0, t]$, and we define on $[0, T]$:*

$$\tilde{W}_s^t(l, j) = \mathbb{1}_{\{l=N_{s-}^t\}} W_s^t(j) \quad \text{for } l, j \in I, \quad \tilde{\mathcal{V}}^t = (V^t, \tilde{W}^t). \quad (12.66)$$

Then $\tilde{\mathcal{Y}}^t = (Y^t, Z^t, \tilde{\mathcal{V}}^t, A^t)$ is an (\mathbb{F}, \mathbb{P}) , (B, μ) -solution to the R2BSDE on $[0, T]$ with the data

$$\mathbb{1}_{\{s>t\}} \tilde{g}(s, \mathcal{X}_s^t, y, z, \tilde{v}), \quad \Phi(\mathcal{X}_T^t), \quad \ell(s \vee t, \mathcal{X}_{s \vee t}^t), \quad h(s \vee t, \mathcal{X}_{s \vee t}^t), \quad (12.67)$$

where

$$\widehat{g}(s, \mathcal{X}_s^t, y, z, \widehat{v}) := g(s, \mathcal{X}_s^t, \widehat{u}_s^t, z, \widetilde{r}_s^t) + \widetilde{r}_s^t(v) - \sum_{l,j \in I} \widehat{w}_{l,j} \widehat{n}^{l,j}, \quad (12.68)$$

with

$$\widetilde{r}_s^t(v) = v \cdot (m(s, \mathcal{X}_s^t) - \widehat{m}), \quad (\widehat{u}_s^t)^j(y, \widehat{w}) = \begin{cases} y, & j = N_s^t \\ y + \sum_{l \in I} \widehat{w}_{l,j}, & j \neq N_s^t. \end{cases} \quad (12.69)$$

Proof (i) Given assumptions (M.0)–(M.2) and the bound estimates (12.43) on X^t and (12.58) on Γ^t , we have that

$$(H.0)^t \quad \Phi(\mathcal{X}_T^t) \in L^2;$$

(H.1.i)^t $\widetilde{g}(\cdot, \mathcal{X}_\cdot^t, y, z, \widetilde{v})$ is a progressively measurable process on $[t, T]$ such that

$$\mathbb{E}^t \left[\int_t^T \widetilde{g}(\cdot, \mathcal{X}_\cdot^t, y, z, \widetilde{v})^2 dt \right] < +\infty,$$

for every $y \in \mathbb{R}$, $z \in \mathbb{R}^d$, $\widetilde{v} \in \mathcal{M}_\varrho$ (where \mathbb{E}^t denotes \mathbb{P}^t -expectation);

(H.1.ii)^t $\widetilde{g}(\cdot, \mathcal{X}_\cdot^t, y, z, \widetilde{v})$ is uniformly Λ -Lipschitz continuous with respect to (y, z, \widetilde{v}) , in the sense that for any $s \in [t, T]$, $y, y' \in \mathbb{R}$, $z, z' \in \mathbb{R}^d$, $\widetilde{v}, \widetilde{v}' \in \mathcal{M}_\varrho$:¹⁷

$$|\widetilde{g}(s, \mathcal{X}_s^t, y, z, \widetilde{v}) - \widetilde{g}(s, \mathcal{X}_s^t, y', z', \widetilde{v}')| \leq \Lambda(|y - y'| + |z - z'| + |\widetilde{v} - \widetilde{v}'|_s^t);$$

(H.2.i)^t $\ell(s, \mathcal{X}_s^t)$ and $h(s, \mathcal{X}_s^t)$ are càdlàg quasi-left continuous processes in S^2 ;

(H.2.ii)^t $\ell(\cdot, \mathcal{X}_\cdot^t) \leq h(\cdot, \mathcal{X}_\cdot^t)$ on $[t, T]$, and $\ell(T, \mathcal{X}_T^t) \leq \Phi(\mathcal{X}_T^t) \leq h(T, \mathcal{X}_T^t)$.

Given the local martingale predictable representation property of Proposition 12.3.7(ii) and the form postulated in (M.3) for ℓ , so that, in particular, the Mokobodski condition holds, the assumption (H) is thus satisfied by the data (12.65) with respect to $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, μ^t) . Proposition 12.1.12 then ensures existence and uniqueness for an $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, μ^t) -solution $\mathcal{Y}^t = (Y^t, Z^t, \mathcal{V}^t, A^t)$ to the R2BSDE with the data (12.65) on $[t, T]$.

(ii) One thus has, using the definition (12.28) of \widetilde{g} , that for $s \in [t, T]$:

$$\begin{aligned} -dY_s^t &= \widetilde{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \mathcal{V}_s^t) ds + dA_s^t - Z_s^t dB_s - V_s^t \cdot d\widetilde{\chi}_s^t - W_s^t \cdot d\widetilde{v}_s^t \\ &= g(s, \mathcal{X}_s^t, \widehat{u}_s^t, Z_s^t, \widetilde{r}_s^t) ds + dA_s^t - Z_s^t dB_s + V_s^t \cdot (d\widetilde{\chi}_s^t - d\widetilde{\chi}_s^t) \\ &\quad - V_s^t \cdot d\widetilde{\chi}_s^t - W_s^t \cdot dv_s^t. \end{aligned}$$

Given (12.59) and that, in view of (12.66), (12.29) and (12.68),

$$W_s^t \cdot dv_s^t = \sum_{l,j \in I} \widetilde{W}_s^t(l, j) dv_s(l, j), \quad \widetilde{u}_s^t(Y_s^t, W_s^t) = \widehat{u}_s^t(Y_s^t, \widetilde{W}_s^t),$$

¹⁷See (12.27) for the definition of $|\widetilde{v} - \widetilde{v}'|_s^t$.

from (12.56) (where $\widehat{f} \equiv 1$) we have that, for $s \geq t$:

$$\begin{aligned} -dY_s^t &= g(s, \mathcal{X}_s^t, \widehat{u}_s^t, Z_s^t, \widetilde{r}_s^t) ds + dA_s^t - Z_s^t dB_s + V_s^t \cdot (m(s, \mathcal{X}_s^t) - \widehat{m}) ds \\ &\quad - V_s^t(y) \cdot d\widetilde{\chi}_s - \sum_{l,j \in I} \widetilde{W}_s^t(l, j) d\nu_s(j, l). \end{aligned}$$

We can then immediately check that $\widetilde{\mathcal{Y}}^t$ is an $(\mathbb{F}, \mathbb{P}), (B, \mu)$ -solution to the R2BSDE with the data (12.67) on $[0, T]$. \square

By application of the a priori estimates of Proposition 12.1.9 to $\widetilde{\mathcal{Y}}^t$ in Proposition 12.4.1(ii), where the $\widetilde{\mathcal{Y}}^t$ are defined for varying (t, x, i) with respect to the common stochastic basis $(\mathcal{Q}, \mathbb{F}, \mathbb{P})$, we then have the following stability result whose detailed proof is given in Chap. 15.1.2.

Proposition 12.4.2

(i) *The following bound estimate holds:*

$$\|Y^t\|_{\mathcal{S}^2}^2 + \|Z^t\|_{\mathcal{H}_d^2}^2 + \|\widetilde{V}^t\|_{\mathcal{H}_\mu^2}^2 + \|A^{t,+}\|_{\mathcal{S}^2}^2 + \|A^{t,-}\|_{\mathcal{S}^2}^2 \leq C(1 + |x|^{2\rho}). \quad (12.70)$$

(ii) *Moreover, letting t_n refer to a perturbed initial condition $(t_n, x_n, i) \in \mathcal{E}$ with $(t_n, x_n) \rightarrow (t, x)$ as $n \rightarrow \infty$, we have that $\widetilde{\mathcal{Y}}^{t_n}$ converges on $\mathcal{S}^2 \times \mathcal{H}_d^2 \times \mathcal{H}_\mu^2 \times \mathcal{A}^2$ to $\widetilde{\mathcal{Y}}^t$ as $n \rightarrow \infty$.*

12.4.1 Semigroup Properties

Let t refer to a constant initial condition (t, x, i) . Let $\mathcal{X}^t = (X^t, N^t)$ and \mathcal{Y}^t be defined as in Proposition 12.3.3 and Proposition 12.4.1. Given $t' \geq t$, let $\widetilde{\mathbb{F}}^{t'}$ represent $(\widetilde{\mathcal{F}}_r^{t'})_{r \geq t'}$,

$$\widetilde{\mathcal{F}}_r^{t'} = \sigma(\mathcal{X}_{t'}^t) \bigvee \mathcal{F}_r,$$

for $r \geq t'$. As for $\mathbb{F}^{t'} = (\mathcal{F}_r^{t'})_{r \geq t'}$, $\mathbb{P}^{t'}$, $B^{t'}$ and $\mu^{t'}$, they are still defined as in Sects. 12.3.1–12.3.2, with t there replaced by t' . Note, in particular, that $\widetilde{\mathbb{F}}^{t'} \subseteq \mathbb{F}_{|[t', T]}^t$, where $\mathbb{F}_{|[t', T]}^t$ denotes the restriction of \mathbb{F}^t to $[t', T]$. The following result is proved in Chap. 15.1.3.

Proposition 12.4.3

(i) *The stochastic differential equation (12.42), with initial condition $\mathcal{X}_{t'}^t$ at t' , admits a unique $(\widetilde{\mathbb{F}}^{t'}, \mathbb{P})$ -solution $\mathcal{X}^{t'} = (X^{t'}, N^{t'})$ that coincides with the restriction of \mathcal{X}^t to $[t', T]$, so that*

$$\mathcal{X}^{t'} = (X_r^{t'}, N_r^{t'})_{t' \leq r \leq T} = (\mathcal{X}_r^t)_{t' \leq r \leq T}.$$

(ii) *The R2BSDE on $[t', T]$ with the data*

$$\tilde{g}(s, \mathcal{X}_s^{t'}, y, z, \tilde{v}), \quad \Phi(\mathcal{X}_T^{t'}), \quad \ell(s, \mathcal{X}_s^{t'}), \quad h(s, \mathcal{X}_s^{t'}) \quad (12.71)$$

has a unique $(\tilde{\mathbb{F}}^{t'}, \mathbb{P}^{t'}), (B^{t'}, \mu^{t'})$ -solution $\mathcal{Y}^{t'} = (Y_r^{t'}, Z_r^{t'}, \mathcal{V}_r^{t'}, A_r^{t'})_{t' \leq r \leq T}$, which is given by

$$\mathcal{Y}^{t'} = (Y_r^{t'}, Z_r^{t'}, \mathcal{V}_r^{t'}, A_r^{t'})_{t' \leq r \leq T} = (Y_r^t, Z_r^t, \mathcal{V}_r^t, A_r^t - A_{t'}^t)_{t' \leq r \leq T}. \quad (12.72)$$

12.4.2 Stopped Problem

Let ϑ^t denote a stopping time in \mathcal{T}_t , parameterized by the initial condition (t, x, i) of \mathcal{X} .

Proposition 12.4.4

(i) *The RDBSDE on $[t, T]$ with the data*

$$\tilde{g}(s, \mathcal{X}_s^t, y, z, \tilde{v}), \quad \Phi(\mathcal{X}_T^t), \quad \ell(s, \mathcal{X}_s^t), \quad h(s, \mathcal{X}_s^t), \quad \vartheta^t \quad (12.73)$$

has a unique $(\mathbb{F}^t, \mathbb{P}^t), (B^t, \mu^t)$ -solution $\check{\mathcal{Y}}^t = (\check{Y}^t, \check{Z}^t, \check{\mathcal{V}}^t, \check{A}^t)$, such that, in particular, $\check{\mathcal{V}}^t = (\check{V}^t, \check{W}^t)$, for some $\check{V}^t \in \mathcal{H}_{x,t}^2, \check{W}^t \in \mathcal{H}_{v,t}^2$. Moreover, $\check{Y}^t = Y^t$ on $[\vartheta^t, T]$, where Y^t is the value process in the solution \mathcal{Y}^t defined at Proposition 12.4.1(i).

(ii) *We extend \check{Y}^t by \check{Y}_t^t and $\check{A}^t, \check{Z}^t, \check{V}^t, \check{W}^t$ and $\check{\mathcal{V}}^t$ by 0 on $[0, t]$, and we define on $[0, T]$:*

$$\begin{aligned} \overline{Y}^t &= \check{Y}_{\cdot \wedge \vartheta^t}^t, & \overline{Z}^t &= \mathbb{1}_{\cdot \leq \vartheta^t} \check{Z}^t, & \overline{V}^t &= \mathbb{1}_{\cdot \leq \vartheta^t} \check{V}^t, & \overline{W}^t &= \mathbb{1}_{\cdot \leq \vartheta^t} \check{W}^t \\ \overline{\mathcal{V}}^t &= \mathbb{1}_{\cdot \leq \vartheta^t} \check{\mathcal{V}}^t, & \overline{A}^t &= \check{A}_{\cdot \wedge \vartheta^t}^t \\ \widehat{W}^t(l, j) &= \mathbb{1}_{\{l=N_{-}^t\}} \check{W}^t(j) \quad \text{for } l, j \in I, & \widehat{\mathcal{V}}^t &= \mathbb{1}_{\cdot \leq \vartheta^t} (\check{V}^t, \widehat{W}^t) \\ \overline{\mathcal{Y}}^t &= (\overline{Y}^t, \overline{Z}^t, \overline{\mathcal{V}}^t, \overline{A}^t), & \widehat{\mathcal{Y}}^t &= (\overline{Y}^t, \overline{Z}^t, \widehat{\mathcal{V}}^t, \overline{A}^t). \end{aligned}$$

*Then:*¹⁸

- $\overline{\mathcal{Y}}^t$ is an $(\mathbb{F}^t, \mathbb{P}^t), (B^t, \mu^t)$ -solution to the stopped RBSDE on $[t, T]$ with the data

$$\tilde{g}(s, \mathcal{X}_s^t, y, z, \tilde{v}), \quad \check{Y}_{\vartheta^t}^t = Y_{\vartheta^t}^t, \quad \ell(s, \mathcal{X}_s^t), \quad \vartheta^t, \quad (12.74)$$

- $\widehat{\mathcal{Y}}^t$ is an $(\mathbb{F}, \mathbb{P}), (B, \mu)$ -solution to the stopped RBSDE on $[0, T]$ with the data

$$\mathbb{1}_{\{s>t\}} \widehat{g}(s, \mathcal{X}_s^t, y, z, \widehat{v}), \quad Y_{\vartheta^t}^t, \quad \ell(s \vee t, \mathcal{X}_{s \vee t}^t), \quad \vartheta^t. \quad (12.75)$$

¹⁸Recall (12.28) and (12.68) for the definitions of \tilde{g} and \widehat{g} .

Proof (i) As demonstrated in [87], existence of an $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, μ^t) -solution \mathcal{Y}^t to the R2BSDE on $[t, T]$ with the data (12.65) in Proposition 12.4.1(i) implies existence of an $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, μ^t) -solution $\check{\mathcal{Y}}^t = (\check{Y}^t, \check{Z}^t, \check{V}^t, \check{A}^t)$ to the RDBSDE on $[t, T]$ with the data (12.73), such that $\check{Y}^t = Y^t$ on $[\vartheta^t, T]$. Uniqueness holds as explained in Sect. 12.1.2.4.

(ii) This implies, as in the proof of Proposition 12.4.1(ii), that $(\check{Y}^t, \check{Z}^t, \check{V}^t, \check{W}^t)$, \check{A}^t , extended to $[0, T]$ as described in the proposition, is an (\mathbb{F}, \mathbb{P}) , (B, μ) -solution to the RDBSDE on $[0, T]$ with the data

$$\mathbb{1}_{\{s > t\}} \widehat{g}(s, \mathcal{X}_s^t, y, z, \widehat{v}), \quad \Phi(\mathcal{X}_T^t), \quad \ell(t \vee s, \mathcal{X}_{s \vee t}^t), \quad h(t \vee s, \mathcal{X}_{s \vee t}^t), \quad \vartheta^t.$$

The results of part (ii) follow in view of Remark 12.1.6(i). \square

Henceforth in this chapter we will work under the following assumption on ϑ^t .

Assumption 12.4.5 ϑ^t is an almost surely continuous random function of (t, x, i) on \mathcal{E} .

Example 12.4.6 Let $\vartheta^t = \inf\{s \geq t; \mathcal{X}_s^t \notin \mathcal{O}\} \wedge T$ denote the minimum of T and the exit time of \mathcal{X}^t from some subset \mathcal{O} of $\mathbb{R}^d \times I$, for a regular and open set \mathcal{O} , in the sense that, for any $i \in I$:

$$\mathcal{O} \cap (\mathbb{R}^d \times \{i\}) = \{\psi^i > 0\} \quad \text{for some } \psi^i \in C^2(\mathbb{R}^d) \text{ with } |\partial \psi^i| > 0 \text{ on } \{\psi^i = 0\}. \quad (12.76)$$

Assumption 12.4.5 is then typically satisfied under a suitable uniform ellipticity condition on the diffusion coefficient σ of X . See [65] for a precise statement and proof in the case of a diffusion X (case where $\chi \equiv 0$ and there is no component N involved in \mathcal{X}). For related results, see also Darling and Pardoux [91], Theorem 13.8 in Dynkin [106], Freidlin [126], or Assumption A2.2 and the related discussion on p. 281 of Kushner and Dupuis [170].

Under Assumption 12.4.5 we have the following stability results on $\widehat{\mathcal{Y}}^t = (\widehat{Y}^t, \widehat{Z}^t, \widehat{V}^t, \widehat{A}^t)$ in Proposition 12.4.4(ii); see Sect. 15.1.4 for the proof.

Proposition 12.4.7

(i) *The following bound estimate holds:*

$$\|\overline{Y}^t\|_{\mathcal{S}^2}^2 + \|\overline{Z}^t\|_{\mathcal{H}_d^2}^2 + \|\widehat{V}^t\|_{\mathcal{H}_\mu^2}^2 + \|\overline{A}^t\|_{\mathcal{S}^2}^2 \leq C(1 + |x|^{2\rho}). \quad (12.77)$$

(ii) *Moreover, letting t_n refer to a perturbed initial condition $(t_n, x_n, i) \in \mathcal{E}$ with $(t_n, x_n) \rightarrow (t, x)$ as $n \rightarrow \infty$, we have that $\widehat{\mathcal{Y}}^{t_n}$ converges on $\mathcal{S}^2 \times \mathcal{H}_d^2 \times \mathcal{H}_\mu^2 \times \mathcal{A}^2$ to $\widehat{\mathcal{Y}}^t$ as $n \rightarrow \infty$.*

Remark 12.4.8 Assumption 12.4.5 may be relaxed to “sequential continuity of ϑ^t up to extraction of a subsequence”, in the sense that for any $\mathcal{E} \ni (t_n, x_n, k) \rightarrow (t, x, k)$, there exists an extraction $(t_{n'}, x_{n'})_n$ of $(t_n, x_n)_n$ such that, almost surely, $\vartheta^{t_{n'}} \rightarrow \vartheta^t$ as $n \rightarrow \infty$. Then convergence still holds along the extracted subsequence n' in Proposition 12.4.7(ii), which is enough for our later purposes.

12.4.2.1 Semigroup Properties

Let $\mathcal{X}^t = (X^t, N^t)$ and \mathcal{Y}^t be defined as in Sect. 12.4.1, let $\widehat{\mathcal{Y}}^t = (\overline{Y}^t, \overline{Z}^t, \widehat{\mathcal{V}}^t, \overline{A}^t)$ and $\check{\mathcal{V}}^t$ be defined as in Proposition 12.4.4(ii), and let $\overline{\mathcal{X}}^t = (\overline{X}^t, \overline{N}^t)$ represent $\mathcal{X}_{\cdot \wedge \vartheta^t}^t$. Given $t' \geq t$, let $\overline{\mathbb{F}}^{t'} = (\overline{\mathcal{F}}_r^{t'})_{r \geq t'}$ be defined, for $r \in [t', T]$, by

$$\overline{\mathcal{F}}_r^{t'} = \sigma(\overline{\mathcal{X}}_r^{t'}) \vee \mathcal{F}_r^{t'}$$

and let $\vartheta' := t' \vee \vartheta^t$. As for $\mathbb{F}^{t'} = (\mathcal{F}_r^{t'})_{r \geq t'}$, $\mathbb{P}^{t'}$, $B^{t'}$ and $\mu^{t'}$, they are still defined as in Sects. 12.3.1–12.3.2, with t there replaced by t' . In particular, $\overline{\mathbb{F}}^{t'} \subseteq \mathbb{F}_{[t', T]}^t$. We make the following:

Assumption 12.4.9 ϑ' is an $\overline{\mathbb{F}}^{t'}$ -stopping time.

Note that this assumption is satisfied in the situation of Example 12.4.6 (where \mathcal{O} is open). The following result, which is proved in Chap. 15.1.5, is the stopped analog of Proposition 12.4.3.

Proposition 12.4.10

(i) The following forward SDE on $[t', T]$:

$$\begin{cases} d\overline{N}_s^{t'} = \mathbb{1}_{s < \vartheta^t} \left(\sum_{l,j \in I} \mathbb{1}_{\{l = \overline{N}_{s-}^{t'}\}} \Delta^{l,j} v_s(l, j) \right) \\ d\overline{X}_s^{t'} = \mathbb{1}_{s < \vartheta^t} (\widehat{b}(s, \overline{\mathcal{X}}_s^{t'}) ds + \sigma(s, \overline{\mathcal{X}}_s^{t'}) dB_s + \delta(s, \overline{\mathcal{X}}_{s-}^{t'}) \cdot d\chi_s), \end{cases} \quad (12.78)$$

with initial condition $\overline{\mathcal{X}}_{t'}^{t'}$ at t' , admits a unique $(\overline{\mathbb{F}}^{t'}, \mathbb{P})$ -solution $\overline{\mathcal{X}}^{t'} = (\overline{X}^{t'}, \overline{N}^{t'})$, given by the restriction of $\overline{\mathcal{X}}^t$ to $[t', T]$. Thus

$$\overline{\mathcal{X}}^{t'} = (\overline{X}^{t'}, \overline{N}^{t'}) = (\overline{X}_{\cdot \wedge \vartheta^t}^{t'}, \overline{N}_{\cdot \wedge \vartheta^t}^{t'}) = (\overline{\mathcal{X}}_r^{t'})_{t' \leq r \leq T}. \quad (12.79)$$

(ii) The stopped RBSDE on $[t', T]$ with the data

$$\widetilde{g}(s, \overline{\mathcal{X}}_s^{t'}, y, z, \widetilde{v}), \quad Y_{\vartheta^t}^t, \quad \ell(s, \overline{\mathcal{X}}_s^{t'}), \quad \vartheta' \quad (12.80)$$

has a unique $(\bar{\mathbb{F}}^{t'}, \bar{\mathbb{P}}^{t'})$, $(B^{t'}, \mu^{t'})$ -solution $\bar{\mathcal{Y}}^{t'} = (\bar{Y}_r^{t'}, \bar{Z}_r^{t'}, \bar{\mathcal{V}}_r^{t'}, \bar{A}_r^{t'})_{t' \leq r \leq T}$, given by

$$(\bar{Y}_r^{t'}, \bar{Z}_r^{t'}, \bar{\mathcal{V}}_r^{t'}, \bar{A}_r^{t'})_{t' \leq r \leq T} = (\bar{Y}_r^t, \bar{Z}_r^t, \bar{\mathcal{V}}_r^t, \bar{A}_r^t - \bar{A}_{t'}^t)_{t' \leq r \leq T}. \quad (12.81)$$

12.5 Markov Properties

Our next goal is to establish the Markov properties that are expected for the solution \mathcal{X} to the Markovian forward SDE and for the solutions $\mathcal{Y}, \bar{\mathcal{Y}}$ to the Markovian reflected backward SDEs. For any initial condition $(t, x, i) \in \mathcal{E}$, let $\mathcal{Y}^t = (Y^t, Z^t, \mathcal{V}^t, A^t)$, with $\mathcal{V}^t = (V^t, W^t) \in (\mathcal{H}_{\chi^t}^2, \mathcal{H}_{v^t}^2)$, be the $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, μ^t) -solution to the R2BSDE on $[t, T]$ with the data (12.65) of Proposition 12.4.1(i); also let $\bar{\mathcal{Y}}^t = (\bar{Y}^t, \bar{Z}^t, \bar{\mathcal{V}}^t, \bar{A}^t)_{t \leq r \leq T}$ be the unique $(\bar{\mathbb{F}}^t, \bar{\mathbb{P}}^t)$, (B^t, μ^t) -solution to the stopped RBSDE on $[t, T]$ with the data (12.74) of Proposition 12.4.4(ii). The following result is proved in Chap. 15.1.6.

Theorem 12.5.1 As (t, x, i) varies in \mathcal{E} :

- (i) Y_t^t defines a continuous function u of class \mathcal{P} on \mathcal{E} ; the identities of Definition 12.2.4(b)(i) hold between \mathcal{Y}^t and u ;
- (ii) \bar{Y}_t^t defines a continuous function v of class \mathcal{P} on \mathcal{E} ; the identities of Definition 12.2.4(b)(ii) hold between $\bar{\mathcal{Y}}^t, v$ and the function u of part (i).

Moreover, in case the pricing functions u and v are sufficiently regular for an Itô formula to be applicable,¹⁹ one could establish further that

$$\begin{aligned} Z_s^t &= (\partial u \sigma)(s, \mathcal{X}_s^t), \quad s \in [t, T] \\ \bar{Z}_s^t &= (\partial v \sigma)(s, \mathcal{X}_s^t), \quad s \in [t, \vartheta^t]. \end{aligned}$$

To close this chapter, we state the following result concerning the Markovian FBSDE of Definition 12.2.4.

Proposition 12.5.2 Under our standing Markovian assumptions (M.0)–(M.2) and the specific assumptions of Sects. 12.3, 12.4, the Markovian FBSDE with generator \mathcal{G} , cost functions \mathcal{C} and stopping time ϑ has a Markovian solution $\mathcal{Z}^t = (\mathbb{F}^t, \mathbb{P}^t)$, (B^t, χ^t, v^t) , $(\mathcal{X}^t, \mathcal{Y}^t, \bar{\mathcal{Y}}^t)$.

¹⁹Otherwise more general but less constructive representations can be given in terms of Malliavin calculus [96, 214].

Chapter 13

Analytic Approach

In this chapter we derive the companion variational inequality approach to the reflected BSDEs of Chap. 12.

In Sect. 13.1 we introduce systems of partial integro-differential variational inequalities associated with these BSDEs and we state suitable definitions of viscosity solutions for related problems.

Existence, uniqueness and stability issues are dealt with in Sect. 13.2. In Sect. 13.2.1, we show that the value processes (first components) in the solutions of the BSDEs can be characterized in terms of the value functions for related optimal stopping or Dynkin game problems. We then establish in Sect. 13.2.2 a discontinuous viscosity solutions comparison principle that, in particular, implies uniqueness of viscosity solutions for the related obstacle problems. This comparison principle is also used in Sect. 13.2.3 for proving the convergence of stable, monotone and consistent deterministic approximation schemes.

13.1 Viscosity Solutions of Systems of PIDEs with Obstacles

The assumptions of Sects. 12.3, 12.4, including the specification of the Markovian change of probability measure (12.56)–(12.57), are quite involved. In the sequel we give up all these specific assumptions, retaining only the standing Markov setup assumptions (M.0)–(M.2), and postulating in addition that the Markovian FBSDE with the data $\mathcal{G}, \mathcal{C}, \vartheta$ has a Markovian solution

$$\mathcal{Z}^t = (\mathbb{F}^t, \mathbb{P}^t), (B^t, \chi^t, v^t), (\mathcal{X}^t, \mathcal{Y}^t, \bar{\mathcal{Y}}^t).$$

Remark 13.1.1 As illustrated in the previous chapter, this assumption covers various issues such as Lipschitz properties of the forward SDE coefficients b, σ, δ with respect to x , martingale representation properties, some kind of consistency between the drivers B^t, χ^t, v^t as $t \equiv (t, x, i)$ varies in \mathcal{E} , or almost sure continuity of the random function ϑ^t .

Our next goal is to establish the connection between a solution \mathcal{Z}^t under the above-mentioned assumptions, collectively denoted by (M) henceforth, and two systems of obstacle problems denoted by $(\mathcal{V}1)$ and $(\mathcal{V}2)$ below. We will consider this from the point of view of viscosity solutions to these problems. We refer the reader to [16, 17, 22, 27, 28] for alternative approaches in terms of weak Sobolev solutions.

For the sake of existence of solutions to the related obstacle problems, we also postulate in this chapter the following assumption (E), in three parts:

- (E.1) all the coefficients of \mathcal{G} are continuous with respect to (t, x) ;
- (E.2) the functions δ , f and η are locally Lipschitz with respect to (t, x) , uniformly in y ;
- (E.3) ϑ^t is defined in terms of a domain $\mathcal{O} \subseteq \mathbb{R}^d \times I$ as in Example 12.4.6.

Let $\mathcal{D} = [0, T] \times \overline{\mathcal{O}}$, where $\overline{\mathcal{O}}$ denotes the closure¹ of \mathcal{O} in $\mathbb{R}^d \times I$, and let

$$\begin{aligned} \text{Int } \mathcal{E} &= [0, T) \times \mathbb{R}^d \times I, & \partial \mathcal{E} &= \{T\} \times \mathbb{R}^d \times I \\ \text{Int } \mathcal{D} &= [0, T) \times \mathcal{O}, & \partial \mathcal{D} &:= \mathcal{E} \setminus \text{Int } \mathcal{D}. \end{aligned} \quad (13.1)$$

The “thick boundary” $\partial \mathcal{D}$ is motivated by the presence of the jumps in X . Given locally bounded test-functions φ and ϕ on \mathcal{E} with φ of class $C^{1,2}$ about a given point $(t, x, i) \in \mathcal{E}$, we define, in correspondence with the generator \mathcal{G} of \mathcal{X} in (12.21):

$$\begin{aligned} \tilde{\mathcal{G}}(\varphi, \phi)^i(t, x) &= \partial_t \varphi^i(t, x) + \partial \varphi^i(t, x) b^i(t, x) + \frac{1}{2} \partial^2 \varphi^i(t, x) : a^i(t, x) \\ &\quad + (\delta \phi \cdot m)^i(t, x). \end{aligned} \quad (13.2)$$

Also let $\tilde{\mathcal{G}}\varphi$ represent $\tilde{\mathcal{G}}(\varphi, \varphi)$. So, in particular,

$$\tilde{\mathcal{G}}\varphi^i(t, x) = \mathcal{G}\varphi^i(t, x) - (\Delta \varphi \cdot n)^i(t, x) \quad (13.3)$$

Finally, with $J = \{1, \dots, q\}$, let

$$\mathcal{I}^l \phi^i(t, x) = ((\delta \phi \eta^l) \cdot m)^i(t, x), \quad l \in J; \quad \mathcal{I}\phi^i(t, x) = ((\delta \phi \eta) \cdot m)^i(t, x). \quad (13.4)$$

The problems $(\mathcal{V}1)$ and $(\mathcal{V}2)$ that we will now introduce will eventually constitute a cascade of two PDEs, since the boundary condition Ψ in the Cauchy-Dirichlet problem $(\mathcal{V}1)$ will be specified later as the value function u of Definition 12.2.4,² characterized as the unique viscosity solution of class \mathcal{P} to $(\mathcal{V}2)$.

We denote by $(\mathcal{V}2)$ the following variational inequality with double obstacle:

$$\begin{aligned} \max & \left(\min \left(-\tilde{\mathcal{G}}u^i(t, x) - g^i(t, x, u(t, x), (\partial u \sigma)^i(t, x), \mathcal{I}u^i(t, x)), \right. \right. \\ & \left. \left. u^i(t, x) - \ell^i(t, x), u^i(t, x) - h^i(t, x) \right) = 0 \right) \end{aligned}$$

¹In the sense that, for every $i \in I$, $\overline{\mathcal{O}} \cap (\mathbb{R}^d \times \{i\})$ is the closure of $\mathcal{O} \cap (\mathbb{R}^d \times \{i\})$, identified with a subset of \mathbb{R}^d .

²Under the assumption (M).

on $\text{Int } \mathcal{E}$, along with the terminal condition $u = \Phi$ (the terminal cost function Φ in the cost data \mathcal{C}) at T . We also consider the problem $(\mathcal{V}1)$ obtained by formally replacing h by $+\infty$ in $(\mathcal{V}2)$, i.e.

$$\min(-\tilde{\mathcal{G}}u^i(t, x) - g^i(t, x, u(t, x), (\partial u \sigma)^i(t, x), \mathcal{I}u^i(t, x)), u^i(t, x) - \ell^i(t, x)) = 0$$

on $\text{Int } \mathcal{D}$, with a continuous boundary condition Ψ extending Φ on $\partial \mathcal{D}$.

The following lemma is proved in Sect. 15.2.1 (see also p. 297 of Alvarez and Tourin [3]).

Lemma 13.1.2 *The functions $(t, x, i) \rightarrow ((\delta\phi) \cdot m)^i(t, x)$ and $(t, x, i) \rightarrow \mathcal{I}\phi^i(t, x)$ are continuous on \mathcal{E} , for every continuous function ϕ on \mathcal{E} .*

The following definitions are obtained by specifying, for problems $(\mathcal{V}1)$ and $(\mathcal{V}2)$, the standard definitions of viscosity solutions for nonlinear PDEs (see Crandall et al. [75] or Fleming and Soner [122]), adapting the definitions to finite activity jumps and systems of PIDEs, as in [3, 20, 57, 150, 219].

Definition 13.1.3

(a) (i) A locally bounded upper semi-continuous (respectively lower semi-continuous) function u on \mathcal{E} is said to be a viscosity subsolution (respectively supersolution) of $(\mathcal{V}2)$ at $(t, x, i) \in \text{Int } \mathcal{E}$ if and only if one of the following four equivalent conditions holds:

- for any φ in $\mathcal{C}^{1,2}(\mathcal{E})$ such that $u^i - \varphi^i$ is maximum (respectively minimum) at (t, x) , we have

$$\max(\min(-\tilde{\mathcal{G}}(\varphi, u)^i(t, x) - g^i(t, x, u(t, x), (\partial \varphi \sigma)^i(t, x), \mathcal{I}u^i(t, x)), u^i(t, x) - \ell^i(t, x)), u^i(t, x) - h^i(t, x)) \leq 0 \quad (\text{respectively } \geq 0);$$

- $u^i(t, x) \leq h^i(t, x)$ (respectively $u^i(t, x) \geq \ell^i(t, x)$); moreover $u^i(t, x) > \ell^i(t, x)$ (respectively $u^i(t, x) < h^i(t, x)$) implies that

$$\begin{aligned} & -\tilde{\mathcal{G}}(\varphi, u)^i(t, x) - g^i(t, x, u(t, x), (\partial \varphi \sigma)^i(t, x), \\ & \mathcal{I}u^i(t, x)) \leq 0 \quad (\text{respectively } \geq 0) \end{aligned} \tag{13.5}$$

for every φ in $\mathcal{C}^{1,2}(\mathcal{E})$ such that $u^i - \varphi^i$ reaches a null maximum (respectively null minimum) at (t, x) ;

- the analogous statements hold with $\tilde{\mathcal{G}}(\varphi, u)$ and $\mathcal{I}u$ replaced by $\tilde{\mathcal{G}}\varphi$ and $\mathcal{I}\varphi$ in (13.5);
- the analogous statements hold with “maximum” (respectively “minimum”) replaced by “strict maximum” (respectively “strict minimum”).

- (ii) A continuous function u on \mathcal{E} is said to be a viscosity solution of $(\mathcal{V}2)$ at $(t, x, i) \in \text{Int } \mathcal{E}$ if and only if it is both a viscosity subsolution and a viscosity supersolution to $(\mathcal{V}2)$ at (t, x, i) .

- (b) (i) By a \mathcal{P} -viscosity subsolution (respectively supersolution) u of $(\mathcal{V}2)$ on \mathcal{E} for the boundary condition Φ we mean an upper (respectively lower) semi-continuous function of class \mathcal{P} on \mathcal{E} that is a viscosity subsolution (respectively supersolution) of $(\mathcal{V}2)$ on $\text{Int } \mathcal{E}$ and such that $u \leq \Phi$ (respectively $u \geq \Phi$) pointwise at T .
- (ii) By a \mathcal{P} -viscosity solution u of $(\mathcal{V}2)$ on \mathcal{E} we mean a function that is both a \mathcal{P} -subsolution and a \mathcal{P} -supersolution to $(\mathcal{V}2)$ on \mathcal{E} —hence $u = \Phi$ at T .
- (c) The notions of viscosity subsolutions, supersolutions and solutions of $(\mathcal{V}1)$ at $(t, x, i) \in \text{Int } \mathcal{D}$ and, given a continuous boundary condition Ψ extending Φ on $\partial \mathcal{D}$, the notions of \mathcal{P} -viscosity subsolutions, supersolutions and solutions of $(\mathcal{V}1)$ on \mathcal{E} , are defined by immediate adaptation of parts (a) and (b) above, changing there $(\mathcal{V}1)$ to $(\mathcal{V}2)$, $+\infty$ to h , $\text{Int } \mathcal{D}$ to $\text{Int } \mathcal{E}$, $\mathcal{C}^{0,0}(\mathcal{E}) \cap \mathcal{C}^{1,2}(\mathcal{D})$ to $\mathcal{C}^{1,2}(\mathcal{E})$, “on $\partial \mathcal{D}$ ” to “at T ” and Ψ to Φ .

Note that, in these definitions, we consider boundary conditions in the classical rather than in the weak viscosity sense (see the proof of Lemma 13.2.8(ii) and Crandall et al. [75] for more about this).

A classical solution to $(\mathcal{V}1)$ (respectively $(\mathcal{V}2)$) is necessarily a viscosity solution to $(\mathcal{V}1)$ (respectively $(\mathcal{V}2)$).

A viscosity solution u of $(\mathcal{V}2)$ necessarily satisfies $\ell \leq u \leq h$. However, a viscosity subsolution (respectively supersolution) u of $(\mathcal{V}2)$ need not satisfy $u \geq \ell$ (respectively $u \leq h$). Likewise, a viscosity solution v of $(\mathcal{V}1)$ necessarily satisfies $\ell \leq v$, but a viscosity subsolution v of $(\mathcal{V}1)$ need not satisfy $v \geq \ell$.

That $\tilde{\mathcal{G}}(\varphi, u)$ and $\mathcal{I}u$ may equivalently be replaced by $\tilde{\mathcal{G}}\varphi$ and $\mathcal{I}\varphi$ in (13.5), or in the analogous inequalities regarding $(\mathcal{V}1)$, can be shown by adaptation to the present setup of Lemma 3.3 on p. 66 of Barles et al. [20] (see also p. 300 of Alvarez–Tourin [3]), using our monotonicity assumption (M.1.iii) on g .

Since we only consider solutions in the viscosity sense in this book, the notions of subsolution, supersolution and solution are to be understood henceforth as viscosity subsolution, supersolution and solution.

13.2 Study of the PIDEs

13.2.1 Existence

The value functions u and v below are the ones of Definition 12.2.4, assuming (M). The next result, which is proved in Sect. 15.2.2, establishes that u and v are viscosity solutions of the related obstacle problems, u playing the role of a boundary Dirichlet condition for v on $\partial \mathcal{D}$.

Theorem 13.2.1 *Assuming (M) and (E):*

- (i) *the value function u is a \mathcal{P} -solution to $(\mathcal{V}2)$ on \mathcal{E} for the terminal condition Φ at T ;*
- (ii) *the value function v is a \mathcal{P} -solution to $(\mathcal{V}1)$ on \mathcal{E} for the boundary condition u on $\partial \mathcal{D}$.*

13.2.2 Uniqueness

We now consider the issue of uniqueness for a solution to $(\mathcal{V}1)$ or $(\mathcal{V}2)$. We prove a discontinuous comparison principle for these problems, which implies uniqueness for a \mathcal{P} -solution in each case. For related comparison and uniqueness results we refer the reader to Alvarez and Tourin [3], Barles et al. [20, 21], Pardoux et al. [219], Pham [223], Amadori [4, 5] or Ma and Cvitanic [196]. With uniqueness in view, we postulate the following assumption (U) in three parts:

- (U.1) The functions b and σ are locally Lipschitz in (t, x) .
- (U.2) For every $R > 0$ there exists a nonnegative function η_R ,³ continuous and null at 0, such that

$$|g^i(t, x, u, z, r) - g^i(t, x', u, z, r)| \leq \eta_R(|x - x'|)(1 + |z|))$$

for every $t \in [0, T]$, $i \in I$, $z \in \mathbb{R}^d$, $r \in \mathbb{R}^q$ and $x, x' \in \mathbb{R}^d$, $u \in \mathbb{R}^k$, with $|x|, |x'|, |u| \leq R$,

- (U.3) The function g^i is nondecreasing with respect to u^j , for every $(i, j) \in I^2$ with $i \neq j$.

In particular, by (U.1) we have that

$$|b| \vee |\sigma| \vee |\delta| < C(1 + |x|) \quad (13.6)$$

on \mathcal{E} . We are now in a position to establish the following:

Theorem 13.2.2 Assuming (M) , (E) and (U) , we have $\mu \leq v$ on \mathcal{E} for every \mathcal{P} -subsolution μ and \mathcal{P} -supersolution v of $(\mathcal{V}2)$ on \mathcal{E} with terminal condition Φ at T , as well as for every \mathcal{P} -subsolution μ and \mathcal{P} -supersolution v of $(\mathcal{V}1)$ on \mathcal{E} with boundary condition u on $\partial\mathcal{D}$.

We first show that, in the proof, we can restrict our attention to the special case where g^i is nondecreasing with respect to u^j for every $(i, j) \in I^2$ (“ g nondecreasing with respect to u ”), rather than g^i nonincreasing with respect to u^j for $i \neq j$ in (U.3). Thus:

Lemma 13.2.3 If Theorem 13.2.2 holds in the special case where g^i is nondecreasing with respect to u^j for every $(i, j) \in I^2$, then it holds in general.

Proof This is established by application of the special case to the transformed functions $e^{-Rt}\mu^i(t, x)$ and $e^{-Rt}v^i(t, x)$, for sufficiently large R . Indeed, under the general assumptions of Theorem 13.2.2, the functions $e^{-Rt}\mu$ and $e^{-Rt}v$ are, respectively, a \mathcal{P} -subsolution and a \mathcal{P} -supersolution to the following transformed problems:

³Modulus of continuity of g .

- For $(\mathcal{V}2)$: $u = e^{-Rt}\phi$ at T and

$$\max(\min(-\tilde{\mathcal{G}}u^i(t, x) - e^{-Rt}g^i(t, x, e^{Rt}u(t, x), e^{Rt}(\partial u\sigma)^i(t, x), e^{Rt}\mathcal{I}u^i(t, x)) \\ - Ru^i(t, x), u^i(t, x) - e^{-Rt}\ell^i(t, x)), u^i(t, x) - e^{-Rt}h^i(t, x)) = 0$$

on $\text{Int } \mathcal{E}$.

- For $(\mathcal{V}1)$: the similar problem with $h = +\infty$ on $\text{Int } \mathcal{D}$ and with the boundary condition $u = e^{-Rt}\psi$ on $\partial\mathcal{D}$.

For sufficient large R , the assumption (U.3) and the Lipschitz property (M.1.ii) of g imply that $e^{-Rt}g(t, x, e^{Rt}u, e^{Rt}z, e^{Rt}r) + Ru^i$ is nondecreasing with respect to u . We thus conclude, by an application of the assumed special case of Theorem 13.2.2, that $e^{-Rt}\mu \leq e^{-Rt}\nu$, so that $\mu \leq \nu$. \square

Having Lemma 13.2.3, we may now restrict our attention, in the proof of Theorem 13.2.2, to the case where the function g is nondecreasing with respect to u . The proof of the next lemma, given in Sect. 15.2.3, is an adaptation to our circumstances of results from Barles et al. [20] or Pardoux et al. [219]. Let $\widehat{\Lambda} = (k + q)\Lambda$, where we recall that Λ is the Lipschitz constant of g , where k is the cardinality of the index set I of the argument i of g and where q is cardinality of the index set J of the argument $r = (r^l)_{l \in J}$ of g (so that q is the dimension of r).

Lemma 13.2.4 *Under (M), (E), (U), if g is nondecreasing with respect to u then, given a \mathcal{P} -subsolution μ and a \mathcal{P} -supersolution ν to $(\mathcal{V}2)$ on \mathcal{E} :*

- (i) $\omega = \mu - \nu$ is a \mathcal{P} -subsolution to

$$\min(w, -\tilde{\mathcal{G}}\omega - \widehat{\Lambda}(|\partial\omega\sigma| + \max_{j \in I}(\omega^j)^+ + \max_{l \in J}(\mathcal{I}^l\omega)^+)) = 0$$

on \mathcal{E} with null boundary condition at T , in the sense that:

- $\omega \leq 0$ at T , and
- $\omega^i(t, x) > 0$ implies

$$-\tilde{\mathcal{G}}\varphi^i(t, x) - \widehat{\Lambda}(|\partial\varphi^i(t, x)\sigma^i(t, x)| + \max_{j \in I}(\omega^j(t, x))^+ + \max_{l \in J}(\mathcal{I}^l\varphi^i(t, x))^+) \leq 0 \quad (13.7)$$

for every $(t, x, i) \in \text{Int } \mathcal{E}$ and $\varphi \in \mathcal{C}^{1,2}(\mathcal{E})$ such that $\omega^i - \varphi^i$ reaches a null maximum at (t, x) ;

- (ii) for every $p > 0$, there exists $c > 0$ such that the regular function

$$\chi^i(t, x) = (1 + |x|^p)e^{c(T-t)}$$

is a strict \mathcal{P} -supersolution of

$$\min(\chi, -\tilde{\mathcal{G}}\chi - \widehat{\Lambda}(\chi + |\partial\chi\sigma| + \max_{l \in J}(\mathcal{I}^l\chi)^+)) = 0$$

on \mathcal{E} , in the sense that $\chi > 0$ and

$$-\tilde{\mathcal{G}}\chi - \tilde{\Lambda}\left(\chi + |\partial\chi\sigma| + \max_{l \in J}(\mathcal{I}^l\chi)^+\right) > 0 \quad (13.8)$$

on \mathcal{E} ;

- (iii) for p in part (ii) greater than the polynomial order of μ and ν in x (recall that μ, ν are in \mathcal{P}), we have $\max_{i \in I}(\omega^i)^+ \leq \alpha\chi$ on $[0, T] \times \mathbb{R}^d$ for every $\alpha > 0$.

The statement regarding $(V2)$ in Theorem 13.2.2 is obtained by letting α go to 0 in Lemma 13.2.4(iii). The proof of the statement regarding $(V1)$ in Theorem 13.2.2 is similarly obtained by changing $(V1)$ to $(V2)$, $+\infty$ to h , $\text{Int } \mathcal{D}$ to $\text{Int } \mathcal{E}$ and $\mathcal{C}^0(\mathcal{E}) \cap \mathcal{C}^{1,2}(\mathcal{D})$ to $\mathcal{C}^{1,2}(\mathcal{E})$ in Lemma 13.2.4 and its proof.

We can synthesize the findings of Theorems 13.2.1 and 13.2.2 in the form of the following:

Proposition 13.2.5 Assuming (M) , (E) and (U) :

- (i) the value function u is the unique \mathcal{P} -solution, the maximal \mathcal{P} -subsolution and the minimal \mathcal{P} -supersolution of $(V2)$ on \mathcal{E} with terminal condition Φ at T ;
- (ii) the value function v is the unique \mathcal{P} -solution, the maximal \mathcal{P} -subsolution and the minimal \mathcal{P} -supersolution of $(V1)$ on \mathcal{E} with boundary condition u on $\partial\mathcal{D}$.

13.2.3 Approximation

Comparison principles such as Theorem 13.2.2 above ensure well-posedness of the related PIDE problem and the convergence of a wide family of deterministic approximation schemes. These are the so-called stable, monotone and consistent schemes originally introduced for PDEs by Barles and Souganidis [23]; see also Briani, La Chioma and Natalini [57], Cont and Voltchkova [72] and Jakobsen et al. [157] for various extensions to PIDEs. The next lemma is standard and elementary, it is thus stated without proof.

Lemma 13.2.6 Let $(\mathcal{E}_h)_{h>0}$ denote a family of rectangular time-space meshes of step h over \mathcal{E} . Let $(u_h)_{h>0}$ be a family of uniformly locally bounded real functions with u_h defined on the mesh \mathcal{E}_h .

- (i) For any $(t, x, i) \in \mathcal{E}$, the set of limits of the form

$$\lim_{n \rightarrow +\infty} u_{h_n}^i(t_n, x_n) \quad \text{with } h_n \rightarrow 0 \quad \text{and} \\ \mathcal{E}^{h_n} \ni (t_n, x_n, i) \rightarrow (t, x, i) \quad \text{as } n \rightarrow \infty \quad (13.9)$$

is nonempty and compact in \mathbb{R} . It thus admits a smallest element $\underline{u}^i(t, x)$ and a greatest element $\bar{u}^i(t, x)$ in \mathbb{R} .

- (ii) The function \underline{u} (respectively \bar{u}) is locally bounded and lower (respectively upper) semi-continuous on \mathcal{E} . We call it the lower (respectively upper) limit

of $(u_h)_{h>0}$ at (t, x, i) as $h \rightarrow 0$. We say that u_h converges to the limit l at $(t, x, i) \in \mathcal{E}$ as $h \rightarrow 0$ and we write :

$$\lim_{\substack{h \rightarrow 0 \\ \mathcal{E}_h \ni (t_h, x_h, i) \rightarrow (t, x, i)}} u_h^i(t_h, x_h) = l$$

if and only if $\underline{u}^i(t, x) = \bar{u}^i(t, x) = l$ or, equivalently,

$$\lim_{n \rightarrow +\infty} u_{h_n}^i(t_n, x_n) = l$$

for every $h_n \rightarrow 0$ and $\mathcal{E}^{h_n} \ni (t_n, x_n, i) \rightarrow (t, x, i)$ as $n \rightarrow \infty$.

- (iii) If u_h converges pointwise everywhere to a continuous function u on \mathcal{E} , then this convergence is locally uniform, i.e.

$$\max_{\mathcal{E}_h \cap \mathcal{C}} |u_h - u| \rightarrow 0$$

as $h \rightarrow 0$, for every compact set \mathcal{C} of \mathcal{E} .

Definition 13.2.7 Let there be given families of discretized operators

$$\tilde{\mathcal{G}}_h = \tilde{\mathcal{G}}_h u^i(t_h, x_h), \quad \partial_h = \partial_h u^i(t_h, x_h), \quad \mathcal{I}_h = \mathcal{I}_h u^i(t_h, x_h),$$

respectively devoted to approximate $\tilde{\mathcal{G}}u^i(t_h, x_h)$, $\partial u^i(t_h, x_h)$ and $\mathcal{I}u^i(t_h, x_h)$ on \mathcal{E}_h for real functions u on \mathcal{E} . For $\mathcal{L} = \partial$, \mathcal{I} or $\tilde{\mathcal{G}}$, we say that:

- (i) the discretized operator $\mathcal{L}_h = \partial_h$, \mathcal{I}_h or $\tilde{\mathcal{G}}_h$ is monotone if, for every $(t_h, x_h, i) \in \mathcal{E}_h$,

$$\mathcal{L}_h u_1^i(t_h, x_h) \leq \mathcal{L}_h u_2^i(t_h, x_h) \tag{13.10}$$

for all functions $u_1 \leq u_2$ on \mathcal{E}_h with $u_1^i(t_h, x_h) = u_2^i(t_h, x_h)$;

- (ii) the discretization scheme $(\mathcal{L}_h)_{h>0}$ is consistent with \mathcal{L} if and only if, for every continuous function φ on \mathcal{E} of class $\mathcal{C}^{1,2}$ about (t, x, i) , we have

$$\mathcal{L}_h(\varphi + \xi_h)^i(t_h, x_h) \rightarrow \mathcal{L}\varphi^i(t, x) \tag{13.11}$$

whenever $h \rightarrow 0$, $\mathcal{E}_h \ni (t_h, x_h, i) \rightarrow (t, x, i) \in \mathcal{E}$ and $\mathbb{R} \ni \xi_h \rightarrow 0$.

We refer the reader to Cont and Voltchkova [72] for the analysis of specific schemes (verification of the monotonicity and consistency conditions). For the sake of well-posedness and approximation we now assume that g is monotone (jointly nondecreasing) in its last three arguments, as follows:⁴

- (A) Whenever $(u, p, r) \leq (u', p', r')$ coordinate by coordinate in $\mathbb{R}^k \times \mathbb{R}^d \times \mathbb{R}^q$, then

$$g^i(t, x, u, p\sigma^i(t, x), r) \leq g^i(t, x, u', p'\sigma^i(t, x), r')$$

for every $(t, x, i) \in \mathcal{E}$.

⁴(A) suggests “approximation”, for which this extended monotonicity of g is intended.

Under the other prevailing monotonicity assumptions on g in (u, r) already contained in (M.1.iii) and (U.3), the assumption (A) is obviously satisfied in each of the following cases:

- the function $g = g^i(t, x, u, z, r)$ doesn't depend on z , typical in risk-neutral pricing problems (see Sect. 12.2.5);
- in pure jump models with $\sigma = 0$ (but note that in this case Assumption 12.4.5 on ϑ^t fails to be satisfied for ϑ^t defined as in Example 12.4.6 in terms of a domain \mathcal{O} of the form $\{|x| < R\} \times I$);
- the component X of \mathcal{X} is univariate ($d = 1$) and g is nondecreasing in z .

Under the weaker assumption that $g^i(t, x, u, p\sigma^i(t, x), r)$ is nondecreasing with respect to (p, r) and nondecreasing with respect to u^j for $j \neq i$, we then have that, for sufficiently large R , the transformed function $\tilde{u}^i(t, x) := e^{-Rt}u^i(t, x)$ satisfies the assumption (A) (see the proof of Lemma 13.2.3). Suitable approximation schemes $(\tilde{u}_h^i)_{h>0}$ may then be applied to \tilde{u} , resulting in convergent approximation schemes $u_h^i := e^{Rt}\tilde{u}_h^i$ for u .

The following lemma is proved in Sect. 15.2.4. In part (a) we say that $(u_h)_{h>0}$ is uniformly polynomially bounded if u_h is bounded by $C(1 + |x|^p)$ for constants C and p , independent of h .

Lemma 13.2.8 Assuming (M), (E), (U) and (A), let there be given monotone and consistent approximation schemes

$$(\tilde{\mathcal{G}}_h)_{h>0}, \quad (\partial_h)_{h>0} \quad \text{and} \quad (\mathcal{I}_h)_{h>0}$$

for $\tilde{\mathcal{G}}$, ∂ and \mathcal{I} , respectively.

- (a) Let $(u_h)_{h>0}$ be uniformly polynomially bounded and such that

$$\max(\min(-\tilde{\mathcal{G}}_h u_h^i(t_h, x_h) - g^i(t_h, x_h, u_h(t_h, x_h), (\partial_h u_h \sigma)^i(t_h, x_h), \\ \mathcal{I}_h u_h^i(t_h, x_h)), u_h^i(t_h, x_h) - \ell^i(t_h, x_h)), u_h^i(t_h, x_h) - h^i(t_h, x_h)) = 0 \quad (13.12)$$

on $\text{Int } \mathcal{E} \cap \mathcal{E}_h$, and $u_h = \Phi$ on $\partial \mathcal{E} \cap \mathcal{E}_h$, for every $h > 0$. Then:

- (i) the upper and lower limits \bar{u} and \underline{u} of u_h as $h \rightarrow 0$ are respectively viscosity subsolutions and supersolutions of (V2) on $\text{Int } \mathcal{E}$;
- (ii) we have $\bar{u} \leq \Phi \leq \underline{u}$ pointwise at T .

- (b) Let $(v_h)_{h>0}$ be uniformly polynomially bounded and such that

$$\min(-\tilde{\mathcal{G}}_h v_h^i(t_h, x_h) - g^i(t_h, x_h, v_h(t_h, x_h), (\partial_h v_h \sigma)^i(t_h, x_h), \mathcal{I}_h v_h^i(t_h, x_h)), \\ v_h^i(t_h, x_h) - \ell^i(t_h, x_h)) = 0 \quad (13.13)$$

on $\text{Int } \mathcal{D} \cap \mathcal{E}_h$ and $v_h = u$ on $\partial \mathcal{D} \cap \mathcal{E}_h$, for every $h > 0$.

- (i) the upper and lower limits \bar{v} and \underline{v} of v_h as $h \rightarrow 0$, are respectively viscosity subsolutions and supersolutions of (V1) on $\text{Int } \mathcal{D}$;
- (ii) we have $\bar{v} \leq u = \Phi \leq \underline{v}$ pointwise at T .

The following result adapts to models with regimes, thus to systems of PIDEs, the convergence results of [23, 57].

Proposition 13.2.9 *Assuming (M), (E), (U) and (A), let $(u_h)_{h>0}$ (respectively $(v)_{h>0}$) denote a stable, monotone and consistent approximation scheme, in the sense that all conditions in Lemma 13.2.8(a) (respectively (b)) are satisfied for the value function u (respectively v). Then:*

- (a) $u_h \rightarrow u$ locally uniformly on \mathcal{E} as $h \rightarrow 0$;
- (b) $v_h \rightarrow v$ locally uniformly on \mathcal{E} as $h \rightarrow 0$, provided $v_h \rightarrow v (= u)$ on $\partial\mathcal{D} \cap \{t < T\}$.

Proof (a) By Lemma 13.2.8(a), the upper and lower limits \bar{u} and \underline{u} are \mathcal{P} -subolutions and \mathcal{P} -supersolutions of $(V2)$ on \mathcal{E} . Thus $\bar{u} \leq \underline{u}$, by Theorem 13.2.2. Conversely $\underline{u} \leq \bar{u}$ by Lemma 13.2.6(i). Thus $\underline{u} = \bar{u}$, which implies that $u_h \rightarrow u$ locally uniformly on \mathcal{E} as $h \rightarrow 0$, by Lemma 13.2.6(iii).

(b) By Lemma 13.2.8(b)(i), \bar{v} and \underline{v} are respectively viscosity subsolutions and supersolutions of $(V1)$ on $\text{Int}\mathcal{D}$. Moreover, they satisfy $\bar{v} \leq u \leq \underline{v}$ at T by Lemma 13.2.8(b)(ii). If, in addition, $v_h \rightarrow v (= u)$ on $\partial\mathcal{D} \cap \{s < T\}$, then $\bar{v} \leq u \leq \underline{v}$ on $\partial\mathcal{D}$, and \bar{v} and \underline{v} are \mathcal{P} -subolutions and \mathcal{P} -supersolutions of $(V1)$ on \mathcal{E} . The proof is concluded as in part (a). \square

Remark 13.2.10 The convergence result for v in Proposition 13.2.9(b) can only be considered as partial, since we only get conditional convergence on $\partial\mathcal{D} \cap \{t < T\}$, for which no explicit criterion is given. Moreover, the approximation scheme v_h is written under the assumption that the true value for u is substituted for v_h on $\partial\mathcal{D}$ (cf. the boundary condition “ $v_h = u$ on $\partial\mathcal{D} \cap \mathcal{D}_h$ ” in Lemma 13.2.8(b)).

Chapter 14

Extensions

In this chapter we provide various extensions to the previous BSDE and PDE results. These are needed for dealing with practical issues such as discrete dividends or discrete path-dependence. Thus let there be given a set $\mathcal{T} = \{T_0, T_1, \dots, T_m\}$ of fixed times with $0 = T_0 < T_1 < \dots < T_{m-1} < T_m = T$ representing, in the financial interpretation, discrete dividends dates, or monitoring dates in the case of a discretely path-dependent payoff. For $l = 1, \dots, m$, let

$$\mathcal{E}_l = [T_{l-1}, T_l] \times \mathbb{R}^d \times I, \quad \mathcal{D}_l = [T_{l-1}, T_l] \times \mathcal{O},$$

where the set \mathcal{O} was introduced in Example 12.4.6. We define $\text{Int } \mathcal{E}_l$, $\partial \mathcal{E}_l$ and $\text{Int } \mathcal{D}$ as the parabolic interiors and boundaries of \mathcal{E}_l and \mathcal{D}_l in analogy with (13.1). Note that the sets $\text{Int } \mathcal{E}_l$ and $\partial \mathcal{E} = \{T\} \times \mathbb{R}^d \times I$ define a partition of \mathcal{E} .

Section 14.1 deals with discrete dividends on a financial derivative or an underlying asset (and also with discrete path-dependence; see Remark 14.1.5). Section 14.2 extends our previous RDBSDE results to more general RIBSDEs that appear in the case of intermittent call protection (also possibly involving some discrete path-dependence), as opposed to call protection before a stopping time in an RDBSDE.

14.1 Discrete Dividends

14.1.1 Discrete Dividends on a Derivative

Practical derivative payoffs entail discrete coupon tenors, i.e. coupons paid at discrete coupon dates T_l . Now, discrete coupons imply predictable jumps, by the coupon amounts, of the related price process at the T_l . But the value process Y in the solution to a BSDE, which is intended to represent the price process of a financial derivative, jumps at totally unpredictable stopping times. One might thus think that pricing problems with discrete coupons are not amenable to BSDE methods.

However, as demonstrated in [37–40], this can be handled by working with a suitable notion of clean price process of a derivative. Here clean price means the

usual ex-dividend price less the accrued interest at time t , a notion commonly used by market practitioners. Passing to clean prices allows us to restore the continuity in time of the price processes, except for totally unpredictable jumps. However, a side-effect of this transformation is that the resulting running cost function g is no longer continuous, but presents left-discontinuities in time at the T_l . This motivates an extension of the former results to the case of a running cost function g defined by concatenation on the $\text{Int } \mathcal{E}_l$ of functions g_l satisfying our previous assumptions with respect to \mathcal{E}_l . Definition 13.1.3 for viscosity solutions of (V1) and (V2) then needs to be amended as follows:

Definition 14.1.1

- (i) A locally bounded upper semi-continuous (resp. lower semi-continuous, continuous) function u on \mathcal{E} , is said to be a viscosity subsolution (resp. supersolution, solution) of (V2) at (t, x, i) in $\text{Int } \mathcal{E}$, if and only if the restriction of u to \mathcal{E}_l with $(t, x, i) \in \text{Int } \mathcal{E}_l$ is a viscosity subsolution (resp. supersolution, solution) of (V2) at (t, x, i) , with respect to \mathcal{E}_l (cf. Definition 13.1.3(a)).
- (ii) A \mathcal{P} -viscosity subsolution (resp. supersolution, solution) u to (V2) on \mathcal{E} for the boundary condition Φ at T , is formally defined as in Definition 13.1.3(b), with the embedded notions of viscosity subsolution (resp. supersolution, solution) of (V2) at any (t, x, i) in $\text{Int } \mathcal{E}$, defined as in part (i) above.
- (iii) The notions of viscosity subsolutions, supersolutions and solutions of (V1) at $(t, x, i) \in \text{Int } \mathcal{D}$ and, given a further continuous boundary condition Ψ on $\partial \mathcal{D}$ such that $\Psi = \Phi$ at T , those of \mathcal{P} -viscosity subsolutions, supersolutions and solutions of (V1) on \mathcal{E} , are defined similarly (cf. Definition 13.1.3(c)).

Proposition 14.1.2 *Using Definition 14.1.1 for the notions of the viscosity solutions involved, the results of Chaps. 12 and 13 remain true under the currently relaxed assumption on g .*

Proof In Chap. 12, the continuity of g was used, first to ensure well-definedness of process $\tilde{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \mathcal{V}_s^t)$ (cf. (12.28)) for every $(Y^t, Z^t, \mathcal{V}^t) \in \mathcal{S}^2 \times \mathcal{H}_d^2 \times \mathcal{H}_{\mu^t}^2$, and then for deriving the stability results of Propositions 12.4.2(ii) and 12.4.7(ii). One can check by inspection of the related proofs that these stability results still hold under the currently relaxed assumption on g . Moreover the process $\tilde{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \mathcal{V}_s^t)$ is obviously still well-defined under the current assumption on g , for every $(Y^t, Z^t, \mathcal{V}^t) \in \mathcal{S}^2 \times \mathcal{H}_d^2 \times \mathcal{H}_{\mu^t}^2$.

Theorem 13.2.1 of Chap. 13 still holds true, by inspection of its proof. Moreover, under an “ l by l ” version of the assumption (U.2) on the g_l , Lemma 13.2.4 and Theorem 13.2.2 can be proved together iteratively on l as follows. Let μ and ν denote a \mathcal{P} -subsolution and a \mathcal{P} -supersolution v of (V2) on \mathcal{E} (the proof would be analogous for (V1)). Lemma 13.2.4 with respect to \mathcal{E}_m is proved as before. We thus have (cf. Theorem 13.2.2) $\mu \leq v$ on \mathcal{E}_m , and in particular at T_m . Thus $\mu - v \leq 0$ on $\partial \mathcal{E}_{m-1}$ by Lemma 13.2.4 with respect to \mathcal{E}_{m-1} , and so on until $l = 1$. Lemma 13.2.6 isn’t affected by the relaxation of the assumption on g . Finally, given Definition 14.1.1, Lemma 13.2.8(a)(i) can be proved exactly as before, on each $\text{Int } \mathcal{E}_l$, and the proof

of Lemma 13.2.8(a)(ii) doesn't change. Lemma 13.2.8(a) is thus still true, and so likewise is Lemma 13.2.8(b), hence Proposition 13.2.9 holds as before. \square

14.1.2 Discrete Dividends on Underlying Assets

Having just considered dividends on a financial derivative with factor process $\mathcal{X} = (X, N)$, we now want to deal with discrete dividends at times T_l on an underlying asset, with price process given as a component of X . Note that so far the process \mathcal{X} cannot jump at a fixed time T_l , since the jump times of the driving random measures χ and ν are totally inaccessible. We thus enrich \mathcal{X} by the introduction of deterministic jumps in X at the T_l of the form

$$X_{T_l} = \theta_l(\mathcal{X}_{T_l-}), \quad (14.1)$$

where a jump function θ is given as a Lipschitz function $y \mapsto \theta_l^j(y)$ from \mathbb{R}^d into itself, for every $i \in I$ and $l = 1, \dots, m$.

Definition 14.1.3

- (i) A Cauchy cascade $(\Phi, (u_l)_{1 \leq l \leq m})$ on \mathcal{E} is a pair formed of a terminal condition Φ of class \mathcal{P} at T , along with a sequence $(u_l)_{1 \leq l \leq m}$ of functions u_l of class \mathcal{P} on the \mathcal{E}_l , satisfying the following jump condition on $\mathbb{R}^d \times I$, for every $l = 1, \dots, m$:

$$u_l^i(T_l, x) = u_{l+1}^i(T_l, \theta_l^i(x)) \quad (14.2)$$

where, in the case $l = m$, u_{l+1}^i is set equal to Φ on the right-hand side of (14.2). A continuous Cauchy cascade is a Cauchy cascade with continuous ingredients Φ, u_l .

- (ii) The function defined by a Cauchy cascade $(\Phi, (u_l)_{1 \leq l \leq m})$ is the function u on \mathcal{E} given as the concatenation of the u_l on the $\text{Int } \mathcal{E}_l$, along with the terminal condition Φ at T .

The formal analogue of Definition 12.2.4 for a Markovian solution to the Markovian BSDE with the data \mathcal{G} ,¹ \mathcal{C} and ϑ may then be formulated, where :

- “A model \mathcal{X} with generator \mathcal{G} ” in Definition 12.2.4(a) is meant here in the sense that, for every initial condition² $\mathcal{X}_t^t = (x, i)$ and for every $l = 1, \dots, m$, with $t \leq T_l$:
 - \mathcal{X}^t obeys the dynamics (12.60) on the time interval $[T_{l-1} \vee t, T_l]$,
 - $X_{T_l}^t = \theta_l(\mathcal{X}_{T_l-}^t)$ and $N_{T_l}^t = N_{T_l-}^t$.

¹Including here the jumps defined by θ in X .

²A superscript t refers, as usual, to a constant initial condition (t, x, i) of \mathcal{X} .

- In Definition 12.2.4(b):

- the deterministic value function u in Definition 12.2.4(b)(i) is no longer continuous on \mathcal{E} , but defined by a continuous Cauchy cascade $(\Phi, (u_l)_{1 \leq l \leq m})$;
- the deterministic value function v in Definition 12.2.4(b)(ii) is defined likewise by a continuous Cauchy cascade $(\Phi, (v_l)_{1 \leq l \leq m})$.

Regarding now the lower and upper cost functions ℓ and h in this subsection, we relax the assumptions (M2) and (M3) into:

(M.2)' the functions ℓ and h are defined by continuous Cauchy cascades $(\Gamma, (\ell_l)_{1 \leq l \leq m})$ and $(\Upsilon, (h_l)_{1 \leq l \leq m})$ such that $\ell_l \leq h_l$ for every $l = 1, \dots, m$, and $\Gamma \leq \Phi \leq \Upsilon$. In particular, it follows that, for every $x \in \mathbb{R}^d$:

$$\ell_m^i(T, x) = \Gamma^i(T, \theta_m^i(x)) \leq \Phi^i(T, \theta_m^i(x)) \leq \Upsilon^i(T, \theta_m^i(x)) = h_m^i(T, x); \quad (14.3)$$

(M.3)' the processes $\ell(s, \mathcal{X}_s^t)$ and $h(s, \mathcal{X}_s^t)$ satisfy the Mokobodski condition.

Therefore these processes still satisfy the assumption (H.2)' of Sect. 12.1.2.4. Semigroup properties analogous to Propositions 12.4.3 and 12.4.10 in Chap. 12 can then be established as in Chap. 12. The existence of a Markovian solution in the above sense to the Markovian BSDE with the data $(\mathcal{G}, \mathcal{C}, \vartheta)$ (cf. Theorems 12.5.1 and Proposition 12.5.2) follows as before. That the value functions u and v are defined by continuous Cauchy cascades can in turn be established much like Theorem 14.2.18 below (see also [65]). Since the proof is simpler here, we omit it and refer the reader to the proof of Theorem 14.2.18 for similar arguments.

The next step consists in deriving analytic characterizations of the value functions u and v in terms of viscosity solutions of related partial integro-differential obstacle problems, under a set of assumptions (M)' including (M0), (M1), (M2)' and existence of a Markovian solution, in the relaxed sense of this section, to the decoupled FBSDE with the data $(\mathcal{G}, \mathcal{C}, \vartheta)$. The assumptions (E) and (U) are defined as in Chap. 13. Reasoning as in Chaps. 12 and 13 (cf. the proof of Proposition 14.1.2 for a review of the main points), one can show:

Proposition 14.1.4 *Under the currently extended model dynamics for \mathcal{X} (with deterministic jumps of X as specified by θ):*

- (i) *The results of Chap. 12 still hold true, using the previously amended notions of solutions to the related Markovian BSDEs;*
- (ii) *assuming (M)', (E) and (U) then, for every $l = m, \dots, 1$:*
 - *u_l is the unique \mathcal{P} -solution, the maximal \mathcal{P} -subsolution and the minimal \mathcal{P} -supersolution of (V2) on \mathcal{E}_l with terminal condition $u_{l+1}^i(T_l, \theta_l^i(x))$ on $\partial\mathcal{E}_l$ —with u_{l+1} set equal to Φ , in the case $l = m$;*
 - *v_l is the unique \mathcal{P} -solution, the maximal \mathcal{P} -subsolution and the minimal \mathcal{P} -supersolution of (V1) on \mathcal{E}_l with boundary condition u_l on $\partial\mathcal{D}_l$.*

Regarding approximation for the value functions, the arguments of Sect. 13.2.3 can only be used in the present setup for establishing, under the additional assumption (A) of Chap. 13, that for l decreasing from m to 1:

- $u_{l,h} \rightarrow u_l$ locally uniformly on \mathcal{E}_l as $h \rightarrow 0$, under the theoretical assumption that the true value for $u_l^i(T_l, x) = u_{l+1}^i(T_l, \theta_l^i(x))$ is substituted at T_l in the approximation scheme for u_l ;
- $v_{l,h} \rightarrow v_l$ locally uniformly on \mathcal{E}_l as $h \rightarrow 0$, under the theoretical assumption that the true value for u_l is substituted on $\partial\mathcal{D}_l$ in the approximation scheme for v_l , and provided that $v_{l,h} \rightarrow v_l (= u_l)$ on $\partial\mathcal{D}_l \cap \{t < T_l\}$.

Remark 14.1.5 (Discrete Path-Dependence) We motivated the introduction of deterministic jumps (14.1) in X by the presence of discrete dividends on an underlying asset, given as one of the components of X in $\mathcal{X} = (X, N)$. Another motivation is related to the discrete path-dependence issue. Thus consider a European option with payoff $\phi(S_{T_0}, S_{T_1}, \dots, S_{T_m})$ at maturity time $T_m = T$, where a jump-diffusion S is used to model an underlying stock price process. Such payoffs are found in cliquet options, volatility and variance swaps and discretely monitored Asian options (see Sect. 8.6.4). Provided we work in a suitably enlarged state space, the previous methods are applicable. In general we can work with the augmented factor process $X_t = (S_t, S_t^0, \dots, S_t^{m-1})$, where the auxiliary factor processes S^l are equal to 0 before T_l and equal to S_{T_l} on $[T_l, T]$. This augmented factor process X exhibits deterministic jumps at the times T_l of the form (14.1) (case of a degenerate model $\mathcal{X} = (X, N) = X$ there), which provides a second motivation for the developments of this subsection. But this state-enlargement isn't the only possible one. Exploiting the specific nature of a payoff function ϕ , more parsimonious alternatives in “smaller” state spaces can often be found; see Sect. 8.6.4 for a first set of examples, or Example 14.2.19 for other applications in which it is enough to know whether the values of S at the T_l are above or below some trigger level.

Note that in general the issue of path-dependence (discrete or not) can also be addressed in the setup of the functional Itô calculus (see [69]).

14.2 Intermittent Call Protection

The “intermittent” (or “Bermudan”) form of call protection corresponds to an upper payoff process of the form

$$\overline{U}_t = \Omega_t^c \infty + \Omega_t U_t \quad (14.4)$$

for more general càdlàg event-processes Ω_t , $\Omega_t^c = 1 - \Omega_t$ than

$$\Omega_t = \mathbb{1}_{\{t \geq \vartheta\}}, \quad \Omega_t^c = \mathbb{1}_{\{t < \vartheta\}} \infty, \quad (14.5)$$

for a stopping time ϑ in (12.11).

Let ϑ represent a nondecreasing sequence of $[0, T]$ -valued predictable stopping times ϑ_l , with $\vartheta_0 = 0$ and $\vartheta_l = T$ for sufficiently large l , almost surely. We assume that a call protection is active at time 0, and that every subsequent time ϑ_l is a time of switching between call protection and no protection. We thus fix in (14.4), for $t \in [0, T]$,

$$\Omega_t = \mathbb{1}_{\{l_t \text{ odd}\}}, \quad (14.6)$$

so that

$$\overline{U}_t = \mathbb{1}_{\{l_t \text{ even}\}}\infty + \mathbb{1}_{\{l_t \text{ odd}\}}U_t, \quad (14.7)$$

where l_t is the index l of the random time interval $[\vartheta_l, \vartheta_{l+1})$ containing t , thus $\vartheta_{l_t} \leq t < \vartheta_{l_t+1}$.

Remark 14.2.1 Sequences ϑ such that $\vartheta_0 = \vartheta_1 = 0$ and $\vartheta_2 > 0$ allow us to account for the case where the protection is inactive on the first nonempty time interval.

The RIBSDE with the data $(g, \xi, L, U, \vartheta)$, where the “I” in RIBSDE stands for “intermittent”, is the generalization of an R2BSDE in which the upper barrier U is only active on the “odd” random time intervals $[\vartheta_{2l+1}, \vartheta_{2l+2})$. Essentially, we replace U by \overline{U} of the form (14.7) in Definition 12.1.4(c)(iii). As we will see in Definition 14.2.2 below, the possibility of jumps from finite to infinite values in \overline{U} leads us to relax the continuity condition on the reflecting process A in the definition of a solution.

14.2.1 General Setup

The results of this subsection extend to RIBSDEs the well-posedness and comparison RDBSDE results of Sects. 12.1.2.2 to 12.1.2.3. We are back at the general setup and assumptions of Sect. 12.1 (also with $\vartheta = (\vartheta_l)_{l \geq 0}$ as defined above). Let \mathcal{A}_j^2 represent the space of finite variation but not necessarily continuous processes A vanishing at time 0, with (possibly discontinuous) Jordan components denoted by A^\pm .

Definition 14.2.2 An $(\mathbb{F}, \mathbb{P}), (B, \mu)$ -solution \mathcal{Y} to the RIBSDE with the data $(g, \xi, L, U, \vartheta)$ is a quadruple $\mathcal{Y} = (Y, Z, V, A)$ such that:

- (i) $Y \in \mathcal{S}^2, Z \in \mathcal{H}_d^2, V \in \mathcal{H}_\mu^2, A \in \mathcal{A}_j^2$,
- (ii) $Y_t = \xi + \int_t^T g_s(Y_s, Z_s, V_s) ds + A_T - A_t - \int_t^T Z_s dB_s - \int_t^T V_s \cdot d\tilde{\mu}_s, t \in [0, T]$,
- (iii) $L \leq Y \leq \overline{U}$ on $[0, T]$ and $\int_0^T (Y_{t-} - L_{t-}) dA_t^+ = \int_0^T (\overline{U}_{t-} - Y_{t-}) dA_t^- = 0$,

where \overline{U} is defined by (14.7) and with the convention that $0 \times \infty = 0$ in (iii).

Remark 14.2.3 If $\vartheta_2 = T$ almost surely (so that $\vartheta_l = \vartheta_2 = T$ for $l \geq 2$), an RIBSDE with the data $(g, \xi, L, U, \vartheta)$ reduces to the RDBSDE with the data $(g, \xi, L, U, \vartheta_1)$ of Definition 12.1.5(a). If, moreover, $\vartheta_1 = 0$, then we are dealing with an R2BSDE.

14.2.1.1 Verification Principle

Given $t \in [0, T]$, let \mathcal{T}_t denote the set of $[t, T]$ -valued stopping times. The following verification principle, stated without proof, is a straightforward generalization of Proposition 12.1.8.

Proposition 14.2.4 *If $\mathcal{Y} = (Y, Z, V, A)$ satisfies the RIBSDE with the data $(g, \xi, L, U, \vartheta)$, then the process Y is the value process of the Dynkin game with payoff functional given, for every $t \in [0, T]$ and $\tau, \theta \in \mathcal{T}_t$, by*

$$J^t(\tau, \theta) = \int_t^{\tau \wedge \theta} g_s(Y_s, Z_s, V_s) ds + \mathbb{1}_{\{\tau \wedge \theta = \tau < T\}} L_\tau + \mathbb{1}_{\{\theta < \tau\}} \overline{U}_\theta + \mathbb{1}_{\{\tau \wedge \theta = T\}} \xi.$$

More precisely, for every $\epsilon > 0$, an ϵ -saddle-point of the game at time t is given by:

$$\tau_\epsilon^t = \inf \{s \in [t, T]; Y_s \leq L_u + \epsilon\} \wedge T, \quad \theta_\epsilon^t = \inf \{s \in [t, T]; Y_s \geq \overline{U}_u - \epsilon\} \wedge T.$$

Thus, for all $\tau, \theta \in \mathcal{T}_t$,

$$\mathbb{E}[J^t(\tau, \theta_\epsilon^t) | \mathcal{F}_t] - \epsilon \leq Y_t \leq \mathbb{E}[J^t(\tau_\epsilon^t, \theta) | \mathcal{F}_t] + \epsilon. \quad (14.8)$$

In view of the definition of \overline{U} in (14.7), this Dynkin game effectively reduces to a ‘‘constrained Dynkin game’’ with upper payoff process U (instead of \overline{U} in Proposition 14.2.4), posed over the constrained set of stopping policies $(\tau, \theta) \in \mathcal{T}_t \times \mathcal{T}_t^\vartheta$, where \mathcal{T}_t^ϑ denotes the set of the $\bigcup_{l \geq 0} [\vartheta_{2l+1} \vee t, \vartheta_{2l+2} \vee t] \cup \{T\}$ -valued stopping times. In particular,

$$\theta_\epsilon^t = \inf \left\{ s \in \bigcup_{l \geq 0} [\vartheta_{2l+1} \vee t, \vartheta_{2l+2} \vee t]; Y_s \geq U_u - \epsilon \right\} \wedge T.$$

14.2.1.2 Well-Posedness and Comparison

By inspection of the proofs in [87], it appears that the RDBSDE a priori bound and error estimates of Proposition 12.1.9 and the comparison in Proposition 12.1.10 (under the strengthened assumption (H.1)' of Sect. 12.1.2.3 on g) are still valid for more general RIBSDEs, provided they share a common sequence ϑ of stopping times. As with RDBSDEs, the a priori error estimate also proves the uniqueness for a solution to an RIBSDE satisfying the standing assumptions (H.0)–(H.2).

Regarding existence, as we already saw for RDBSDEs in Sect. 12.1.2.4, we need only work under the assumption (H). The following two lemmas then establish existence of a solution to an RIBSDE with only one call protection switching time involved. We first consider an RIBSDE with $\vartheta_2 = T$, namely an RDBSDE (see Remark 14.2.3). We therefore have, as a special case of Proposition 12.1.12 (where continuity of the reflecting process A was part of the definition of a solution to an RDBSDE):

Lemma 14.2.5 *Assuming (H) in the special case where $\vartheta_2 = T$ almost surely, the RIBSDE with the data $(g, \xi, L, U, \vartheta)$ has a (unique) solution (Y, Z, V, A) . Moreover, the reflecting process A is continuous.*

We now consider the case where $\vartheta_1 = 0$ and $\vartheta_3 = T$ almost surely, so that the upper barrier is active on $[0, \vartheta_2)$ and inactive on $[\vartheta_2, T]$. Let

$$[\theta] = \{(t, \omega) \in [0, T] \times \Omega; t = \theta(\omega)\}$$

denote the graph of a stopping time θ .

Lemma 14.2.6 *Assuming (H) in the special case where $0 = \vartheta_1 \leq \vartheta_2 \leq \vartheta_3 = T$ almost surely, the RIBSDE with the data $(g, \xi, L, U, \vartheta)$ has a solution (Y, Z, V, A) . Moreover, A^+ is a continuous process, and*

$$\{(t, \omega); \Delta A_t^- \neq 0\} \subseteq [\vartheta_2], \quad \Delta Y_{\vartheta_2} = \Delta A_{\vartheta_2}^- = (Y_{\vartheta_2} - U_{\vartheta_2})^+.$$

Proof The solution (Y, Z, V, A) is obtained by a two-stage construction similar to the one used for establishing existence of a solution to an RDBSDE in [87], i.e. by “pasting” appropriately the solution $(\widehat{Y}, \widehat{Z}, \widehat{V}, \widehat{A})$ of an RBSDE over the random time interval $[\vartheta_2, T]$, with the solution $(\overline{Y}, \overline{Z}, \overline{V}, \overline{A})$ of an R2BSDE over the random time interval $[0, \vartheta_2]$, subject to the terminal condition $\overline{Y}_{\vartheta_2} = \min(Y_{\vartheta_2}, U_{\vartheta_2})$. The detail of this construction is apparent in the statement of Proposition 14.2.7(i) below. In particular, if $Y_{\vartheta_2} > U_{\vartheta_2}$, the jump $\Delta A_{\vartheta_2}^-$ of the reflecting process A^- at time ϑ_2 is chosen so that

$$Y_{\vartheta_2-} = U_{\vartheta_2} = U_{\vartheta_2-} = \overline{U}_{\vartheta_2-},$$

and hence the upper obstacle minimality condition is satisfied in the second line of Definition 14.2.2(iii). Note that here the process U cannot jump at ϑ_2 , by the assumption (H.2,i)' and the fact that the ϑ_l are predictable stopping times. The jump measure μ cannot jump at ϑ_2 either. \square

Iterated and alternate applications of Lemmas 14.2.5 and 14.2.6 finally yield the following RIBSDE existence result.

Proposition 14.2.7 *Let there be given an RIBSDE with the data $(g, \xi, L, U, \vartheta)$. We assume (H) and that $\vartheta_{m+1} = T$ almost surely for some fixed index m .*

- (i) The following iterative construction is well-defined for l decreasing from m to 0: $\mathcal{Y}^l = (Y^l, Z^l, V^l, A^l)$ is the $(\mathbb{F}, \mathbb{P}), (B, \mu)$ -solution, with A^l continuous, to the stopped RBSDE (for l even) or to the R2BSDE (for l odd) on $[0, T]$ with the data

$$\begin{cases} g, Y_{\vartheta_{l+1}}^{l+1}, L, \vartheta_{l+1} & (l \text{ even}) \\ g, \min(Y_{\vartheta_{l+1}}^{l+1}, U_{\vartheta_{l+1}}), L, U, \vartheta_{l+1} & (l \text{ odd}) \end{cases} \quad (14.9)$$

where, if $l = m$, $Y_{\vartheta_{l+1}}^{l+1}$ is set equal to ξ (so that $\min(Y_{\vartheta_{m+1}}^{m+1}, U_{\vartheta_{m+1}}) = \min(\xi, U_T) = \xi$).

- (ii) Let $\mathcal{Y} = (Y, Z, V, A)$ be defined on $[0, T]$, for every $l = 0, \dots, m$, by:

- $(Y, Z, V) = (Y^l, Z^l, V^l)$ on $[\vartheta_l, \vartheta_{l+1}]$, and also at $\vartheta_{m+1} = T$ if $l = m$,
- $dA = dA^l$ on $(\vartheta_l, \vartheta_{l+1})$, $\Delta A_{\vartheta_l} = (Y_{\vartheta_l}^l - U_{\vartheta_l})^+ = \Delta Y_{\vartheta_l}$ ($=0$ for l odd) and $\Delta A_T = \Delta Y_T = 0$.

Then $\mathcal{Y} = (Y, Z, V, A)$ is the $(\mathbb{F}, \mathbb{P}), (B, \mu)$ -solution to the RIBSDE with the data $(g, \xi, L, U, \vartheta)$. Moreover, A^+ is a continuous process and

$$\{(t, \omega); \Delta A^- \neq 0\} \subseteq \bigcup_{\{l \text{ even}\}} [\vartheta_l], \quad \Delta Y = \Delta A^- = (Y - U)^+ \quad \text{on } \bigcup_{\{l \text{ even}\}} [\vartheta_l].$$

Remark 14.2.8 We conjecture that Proposition 14.2.7 holds true without the condition that $\vartheta_{m+1} = T$ for some fixed index m . In the case of a Brownian filtration (so $\mathbb{F} = \mathbb{F}_B$ and no jump measure μ is involved), this actually follows from the results of Peng and Xu [222]. More precisely, it follows from an immediate extension of their results to the case of an $\mathbb{R} \cup \{+\infty\}$ -valued upper barrier \overline{U} (the results of Peng and Xu [222], based on Peng [221], even if stated for real-valued barriers, only use the fact that $\overline{U}^- = U^-$ lies in \mathcal{S}^2 , which is satisfied under our standing assumption (H.2.i)). Moreover it is apparent that the penalization approach and the related results of Peng [221] and Peng and Xu [222] can be extended in a rather straightforward way to the more general case of a filtration $\mathbb{F} = \mathbb{F}_B \vee \mathbb{F}_\mu$, which would then establish the above conjecture. However, since Proposition 14.2.7 is enough for our later purposes, we do not push this further.

14.2.2 Marked Jump-Diffusion Setup

We now postulate the assumptions of Sects. 12.3, 12.4 on an $(\mathbb{F}^t, \mathbb{P}^t), (B^t, \chi^t)$ -factor process $\mathcal{X}^t = (X^t, N^t)$. More precisely, we consider, under these assumptions, a “marked jump-diffusion” variant of this setup so that X^t is a Markovian jump-diffusion not depending on N^t and the I -valued pure jump marker process N^t is constant except for deterministic jumps, at the times T_l , from $N_{T_l-}^t$ to

$$N_{T_l}^t = \theta_l(\mathcal{X}_{T_l-}^t), \quad (14.10)$$

for some jump function θ .

Remark 14.2.9 In this setup:

- (i) If we use the notation of Sect. 12.3.1, we have that $\mathbb{F}_{v^t} \subseteq \mathbb{F}_{X^t}$. Therefore $\mathbb{F}^t = \mathbb{F}_{B^t} \vee \mathbb{F}_{\chi^t} \vee \mathbb{F}_{v^t} = \mathbb{F}_{B^t} \vee \mathbb{F}_{\chi^t}$, where $(\mathbb{F}_{B^t} \vee \mathbb{F}_{\chi^t}, \mathbb{P}^t)$, (B^t, χ^t) has the local martingale predictable representation property (same proof as Proposition 12.3.4(ii)). Consequently, there are no v^t -martingale components in any of the related forward or backward SDEs.
- (ii) The error estimate (12.45) on N no longer holds, since N now depends on X via (14.10), even under the original measure \mathbb{P} (before the change of measure to \mathbb{P}^t). However, since X doesn't depend on N , the error estimate (12.46) on X and the estimates for $\tilde{\mathcal{Y}}$ in Proposition 12.4.2 hold independent of (12.45).

For a regular function u over $[0, T] \times \mathbb{R}^d$, let

$$\mathcal{G}u(t, x) = \partial_t u(t, x) + \partial u(t, x)b(t, x) + \frac{1}{2}\partial^2 u(t, x) : a(t, x) + (\delta u \cdot m)(t, x), \quad (14.11)$$

with $a = \sigma\sigma^\top$. In the present setup, \mathcal{G} defined by (14.11) is the generator of the Markov process X . We now consider a Markovian RIBSDE with underlying factor process $\mathcal{X} = (X, N)$. More precisely, let there be given a family of RIBSDEs parameterized by an initial condition ${}^t \equiv (t, x, i)$ of \mathcal{X} , with the following data:

- the generator \mathcal{G} of X defined by (14.11) and the specification of the jump size function θ of N in (14.10),
- cost data \mathcal{C} as in Sect. 12.2.3, assumed here not to depend on $i \in I$,
- a parameterized sequence of stopping times ϑ^t defined by $\vartheta_0^t = t$ and, for every $l \geq 0$,

$$\vartheta_{2l+1}^t = \inf\{s > \vartheta_{2l}^t; N_s^t \notin K\} \wedge T, \quad \vartheta_{2l+2}^t = \inf\{s > \vartheta_{2l+1}^t; N_s^t \in K\} \wedge T \quad (14.12)$$

for a given subset K of I , resulting in an effective upper payoff process \bar{U} of the form (14.4) with

$$\mathcal{Q}_s^t = \mathbb{1}_{N_s^t \notin K}. \quad (14.13)$$

Since the cost data don't depend on i , the only impact of the mark N^t is via its impact on ϑ^t . Also note that the ϑ_i^t effectively reduce to \mathfrak{T} -valued stopping times, and that one almost surely has $\vartheta_{m+1}^t = T$.

This setup allows us to account for various forms of intermittent path-dependent call protection. Denoting by S_s^t the first component of the \mathbb{R}^d -valued process X_s^t and by S the first component of $x \in \mathbb{R}^d$, this includes, in particular, the following clauses of call protection, introduced in Example 4.3.16 and dealt with numerically in Chap. 10.

Example 14.2.10 Assume a constant trigger level \bar{S} , an integer $\iota \leq m$ and ϑ^t of the form (14.12), with:

(i) $I = \{0, \dots, \iota\}$, $K = \{0, \dots, \iota - 1\}$ and θ defined by

$$\theta_l^i(x) = \begin{cases} (i+1) \wedge \iota, & S \geq \bar{S} \\ 0, & S < \bar{S} \end{cases}$$

(independent of l). With the initial condition $N_t^t = 0$, N_s^t then represents the minimum of ι and of the consecutive monitoring dates T_l with $S_{T_l}^t \geq \bar{S}$ from time s backwards since the initial time t . Call is possible whenever $N_s^t = \iota$, which means that S_s^t was $\geq \bar{S}$ at the last ι monitoring times since the initial time t .

Or, more generally, with:

(ii) $I = \{0, 1\}^J$ for some given integer $J \in \{\iota, \dots, m\}$, $K = \{i \in I; |i| < \iota\}$ with $|i| = \sum_{1 \leq n \leq J} i_n$, and θ defined by

$$\theta_l^i(x) = (\mathbb{1}_{S \geq \bar{S}}, i_1, \dots, i_{d-1})$$

(independent of l again). With the initial condition $N_t^t = 0_J$, N_s^t then represents the vector of the indicator functions of the events $S_{T_l}^t \geq \bar{S}$ at the last J monitoring dates preceding time s since the initial time t . Call is possible whenever $|N_s^t| \geq \iota$, which means that S_s^t has been $\geq \bar{S}$ on at least ι of the last J monitoring times since the initial time t .

14.2.3 Well-Posedness of the Markovian RIBSDE

In the present setup, where $\mathbb{F}^t = \mathbb{F}_{B^t} \vee \mathbb{F}_{\chi^t}$, there are no v^t -martingale components in any of the related forward or backward SDEs, and the expressions of \tilde{g} in (12.28) and \widehat{g} in (12.68) reduce to the following expressions:

$$\begin{aligned} \tilde{g}(s, \mathcal{X}_s^t, y, z, v) &= g(s, X_s^t, y, z, \tilde{r}_s^t) \quad \text{with } \tilde{r}_s^t = \tilde{r}_s^t(v) = (v\eta)(s, X_s^t) \cdot m(s, X_s^t) \\ \widehat{g}(s, \mathcal{X}_s^t, y, z, \widehat{v}) &= g(s, X_s^t, y, z, \tilde{r}_s^t) + \tilde{r}_s^t \quad \text{with } \tilde{r}_s^t = v(s, X_s^t) \cdot (m(s, X_s^t) - \widehat{m}), \end{aligned} \tag{14.14}$$

where v denotes a generic element of $\mathcal{M}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \widehat{m}(dy); \mathbb{R})$. Accordingly, the V^t -component of a solution to any Markovian BSDE in Proposition 14.2.11 lives in $\mathcal{H}_{\mu^t}^2 = \mathcal{H}_{\chi^t}^2$.

Proposition 14.2.11

(i) *The following iterative construction is well-defined, for l decreasing from m to 0 : $\mathcal{Y}^{l,t} = (Y^{l,t}, Z^{l,t}, V^{l,t}, A^{l,t})$ is the $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, μ^t) -solution, with $A^{l,t}$*

continuous, to the stopped RBSDE (for l even) or to the R2BSDE (for l odd) on $[t, T]$ with the data:

$$\begin{aligned} & \mathbb{1}_{\{s>t\}} \tilde{g}(s, X_s^t, y, z, v), \quad Y_{\vartheta_{l+1}^t}^{l+1,t}, \quad \ell(s \vee t, X_{s \vee t}^t), \quad \vartheta_{l+1}^t \text{ (} l \text{ even)} \\ & \mathbb{1}_{\{s>t\}} \tilde{g}(s, X_s^t, y, z, v), \quad \min(Y_{\vartheta_{l+1}^t}^{l+1,t}, h(\vartheta_{l+1}^t, X_{\vartheta_{l+1}^t}^t)) \quad (14.15) \\ & \ell(s \vee t, X_{s \vee t}^t), \quad h(s \vee t, X_{s \vee t}^t), \quad \vartheta_{l+1}^t \text{ (} l \text{ odd)} \end{aligned}$$

where, if $l = m$, $Y_{\vartheta_{l+1}^t}^{l+1,t}$ is set equal to $\Phi(X_T^t)$.

Let $\mathcal{Y}^t = (Y^t, Z^t, V^t, A^t)$ be defined in terms of the $\mathcal{Y}^{l,t}$ as \mathcal{Y} in terms of the \mathcal{Y}^l in Proposition 14.2.7(ii). So, in particular, $Y^t = Y^{l,t}$ on $[\vartheta_l^t, \vartheta_{l+1}^t]$ for every $l = 0, \dots, m$, and

$$Y_t^t = \begin{cases} Y_t^{0,t}, & i \in K \\ Y_t^{1,t}, & i \notin K. \end{cases} \quad (14.16)$$

Then \mathcal{Y}^t is the $(\mathbb{F}^t, \mathbb{P}^t), (B^t, \mu^t)$ -solution to the RIBSDE on $[t, T]$ with the data

$$\tilde{g}(s, X_s^t, y, z, v), \quad \Phi(X_T^t), \quad \ell(s, X_s^t), \quad h(s, X_s^t), \quad \vartheta^t. \quad (14.17)$$

- (ii) For every $l = 0, \dots, m$, we extend $Y^{l,t}$ by $Y_t^{l,t}$, and $A_t^{l,t}$, $Z^{l,t}$ and $V^{l,t}$ by 0 on $[0, t]$. Then, for every $l = m, \dots, 0$, $\mathcal{Y}^{l,t} = (Y^{l,t}, Z^{l,t}, V^{l,t}, A^{l,t})$ is the $(\mathbb{F}, \mathbb{P}), (B, \mu)$ -solution, with $A^{l,t}$ continuous, to the stopped RBSDE (for l even) or R2BSDE (for l odd) on $[0, T]$ with the data of the form (14.15), but with \tilde{g} replaced by \hat{g} there.

Proof It is easy to check that the assumption (H) is satisfied (recalling, in particular, regarding (H.2)' that the cost functions do not depend on i in this section), so that part (i) follows by application of Proposition 14.2.7. The identity (14.16) results from the fact that, since $Y^t = Y^{l,t}$ on $[\vartheta_l^t, \vartheta_{l+1}^t]$, we have

$$\begin{cases} Y_t^t = Y_t^{0,t}, & N_t^t \in K \\ Y_t^t = Y_t^{1,t}, & N_t^t \notin K, \end{cases} \quad (14.18)$$

with $N_t^t = i$. Part (ii) then follows from part (i) as in the proof of Proposition 12.4.4. \square

Our next goal is to derive stability results on \mathcal{Y}^t or, more precisely, on the $\mathcal{Y}^{l,t}$. Toward this end, a stability condition on ϑ^t is required. We consider the following weak form of Assumption 12.4.5 (the latter doesn't hold in the situations of Example 14.2.10).

Assumption 14.2.12 Viewed as a random function of the initial condition (t, x, i) of \mathcal{X} , at every (t, x, i) in \mathcal{E} , ϑ is almost surely:

- (i) continuous at (t, x, i) if $t \notin \mathfrak{T}$, and right-continuous at (t, x, i) if $t \in \mathfrak{T}$,
- (ii) left-limited at (t, x, i) if $t = T_l \in \mathfrak{T}$ and if θ_l is continuous at (x, i) .

By this, we mean that, respectively:

- (i) $\vartheta^{t_n} \rightarrow \vartheta^t$ if $(t_n, x_n, i) \rightarrow (t, x, i)$ with $t \notin \mathfrak{T}$, or, for $t = T_l \in \mathfrak{T}$, if $\mathcal{E}_{l+1} \ni (t_n, x_n, i) \rightarrow (T_l, x, i)$.
- (ii) if $\text{Int } \mathcal{E}_l \ni (t_n, x_n, i) \rightarrow (t = T_l, x, i)$ with θ_l continuous at (x, i) , then ϑ^{t_n} converges to some nondecreasing sequence, denoted by $\tilde{\vartheta}^t$, of predictable stopping times (thus, in particular, $\tilde{\vartheta}_l^t = T$ for $l \geq m + 1$).

Observe that since the ϑ_l^t are \mathfrak{T} -valued stopping times:

- The continuity assumption on ϑ^t effectively means that, for sufficiently large n , $\vartheta_l^{t_n} = \vartheta_l^t$ almost surely, for every $l = 1, \dots, m + 1$ and $\mathcal{E} \ni (t_n, x_n, i) \rightarrow (t, x, i) \in \mathcal{E}$ with $t \notin \mathfrak{T}$.
- The right-continuity (respectively left-limit) assumption effectively means that for n large enough $\vartheta_l^{t_n} = \vartheta_l^t$ (respectively $\tilde{\vartheta}_l^t$) almost surely, for every $l = 1, \dots, m + 1$ and $\mathcal{E}_{l'+1} \ni (t_n, x_n, i) \rightarrow (T_{l'}, x, i) \in \mathcal{E}$.

It is intuitively clear that Assumption 14.2.12 holds in the situations of Example 14.2.10, provided the jump-diffusion X is uniformly elliptic in the direction of its first component S ; see [65] for a precise statement and proof in a diffusion setup.

We also make the following additional hypothesis on the upper payoff function h , whereas the lower payoff function ℓ satisfies (M.3). Also recall that in this section the cost data \mathcal{C} , including the barrier functions ℓ and h , do not depend on $i \in I$.

Assumption 14.2.13 h is Lipschitz in (t, x) .

We denote by $\tilde{\mathcal{Y}}^t = (\tilde{\mathcal{Y}}^{l,t})_{0 \leq l \leq m}$, with $\tilde{\mathcal{Y}}^{l,t} = (\tilde{Y}^{l,t}, \tilde{Z}^{l,t}, \tilde{V}^{l,t}, \tilde{A}^{l,t})$ and $\tilde{A}^{l,t}$ continuous for every $l = 0, \dots, m$, the sequence of solutions of stopped RBSDEs (for l even) or R2BSDEs (for l odd) that is obtained by substituting $\tilde{\vartheta}^t$ for ϑ^t in the construction of \mathcal{Y}^t in Proposition 14.2.11(i).

Proposition 14.2.14 For every $l = m, \dots, 0$:

- (i) We have the following estimate on $\mathcal{Y}^{l,t}$:

$$\|Y^{l,t}\|_{\mathcal{S}^2}^2 + \|Z^{l,t}\|_{\mathcal{H}_d^2}^2 + \|V^{l,t}\|_{\mathcal{H}_\mu^2}^2 + \|A^{l,t}\|_{\mathcal{S}^2}^2 \leq C(1 + |x|^{2\rho}). \quad (14.19)$$

Moreover, an analogous bound estimate is satisfied by $\tilde{\mathcal{Y}}^{l,t}$.

- (ii) Letting t_n denote a perturbed initial condition (t_n, x_n, i) of \mathcal{X} :

- if $t \notin \mathfrak{T}$, then \mathcal{Y}^{l,t_n} converges on $\mathcal{S}^2 \times \mathcal{H}_d^2 \times \mathcal{H}_\mu^2 \times \mathcal{A}_j^2$ to $\mathcal{Y}^{l,t}$ as $\mathcal{E} \ni (t_n, x_n, i) \rightarrow (t, x, i)$;

- in the case $t = T_{l'} \in \mathfrak{T}$,
 - \mathcal{Y}^{l,t_n} converges on $\mathcal{S}^2 \times \mathcal{H}_d^2 \times \mathcal{H}_\mu^2 \times \mathcal{A}_j^2$ to $\mathcal{Y}^{l,t}$ as $\mathcal{E}_{l'+1} \ni (t_n, x_n, i) \rightarrow (t, x, i)$;
 - if $\theta_{l'}$ is continuous at (x, i) , then \mathcal{Y}^{l,t_n} converges on $\mathcal{S}^2 \times \mathcal{H}_d^2 \times \mathcal{H}_\mu^2 \times \mathcal{A}_j^2$ to $\tilde{\mathcal{Y}}^{l,t}$ as $\text{Int } \mathcal{E}_{l'} \ni (t_n, x_n, i) \rightarrow (t, x, i)$.

Proof Under Assumption 14.2.12, these results can be established recursively for l decreasing from m to 0, by routine amendments to the proof of Proposition 12.4.7. Assumption 14.2.13 is needed for controlling new terms in $\|h(t \vee \cdot \wedge \vartheta_{l+1}^t, \mathcal{X}_{t \cdot \wedge \vartheta_{l+1}^t}^t) - h(t_n \cdot \wedge \vartheta_{l+1}^{t_n}, \mathcal{X}_{t_n \cdot \wedge \vartheta_{l+1}^{t_n}}^{t_n})\|_{\mathcal{S}^2}$ and $\|h(t \vee \cdot \wedge \tilde{\vartheta}_{l+1}^t, \mathcal{X}_{t \cdot \wedge \tilde{\vartheta}_{l+1}^t}^t) - h(t_n \cdot \wedge \tilde{\vartheta}_{l+1}^{t_n}, \mathcal{X}_{t_n \cdot \wedge \tilde{\vartheta}_{l+1}^{t_n}}^{t_n})\|_{\mathcal{S}^2}$ that arise, for l odd, on the right-hand side of analogs to the inequality (15.5). \square

The analog of Remark 12.4.8 holds (regarding the possibility of considering the “sequentially relaxed” form of Assumption 14.2.12 for deducing convergences along suitable subsequences in Proposition 14.2.14(ii)), which is enough for all our purposes.

14.2.4 Semigroup and Markov Properties

Let t' refer to a constant initial condition (t, x, i) . Let $\mathcal{X}^t = (X^t, N^t)$ and $\mathcal{Y}^t = (Y^t, Z^t, V^t, A^t)$ be, respectively, defined as in Sect. 14.2.2 and Proposition 14.2.11. Given $t' \geq t$, let $\tilde{\mathbb{F}}^{t'}$ represent $(\tilde{\mathcal{F}}_r^{t'})_{r \geq t'}$ where, for $r \geq t'$,

$$\tilde{\mathcal{F}}_r^{t'} = \sigma(\mathcal{X}_r^t) \bigvee \mathcal{F}_r^{t'}.$$

Let $\vartheta' = t' \vee \vartheta^t$ in the sense that $\vartheta'_l = t' \vee \vartheta_l^t$ for $l = 1, \dots, m + 1$. As for $\mathbb{F}^t = (\mathcal{F}_r^t)_{r \geq t}$, \mathbb{P}^t , B^t and μ^t , they are defined in the usual way as in Sects. 12.3.1–12.3.2, with t' replaced by t . Note, in particular, that $\tilde{\mathbb{F}}^{t'} \subseteq \mathbb{F}_{[t', T]}^t$. We then have the following semigroup properties, which are the analogs, in the present setup, of Propositions 12.4.3 and 12.4.10.

Proposition 14.2.15

- (i) On $[t', T]$, the marked jump-diffusion with initial condition \mathcal{X}_t^t at t' admits a unique $(\tilde{\mathbb{F}}^{t'}, \mathbb{P})$ -solution $\mathcal{X}^{t'} = (X^{t'}, N^{t'})$. This solution coincides with the restriction of \mathcal{X}^t to $[t', T]$, so that:

$$\mathcal{X}^{t'} = (X_r^{t'}, N_r^{t'})_{t' \leq r \leq T} = (\mathcal{X}_r^t)_{t' \leq r \leq T}.$$

- (ii) For t' in the same monitoring time strip as t , i.e. $T_{l-1} \leq t < t' < T_l$ for some $l = 1, \dots, m$, we have that $\vartheta' = t' \vee \vartheta^t$ is an $\widetilde{\mathbb{F}}^{t'}$ -stopping time, and the RIBSDE on $[t', T]$ with the data

$$\widetilde{g}(s, X_s^{t'}, y, z, \widetilde{v}), \quad \Phi(X_T^{t'}), \quad \ell(s, X_s^{t'}), \quad h(s, X_s^{t'}), \quad \vartheta' \quad (14.20)$$

has a unique $(\widetilde{\mathbb{F}}^{t'}, \mathbb{P}^{t'}), (\mathcal{B}^{t'}, \mu^{t'})$ -solution $\mathcal{Y}^{t'} = (Y_r^{t'}, Z_r^{t'}, V_r^{t'}, A_r^{t'})$. With $\mathcal{Y}^t = (Y_r^t, Z_r^t, V_r^t, A_r^t)_{t \leq r \leq T}$ defined as in Proposition 14.2.11, we have

$$\mathcal{Y}^{t'} = (Y_r^{t'}, Z_r^{t'}, V_r^{t'}, A_r^{t'})_{t' \leq r \leq T} = (Y_r^t, Z_r^t, V_r^t, A_r^t - A_{t'}^t)_{t' \leq r \leq T}. \quad (14.21)$$

Proof Part (i) can be shown much like Proposition 12.4.3(i), which implies, in particular, that whenever $T_{l-1} \leq t < t' < T_l$ for some $l = 1, \dots, m$, then $N_r^t = N_r^{t'} = i$ for $r \in [t', T_l]$. In view of (14.12) we thus have $\vartheta'_0 = t'$ and, for every $l \geq 0$:

$$\vartheta'_{2l+1} = \inf\{s > \vartheta'_{2l}; N_s^{t'} \notin K\} \wedge T, \quad \vartheta'_{2l+2} = \inf\{s > \vartheta'_{2l+1}; N_s^{t'} \in K\} \wedge T. \quad (14.22)$$

This shows that ϑ' is an $\widetilde{\mathbb{F}}^{t'}$ -stopping time, namely the analog of ϑ^t regarding $N^{t'}$. Part (ii) can then be established much like Propositions 12.4.3(ii) or 12.4.10(ii). \square

In the present setup, the appropriate notion³ of a Cauchy cascade of functions over \mathcal{E} takes the form of the following:

Definition 14.2.16

- (i) A Cauchy cascade $(\Phi, (v_l)_{1 \leq l \leq m})$ on \mathcal{E} is a pair formed of a terminal condition Φ of class \mathcal{P} at T , along with a sequence $(v_l)_{1 \leq l \leq m}$ of functions v_l of class \mathcal{P} on the \mathcal{E}_l , satisfying the following jump condition at every point of continuity of θ_l^i in x :

$$v_l^i(T_l, x) = \begin{cases} \min(v_{l+1}(T_l, x, \theta_l^i(x)), h(T_l, x)) & \text{if } i \notin K \text{ and } \theta_l^i(x) \in K, \\ v_{l+1}(T_l, x, \theta_l^i(x)) & \text{otherwise,} \end{cases} \quad (14.23)$$

where, if $l = m$, v_{l+1} is set equal to Φ .

A continuous Cauchy cascade is a Cauchy cascade with continuous ingredients Φ at T and v_l on the \mathcal{E}_l , except perhaps for discontinuities of the v_l at the points (T_l, x, i) of discontinuity of θ_l^i in x .

- (ii) The function defined by a Cauchy cascade is the function on \mathcal{E} given by the concatenation of the v_l on the $\text{Int } \mathcal{E}_l$, and by the terminal condition Φ at T .

Remark 14.2.17 At a point (T_l, x, i) of discontinuity of θ_l^i in x , $v_l^i(t_n, x_n)$ may fail to converge to $v_l^i(T_l, x)$ as $\mathcal{E}_l \ni (t_n, x_n, i) \rightarrow (T_l, x, i)$. Note that in the specific

³A variant of that of Definition 14.1.3.

situations of Example 14.2.10, the set of discontinuities x of θ_l^i is given by the hyperplane $\{x_1 = \bar{S}\}$ of \mathbb{R}^d , for every l, i .

The Markov properties of \mathcal{Y} are gathered together in the following result, whose proof is presented in Sect. 15.1.7. The notion of ϵ -saddle-point in part (iii) was introduced in the RIBSDE verification principle of Proposition 14.2.4.

Theorem 14.2.18

- (i) Given $(t, x, i) \in \mathcal{E}$, let $\mathcal{Y}^t = (Y^t, Z^t, V^t, A^t)$ be defined as in Proposition 14.2.11. As (t, x, i) varies in \mathcal{E} , Y_t^i yields a function v , defined by a continuous Cauchy cascade $(\Phi, (v_l)_{1 \leq l \leq m})$ on \mathcal{E} .
- (ii) We have \mathbb{P}^t -almost surely that

$$Y_s^t = v(s, \mathcal{X}_s^t), \quad s \in [t, T]. \quad (14.24)$$

- (iii) For every $\epsilon > 0$, an ϵ -saddle-point of the related Dynkin game at time t is given by

$$\begin{aligned} \tau_\epsilon^t &= \inf\{s \in [t, T]; (s, \mathcal{X}_s^t) \in \mathcal{E}_\epsilon^+\} \wedge T \\ \vartheta_\epsilon^t &= \inf\{s \in \cup_{l \geq 0} [\vartheta_{2l+1}^t, \vartheta_{2l+2}^t); (s, \mathcal{X}_s^t) \in \mathcal{E}_\epsilon^-\} \wedge T, \end{aligned}$$

with

$$\begin{aligned} \mathcal{E}_\epsilon^+ &= \{(t, x, i) \in \mathcal{E}; v^i(t, x) \geq \ell^i(t, x) + \epsilon\} \\ \mathcal{E}_\epsilon^- &= \{(t, x, i) \in \mathcal{E}; v^i(t, x) \geq h^i(t, x) - \epsilon\}. \end{aligned}$$

14.2.5 Viscosity Solutions Approach

The next step consists in deriving an analytic characterization of the value function v or, more precisely, of the corresponding Cauchy cascade of functions $(v_l)_{1 \leq l \leq m}$, in terms of viscosity solutions to the following cascade of variational inequalities:

For l decreasing from m to 1:

- at $t = T_l$, for every $i \in I$ and $x \in \mathbb{R}^d$, if θ_l^i is continuous at x , then

$$v_l^i(T_l, x) = \begin{cases} \min(v_{l+1}(T_l, x, \theta_l^i(x)), h(T_l, x)), & i \notin K \text{ and } \theta_l^i(x) \in K \\ v_{l+1}(T_l, x, \theta_l^i(x)), & \text{otherwise,} \end{cases} \quad (14.25)$$

with v_{l+1} set equal to Φ if $l = m$;

- on the time interval $[T_{l-1}, T_l]$, for every $i \in I$,

$$\begin{cases} \min(-\mathcal{G}v_l^i - g^{v_l^i}, v_l^i - \ell) = 0, & i \in K \\ \max(\min(-\mathcal{G}v_l^i - g^{v_l^i}, v_l^i - \ell), v_l^i - h) = 0, & i \notin K, \end{cases} \quad (14.26)$$

where \mathcal{G} is given by (14.11) and where we have set, for every function $\phi = \phi(t, x)$,

$$g^\phi = g^\phi(t, x) = g\left(t, x, \phi(t, x), (\partial\phi\sigma)(t, x), \mathcal{I}\phi(t, x)\right). \quad (14.27)$$

In the special case of a jump size function θ independent of x , i.e. $\theta_l^i(x) = \theta_l^i$, the v_l are in fact continuous over the \mathcal{E}^l . This can be shown by a simplified version of the proof of Theorem 14.2.18. Using the notions of viscosity solutions introduced in Definition 14.1.1, then by virtue of arguments already used in Chap. 13 (see also Proposition 14.1.4(ii)) we have that, for every $l = 1, \dots, m$ and $i \in I$, the function v_l^i is the unique \mathcal{P} -solution, the maximal \mathcal{P} -subsolution and the minimal \mathcal{P} -supersolution of the related problem $(\mathcal{V}1)$ or $(\mathcal{V}2)$ on \mathcal{E}_l that is visible in (14.25)–(14.26), with terminal condition at T_{l+1} dictated by v_{l+1}, h and/or Φ . Moreover, under the working assumption that the true value for v_{l+1} is substituted at T_{l+1} in an approximation scheme for v_l , we have that $v_{l,h} \rightarrow v_l$ locally uniformly on \mathcal{E}_l as $h \rightarrow 0$.

But, thinking for instance of the situations of Example 14.2.10, the case of θ not depending on x is of course too restrictive. Now, as soon as θ depends on x , θ presents some discontinuities in x and, under Assumption 14.2.12, the functions v_l present discontinuities at the points (T_l, x, i) of discontinuity of the θ_l^i . There is then no chance that one could any longer characterize the v_l in terms of continuous viscosity solutions to (14.25)–(14.26). It is possible however, though we will not develop it further here, to characterize the upper semicontinuous envelope \bar{v}_l of v_l as the maximal \mathcal{P} -subsolution of (14.25)–(14.26) on \mathcal{E}_l , for every $l = 1, \dots, m$; see [65] for a precise statement and proof in the case of a diffusion X .

The Cauchy cascade (14.25)–(14.26) involves k equations, where k is the cardinality of the index set I of i . In some applications, such as that of Example 14.2.10(ii), k may be very large, so that deterministic numerical schemes for (14.25)–(14.26) are not practical. For very large k simulation schemes such as the ones presented in Chap. 10 (or perhaps Monte Carlo branching particle schemes [135]) are the only viable alternative.

14.2.6 Protection Before a Stopping Time Again

We finally consider the special case where the marker process N is stopped at its first exit time from K , which corresponds to jump functions $\theta_l^i(x)$ such that $\theta_l^i(x) = i$ for $i \notin K$. The sequence $\vartheta^t = (\vartheta_l^t)_{l \geq 0}$ is then stopped at rank $l = 2$, so $\vartheta_l^t = T$ for $l \geq 2$. In this case, (14.13) reduces to

$$\Omega_s^t = \mathbb{1}_{N_s^t \notin K} = \mathbb{1}_{s \geq \vartheta_1^t}. \quad (14.28)$$

We thus recover a case of call protection before a stopping time as in Chaps. 12 and 13. If $N_t^t = i \notin K$, we have $\vartheta_1^t = t$, and call protection on $[0, \vartheta_1^t)$ reduces to no protection. For nontrivial examples with $N_t^t = i \in K$, we can, for instance, consider the following “stopped version” of Example 14.2.10.

Example 14.2.19 Given a constant trigger level \bar{S} and an integer i :

- (i) call possible from the first time ϑ that S has been $\geq \bar{S}$ at the last i monitoring times;
- or, more generally,
- (ii) given a further integer $j \geq i$, call possible from the first time ϑ that S has been $\geq \bar{S}$ on at least i of the last j monitoring times.

From a mathematical point of view, with (14.28), one is back at an RDBSDE in the sense of Definition 12.1.5(a). But this is for a stopping time ϑ_1^t that falls outside the scope of Example 12.4.6 / Assumption (E.3), so that the PDE results of Chap. 13 cannot be applied directly. However, these results are still valid. Indeed, one can check by inspection of the arguments of Sects. 14.2.3–14.2.5 that, in the case of (14.28):

- for $i \notin K$, the $\mathcal{Y}^{l,t}$ do not depend on i , and \mathcal{Y}^t in Proposition 14.2.11 coincides with \mathcal{Y}^t in Proposition 12.4.1(i) (special case of \mathcal{X}^t there given as X^t here);
- since discontinuities of the $A^{l,t}$ only occur at times of switching from no call protection to call protection, since in addition such switchings are not possible in the case of (14.28), therefore the $\mathcal{Y}^{l,t}$ have in this case continuous reflecting processes $A^{l,t}$;
- since Assumption 14.2.13 of this section was only used for taking care of the case where a call protection period follows a no call protection period, Proposition 14.2.14 is true here independent of Assumption 14.2.13, which is not needed in the case of (14.28).

14.2.6.1 No-Protection Price

Regarding the no-protection period $[\vartheta_1^t, T]$ we thus have, under the assumptions (E) and (U) for (ii) and, in addition, (A) for (iii):

Proposition 14.2.20

- (i) *For $i \notin K$, $Y_t^{1,t} =: u(t, x)$ defines a continuous function u on $[0, T] \times \mathbb{R}^d$. This function u corresponds to a no call protection pricing function in the sense that, for every initial condition $(t, x, i) \in \mathcal{E}$, we have*

$$Y_s^t = u(s, X_s^t) \quad \text{on } [\vartheta_1^t, T],$$

with $\vartheta_1^t = \inf\{s > t; N_s^t \notin K\}$.

- (ii) *The no protection value function u thus defined is the unique \mathcal{P} -solution, the maximal \mathcal{P} -subsolution, and the minimal \mathcal{P} -supersolution, of*

$$\max(\min(-\mathcal{G}u - g^u, u - \ell), u - h) = 0 \tag{14.29}$$

on \mathcal{E} with boundary condition Φ at T , where \mathcal{G} is given by (14.11) and g^u is defined by (14.27).

- (iii) *Stable, monotone and consistent approximation schemes u_h for u converge to u locally uniformly on \mathcal{E} as $h \rightarrow 0$.*

Note that the no-protection pricing function u coincides with the function v^i of Theorem 14.2.18 for $i \notin K$ (v^i doesn't depend on i outside K , in the case of (14.28)).

14.2.6.2 Protection Price

As for the protection period $[0, \vartheta_1^t]$, since the v_l^i for $i \notin K$ all reduce to u , the Cauchy cascade (14.25)–(14.26) to be solved for $(v_l)_{1 \leq l \leq m} = (v_l^i)_{1 \leq l \leq m}^{i \in I}$ reduces to the following Cauchy-Dirichlet cascade in $(v_l^i)_{1 \leq l \leq m}^{i \in K}$, with the function u as boundary condition and where, in view of the identity (14.24) in Theorem 14.2.18, $(v_l^i)_{1 \leq l \leq m}^{i \in K}$ can be interpreted as the protection pricing function.

For l decreasing from m to 1:

- at $t = T_l$, for every $i \in K$ and $x \in \mathbb{R}^d$ with θ_l^i continuous at x ,

$$v_l^i(T_l, x) = \begin{cases} u(T_l, x), & l = m \text{ or } \theta_l^i(x) \notin K \\ v_{l+1}(T_l, x, \theta_l^i(x)), & \text{otherwise;} \end{cases} \quad (14.30)$$

- on the time interval $[T_{l-1}, T_l]$, for every $i \in K$,

$$\min(-\mathcal{G}v_l^i - g^{v_l^i}, v_l^i - \ell) = 0. \quad (14.31)$$

Proceeding as in [65], the upper semicontinuous envelope $(\bar{v}_l^i)_{l=1}^m$ of $(v_l^i)_{l=1}^m$ could then be characterized as the maximal \mathcal{P} -subsolution of (14.30)–(14.31) on \mathcal{E}_l , for every $l = 1, \dots, m$.

The Cauchy-Dirichlet cascade (14.29)–(14.31) involves fewer equations than the Cauchy cascade (14.25)–(14.26). However “less” here is still often far too much, as for instance in the situation of Example 14.2.19(ii), from the point of view of a practical solution by deterministic numerical schemes. For large sets K the simulation schemes of Chap. 10 (or perhaps Monte Carlo branching particle schemes [135]) are again the only viable alternative.

Part VI

Appendix

Chapter 15

Technical Proofs (**)

In this chapter we provide the proofs of the most demanding results of Part V. In these proofs we do not use the “.” notation for jump space integrals, which we instead write explicitly. Likewise, we do not use the $A : B$ notation for the trace (sum of the diagonal elements) of the product of the square matrices A and B , but write instead the more explicit expression $\text{Tr}(AB)$.

15.1 Proofs of BSDE Results

15.1.1 Proof of Lemma 12.3.6

Recall that a càdlàg process M^t is a \mathbb{P}^t -local martingale if and only if $(\Gamma^t M^t)$ is a \mathbb{P} -local martingale (see, e.g., Proposition III.3.8 in Jacod and Shiryaev [153]). Now, for the three quantities

$$M^t = B^t, \quad \int_t^\cdot \int_{\mathbb{R}^d} V_s^t(y) \tilde{\chi}^t(ds, dy), \quad \sum_{j \in I} \int_t^\cdot W_s^t(j) d\tilde{v}_s^t(j),$$

with V^t , W^t in the related spaces of predictable integrands and with “ \doteq ” standing for “equality up to an $(\mathbb{F}^t, \mathbb{P})$ -local martingale”, we have

$$d(\Gamma_s^t M_s^t) \doteq \Gamma_{s-}^t dM_s^t + \Delta \Gamma_s^t \Delta M_s^t$$

where (in regard to the above three respective quantities for M^t):

$$\Delta M_s^t = 0, \quad \text{resp. } \int_{\mathbb{R}^d} V_s^t(y) \chi(ds, dy), \quad \text{resp. resp. } \sum_{j \in I} W_s^t(j) d\nu_s^t(j).$$

- For $M^t = B^t$, $(\Gamma^t M^t)$ is obviously a \mathbb{P} -local martingale. Process B^t is thus a continuous \mathbb{P}^t -local martingale, null at time t with $\langle B^t, B^t \rangle_s = (s - t)I_d$. Therefore B^t is a d -variate standard \mathbb{P}^t -Brownian motion starting at time t .

- For $M^t = \int_t^\cdot \int_{\mathbb{R}^d} V_s^t(y) \tilde{\chi}^t(ds, dy)$, since χ and v cannot jump together, we have by (12.57):

$$\Delta \Gamma_s^t \Delta M_s^t = \Gamma_{s-}^t \int_{\mathbb{R}^d} V_s^t(y) \left(\frac{f(s, \mathcal{X}_{s-}^t, y)}{\widehat{f}(s, \mathcal{X}_{s-}^t, y)} - 1 \right) \chi(ds, dy).$$

Thus

$$\begin{aligned} d(\Gamma_s^t M_s^t) &\doteq \Gamma_{s-}^t \int_{\mathbb{R}^d} V_s^t(y) \tilde{\chi}^t(ds, dy) \\ &\quad + \Gamma_{s-}^t \int_{\mathbb{R}^d} V_s^t(y) \left(\frac{f(s, \mathcal{X}_{s-}^t, y)}{\widehat{f}(s, \mathcal{X}_{s-}^t, y)} - 1 \right) \chi(ds, dy) \\ &= -\Gamma_{s-}^t \int_{\mathbb{R}^d} V_s^t(y) f(s, \mathcal{X}_s^t, y) \widehat{m}(dy) ds \\ &\quad + \Gamma_{s-}^t \int_{\mathbb{R}^d} V_s^t(y) \frac{f(s, \mathcal{X}_{s-}^t, y)}{\widehat{f}(s, \mathcal{X}_{s-}^t, y)} \chi(ds, dy) \\ &= \Gamma_{s-}^t \int_{\mathbb{R}^d} V_s^t(y) \frac{f(s, \mathcal{X}_{s-}^t, y)}{\widehat{f}(s, \mathcal{X}_{s-}^t, y)} \tilde{\chi}(ds, dy) \end{aligned}$$

and $(\Gamma^t M^t)$ is also a \mathbb{P} -local martingale.

- For $M^t = \sum_{j \in I} \int_t^\cdot W_s^t(j) d\tilde{v}_s^t(j)$ we likewise obtain

$$d(\Gamma_s^t M_s^t) \doteq \Gamma_{s-}^t \sum_{j \in I} W_s^t(j) \frac{n^j(s, \mathcal{X}_{s-}^t)}{\widehat{n}^j(N_{s-}^t)} d\tilde{v}_s(N_s^t, j)$$

and $(\Gamma^t M^t)$ is again a \mathbb{P} -local martingale.

15.1.2 Proof of Proposition 12.4.2

Since f (cf. Lemma 12.3.5(i)) and \widehat{n} are bounded, with f positively bounded in order to obtain (H.1.ii), we have:

(H.1.i) $\mathbb{1}_{\{\cdot > t\}} \widehat{g}(\cdot, \mathcal{X}_\cdot^t, y, z, \widehat{v})$ is a progressively measurable process, with

$$\|\mathbb{1}_{\{\cdot > t\}} \widehat{g}(\cdot, \mathcal{X}_\cdot^t, y, z, \widehat{v})\|_{\mathcal{H}^2} < +\infty \quad \text{for all } y \in \mathbb{R}, z \in \mathbb{R}^d \text{ and } \widehat{v} \in \mathcal{M}_{\widehat{\varrho}};$$

(H.1.ii) $\mathbb{1}_{\{\cdot > t\}} \widehat{g}(\cdot, \mathcal{X}_\cdot^t, y, z, \widehat{v})$ is uniformly Λ -Lipschitz continuous with respect to (y, z, \widehat{v}) , in the sense that for any $s \in [0, T]$, $y, y' \in \mathbb{R}$, $z, z' \in \mathbb{R}^d$ and $\widehat{v}, \widehat{v}' \in \mathcal{M}_{\widehat{\varrho}}$:¹

$$|\widehat{g}(s, \mathcal{X}_s^t, y, z, \widehat{v}) - \widehat{g}(s, \mathcal{X}_s^t, y', z', \widehat{v}')| \leq \Lambda(|y - y'| + |z - z'| + |\widehat{v} - \widehat{v}'|).$$

¹Cf. (12.63) for the definition of $|\widehat{v} - \widehat{v}'|$.

So the driver $\mathbb{1}_{\{\cdot > t\}} \widehat{g}$ satisfies the assumption (H.1), hence the data (12.67) satisfy the assumptions (H.0)–(H.2) with respect to $(\mathbb{F}, \mathbb{P}), (B, \mu)$.

(i) Under the assumption (M.3) on ℓ , an application of (12.16) yields the following bound estimate on $\widetilde{\mathcal{Y}}^t$:

$$\|Y^t\|_{\mathcal{S}^2}^2 + \|Z^t\|_{\mathcal{H}_d^2}^2 + \|\widetilde{V}^t\|_{\mathcal{H}_\mu^2}^2 + \|A^{t,+}\|_{\mathcal{S}^2}^2 + \|A^{t,-}\|_{\mathcal{S}^2}^2 \leq c(\Lambda)c,$$

with

$$\begin{aligned} c := & \|\Phi(\mathcal{X}_T^t)\|_{L^2}^2 + \|\mathbb{1}_{\{\cdot > t\}} \widehat{g}(\cdot, \mathcal{X}_\cdot^t, 0, 0, 0)\|_{\mathcal{H}^2}^2 \\ & + \|\ell(\cdot \vee t, \mathcal{X}_{\cdot \vee t}^t)\|_{\mathcal{S}^2}^2 + \|h(\cdot \vee t, \mathcal{X}_{\cdot \vee t}^t)\|_{\mathcal{S}^2}^2 + \left\| \int_{\cdot \vee t}^{\cdot} \mathcal{G}\phi(r, \mathcal{X}_r^t) dr \right\|_{\mathcal{S}^2}^2, \end{aligned}$$

where ϕ is the function of the assumption (M.3). The estimate (12.70) then follows by standard computations, given the Lipschitz and growth assumptions on the data and the bound estimate (12.43) on X^t .

(ii) An application of (12.17) yields the following error estimate on $\widetilde{\mathcal{Y}}^t$, in which c is as above:

$$\begin{aligned} & \|Y^t - Y^{t_n}\|_{\mathcal{S}^2}^2 + \|Z^t - Z^{t_n}\|_{\mathcal{H}_d^2}^2 + \|\widetilde{V}^t - \widetilde{V}^{t_n}\|_{\mathcal{H}_\mu^2}^2 + \|A^t - A^{t_n}\|_{\mathcal{S}^2}^2 \leq c(\Lambda)c \\ & \times (\|\Phi(\mathcal{X}_T^t) - \Phi(\mathcal{X}_T^{t_n})\|_{L^2}^2 + \|\mathbb{1}_{\{\cdot > t\}} \widehat{g}(\cdot, \mathcal{X}_\cdot^t, Y_\cdot^t, Z_\cdot^t, \widetilde{V}_\cdot^t) \\ & - \mathbb{1}_{\{\cdot > t_n\}} \widehat{g}(\cdot, \mathcal{X}_{\cdot \vee t_n}^{t_n}, Y_\cdot^t, Z_\cdot^t, \widetilde{V}_\cdot^t)\|_{\mathcal{H}^2}^2 + \|\ell(\cdot \vee t, \mathcal{X}_{\cdot \vee t}^t) - \ell(\cdot \vee t_n, \mathcal{X}_{\cdot \vee t_n}^{t_n})\|_{\mathcal{S}^2} \\ & + \|h(\cdot \vee t, \mathcal{X}_{\cdot \vee t}^t) - h(\cdot \vee t_n, \mathcal{X}_{\cdot \vee t_n}^{t_n})\|_{\mathcal{S}^2}). \end{aligned} \quad (15.1)$$

First note that $c(\Lambda)c \leq C(1 + |x|^{2\rho})$, by the results of part (i). We only then need show that each term goes to 0 as $n \rightarrow \infty$ on the right-hand side of (15.1). We give the details for the term

$$\|\mathbb{1}_{\{\cdot > t\}} \widehat{g}(\cdot, \mathcal{X}_\cdot^t, Y_\cdot^t, Z_\cdot^t, \widetilde{V}_\cdot^t) - \mathbb{1}_{\{\cdot > t_n\}} \widehat{g}(\cdot, \mathcal{X}_{\cdot \vee t_n}^{t_n}, Y_\cdot^t, Z_\cdot^t, \widetilde{V}_\cdot^t)\|_{\mathcal{H}^2}^2.$$

The other terms can be treated similarly. Introducing a sequence (R_m) of positive numbers going to infinity as $m \rightarrow \infty$, let

$$\mathcal{Q}_s^{m,n} := \{s \geq t \vee t_n\} \cap \{N_s^t = N_s^{t_n}\} \cap \{|X_s^t| \vee |X_s^{t_n}| \vee |Y_s^t| \vee |Z_s^t| \vee r_s^t \leq R_m\},$$

with $r_s^t := |\widehat{r}_s^t| \vee |\widetilde{r}_s^{t_n}| \vee |\widetilde{r}_s^t| \vee |\widetilde{r}_s^{t_n}|$, where

$$\begin{aligned}
\tilde{r}_s^t &= \int_{\mathbb{R}^d} V_s^t(y) \eta(s, \mathcal{X}_s^t) f(s, \mathcal{X}_s^t, y) \hat{m}(dy) \\
\tilde{r}_s^{t_n} &= \int_{\mathbb{R}^d} V_s^t(y) \eta(s, X_s^{t_n}, N_s^t) f(s, X_s^{t_n}, N_s^t, y) \hat{m}(dy) \\
\tilde{r}_s^t &= \int_{\mathbb{R}^d} V_s^t(y) (m(s, \mathcal{X}_s^t) - \hat{m}(dy)) \\
\tilde{r}_s^{t_n} &= \int_{\mathbb{R}^d} V_s^t(y) (m(s, X_s^{t_n}, N_s^t) - \hat{m}(dy)),
\end{aligned} \tag{15.2}$$

and let $\bar{\Omega}_s^{m,n}$ denote the complement of the set $\Omega_s^{m,n}$. For all m, n , we have:

$$\begin{aligned}
&\| \mathbb{1}_{\{\cdot > t\}} \widehat{g}(\cdot, \mathcal{X}_s^t, Y_s^t, Z_s^t, \tilde{V}_s^t) - \mathbb{1}_{\{\cdot > t_n\}} \widehat{g}(\cdot, \mathcal{X}_s^{t_n}, Y_s^t, Z_s^t, \tilde{V}_s^t) \|_{\mathcal{H}^2}^2 \\
&= \mathbb{E} \int_{t \wedge t_n}^T [\mathbb{1}_{\{s > t\}} \widehat{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \tilde{V}_s^t) - \mathbb{1}_{\{s > t_n\}} \widehat{g}(s, \mathcal{X}_s^{t_n}, Y_s^t, Z_s^t, \tilde{V}_s^t)]^2 ds \\
&= \mathbb{E} \int_{t \wedge t_n}^T [\mathbb{1}_{\{s > t\}} \widehat{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \tilde{V}_s^t) - \mathbb{1}_{\{s > t_n\}} \widehat{g}(s, \mathcal{X}_s^{t_n}, Y_s^t, Z_s^t, \tilde{V}_s^t)]^2 \mathbb{1}_{\bar{\Omega}_s^{m,n}} ds \\
&\quad + \mathbb{E} \int_{t \wedge t_n}^T [\mathbb{1}_{\{s > t\}} \widehat{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \tilde{V}_s^t) - \mathbb{1}_{\{s > t_n\}} \widehat{g}(s, \mathcal{X}_s^{t_n}, Y_s^t, Z_s^t, \tilde{V}_s^t)]^2 \mathbb{1}_{\Omega_s^{m,n}} ds \\
&\leq 2\mathbb{E} \int_{t \wedge t_n}^T [\widehat{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \tilde{V}_s^t)^2 + \widehat{g}(s, \mathcal{X}_s^{t_n}, Y_s^t, Z_s^t, \tilde{V}_s^t)^2] \mathbb{1}_{\bar{\Omega}_s^{m,n}} ds \\
&\quad + \mathbb{E} \int_0^T [\widehat{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \tilde{V}_s^t) - \widehat{g}(s, \mathcal{X}_s^{t_n}, Y_s^t, Z_s^t, \tilde{V}_s^t)]^2 \mathbb{1}_{\Omega_s^{m,n}} ds \\
&=: I_{m,n} + II_{m,n}.
\end{aligned}$$

Now,

$$\begin{aligned}
&\widehat{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \tilde{V}_s^t)^2 + \widehat{g}(s, \mathcal{X}_s^{t_n}, Y_s^t, Z_s^t, \tilde{V}_s^t)^2 \\
&\leq C(1 + |X_s^t|^{2\rho} + |X_s^{t_n}|^{2\rho} + |Y_s^t|^2 + |Z_s^t|^2 + |\tilde{V}_s^t|^2).
\end{aligned} \tag{15.3}$$

Observe that the sequence of processes $|X_s^{t_n}|^{2\rho}$ is equi- $dt \otimes d\mathbb{P}$ -integrable, by the estimate (12.43) on X applied for $p > 2\rho$. So likewise is therefore the right-hand side, and in turn the left-hand side, in (15.3), since $\tilde{Y}^t \in \mathcal{S}^2 \times \mathcal{H}_d^2 \times \mathcal{H}_\mu^2 \times \mathcal{A}^2$. Besides, we have that

$$\mathbb{E} \int_{t \wedge t_n}^T \mathbb{1}_{\bar{\Omega}_s^{m,n}} ds \leq |t - t_n| + \mathbb{E} \int_{t \vee t_n}^T \mathbb{1}_{\bar{\Omega}_s^{m,n}} ds, \tag{15.4}$$

where for $s \geq t \vee t_n$:

$$\bar{\Omega}_s^{m,n} \subseteq \{N_s^t \neq N_s^{t_n}\} \cup \{|X_s^t| \vee |X_s^{t_n}| \vee |Y_s^t| \vee |Z_s^t| \vee |r_s^t| \geq R_m\}.$$

Also, note that $\|r^t\|_{\mathcal{H}^2} < +\infty$. Using these estimates (12.45) on N , (12.43) on X and (12.70) on $\tilde{\mathcal{Y}}$, by the Markov inequality we thus obtain:

$$\begin{aligned} & \mathbb{E} \int_{t \vee t_n}^T \mathbb{1}_{\bar{\Omega}_s^{m,n}} ds \\ & \leq C |t - t_n| \\ & \quad + \mathbb{E} \int_{t \vee t_n}^T (\mathbb{1}_{\{|X_s^t| \geq R_m\}} + \mathbb{1}_{\{|X_s^{t_n}| \geq R_m\}} + \mathbb{1}_{\{|Y_s^t| \geq R_m\}} + \mathbb{1}_{\{|Z_s^t| \geq R_m\}} + \mathbb{1}_{\{|r_s^t| \geq R_m\}}) ds \\ & \leq C \left(|t - t_n| + \frac{1}{R_m^2} \right). \end{aligned}$$

Therefore (15.4) implies that $\mathbb{E} \int_{t \wedge t_n}^T \mathbb{1}_{\bar{\Omega}_s^{m,n}} ds$ goes to 0 as $m, n \rightarrow \infty$.

Now, notice that $\mathbb{E} \int_{t \wedge t_n}^T \mathbb{1}_{\bar{\Omega}_s^{m,n}} ds = \mathbb{E} \int_{t \wedge t_n}^T \mathbb{1}_{\tilde{\Omega}_s^{m,n}} ds$, with $\tilde{\Omega}_s^{m,n} = \bar{\Omega}_s^{m,n} \cap \{s > t \wedge t_n\}$. But, by standard results, the fact that $\mathbb{E} \int_0^T \mathbb{1}_{\tilde{\Omega}_s^{m,n}} ds \rightarrow 0$ as $m, n \rightarrow \infty$ implies that $\mathbb{E} \int_0^T f_s^{n,m} \mathbb{1}_{\tilde{\Omega}_s^{m,n}} ds \rightarrow 0$ as $m, n \rightarrow \infty$ for every equi-dt $\otimes d\mathbb{P}$ -integrable family of nonnegative processes $f = (f_s^{n,m})_{m,n}$. Applying this to

$$f^{n,m} = \hat{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \tilde{\mathcal{V}}_s^t)^2 + \hat{g}(s, \mathcal{X}_s^{t_n}, Y_s^t, Z_s^t, \tilde{\mathcal{V}}_s^t)^2,$$

we conclude that $I_{m,n} \rightarrow 0$ as $m, n \rightarrow \infty$.

On the other hand, since $N_s^t = N_s^{t_n}$ on $\Omega_s^{m,n}$, using the form (12.68) of \hat{g} in which g satisfies (M.1), we have

$$\begin{aligned} II_{m,n} &= \mathbb{E} \int_0^T [\hat{g}(s, X_s^t, N_s^t, Y_s^t, Z_s^t, \tilde{\mathcal{V}}_s^t) - \hat{g}(s, X_s^{t_n}, N_s^t, Y_s^t, Z_s^t, \tilde{\mathcal{V}}_s^t)]^2 \mathbb{1}_{\Omega_s^{m,n}} ds \\ &\leq \mathbb{E} \int_0^T \eta_m(|X_s^t - X_s^{t_n}| + |\tilde{r}_s^t - \tilde{r}_s^{t_n}|) ds \end{aligned}$$

for a modulus of continuity² η_m . Given $\epsilon > 0$, let m_ϵ, n_ϵ be such that $I_{m_\epsilon, n} \leq \frac{\epsilon}{2}$ for $n \geq n_\epsilon$. Further, letting μ_ϵ be such $\eta_{m_\epsilon} \leq \frac{\epsilon}{4T}$ on $[0, \mu_\epsilon]$ and C_ϵ denote an upper bound on η_{m_ϵ} , we obtain that, for every n :

$$\begin{aligned} II_{m_\epsilon, n} &\leq \mathbb{E} \int_0^T \eta_{m_\epsilon}(|X_s^t - X_s^{t_n}| + |\tilde{r}_s^t - \tilde{r}_s^{t_n}|) ds \\ &\leq \mathbb{E} \int_0^T \left(\frac{\epsilon}{4T} + C_\epsilon \mathbb{1}_{\{|X_s^t - X_s^{t_n}| \geq \mu_\epsilon\}} + C_\epsilon \mathbb{1}_{\{|\tilde{r}_s^t - \tilde{r}_s^{t_n}| \geq \mu_\epsilon\}} \right) ds \\ &\leq T \left(\frac{\epsilon}{4T} + C_\epsilon \mathbb{P} \left[\sup_{[0,T]} |X_s^t - X_s^{t_n}| \geq \mu_\epsilon \right] \right) + C_\epsilon \mathbb{E} \int_0^T \mathbb{1}_{\{|\tilde{r}_s^t - \tilde{r}_s^{t_n}| \geq \mu_\epsilon\}} ds. \end{aligned}$$

²Nonnegative function from $[0, \infty]$ to itself, continuous and null at 0.

Now, given the estimate (12.46), we have that $\mathbb{P}[\sup_{[0,T]} |X_s^t - X_s^{t_n}| \geq \mu_\epsilon] \rightarrow 0$ as $n \rightarrow \infty$, by the Markov inequality. Moreover (cf. (15.2)),

$$|\widehat{r}_s^t - \widehat{r}_s^{t_n}| \leq \int_{\mathbb{R}^d} |V_s^t(y)| |f(s, X_s^t, N_s^t, y) - f(s, X_s^{t_n}, N_s^t, y)| \widehat{m}(dy),$$

so that $\|\widehat{r}^t - \widehat{r}^{t_n}\|_{\mathcal{H}^2} \rightarrow 0$ as $n \rightarrow \infty$ by dominated convergence, using the Lipschitz continuity property of f in Lemma 12.3.5(i). Thus, by the Markov inequality,

$$\mathbb{E} \int_0^T \mathbb{1}_{\{|\widehat{r}_s^t - \widehat{r}_s^{t_n}| \geq \mu_\epsilon\}} ds \leq \frac{\|\widehat{r}^t - \widehat{r}^{t_n}\|_{\mathcal{H}^2}^2}{\mu_\epsilon^2},$$

which converges to 0 as $n \rightarrow \infty$.

Finally, for every $\epsilon > 0$, we have that $I_{m_\epsilon, n} + II_{m_\epsilon, n} \leq \epsilon$ for $n \geq n_\epsilon \vee n'_\epsilon$, which proves that

$$\|\mathbb{1}_{\{\cdot > t\}} \widehat{g}(\cdot, \mathcal{X}_\cdot^t, Y_\cdot^t, Z_\cdot^t, \widetilde{\mathcal{V}}_\cdot^t) - \mathbb{1}_{\{\cdot > t_n\}} \widehat{g}(\cdot, \mathcal{X}_\cdot^{t_n}, Y_\cdot^t, Z_\cdot^t, \widetilde{\mathcal{V}}_\cdot^t)\|_{\mathcal{H}^2}^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

15.1.3 Proof of Proposition 12.4.3

(i) As a diffusion SDE with Lipschitz and bounded coefficients between jump times with explicit jumps at the jump times, the stochastic differential equation (12.42) with initial condition $(t', \mathcal{X}_{t'}^t)$ admits a unique $(\mathbb{F}^{t'}, \mathbb{P})$ -solution $\mathcal{X}^{t'} = (X^{t'}, N^{t'})$ (see e.g. Ikeda and Watanabe [149]). By the same arguments this SDE also admits a unique $(\mathbb{F}_{|[t', T]}, \mathbb{P})$ -solution which, by uniqueness is given by $\mathcal{X}^{t'}$ as well, since $\widetilde{\mathbb{F}}^{t'} \subseteq \mathbb{F}_{|[t', T]}$. Now, $(N_r^t)_{t' \leq r \leq T}$ is an $\mathbb{F}_{|[t', T]}$ -adapted process satisfying the first line of (12.42) on $[t', T]$, so that $(X_r^t)_{t' \leq r \leq T}$ is in turn an $\mathbb{F}_{|[t', T]}$ -adapted process satisfying the second line of (12.42) on $[t', T]$. Therefore $\mathcal{X}^{t'} = (\mathcal{X}_r^t)_{t' \leq r \leq T}$, by uniqueness with respect to $(\mathcal{Q}, \mathbb{F}_{|[t', T]}, \mathbb{P})$.

(ii) Note that the bound estimate (12.43) on X^t is also valid for solutions of stochastic differential equations with random initial condition such as $X^{t'}$ in part (i), by application of Proposition 12.1.3 (cf. the proof of Proposition 12.3.3). With $X^{t'}$ extended by $X^{t'} = X_{t'}^t$ on $[0, t']$, we thus have for every $p \in [2, +\infty)$:

$$\|X^{t'}\|_{\mathcal{S}_d^p}^p \leq C_p (1 + \mathbb{E}|X_{t'}^t|^p) \leq C'_p (1 + |x|^p),$$

where the last inequality comes from (12.43). Consequently, (H.0) ^{t'} –(H.2) ^{t'} in the proof of Proposition 12.4.1(i) still hold, with t' (in the sense of the initial condition $(t', \mathcal{X}_{t'}^t)$ for \mathcal{X}) instead of t there. Also, given the local martingale predictable representation property of Proposition 12.3.7(ii) applied with t' and $\mathcal{X}_{t'}^t$ here replacing t and \tilde{M}_t there, in view of the form postulated in (M.3) for ℓ , existence and uniqueness for an $(\mathbb{F}^{t'}, \mathbb{P}^{t'})$, $(B^{t'}, \mu^{t'})$ -solution $\mathcal{Y}^{t'} = (Y^{t'}, Z^{t'}, \mathcal{V}^{t'}, A^{t'})$ of the R2BSDE with

the data (12.71) on $[t', T]$ follows from Proposition 12.1.12. Proposition 12.1.12 also implies uniqueness for an $(\mathbb{F}_{|[t', T]}^t, \mathbb{P}^{t'})$, $(B^{t'}, \mu^{t'})$ -solution to the R2BSDE with the data (12.71) on $[t', T]$, given “(H.0) t –(H.2) t with t' replacing t ”. Since $\widetilde{\mathbb{F}}^{t'} \subseteq \mathbb{F}_{|[t', T]}^t$, $\mathcal{Y}^{t'} = (Y^{t'}, Z^{t'}, \mathcal{V}^{t'}, A^{t'})$ is thus the unique $(\mathbb{F}_{|[t', T]}^t, \mathbb{P}^{t'})$, $(B^{t'}, \mu^{t'})$ -solution to the R2BSDE with the data (12.71) on $[t', T]$. Finally, given part (i), it can be immediately verified that $(Y_r^t, Z_r^t, \mathcal{V}_r^t, A_r^t - A_{t'}^{t'})_{t' \leq r \leq T}$ is an $(\mathbb{F}_{|[t', T]}^t, \mathbb{P}^{t'})$, $(B^{t'}, \mu^{t'})$ -solution to the R2BSDE with the data (12.71) on $[t', T]$. We conclude the proof by the uniqueness with respect to $(\mathbb{F}_{|[t', T]}^t, \mathbb{P}^{t'})$, $(B^{t'}, \mu^{t'})$.

15.1.4 Proof of Proposition 12.4.7

By the bound estimate (12.70) on $\widetilde{\mathcal{Y}}^t$, we have that $Y_{\vartheta^t}^t \in L^2$. Moreover, we can check that, as in the proof of Proposition 12.4.2, the driver $\mathbb{1}_{\{t < \cdot < \vartheta^t\}} \widehat{g}(\cdot, \mathcal{X}_s^t, y, z, \widehat{v})$ satisfies (H.1). Hence the data

$$\mathbb{1}_{\{t < s < \vartheta^t\}} \widehat{g}(s, \mathcal{X}_s^t, y, z, \widehat{v}), \quad Y_{\vartheta^t}^t, \quad \ell(t \vee s \wedge \vartheta^t, \mathcal{X}_{s \vee t \wedge \vartheta^t}^t)$$

satisfy (H.0), (H.1) and the assumptions regarding L in (H.2) with respect to (\mathbb{F}, \mathbb{P}) , (B, μ) .

(i) An application of (12.16) yields the following bound estimate on $\widetilde{\mathcal{Y}}^t$:

$$\|\overline{Y}^t\|_{\mathcal{S}^2}^2 + \|\overline{Z}^t\|_{\mathcal{H}_d^2}^2 + \|\widehat{\mathcal{V}}^t\|_{\mathcal{H}_\mu^2}^2 + \|\overline{A}^t\|_{\mathcal{S}^2}^2 \leq c(\Lambda)c,$$

with

$$c := \|Y_{\vartheta^t}^t\|_{L^2}^2 + \|\widehat{g}(\cdot, \mathcal{X}_s^t, 0, 0, 0)\|_{\mathcal{H}^2}^2 + \|\ell(t \vee \cdot \wedge \vartheta^t, \mathcal{X}_{t \vee \cdot \wedge \vartheta^t}^t)\|_{\mathcal{S}^2}^2.$$

The estimate (12.77) follows by standard computations, given the Lipschitz and growth assumptions on the data and the estimate (12.43) on X^t .

(ii) Given the assumptions made on ℓ , an application of (12.17) (or more precisely its variant addressed in the lines following it) yields the following error estimate, in which c is as above:

$$\begin{aligned} & \|\overline{Y}^t - \overline{Y}^{t_n}\|_{\mathcal{S}^2}^2 + \|\overline{Z}^t - \overline{Z}^{t_n}\|_{\mathcal{H}_d^2}^2 + \|\widehat{\mathcal{V}}^t - \widehat{\mathcal{V}}^{t_n}\|_{\mathcal{H}_\mu^2}^2 + \|\overline{A}^t - \overline{A}^{t_n}\|_{\mathcal{S}^2}^2 \leq c(\Lambda)c \\ & \times (\|Y_{\vartheta^t}^t - Y_{\vartheta^{t_n}}^{t_n}\|_{L^2}^2 \\ & + \|\mathbb{1}_{\{t < \cdot < \vartheta^t\}} \widehat{g}(\cdot, \mathcal{X}_s^t, \overline{Y}_s^t, \overline{Z}_s^t, \widehat{\mathcal{V}}_s^t) - \mathbb{1}_{\{t_n < \cdot < \vartheta^{t_n}\}} \widehat{g}(\cdot, \mathcal{X}_s^{t_n}, \overline{Y}_s^t, \overline{Z}_s^t, \widehat{\mathcal{V}}_s^t)\|_{\mathcal{H}^2}^2 \\ & + \|\ell(t \vee \cdot \wedge \vartheta^t, \mathcal{X}_{t \vee \cdot \wedge \vartheta^t}^t) - \ell(t_n \vee \cdot \wedge \vartheta^{t_n}, \mathcal{X}_{t_n \vee \cdot \wedge \vartheta^{t_n}}^{t_n})\|_{\mathcal{H}^2}) \end{aligned} \tag{15.5}$$

(with, in particular, $\|\cdot\|_{\mathcal{H}^2}$, better than $\|\cdot\|_{\mathcal{S}^2}$ in the last term). Using the result of part (i), we already have that $c(\Lambda)c \leq C(1 + |x|^{2\rho})$ by (i). We need only show

that each term goes to 0 as $n \rightarrow \infty$ on the right-hand side of (15.5). We provide a detailed proof for the term

$$\|\mathbb{1}_{\{t < \cdot < \vartheta^t\}} \widehat{g}(\cdot, \mathcal{X}^t, \bar{Y}^t, \bar{Z}^t, \widehat{\mathcal{V}}^t) - \mathbb{1}_{\{t_n < \cdot < \vartheta^{t_n}\}} \widehat{g}(\cdot, \mathcal{X}^{t_n}, \bar{Y}^t, \bar{Z}^t, \widehat{\mathcal{V}}^t)\|_{\mathcal{H}^2}^2;$$

the other terms can be treated along the same lines. Introducing a sequence (R_m) of positive numbers going to infinity as $m \rightarrow \infty$, let $\Omega_s^{m,n}$ and $\bar{\Omega}_s^{m,n}$ be defined as in the proof of Proposition 12.4.2(ii), with $(\bar{Y}^t, \bar{Z}^t, \widehat{\mathcal{V}}^t)$ instead of $(Y^t, Z^t, \mathcal{V}^t)$ there. For all m, n we have:

$$\begin{aligned} & \|\mathbb{1}_{\{t < s < \vartheta^t\}} \widehat{g}(\cdot, \mathcal{X}^t, \bar{Y}^t, \bar{Z}^t, \widehat{\mathcal{V}}^t) - \mathbb{1}_{\{t_n < s < \vartheta^{t_n}\}} \widehat{g}(\cdot, \mathcal{X}^{t_n}, \bar{Y}^t, \bar{Z}^t, \widehat{\mathcal{V}}^t)\|_{\mathcal{H}^2}^2 \\ &= \mathbb{E} \int_0^T [\mathbb{1}_{\{t < s < \vartheta^t\}} \widehat{g}(s, \mathcal{X}_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t) - \mathbb{1}_{\{t_n < s < \vartheta^{t_n}\}} \widehat{g}(s, \mathcal{X}_s^{t_n}, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t)]^2 ds \\ &\leq 2\mathbb{E} \int_0^T [\widehat{g}(s, \mathcal{X}_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t)^2 + \widehat{g}(s, \mathcal{X}_s^{t_n}, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t)^2] \mathbb{1}_{\bar{\Omega}_s^{m,n}} ds \\ &\quad + \mathbb{E} \int_0^T [\mathbb{1}_{\{t < s < \vartheta^t\}} \widehat{g}(s, \mathcal{X}_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t) \\ &\quad - \mathbb{1}_{\{t_n < s < \vartheta^{t_n}\}} \widehat{g}(s, \mathcal{X}_s^{t_n}, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t)]^2 \mathbb{1}_{\Omega_s^{m,n}} ds \\ &=: I_{m,n} + II_{m,n}. \end{aligned}$$

As in the proof Proposition 12.4.2(ii) (using the fact that $\widehat{\mathcal{Y}}^t \in \mathcal{S}^2 \times \mathcal{H}_d^2 \times \mathcal{H}_\mu^2 \times \mathcal{A}^2$ instead of $\widetilde{\mathcal{Y}}^t$ there), $I_{m,n} \rightarrow 0$ as $m, n \rightarrow \infty$. Moreover, since $N_s^t = N_s^{t_n}$ on $\Omega_s^{m,n}$, we have that

$$\begin{aligned} II_{m,n} &= \mathbb{E} \int_0^T [\mathbb{1}_{\{t < s < \vartheta^t\}} \widehat{g}(s, X_s^t, N_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t) \\ &\quad - \mathbb{1}_{\{t_n < s < \vartheta^{t_n}\}} \widehat{g}(s, X_s^{t_n}, N_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t)]^2 \mathbb{1}_{\Omega_s^{m,n}} ds \\ &\leq 2\mathbb{E} \int_0^T [\widehat{g}(s, X_s^t, N_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t) - \widehat{g}(s, X_s^{t_n}, N_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t)]^2 \mathbb{1}_{\Omega_s^{m,n}} ds \\ &\quad + 2\mathbb{E} \int_0^T |\mathbb{1}_{\{t < s < \vartheta^t\}} - \mathbb{1}_{\{t_n < s < \vartheta^{t_n}\}}| \widehat{g}(s, X_s^t, N_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t)^2 ds, \end{aligned}$$

where in the last inequality:

- $\mathbb{E} \int_0^T [\widehat{g}(s, X_s^t, N_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t) - \widehat{g}(s, X_s^{t_n}, N_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{\mathcal{V}}_s^t)]^2 \mathbb{1}_{\Omega_s^{m,n}} ds \leq \mathbb{E} \int_0^T \eta_m(|X_s^t - X_s^{t_n}|) ds$ for a bounded modulus of continuity η_m (cf. the proof of Proposition 12.4.2(ii));

- $\mathbb{E} \int_0^T |\mathbb{1}_{\{t < s < \vartheta^t\}} - \mathbb{1}_{\{t_n < s < \vartheta^{t_n}\}}| \widehat{g}(s, X_s^t, N_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{V}_s^t)^2 ds$ goes to 0 as $n, m \rightarrow \infty$, by $dt \otimes d\mathbb{P}$ -integrability of $\widehat{g}(s, X_s^t, N_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \widehat{V}_s^t)^2$ joint with the fact that

$$\mathbb{E} \int_0^T |\mathbb{1}_{\{t < s < \vartheta^t\}} - \mathbb{1}_{\{t_n < s < \vartheta^{t_n}\}}| ds = \mathbb{E} |\vartheta^t - \vartheta^{t_n}| + |t - t_n| \rightarrow 0 \quad \text{as } n, m \rightarrow \infty,$$

by dominated convergence (under Assumption 12.4.5). We conclude as for the proof of Proposition 12.4.2(ii).

15.1.5 Proof of Proposition 12.4.10

(i) As a diffusion SDE with Lipschitz and bounded coefficients between jump times, with explicit jumps at jump times, the stochastic differential equation (12.78), with initial condition $(t', \bar{\mathcal{X}}_{t'}')$, admits a unique $(\bar{\mathbb{F}}^t, \mathbb{P})$ -solution $\bar{\mathcal{X}}^{t'} = (\bar{X}^{t'}, \bar{N}^{t'})$ (see Ikeda and Watanabe [149]). This SDE also admits a unique $(\mathbb{F}_{|[t', T]}, \mathbb{P})$ -solution, which by uniqueness is given by $\bar{\mathcal{X}}^{t'}$ as well, since $\bar{\mathbb{F}}^t \subseteq \mathbb{F}_{|[t', T]}$. Now, $(\bar{N}_r^t)_{t' \leq r \leq T}$ is an $\mathbb{F}_{|[t', T]}$ -adapted process that satisfies the first line of (12.78) on $[t', T]$, and $(\bar{X}_r^t)_{t' \leq r \leq T}$ is then in turn an $\mathbb{F}_{|[t', T]}$ -adapted process that satisfies the second line of (12.78) on $[t', T]$. Therefore $\bar{\mathcal{X}}^{t'} = (\bar{X}_r^t)_{t' \leq r \leq T}$, by uniqueness with respect to $(\Omega, \mathbb{F}_{|[t', T]}, \mathbb{P})$.

(ii) We have, as in the proof of Proposition 12.4.3(ii):

$$\|\bar{X}^{t'}\|_{\mathcal{S}_d^p}^p \leq C_p (1 + \mathbb{E} |\bar{X}_{t'}^t|^p) \leq C'_p (1 + |x|^p).$$

Consequently, the data

$$\mathbb{1}_{\{s < \vartheta'\}} \widetilde{g}(s, \bar{\mathcal{X}}_s^{t'}, y, z, \widetilde{v}), \quad Y_{\vartheta^t}^t, \quad \ell(s \wedge \vartheta', \bar{\mathcal{X}}_{s \wedge \vartheta'}^{t'}) \quad (15.6)$$

satisfy the general assumptions (H.0), (H.1) and the assumptions regarding L in (H.2) with respect to $(\bar{\mathbb{F}}^t, \mathbb{P}^t)$, (B^t, μ^t) or $(\mathbb{F}_{|[t', T]}, \mathbb{P}^t)$, (B^t, μ^t) . Given the local martingale predictable representation property of $(\mathbb{F}^t, \mathbb{P}^t; B^t, \chi^t, v^t)$ (cf. Proposition 12.3.7(ii)) and the form postulated in (M.3) for ℓ , Proposition 12.1.12 implies existence and uniqueness for an $(\bar{\mathbb{F}}^t, \mathbb{P}^t)$, (B^t, μ^t) -solution $\bar{\mathcal{Y}}^{t'} = (\bar{Y}^{t'}, \bar{Z}^{t'}, \bar{V}^{t'}, \bar{A}^{t'})$ to the stopped RBSDE with the data (12.80) on $[t', T]$, which is also the unique $(\mathbb{F}_{|[t', T]}, \mathbb{P}^t)$, (B^t, μ^t) -solution to the stopped RBSDE with the data (12.80) on $[t', T]$. Moreover, by Proposition 12.4.4(ii), $(\bar{Y}_r^t, \bar{Z}_r^t, \bar{V}_r^t, \bar{A}_r^t)_{t \leq r \leq T}$ is an $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, μ^t) -solution to the stopped RBSDE on $[t, T]$ with the data (12.74), where, in particular, $\bar{V}^t = \mathbb{1}_{s \leq \vartheta^t} \check{V}^t$ and $\check{V}^t = (\check{V}^t, \check{W}^t)$ for some $\check{V}^t \in \mathcal{H}_{\chi^t}^2$, $\check{W}^t \in \mathcal{H}_{v^t}^2$. So, by Definition 12.1.5(b)(ii):

$$\begin{cases} \bar{Y}_s^t = Y_{\vartheta^t}^t + \int_{s \wedge \vartheta^t}^{\vartheta^t} \tilde{g}(r, \bar{\mathcal{X}}_r^t, \bar{Y}_r^t, \bar{Z}_r^t, \bar{V}_r^t) dr + \bar{A}_{\vartheta^t}^t - \bar{A}_{s \wedge \vartheta^t}^t - \int_{s \wedge \vartheta^t}^{\vartheta^t} \bar{Z}_r^t dB_r \\ \quad - \int_{s \wedge \vartheta^t}^{\vartheta^t} \int_{\mathbb{R}^d} \check{V}_r^t \tilde{\chi}^t(dy, dr) - \sum_{j \in I} \int_{s \wedge \vartheta^t}^{\vartheta^t} \check{W}_r^t d\tilde{v}_r^t(j), \quad s \in [t, T] \\ \ell(s, \bar{\mathcal{X}}_s^t) \leq \bar{Y}_s^t \quad \text{for } s \in [t, \vartheta^t], \text{ and } \int_t^{\vartheta^t} (\bar{Y}_s^t - \ell(s, \bar{\mathcal{X}}_s^t)) d\bar{A}_s^t = 0 \\ \bar{Y}^t, \bar{A}^t \quad \text{are constant on } [\vartheta^t, T]. \end{cases}$$

Therefore, given in particular (12.79) in part (i) of the proposition:

$$\begin{cases} \bar{Y}_s^t = Y_{\vartheta'}^t + \int_{s \wedge \vartheta'}^{\vartheta'} \tilde{g}(r, \bar{\mathcal{X}}_r^{\vartheta'}, \bar{Y}_r^t, \bar{Z}_r^t, \bar{V}_r^t) dr + \bar{A}_{\vartheta'}^t - \bar{A}_{s \wedge \vartheta'}^t - \int_{s \wedge \vartheta'}^{\vartheta'} \bar{Z}_r^t dB_r \\ \quad - \int_{s \wedge \vartheta'}^{\vartheta'} \int_{\mathbb{R}^d} \check{V}_r^t \tilde{\chi}^t(dy, dr) - \sum_{j \in I} \int_{s \wedge \vartheta'}^{\vartheta'} \check{W}_r^t d\tilde{v}_r^t(j), \quad s \in [t', T] \\ \ell(s, \bar{\mathcal{X}}_s^{\vartheta'}) \leq \bar{Y}_s^t \quad \text{for } s \text{ in } (t', \vartheta'], \text{ and } \int_{t'}^{\vartheta'} (\bar{Y}_s^t - \ell(s, \bar{\mathcal{X}}_s^{\vartheta'})) d(\bar{A}_s^t - \bar{A}_{t'}) = 0 \\ \bar{Y}^t, \bar{A}^t - \bar{A}_{t'}^t \quad \text{are constant on } [\vartheta', T], \end{cases}$$

where $\ell(s, \bar{\mathcal{X}}_s^{\vartheta'}) \leq \bar{Y}_s^t$ for s in $(t', \vartheta']$ in the third line implies that the last inequality also holds at $s = t'$, by right-continuity. So $(\bar{Y}_r^t, \bar{Z}_r^t, \bar{V}_r^t, \bar{A}_r^t - \bar{A}_{t'}^t)_{t' \leq r \leq T}$ is an $(\mathbb{F}_{[t', T]}, \mathbb{P}^t)$, (B^t, μ^t) -solution to the stopped RBSDE with the data (12.80) on $[t', T]$ (see Definition 12.1.5(b)(ii)). This implies (12.81), by the uniqueness established above for an $(\mathbb{F}_{[t', T]}, \mathbb{P}^t)$, (B^t, μ^t) -solution to the stopped RBSDE with the data (12.80) on $[t', T]$.

15.1.6 Proof of Theorem 12.5.1

(i) Letting $r = t' = s$ in the semigroup property (12.72) of \mathcal{Y} yields

$$Y_s^t = u(s, \mathcal{X}_s^t), \quad \mathbb{P}^t\text{-a.s.} \quad (15.7)$$

for a function u on \mathcal{E} . In particular:

$$Y_t^t = u^i(t, x), \quad (t, x, i) \in \mathcal{E}. \quad (15.8)$$

That u is in \mathcal{P} follows directly from (15.8), by the bound estimate (12.70) on \tilde{y}^t . Let $\mathcal{E} \ni (t_n, x_n, i) \rightarrow (t, x, i)$ as $n \rightarrow \infty$. We make the decomposition

$$|u^i(t, x) - u^i(t_n, x_n)| = |Y_t^t - Y_{t_n}^{t_n}| \leq |\mathbb{E}(Y_t^t - Y_{t_n}^t)| + \mathbb{E}|Y_{t_n}^t - Y_{t_n}^{t_n}|,$$

where the second term goes to 0 as $n \rightarrow \infty$ by Proposition 12.4.2(ii). As for the first term, since the R2BSDE with the data (12.67) is satisfied by $\tilde{\mathcal{Y}}^t$, we have:

$$\mathbb{E}(Y_t^t - Y_{t_n}^t) \leq \mathbb{E} \int_{t \wedge t_n}^{t \vee t_n} |\hat{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \tilde{\mathcal{V}}_s^t)| ds + \mathbb{E}|A_{t \vee t_n}^t - A_{t \wedge t_n}^t|,$$

in which the second term goes to 0 as $n \rightarrow \infty$ by Proposition 12.4.2(i), and

$$\mathbb{E} \int_{t \wedge t_n}^{t \vee t_n} |\hat{g}(s, \mathcal{X}_s^t, Y_s^t, Z_s^t, \tilde{\mathcal{V}}_s^t)| ds \leq \|\hat{g}(\cdot, \mathcal{X}_\cdot^t, Y_\cdot^t, Z_\cdot^t, \tilde{\mathcal{V}}_\cdot^t)\|_{\mathcal{H}^2} |t - t_n|^{\frac{1}{2}},$$

which also goes to 0 as $n \rightarrow \infty$, by the properties of g and the bound estimate (12.4.2) on $\tilde{\mathcal{Y}}^t$. So $u^i(t_n, x_n) \rightarrow u^i(t, x)$ whenever $\mathcal{E} \ni (t_n, x_n) \rightarrow (t, x)$ as $n \rightarrow \infty$, which establishes the continuity of u on \mathcal{E} . The identity (12.32) then follows from (15.7) by the fact that Y^t and³ $u(\cdot, \mathcal{X}^t)$ are càdlàg processes. On $\{\Delta N^t \neq 0\}$ (thus $\Delta X^t = 0$), also using the continuity of u , we then have:

$$\begin{aligned} \Delta Y_s^t &= u(s, \mathcal{X}_s^t) - u(s, \mathcal{X}_{s-}^t) \\ &= \sum_{j \in I} (u^j(s, X_{s-}^t) - u(s, \mathcal{X}_{s-}^t)) \Delta v_s^t(j) = \sum_{j \in I} W_s(j) \Delta v_s^t(j), \end{aligned}$$

where the last equality comes from the R2BSDE with the data (12.30) satisfied by \mathcal{Y}^t . The last equality also trivially holds on $\{\Delta N^t = 0\}$. Denoting $\mathcal{W}_s^t(j) = u^j(s, X_{s-}^t) - u(s, \mathcal{X}_{s-}^t)$, on $[t, T]$ we thus have:

$$\begin{aligned} 0 &= \sum_{j \in I} (\mathcal{W}_s^t(j) - W_s^t(j)) \Delta v_s^t(j) \\ &= \sum_{j \in I} (\mathcal{W}_s^t(j) - W_s^t(j)) \Delta \tilde{v}_s^t(j) + \sum_{j \in I} (\mathcal{W}_s^t(j) - W_s^t(j)) n^j(s, \mathcal{X}_s^t) ds, \end{aligned}$$

\mathbb{P}^t -almost surely. Therefore $\mathcal{W}_s^t(j) = W_s^t(j)$ on $[t, T]$, \mathbb{P}^t -almost surely, by the uniqueness for the canonical decomposition of a special semimartingale. This proves (12.33). Now note that, for $(y, z, \tilde{v}) = (Y_s^t, Z_s^t, \mathcal{V}_s^t)$ in (12.29):

$$\tilde{u}_s^t(N_s^t) = Y_s^t = u(s, \mathcal{X}_s^t),$$

by (12.32), and that then, for $j \neq N_s^t$:

$$(\tilde{u}_s^t)^j = Y_s^t + W_s^t(j) = u(s, \mathcal{X}_s^t) + (u^j(s, X_{s-}^t) - u(s, \mathcal{X}_{s-}^t))$$

by (12.33). Therefore $\tilde{u}_{s-}^t = u(s, X_{s-}^t)$, so that, by definition (12.28) of \tilde{g} :

³Given the continuity of u .

$$\begin{aligned}
& \int_t^s \tilde{g}(\zeta, \mathcal{X}_\zeta^t, Y_\zeta^t, Z_\zeta^t, \mathcal{V}_\zeta^t) d\zeta \\
&= \int_t^s \left[g(\zeta, \mathcal{X}_\zeta^t, \tilde{u}_\zeta^t, Z_\zeta^t, \tilde{r}_\zeta^t) - \sum_{j \in I} W_\zeta^t(j) n^j(\zeta, \mathcal{X}_\zeta^t) \right] d\zeta \\
&= \int_t^s \left[g(\zeta, \mathcal{X}_\zeta^t, \tilde{u}_{\zeta-}^t, Z_\zeta^t, \tilde{r}_\zeta^t) - \sum_{j \in I} (u^j(s, X_{\zeta-}^t) - u(\zeta, \mathcal{X}_{\zeta-}^t)) n^j(\zeta, \mathcal{X}_\zeta^t) \right] d\zeta \\
&= \int_t^s \left[g(\zeta, \mathcal{X}_\zeta^t, u(\zeta, X_{\zeta-}^t), Z_\zeta^t, \tilde{r}_\zeta^t) - \sum_{j \in I} (u^j(\zeta, X_\zeta^t) - u(\zeta, \mathcal{X}_\zeta^t)) n^j(\zeta, \mathcal{X}_\zeta^t) \right] d\zeta \\
&= \int_t^s \left[g(\zeta, \mathcal{X}_\zeta^t, u(\zeta, X_\zeta^t), Z_\zeta^t, \tilde{r}_\zeta^t) - \sum_{j \in I} (u^j(\zeta, X_\zeta^t) - u(\zeta, \mathcal{X}_\zeta^t)) n^j(\zeta, \mathcal{X}_\zeta^t) \right] d\zeta,
\end{aligned}$$

which is (12.34).

(ii) Letting $r = t' = s$ in the semigroup property (12.81) of $\bar{\mathcal{Y}}$ yields:

$$\bar{Y}_s^t = v(s, \bar{\mathcal{X}}_s^t), \quad \mathbb{P}^t\text{-a.s.} \quad (15.9)$$

for a function v on \mathcal{E} . In particular,

$$\bar{Y}_t^t = v^i(t, x), \quad (t, x, i) \in \mathcal{E}. \quad (15.10)$$

That v is in \mathcal{P} then directly follows from the bound estimate (12.77) on $\hat{\mathcal{Y}}^t$. Moreover, given $\mathcal{E} \ni (t_n, x_n, i) \rightarrow (t, x, i)$ as $n \rightarrow \infty$, we make the decomposition

$$|v^i(t, x) - v^i(t_n, x_n)| = |\bar{Y}_t^t - \bar{Y}_{t_n}^{t_n}| \leq |\mathbb{E}(\bar{Y}_t^t - \bar{Y}_{t_n}^{t_n})| + \mathbb{E}|\bar{Y}_{t_n}^{t_n} - \bar{Y}_{t_n}^{t_n}|,$$

where the second term goes to 0 as $n \rightarrow \infty$ by Proposition 12.4.7(ii). As for the first term, we have, by the stopped RBSDE with the data (12.75) satisfied by $\hat{\mathcal{Y}}^t$:

$$|\mathbb{E}(\bar{Y}_t^t - \bar{Y}_{t_n}^{t_n})| \leq \mathbb{E} \int_{t \wedge t_n}^{t \vee t_n} |\hat{g}(s, \bar{\mathcal{X}}_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \hat{\mathcal{V}}_s^t)| ds + \mathbb{E}|\bar{A}_{t \vee t_n}^t - \bar{A}_{t \wedge t_n}^t|,$$

in which the second term goes to 0 as $n \rightarrow \infty$ by Proposition 12.4.7(i), and

$$\mathbb{E} \int_{t \wedge t_n}^{t \vee t_n} |\hat{g}(s, \bar{\mathcal{X}}_s^t, \bar{Y}_s^t, \bar{Z}_s^t, \hat{\mathcal{V}}_s^t)| ds \leq \|\hat{g}(\cdot, \bar{\mathcal{X}}_\cdot^t, \bar{Y}_\cdot^t, \bar{Z}_\cdot^t, \hat{\mathcal{V}}_\cdot^t)\|_{\mathcal{H}^2} |t - t_n|^{\frac{1}{2}},$$

which also goes to 0 as $n \rightarrow \infty$, by the properties of g and the bound estimate (12.77) on $\hat{\mathcal{Y}}^t$. Thus $v^i(t_n, x_n) \rightarrow v^i(t, x)$ whenever $\mathcal{E} \ni (t_n, x_n, i) \rightarrow (t, x, i)$ as $n \rightarrow \infty$, which establishes the continuity of v on \mathcal{E} . The identity (12.36) then follows from (15.9) by the fact that \bar{Y}^t and⁴ $v(\cdot, \mathcal{X}^t)$ are càdlàg processes. Since $\bar{Y}_{\vartheta^t}^t = Y_{\vartheta^t}^t$ (cf. Proposition 12.4.4(ii)), (12.36) and (12.32) in turn imply (12.35).

⁴Given the continuity of v .

Furthermore, on $\{(s, \omega); s \in [t, \vartheta^t], N_s^t \neq N_{s-}^t\}$ (thus $\Delta X^t = 0$), also using the continuity of v , we have:

$$\begin{aligned}\Delta \bar{Y}_s^t &= v(s, \mathcal{X}_s^t) - v(s, \mathcal{X}_{s-}^t) \\ &= \sum_{j \in I} (v^j(s, X_{s-}^t) - v(s, \mathcal{X}_{s-}^t)) \Delta v_s^t(j) = \sum_{j \in I} \check{W}_s(j) \Delta v_s^t(j),\end{aligned}$$

where the last equality comes from the stopped RBSDE on $[t, T]$ with the data (12.74) which is satisfied by \bar{Y}^t . The last equality also holds trivially on $\{(s, \omega); s \in [t, \vartheta^t], N_s^t = N_{s-}^t\}$. Denoting $\mathcal{W}_s^t(j) := v^j(s, X_{s-}^t) - v(s, \mathcal{X}_{s-}^t)$, we thus have, on $[t, \vartheta^t]$:

$$\begin{aligned}0 &= \sum_{j \in I} (\mathcal{W}_s^t(j) - \check{W}_s^t(j)) \Delta v_s^t(j) \\ &= \sum_{j \in I} (\mathcal{W}_s^t(j) - \check{W}_s^t(j)) \Delta \tilde{v}_s^t(j) + \sum_{j \in I} (\mathcal{W}_s^t(j) - \check{W}_s^t(j)) n^j(s, \mathcal{X}_s^t) ds,\end{aligned}$$

\mathbb{P}^t -almost surely. Therefore $\mathcal{W}_s^t(j) = \check{W}_s^t(j)$ on $[t, \vartheta^t]$, by the uniqueness for the canonical decomposition of a special semimartingale. Hence (12.37) results. Finally, (12.38) derives from (12.36) and (12.37), as does (12.34) from (12.32) and (12.33).

15.1.7 Proof of Theorem 14.2.18

We prove parts (i) and (ii), which immediately imply (iii) by an application of Proposition 14.2.4. By taking $r = t'$ in the semigroup property (14.21) of \mathcal{Y} , for all $l = 1, \dots, m$ and $T_{l-1} \leq t \leq r < T_l$, we obtain

$$Y_r^t = v_l(r, \mathcal{X}_r^t), \quad \mathbb{P}^t\text{-a.s.} \quad (15.11)$$

for a function v_l on $\text{Int } \mathcal{E}_l$. In particular,

$$Y_t^t = v^i(t, x), \quad \text{for every } (t, x, i) \in \mathcal{E}, \quad (15.12)$$

where v is the function defined on \mathcal{E} by the concatenation of the v_l in $\text{Int } \mathcal{E}_l$ and the terminal condition Φ at T . In view of (14.16), the fact that v is in \mathcal{P} then follows directly from the bound estimates (14.19) on $\mathcal{Y}^{0,t}$ and $\mathcal{Y}^{1,t}$.

We show that v_l is continuous on $\text{Int } \mathcal{E}_l$. Given $\mathcal{E} \ni (t_n, x_n, i) \rightarrow (t, x, i)$ with $t \notin \mathfrak{T}$ or $t_n \geq T_l = t$, by (14.16) we make the decomposition:

$$\begin{aligned}|u^i(t, x) - u^i(t_n, x_n)| &= |Y_t^t - Y_{t_n}^{t_n}| \\ &\leq \begin{cases} |\mathbb{E}(Y_t^{0,t} - Y_{t_n}^{0,t})| + \mathbb{E}|Y_{t_n}^{0,t} - Y_{t_n}^{0,t_n}|, & i \in K \\ |\mathbb{E}(Y_t^{1,t} - Y_{t_n}^{1,t})| + \mathbb{E}|Y_{t_n}^{1,t} - Y_{t_n}^{1,t_n}|, & i \notin K. \end{cases}\end{aligned}$$

In either case we conclude, as in the proof of Theorem 12.5.1(ii), using Proposition 14.2.14 as our main tool, that $|v^i(t, x) - v^i(t_n, x_n)|$ goes to zero as $n \rightarrow \infty$.

It remains to show that v_l can be extended by continuity over \mathcal{E}_l , and that the jump condition (14.23) is satisfied (except perhaps at the boundary points (T_l, x, i) such that θ_l^i is discontinuous at x). Given $\text{Int } \mathcal{E}_l \ni (t_n, x_n, i) \rightarrow (t = T_l, x, i)$ with θ_l continuous at (x, i) , we need to show that $v_l^i(t_n, x_n) = v^i(t_n, x_n) \rightarrow v_l^i(T_l, x)$, where $v_l^i(T_l, x)$ is defined here by (14.23). We distinguish four cases.

- If $i \notin K$ and $\theta_l^i(x) \in K$, writing $\tilde{v}^j(s, y) = \min(v(s, y, \theta_l^j(y)), h(y))$, $\hat{v}^j(s, y) = \min(v(s, y, j), h(y))$, we have:

$$\begin{aligned} & |\tilde{u}^i(t, x) - u^i(t_n, x_n)|^2 \\ &= |\tilde{u}^i(t, x) - Y_{t_n}^{1, t_n}|^2 \\ &\leq 2\mathbb{E}|\tilde{v}^i(t, x) - \hat{v}(t, \mathcal{X}_t^{t_n})|^2 + 2\mathbb{E}(\hat{v}(t, \mathcal{X}_t^{t_n}) - Y_{t_n}^{1, t_n})|^2. \end{aligned} \quad (15.13)$$

By the continuity of θ_l at (x, i) , we obtain $\theta_l(\mathcal{X}_t^{t_n}) = \theta_l^i(x) \in K$ for $X_t^{t_n}$ close enough to x , say $|X_t^{t_n} - x| \leq c$. In this case $t = \vartheta_2^{t_n}$ and so (cf. (14.15)) $Y_t^{1, t_n} = \hat{v}(t, \mathcal{X}_t^{t_n})$. Thus

$$\mathbb{E}|\mathbb{1}_{|X_t^{t_n} - x| \leq c}(\hat{v}(t, \mathcal{X}_t^{t_n}) - Y_{t_n}^{1, t_n})|^2 \leq \mathbb{E}|Y_t^{1, t_n} - Y_{t_n}^{1, t_n}|^2,$$

which converges to zero as $n \rightarrow \infty$, since the R2BSDE is satisfied by Y^{1, t_n} and since \mathcal{Y}^{1, t_n} converges to $\tilde{\mathcal{Y}}^{1, t}$. Moreover, $\mathbb{E}|\mathbb{1}_{|X_t^{t_n} - x| > c}(\hat{v}(t, \mathcal{X}_t^{t_n}) - Y_{t_n}^{1, t_n})|^2$ goes to zero as $n \rightarrow \infty$ by the a priori estimates for X and Y^{1, t_n} and the continuity of \hat{v} already established over $\text{Int } \mathcal{E}_{l+1}$. Finally, by this continuity and the a priori estimates for X , the first term in (15.13) also goes to zero as $n \rightarrow \infty$. Thus, as $n \rightarrow \infty$,

$$v^i(t_n, x_n) \rightarrow \tilde{v}^i(t, x) = \min(v(t, x, \theta_l^i(x)), h(t, x)) = v_l^i(T_l, x).$$

- If $i \in K$ and $\theta_l^i(x) \notin K$, using $\check{v}^j(s, y) := v(s, y, \theta_l^j(y))$ instead of $\tilde{v}^j(s, y)$, $v(t, \mathcal{X}_t^{t_n})$ instead of $\hat{v}(t, \mathcal{X}_t^{t_n})$ and Y^0 instead of Y^1 above, we can likewise show that

$$v^i(t_n, x_n) \rightarrow \check{v}^i(t, x) = v_l^i(T_l, x) \quad (15.14)$$

as $n \rightarrow \infty$.

- If $i, \theta_l^i(x) \notin K$, we obtain

$$\begin{aligned} & |\check{u}^i(t, x) - u^i(t_n, x_n)|^2 \\ &= |\check{u}^i(t, x) - Y_{t_n}^{1, t_n}|^2 \\ &\leq 2\mathbb{E}|\check{v}^i(t, x) - v(t, \mathcal{X}_t^{t_n})|^2 + 2\mathbb{E}|v(t, \mathcal{X}_t^{t_n}) - Y_{t_n}^{1, t_n})|^2 \\ &\leq 2\mathbb{E}|\check{v}^i(t, x) - v(t, \mathcal{X}_t^{t_n})|^2 + 2\mathbb{E}|Y_{T_l}^{1, t_n} - Y_{t_n}^{1, t_n})|^2, \end{aligned}$$

which goes to zero as $\rightarrow \infty$, by an analysis similar to (and actually simpler than) that of the first case. Hence (15.14) results.

- If $i, \theta_i^j(x) \in K$, (15.14) can be shown as in the above case.

15.2 Proofs of PDE Results

15.2.1 Proof of Lemma 13.1.2

Letting $\varsigma = \eta f$, we make the decomposition

$$\begin{aligned} & \mathcal{I}\phi^i(t_n, x_n) - \mathcal{I}\phi^i(t, x) \\ &= - \int_{\mathbb{R}^d} (\phi^i(t_n, x_n) \varsigma^i(t_n, x_n, y) - \phi^i(t, x) \varsigma^i(t, x, y)) \widehat{m}(dy) \\ &\quad + \int_{\mathbb{R}^d} (\phi^i(t_n, x_n + \delta^i(t_n, x_n, y)) \varsigma^i(t_n, x_n, y) \\ &\quad - \phi^i(t, x + \delta^i(t, x, y)) \varsigma^i(t, x, y)) \widehat{m}(dy), \end{aligned}$$

where

$$\begin{aligned} & \int_{\mathbb{R}^d} (\phi^i(t_n, x_n + \delta^i(t_n, x_n, y)) \varsigma^i(t_n, x_n, y) - \phi^i(t, x + \delta^i(t, x, y)) \varsigma^i(t, x, y)) \widehat{m}(dy) \\ &= \int_{\mathbb{R}^d} (\phi^i(t_n, x_n + \delta^i(t_n, x_n, y)) - \phi^i(t, x + \delta^i(t, x, y))) \varsigma^i(t_n, x_n, y) \widehat{m}(dy) \\ &\quad + \int_{\mathbb{R}^d} \phi^i(t, x + \delta^i(t, x, y)) (\varsigma^i(t_n, x_n, y) - \varsigma^i(t, x, y)) \widehat{m}(dy) \end{aligned} \quad (15.15)$$

goes to 0 as $\mathcal{E} \ni (t_n, x_n) \rightarrow (t, x)$, by the assumption (E.2), and likewise for

$$\int_{\mathbb{R}^d} (\phi^i(t_n, x_n) \varsigma^i(t_n, x_n, y) - \phi^i(t, x) \varsigma^i(t, x, y)) \widehat{m}(dy).$$

The proof for $\delta\phi \cdot m$ is similar.

15.2.2 Proof of Theorem 13.2.1

(i) By definition, u is a continuous function in \mathcal{P} on \mathcal{E} . Moreover, by definition of u and \mathcal{Y} and letting the superscript T refer to an initial condition (T, x, i) for \mathcal{X} , we have:

$$\begin{aligned} u^i(T, x) &= Y_T^T = \Phi^i(x) \\ \ell^i(t, x) &\leq Y_t^t = u^i(t, x) \leq h^i(t, x). \end{aligned}$$

Thus $u = \Phi$ pointwise at T and $\ell \leq u \leq h$ on \mathcal{E} . We show that u is a subsolution to (V2) on $\text{Int } \mathcal{E}$. Readers may likewise check that u is a supersolution to (V2) on $\text{Int } \mathcal{E}$. Thus let $(t, x, i) \in \text{Int } \mathcal{E}$ and $\varphi \in \mathcal{C}^{1,2}(\mathcal{E})$ be such that $u^i - \varphi^i$ attains its maximum at (t, x) . Given that $u \leq h$, it suffices to prove that

$$-\tilde{\mathcal{G}}\varphi^i(t, x) - g^i(t, x, u(t, x), (\partial\varphi\sigma)^i(t, x), \mathcal{I}\varphi^i(t, x)) \leq 0, \quad (15.16)$$

where it is further assumed that $u^i(t, x) > \ell^i(t, x)$ and $u^i(t, x) = \varphi^i(t, x)$ (cf. Definition 13.1.3(a)(i)). Suppose to the contrary that (15.16) doesn't hold. Then by a continuity argument, using in particular Lemma 13.1.2, we have that

$$\psi(s, y) := \tilde{\mathcal{G}}\varphi^i(s, y) + g^i(s, y, u(s, y), (\partial\varphi\sigma)^i(s, y), \mathcal{I}\varphi^i(s, y)) < 0 \quad (15.17)$$

for every (s, y) such that $s \in [t, t + \alpha]$ and $|y - x| \leq \alpha$, for some sufficiently small $\alpha > 0$ with $t + \alpha < T$. Let

$$\tau = \inf\{s \geq t; |X_s^t - x| \geq \alpha, N_s^t \neq i, Y_s^t = \ell^i(s, X_s^t)\} \wedge (t + \alpha) \quad (15.18)$$

$$(\widehat{Y}^t, \widehat{Z}^t, \widehat{V}^t, \widehat{A}^t) = (\mathbb{1}_{\cdot < \tau} Y^t + \mathbb{1}_{\cdot \geq \tau} u^i(\tau, X_\tau^t), \mathbb{1}_{\cdot \leq \tau} Z^t, \mathbb{1}_{\cdot \leq \tau} V^t, A_{\cdot \wedge \tau}^t) \quad (15.19)$$

$$\begin{aligned} (\widetilde{Y}^t, \widetilde{Z}^t, \widetilde{V}^t) &= (\varphi^i(\cdot, X_{\cdot \wedge \tau}^t), \mathbb{1}_{\cdot \leq \tau} (\partial\varphi\sigma)^i(\cdot, X_\cdot^t)), \\ \mathbb{1}_{\cdot \leq \tau} ([\varphi^i(\cdot, X_{\cdot -}^t) + \delta^i(\cdot, X_{\cdot -}^t, y)] - \varphi^i(\cdot, X_{\cdot -}^t))]_{y \in \mathbb{R}^d}. \end{aligned} \quad (15.20)$$

Note that $\tau > t$, \mathbb{P}^t -almost surely. Thus, also using the continuity of u^i :

$$\widehat{Y}_t^t = Y_t^t = u^i(t, x) = \varphi^i(t, x) = \widetilde{Y}_t^t. \quad (15.21)$$

Also observe that $A^{t,+} = 0$ on $[t, \tau]$, by the related minimality condition in the R2BSDE equation for \mathcal{Y}^t , given that $\ell^i(s, X_s^t) < Y_s^t$ on $[t, \tau)$. We now show that, for $s \in [t, \tau]$, we have:

$$\begin{aligned} \widehat{Y}_s^t &= u^i(\tau, X_\tau^t) + \int_s^\tau g^i(\zeta, X_\zeta^t, u(\zeta, X_\zeta^t), \widehat{Z}_\zeta^t, \widehat{r}_\zeta^t) d\zeta - (\widehat{A}_\tau^{t,-} - \widehat{A}_s^{t,-}) \\ &\quad - \int_s^\tau \widehat{Z}_\zeta^t dB_\zeta^t - \int_s^\tau \int_{\mathbb{R}^d} \widehat{V}_\zeta^t(y) \widetilde{\chi}^t(d\zeta, dy). \end{aligned} \quad (15.22)$$

Indeed, this is true on $\{s = \tau\}$ by the definition of \widehat{Y}^t in (15.19). Furthermore, on $\{s < \tau\}$:

- either χ^t , and whence also X^t , do not jump at τ and the identity (15.22), with τ replaced by $r < \tau$, follows from the R2BSDE equation for \mathcal{Y}^t (in which $A^{t,+} = 0$ on $[t, \tau]$), so that (15.22) holds by passage to the limit as $r \rightarrow \tau-$;
- or N^t doesn't jump at τ , in which case the R2BSDE equation for \mathcal{Y}^t , integrated between s and τ , gives (15.22) directly.

Moreover, by application of the Itô formula (12.26) to the function $\widetilde{\varphi}$ defined by $\widetilde{\varphi}^j = \varphi^i$ for every $j \in I$, for every $s \in [t, \tau]$ we obtain:

$$\begin{aligned}
d\varphi^i(s, X_s^t) &= \mathcal{G}\tilde{\varphi}(s, \mathcal{X}_s^t) ds + (\partial\varphi\sigma)(s, \mathcal{X}_s^t) dB_s^t \\
&\quad + \int_{\mathbb{R}^d} (\varphi^i(s, X_{s-}^t + \delta(s, \mathcal{X}_{s-}^t, y)) - \varphi^i(s, \mathcal{X}_{s-}^t)) \tilde{\chi}^t(ds, dy) \\
&= \tilde{\mathcal{G}}\tilde{\varphi}(s, \mathcal{X}_s^t) ds + (\partial\varphi\sigma)(s, \mathcal{X}_s^t) dB_s^t \\
&\quad + \int_{\mathbb{R}^d} (\varphi^i(s, X_{s-}^t + \delta(s, \mathcal{X}_{s-}^t, y)) - \varphi^i(s, \mathcal{X}_{s-}^t)) \tilde{\chi}^t(ds, dy) \\
&= \tilde{\mathcal{G}}\varphi^i(s, X_s^t) ds + (\partial\varphi\sigma)^i(s, X_s^t) dB_s^t \\
&\quad + \int_{\mathbb{R}^d} (\varphi^i(s, X_{s-}^t + \delta^i(s, X_{s-}^t, y)) - \varphi^i(s, \mathcal{X}_{s-}^t)) \tilde{\chi}^t(ds, dy),
\end{aligned}$$

where the second equality uses (13.3), applied to $\tilde{\varphi}$, and the third exploits the facts that N^t cannot jump before τ and that $\tilde{\chi}^t$ cannot jump at τ if N^t does. Hence (cf. (15.20)):

$$\begin{aligned}
\tilde{Y}_s^t &= \varphi^i(\tau, X_\tau^t) - \int_s^\tau \tilde{\mathcal{G}}\varphi^i(r, X_r^t) dr - \int_s^\tau \tilde{Z}_r^t dB_r^t - \int_s^\tau \int_{\mathbb{R}^d} \tilde{V}_r^t(y) \tilde{\chi}^t(dr, dy) \\
&= \varphi^i(\tau, X_\tau^t) - \int_s^\tau (\psi(r, X_r^t) - g^i(r, X_r^t, u(r, X_r^t), (\partial\varphi\sigma)^i(r, X_r^t), \mathcal{I}\varphi^i(r, X_r^t))) dr \\
&\quad - \int_s^\tau \tilde{Z}_r^t dB_r^t - \int_s^\tau \int_{\mathbb{R}^d} \tilde{V}_r^t(y) \tilde{\chi}^t(dr, dy),
\end{aligned}$$

by definition (15.17) of ψ . For $s \in [t, \tau]$ we thus have:

$$\begin{aligned}
\tilde{Y}_s^t &= \varphi^i(\tau, X_\tau^t) - \int_s^\tau (\psi(\zeta, X_\zeta^t) - g^i(\zeta, X_\zeta^t, u(\zeta, X_\zeta^t), \tilde{Z}_\zeta^t, \mathcal{I}\varphi^i(\zeta, X_\zeta^t))) d\zeta \\
&\quad - \int_s^\tau \tilde{Z}_\zeta^t dB_\zeta^t - \int_s^\tau \int_{\mathbb{R}^d} \tilde{V}_\zeta^t(y) \tilde{\chi}^t(d\zeta, dy). \tag{15.23}
\end{aligned}$$

Let $\varsigma = \eta f$. Note that in (15.22)–(15.23), by definitions (12.29) of $\tilde{r}_\zeta^t = \tilde{r}_\zeta^t(V_\zeta^t)$, (13.4) of \mathcal{I} and (15.20) of \tilde{V} we have:

$$\begin{aligned}
&\int_s^\tau g^i(\zeta, X_\zeta^t, u(\zeta, X_\zeta^t), \tilde{Z}_\zeta^t, \tilde{r}_\zeta^t) d\zeta \\
&= \int_s^\tau g^i\left(\zeta, X_\zeta^t, u(\zeta, X_\zeta^t), \tilde{Z}_\zeta^t, \int_{\mathbb{R}^d} \widehat{V}_\zeta(y) \varsigma^i(\zeta, X_\zeta^t, y) \widehat{m}(dy)\right) d\zeta \\
&\int_s^\tau g^i(\zeta, X_\zeta^t, u(\zeta, X_\zeta^t), \tilde{Z}_\zeta^t, \mathcal{I}\varphi^i(\zeta, X_\zeta^t)) d\zeta \\
&= \int_s^\tau g^i\left(\zeta, X_\zeta^t, u(\zeta, X_\zeta^t), \tilde{Z}_\zeta^t, \int_{\mathbb{R}^d} \tilde{V}_\zeta(y) \varsigma^i(\zeta, X_\zeta^t, y) \widehat{m}(dy)\right) d\zeta.
\end{aligned}$$

Equations (15.22) and (15.23) respectively imply that:

- $(\widehat{Y}^t, \widehat{Z}^t, \widehat{V}^t)$ is an $(\mathbb{F}^t, \mathbb{P}^t), (B^t, \chi^t)$ -solution to the stopped BSDE on $[t, t + \alpha]$, with driver (cf. Definition 12.1.4(d) and Remark 12.1.7)

$$g^i\left(s, X_s^t, u(s, X_s^t), z, \int_{\mathbb{R}^d} v(y)\varsigma^i(s, X_s^t, y)\widehat{m}(dy)\right)ds - d\widehat{A}_s^{t,-}$$

and terminal condition $u^i(\tau, X_\tau^t)$ at τ ;

- $(\widetilde{Y}^t, \widetilde{Z}^t, \widetilde{V}^t)$ is an $(\mathbb{F}^t, \mathbb{P}^t), (B^t, \chi^t)$ -solution to the stopped BSDE on $[t, t + \alpha]$, with driver

$$g^i\left(s, X_s^t, u(s, X_s^t), z, \int_{\mathbb{R}^d} v(y)\varsigma^i(s, X_s^t, y)\widehat{m}(dy)\right)ds - \psi(s, X_s^t)ds$$

and terminal condition $\varphi^i(\tau, X_\tau^t)$ at τ .

Setting $\delta Y^t = \widehat{Y}^t - \widetilde{Y}^t$, we deduce by standard computations⁵ that

$$\Gamma_t^t \delta Y_t^t = \mathbb{E}^t \left[\Gamma_\tau^t \delta Y_\tau^t + \int_t^\tau \Gamma_s^t dD_s^t \right], \quad (15.24)$$

where:

- $\delta Y_\tau^t = \widehat{Y}_\tau^t - \widetilde{Y}_\tau^t = u^i(\tau, X_\tau^t) - \varphi^i(\tau, X_\tau^t) \leq 0$, by setting $s = \tau$ in (15.22)–(15.23) and since $u^i \leq \varphi^i$;
- $dD_s^t = \psi(r, X_s^t)ds - d\widehat{A}_s^{t,-}$, so that D^t is decreasing on $[t, \tau]$, by (15.17);
- Γ^t is a positive process, the so-called adjoint of δY^t (see, for instance, [87]).

Since, furthermore, $\tau > t$ holds \mathbb{P}^t -almost surely, we deduce that $\int_t^\tau \Gamma_s^t dD_s^t < 0$, \mathbb{P}^t -almost surely, whence $\delta Y_t^t < 0$ by (15.24). But this contradicts (15.21).

(ii) By definition, v is a continuous function on \mathcal{E} . Moreover, by the definitions of u , v , \mathcal{Y} and $\overline{\mathcal{Y}}$ (with ϑ defined as in Example 12.4.6), for $(t, x, i) \in \partial\mathcal{D}$ we have:

$$v^i(t, x) = \overline{Y}_t^t = Y_t^t = u^i(t, x)$$

and, for every $(t, x, i) \in \mathcal{E}$:

$$\ell^i(t, x) \leq \overline{Y}_t^t = v^i(t, x).$$

Thus $v = u$ on $\partial\mathcal{D}$ and $\ell \leq v$ on \mathcal{E} . We now show that v is a subsolution to (V1) on $\text{Int } \mathcal{D}$. Readers may likewise check that v is a supersolution to (V1) on $\text{Int } \mathcal{D}$. So, let $(t, x, i) \in \text{Int } \mathcal{D}$ and $\varphi \in \mathcal{C}^0(\mathcal{E}) \cap \mathcal{C}^{1,2}(\mathcal{D})$ such that $v^i - \varphi^i$ reaches its maximum at (t, x) . We need to prove that

$$-\tilde{\mathcal{G}}\varphi^i(t, x) - g^i(t, x, v(t, x), (\partial\varphi\sigma)^i(t, x), \mathcal{I}\varphi^i(t, x)) \leq 0, \quad (15.25)$$

where it is further assumed that $v^i(t, x) > \ell^i(t, x)$ and $v^i(t, x) = \varphi^i(t, x)$ (cf. Definition 13.1.3(a)(i)). Suppose to the contrary that (15.25) doesn't hold. By continuity,

⁵See for instance the proof of the comparison principle of Proposition 12.1.10 in [87].

we then have

$$\psi(s, y) := \tilde{\mathcal{G}}\varphi^i(s, y) + g^i(s, y, v(s, y), (\partial\varphi\sigma)^i(s, y), \mathcal{I}\varphi^i(s, y)) < 0 \quad (15.26)$$

for every (s, y) such that $(s, y, i) \in \text{Int } \mathcal{D}$, $s \in [t, t + \alpha]$ and $|y - x| \leq \alpha$, for some sufficiently small $\alpha > 0$. Let

$$\tau = \inf\{s \geq t; |X_s^t - x| \geq \alpha, N_s^t \neq i, \bar{Y}_s^t = \ell^i(s, X_s^t)\} \wedge (t + \alpha) \wedge \vartheta^t \quad (15.27)$$

$$(\hat{Y}^t, \hat{Z}^t, \hat{V}^t, \hat{A}^t) = (\mathbb{1}_{\cdot < \tau} \bar{Y}^t + \mathbb{1}_{\cdot \geq \tau} v^i(\tau, X_\tau^t), \mathbb{1}_{\cdot < \tau} \bar{Z}^t, \mathbb{1}_{\cdot < \tau} \bar{V}^t, \bar{A}_{\cdot \wedge \tau}^t) \quad (15.28)$$

$$\begin{aligned} (\tilde{Y}^t, \tilde{Z}^t, \tilde{V}^t) &= (\varphi^i(\cdot X_{\cdot \wedge \tau}^t), \mathbb{1}_{\cdot \leq \tau} (\partial\varphi\sigma)^i(\cdot, X_\cdot^t), \\ &\quad \mathbb{1}_{\cdot \leq \tau} ([\varphi^i(\cdot, X_{\cdot -}^t) + \delta^i(\cdot, X_{\cdot -}^t, y)] - \varphi^i(\cdot, X_{\cdot -}^t)])_{y \in \mathbb{R}^d}. \end{aligned} \quad (15.29)$$

Using, in particular, the fact that the set D is open from (12.76), we have that $\tau > t$, \mathbb{P}^t -almost surely. Also, using the continuity of v^i , it follows that

$$\hat{Y}_t^t = \bar{Y}_t^t = v^i(t, x) = \varphi^i(t, x) = \tilde{Y}_t^t. \quad (15.30)$$

By the minimality condition in the stopped RBSDE for \bar{Y}^t , we have that $\bar{A}^t = 0$ on $[t, \tau]$, since $\ell^i(s, X_s^t) < \bar{Y}_s^t$ on $[t, \tau]$ and $\tau \leq \vartheta^t$. Using the stopped RBSDE equation for \bar{Y}^t , we can then show, as for (15.22) in part (i), that for every $s \in [t, \tau]$ we have:

$$\begin{aligned} \hat{Y}_s^t &= v^i(\tau, X_\tau^t) + \int_s^\tau g^i(\zeta, X_\zeta^t, v(\zeta, X_\zeta^t), \hat{Z}_\zeta^t, \bar{r}_\zeta^t) d\zeta \\ &\quad - \int_s^\tau \hat{Z}_\zeta^t dB_\zeta^t - \int_s^\tau \int_{\mathbb{R}^d} \hat{V}_\zeta^t(y) \tilde{\chi}^t(d\zeta, dy), \end{aligned} \quad (15.31)$$

in which, with $\varsigma = \eta f$,

$$\begin{aligned} &\int_s^\tau g^i(\zeta, X_\zeta^t, v(\zeta, X_\zeta^t), \hat{Z}_\zeta^t, \bar{r}_\zeta^t) d\zeta \\ &= \int_s^\tau g^i\left(\zeta, X_\zeta^t, v(\zeta, X_\zeta^t), \hat{Z}_\zeta^t, \int_{\mathbb{R}^d} \hat{V}_\zeta^t(y) \varsigma^i(\zeta, X_\zeta^t, y) \hat{m}(dy)\right) d\zeta. \end{aligned}$$

Expressed otherwise, $(\hat{Y}^t, \hat{Z}^t, \hat{V}^t)$ is an $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, χ^t) -solution to the stopped BSDE on $[t, t + \alpha]$ with terminal condition $v^i(\tau, X_\tau^t)$ at τ and with driver

$$g^i\left(s, X_s^t, v(s, X_s^t), z, \int_{\mathbb{R}^d} v(y) \varsigma^i(s, X_s^t, y) \hat{m}(dy)\right) ds,$$

in which z and $v = v(y)$ refer to generic elements of \mathbb{R}^d and $\mathcal{M}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \hat{m}(dy); \mathbb{R})$. Moreover, we can show, as in part (i), that $(\tilde{Y}^t, \tilde{Z}^t, \tilde{V}^t)$ is an $(\mathbb{F}^t, \mathbb{P}^t)$, (B^t, χ^t) -solution to the stopped BSDE on $[t, t + \alpha]$, with driver

$$g^i \left(s, X_s^t, v(s, X_s^t), z, \int_{\mathbb{R}^d} v(y) \varsigma^i(s, X_s^t, y) \hat{m}(dy) \right) ds - \psi(s, X_s^t) ds$$

and terminal condition $\varphi^i(\tau, X_\tau^t)$ at τ . We conclude as in part (i).

15.2.3 Proof of Lemma 13.2.4

(i) Let $(t^*, x^*, i) \in (0, T) \times \mathbb{R}^d \times I$ be such that $\omega^i(t^*, x^*) > 0$ and (t^*, x^*) maximizes $\omega^i - \varphi^i$ for some function $\varphi \in \mathcal{C}^{1,2}(\mathcal{E})$. We need to show that (13.7) holds at (t^*, x^*, i) . We first assume that $t^* > 0$. By a classical viscosity solutions argument, we may limit our attention to the case where (t^*, x^*) strictly maximizes $\omega^i - \varphi^i$. We thus introduce the function

$$\varphi_{\epsilon,\alpha}^i(t, x, s, y) = \mu^i(t, x) - v^i(s, y) - \frac{|x - y|^2}{\epsilon^2} - \frac{|t - s|^2}{\alpha^2} - \varphi^i(t, x) \quad (15.32)$$

on $[0, T] \times \mathbb{R}^d$, where ϵ, α are positive parameters made to tend to zero later in the proof. By the viscosity solutions Jensen-Ishii Lemma (see Crandall et al. [75] or Fleming and Soner [122]), there exist, for every positive ϵ and α , related points (t, x) and (s, y) in $[0, T] \times \overline{B}_R$ (we omit the dependence of t, x, s, y in ϵ, α , for notational simplicity), where \overline{B}_R is a closed ball around x^* , with a large radius R to be specified later, such that:

- for every positive ϵ, α , the related quadruple (t, x, s, y) maximizes $\varphi_{\epsilon,\alpha}^i$ over $([0, T] \times \overline{B}_R)^2$ and, in particular,

$$\begin{aligned} & \mu^i(t^*, x^*) - v^i(t^*, x^*) - \varphi^i(t^*, x^*) \\ &= \varphi_{\epsilon,\alpha}^i(t^*, x^*, t^*, x^*) \\ &\leq \varphi_{\epsilon,\alpha}^i(t, x, s, y) = \mu^i(t, x) - v^i(s, y) - \frac{|x - y|^2}{\epsilon^2} - \frac{|t - s|^2}{\alpha^2} - \varphi^i(t, x); \end{aligned} \quad (15.33)$$

- $(t, x), (s, y) \rightarrow (t^*, x^*)$ as $\epsilon, \alpha \rightarrow 0$;
- $\frac{|x-y|^2}{\epsilon^2}$ and $\frac{|t-s|^2}{\alpha^2}$ are bounded and tend to zero as $\epsilon, \alpha \rightarrow 0$.

It then follows from Theorem 8.3 of [75] that there exist symmetric matrices $X, Y \in \mathbb{R}^{d \times d}$ such that

$$\begin{aligned} & (p + \partial_t \varphi(t, x), q + \partial \varphi^i(t, x), X) \in \overline{\mathcal{P}}^{2,+} \mu^i(t, x) \\ & (p, q, Y) \in \overline{\mathcal{P}}^{2,-} v^i(s, y) \\ & \begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq \frac{4}{\epsilon^2} \begin{pmatrix} I_d & -I_d \\ -I_d & I_d \end{pmatrix} + \begin{pmatrix} \partial^2 \varphi(t, x) & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned} \quad (15.34)$$

where $\overline{\mathcal{P}}^{2,+}\mu^i(t, x)$, respectively $\overline{\mathcal{P}}^{2,-}v^i(s, y)$, denotes the closure of the parabolic superjet of μ^i at (t, x) , respectively subjet of v^i at (s, y) (see [75, 122]), and where

$$p = \frac{2(t-s)}{\alpha^2}, \quad q = \frac{2(x-y)^\top}{\epsilon^2}. \quad (15.35)$$

Remark 15.2.1 The inequality (15.34) is meant in the sense of the order between symmetric real matrices, so that $A \leq B$ if and only if the real inequality $X^\top AX \leq X^\top BX$ holds for every real vector X .

Modifying, if necessary, $\varphi_{\epsilon,\alpha}^i = \varphi_{\epsilon,\alpha}^i(t', x', s', y')$ by adding terms of the form $\xi(x')$ and $\xi(y')$ with supports in the complement $\overline{B}_{R/2}^c$ of $\overline{B}_{R/2}$, we may assume that (t, x, s, y) is a maximum of $\varphi_{\epsilon,\alpha}^i$ over $([0, T] \times \mathbb{R}^d)^2$. Since $\omega^i(t^\star, x^\star) > 0$, we have that $\mu^i(t, x) - v^i(s, y)$ is positively bounded from below (at least, for ϵ and α sufficiently small), by (15.33). Combining this inequality with the fact that $\ell \leq v$ and $\mu \leq h$, we deduce by continuity of the obstacles ℓ and h that, for sufficiently small (ϵ, α) :

$$\ell^i(t, x) < \mu^i(t, x), \quad v^i(s, y) < \ell^i(s, y),$$

so that the related sub- and super-solution inequalities are satisfied by μ at (t, x, i) and v at (s, y, i) . Thus

$$\begin{aligned} & -p - \partial_t \varphi^i(t, x) - \frac{1}{2} \text{Tr}(a^i(t, x)X) - qb^i(t, x) - \partial \varphi^i(t, x)b^i(t, x) \\ & - \int_{\mathbb{R}^d} (\mu^i(t, x + \delta^i(t, x, z)) - \mu^i(t, x)) f^i(t, x, z) \widehat{m}(dz) \\ & - g^i(t, x, \mu(t, x), (q + \partial \varphi^i(t, x)) \sigma^i(t, x), \mathcal{I}\mu^i(t, x)) \leq 0 \\ & -p - \frac{1}{2} \text{Tr}(a^i(s, y)Y) - qb^i(s, y) \\ & - \int_{\mathbb{R}^d} (v^i(s, y + \delta^i(s, y, z)) - v^i(s, y)) f^i(s, y, z) \widehat{m}(dz) \\ & - g^i(s, y, v(s, y), q\sigma^i(s, y), \mathcal{I}v^i(s, y)) \geq 0. \end{aligned}$$

Note that the ξ terms that have been added to $\varphi_{\epsilon,\alpha}$ to ensure global maxima do not appear in these inequalities, since δ has linear growth in x and is thus locally bounded, whereas these terms have a support which is included in $\overline{B}_{R/2}^c$ with R large.

By subtracting the previous inequalities, we obtain that

$$\begin{aligned} & -\partial_t \varphi^i(t, x) - \frac{1}{2} (\text{Tr}(a^i(t, x)X) - \text{Tr}(a^i(s, y)Y)) \\ & - q(b^i(t, x) - b^i(s, y)) - \partial \varphi^i(t, x)b^i(t, x) \end{aligned}$$

$$\begin{aligned}
& - \int_{\mathbb{R}^d} [(\mu^i(t, x + \delta^i(t, x, z)) - \mu^i(t, x)) \\
& - (v^i(s, y + \delta^i(s, y, z)) - v^i(s, y))] f^i(t, x, z) \hat{m}(dz) \\
& + \int_{\mathbb{R}^d} [v^i(s, y + \delta^i(s, y, z)) - v^i(s, y)] [f^i(t, x, z) - f^i(s, y, z)] \hat{m}(dz) \\
& - (g^i(t, x, \mu(t, x), (q + \partial\varphi^i(t, x))\sigma^i(t, x), \mathcal{I}\mu^i(t, x)) \\
& - g^i(s, y, v(s, y), q\sigma^i(s, y), \mathcal{I}v^i(s, y))) \leq 0.
\end{aligned}$$

Denoting by $(e^l)_{l=1}^d$ an orthonormal basis of \mathbb{R}^d , the matrix inequality (15.34), in the sense explained in Remark 15.2.1, yields that

$$\begin{aligned}
& \text{Tr}(a^i(t, x)X) - \text{Tr}(a^i(s, y)Y) \\
& = \sum_{l=1}^d ((\sigma^i(t, x)e^l)^\top X \sigma^i(t, x)e^l - (\sigma^i(s, y)e^l)^\top Y \sigma^i(s, y)e^l) \\
& = \sum_{l=1}^d ((\sigma^i(t, x)e^l)^\top, (\sigma^i(s, y)e^l)^\top) \begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \begin{pmatrix} \sigma^i(t, x)e^l \\ \sigma^i(s, y)e^l \end{pmatrix} \\
& \leq \frac{4}{\epsilon^2} \sum_{l=1}^d |(\sigma^i(t, x) - \sigma^i(s, y))e^l|^2 + \text{Tr}(a^i(t, x)\partial^2\varphi^i(t, x)) \\
& \leq C \frac{|t-s|^2 + |x-y|^2}{\epsilon^2} + \text{Tr}(a^i(t, x)\partial^2\varphi^i(t, x)),
\end{aligned}$$

where the joint local Lipschitz continuity property of σ was used in the last line. Other computations analogous to those on pp. 76, 77 of [20] (see also [219]), using the maximization property of (t, x, s, y) , the definition of q in (15.35) and the Lipschitz continuity properties of the data yield also that

$$\begin{aligned}
|q|(|t-s| + |x-y|) & \leq C \frac{|t-s|^2 + |x-y|^2}{\epsilon^2} \\
|q(b^i(t, x) - b^i(s, y))| & \leq C \frac{|t-s|^2 + |x-y|^2}{\epsilon^2} \\
& (\mu^i(t, x + \delta^i(t, x, z)) - \mu^i(t, x)) - (v^i(s, y + \delta^i(s, y, z)) - v^i(s, y)) \\
& \leq (\varphi^i(t, x + \delta^i(t, x, z)) - \varphi^i(t, x)) \\
& + \left(\frac{|x + \delta^i(t, x, z) - y - \delta^i(s, y, z)|^2}{\epsilon^2} - \frac{|x-y|^2}{\epsilon^2} \right),
\end{aligned}$$

where in the last inequality

$$\begin{aligned}
& \frac{|x + \delta^i(t, x, z) - y - \delta^i(s, y, z)|^2}{\epsilon^2} - \frac{|x - y|^2}{\epsilon^2} \\
&= \frac{1}{\epsilon^2} [2(x - y)^\top (\delta^i(t, x, z) - \delta^i(s, y, z)) + |\delta^i(t, x, z) - \delta^i(s, y, z)|^2] \\
&= q(\delta^i(t, x, z) - \delta^i(s, y, z)) + \frac{1}{\epsilon^2} |\delta^i(t, x, z) - \delta^i(s, y, z)|^2 \\
&\leq C \frac{|t - s|^2 + |x - y|^2}{\epsilon^2}.
\end{aligned}$$

Therefore, with $\varsigma = \eta f$,

$$\begin{aligned}
& -\partial_t \varphi^i(t, x) - \frac{1}{2} \text{Tr}(a^i(t, x) \partial^2 \varphi^i(t, x)) - \partial \varphi^i(t, x) b^i(t, x) \\
& - \int_{\mathbb{R}^d} (\varphi^i(t, x + \delta^i(t, x, z)) - \varphi^i(t, x)) \varsigma^i(t, x, z) \widehat{m}(dz) \\
& - (g^i(t, x, \mu(t, x), (q + \partial \varphi^i(t, x)) \sigma^i(t, x), \mathcal{I}\mu^i(t, x)) \\
& - g^i(s, y, \nu(s, y), q \sigma^i(s, y), \mathcal{I}\nu^i(s, y))) \\
& \leq C \left(|t - s| + |x - y| + \frac{|t - s|^2 + |x - y|^2}{\epsilon^2} \right), \\
& \mathcal{I}\mu^i(t, x) - \mathcal{I}\nu^i(s, y) \\
&= \int_{\mathbb{R}^d} [v^i(s, y + \delta^i(s, y, z)) - v^i(s, y)] [\varsigma^i(t, x, z) - \varsigma^i(s, y, z)] \widehat{m}(dz) \\
&+ \int_{\mathbb{R}^d} [(\mu^i(t, x + \delta^i(t, x, z)) - \mu^i(t, x)) \\
&- (v^i(s, y + \delta^i(s, y, z)) - v^i(s, y))] \varsigma^i(t, x, z) \widehat{m}(dz) \\
&\leq \int_{\mathbb{R}^d} [(\varphi^i(t, x + \delta^i(t, x, z)) - \varphi^i(t, x))] \varsigma^i(t, x, z) \widehat{m}(dz) \\
&+ C \left(|t - s| + |x - y| + \frac{|t - s|^2 + |x - y|^2}{\epsilon^2} \right), \\
&= \mathcal{I}\varphi^i(t, x) + C \left(|t - s| + |x - y| + \frac{|t - s|^2 + |x - y|^2}{\epsilon^2} \right) \\
& g^i(t, x, \mu(t, x), (q + \partial \varphi^i(t, x)) \sigma^i(t, x), \mathcal{I}\mu^i(t, x)) \\
&- g^i(s, y, \nu(s, y), q \sigma^i(s, y), \mathcal{I}\nu^i(s, y)) \\
&\leq \eta_\epsilon(|t - s|) + \eta_R(|x - y| (1 + |q \sigma^i(s, y)|)) + \widehat{\Lambda} \max_{j \in I} (\mu^j(t, x) - \nu^j(s, y))^+
\end{aligned}$$

$$\begin{aligned}
& + \Lambda |q(\sigma^i(t, x) - \sigma^i(s, y)) + (\partial\varphi\sigma)^i(t, x)| \\
& + \widehat{\Lambda} \max_{l \in J} (\mathcal{I}^l \mu^i(t, x) - \mathcal{I}^l v^i(s, y))^+,
\end{aligned}$$

where the next-to-last inequality holds componentwise, and where in the last inequality:

- η_ϵ is a modulus of continuity of g^i on a compact set parameterized by ϵ , obtained by using the fact that q in (15.35) is bounded independently of α , for fixed ϵ ;
- η_R is the modulus of continuity in the assumption (U.2);
- the three last terms come from the Lipschitz and monotonicity properties of g .

Therefore

$$\begin{aligned}
-\tilde{\mathcal{G}}\varphi^i(t, x) &= -\partial_t\varphi^i(t, x) - \frac{1}{2}\text{Tr}(a^i(t, x)\partial^2\varphi^i(t, x)) - \partial\varphi^i(t, x)b^i(t, x) \\
&\quad - \int_{\mathbb{R}^d} (\varphi^i(t, x + \delta^i(t, x, z)) - \varphi^i(t, x))\varsigma^i(t, x, z)\widehat{m}(dz) \\
&\leq \widehat{\Lambda} \left(\max_{j \in I} (\mu^j(t, x) - v^j(s, y))^+ + |(\partial\varphi\sigma)^i(t, x)| + \max_{l \in J} \mathcal{I}^l \varphi^i(t, x)^+ \right) \\
&\quad + \eta_\epsilon(|t - s|) + \eta_R(|x - y|(1 + |q\sigma^i(s, y)|)) \\
&\quad + C \left(|t - s| + |x - y| + \frac{|t - s|^2 + |x - y|^2}{\epsilon^2} \right).
\end{aligned}$$

Given $\varrho > 0$, and for $\epsilon \leq \epsilon_\varrho$ and $\alpha \leq \epsilon$, using the properties of (t, x, s, y) in the Jensen-Ishii Lemma and the regularity of φ^i , we obtain:

$$\begin{aligned}
& -\tilde{\mathcal{G}}\varphi^i(t^*, x^*) \\
& \leq \widehat{\Lambda} \left(\max_{j \in I} (\mu^j(t, x) - v^j(s, y))^+ + |(\partial\varphi\sigma)^i(t^*, x^*)| + \max_{l \in J} \mathcal{I}^l \varphi^i(t^*, x^*)^+ \right) \\
& \leq \varrho + \eta_\epsilon(|t - s|).
\end{aligned}$$

Note that $t - s \rightarrow 0$ for fixed ϵ as $\alpha \rightarrow 0$, by the boundness of $\frac{|t-s|^2}{\alpha^2}$ in the Jensen-Ishii Lemma. Whence, for $\alpha \leq \alpha_\epsilon (\leq \epsilon)$:

$$\begin{aligned}
& -\tilde{\mathcal{G}}\varphi^i(t^*, x^*) \\
& \leq \widehat{\Lambda} \left(\max_{j \in I} (\mu^j(t, x) - v^j(s, y))^+ + |(\partial\varphi\sigma)^i(t^*, x^*)| + \max_{l \in J} (\mathcal{I}^l \varphi^i(t^*, x^*))^+ \right) \\
& \leq 2\varrho.
\end{aligned}$$

Sending $\varrho, \epsilon, \alpha$ to zero with $\epsilon \leq \epsilon_\varrho$ and $\alpha \leq \alpha_\epsilon$, the inequality (13.7) at (t^*, x^*, i) follows by upper semi-continuity of the function $(t', x', s', y') \mapsto \max_{j \in I} (\mu^j(t', x') - v^j(s', y'))^+$. This completes the proof that (13.7) holds at (t^*, x^*, i) in the case $t^* > 0$.

We deal finally with the case $t^* = 0$, for which we introduce the function

$$\varphi_\epsilon^i(t, x) = \omega^i(t, x) - \left(\varphi^i(t, x) + \frac{\epsilon}{t} \right) \quad (15.36)$$

on $[0, T] \times \overline{B}_R$, where ϵ is a positive parameter that tends to zero. Again assuming, without loss of generality, that $(t^* = 0, x^*)$ strictly maximizes $\omega^i - \varphi^i$, for every $\epsilon > 0$ there exists a point (t, x) in $[0, T] \times \overline{B}_R$ (for notational simplicity we omit the dependence of (t, x) in ϵ), where \overline{B}_R is a closed ball with large radius R around x^* , such that:

- for every $\epsilon > 0$ the related point (t, x) maximizes φ_ϵ^i over $[0, T] \times \overline{B}_R$;
- we have $t > 0$ for sufficiently small ϵ ;
- we have $(t, x) \rightarrow (t^*, x^*)$ as $\epsilon \rightarrow 0$.

By virtue of the part of the result already established in the case $t^* > 0$, we may apply (13.7) to the function $(s, y) \mapsto \varphi^i(s, y) + \frac{\epsilon}{s}$ at (t, x, i) , whence:

$$\begin{aligned} -\tilde{\mathcal{G}}\varphi^i(t, x) - \widehat{\Lambda} \left(\max_{j \in I} (\omega^j(t, x))^+ + |(\partial\varphi\sigma)^i(t, x)| + \max_{l \in J} (\mathcal{I}^l \varphi^i(t, x))^+ \right) \\ \leq -\frac{\epsilon}{t^2} \leq 0. \end{aligned}$$

Sending ϵ to 0 on the left-hand side proves, by the upper semi-continuity of $\max_{j \in I} (\omega^j)^+$, that (13.7) holds at $(t^* = 0, x^*, i)$.

(ii) Straightforward computations show that

$$\begin{aligned} -\partial_t \chi(t, x) &= c \chi(t, x) \\ (1 + |x|) |\partial \chi(t, x)| \vee (1 + |x|^2) |\partial^2 \chi(t, x)| \vee \chi(t, x + \delta^i(t, x, z)) &\leq C |\chi(t, x)| \end{aligned}$$

on \mathcal{E} , for a constant C independent of c . Therefore, for sufficiently large positive c ,

$$-\tilde{\mathcal{G}}\chi - \widehat{\Lambda} \left(\chi + |\partial \chi \sigma| + \max_{l \in J} (\mathcal{I}^l \chi)^+ \right) > 0$$

on \mathcal{E} .

(iii) By assumption on p , $\frac{|\omega|}{\chi}$ goes to 0 uniformly in t as $|x| \rightarrow \infty$. Given $\alpha > 0$, we prove that

$$\sup_{(t, x, i) \in \mathcal{E}} (\omega^i(t, x)^+ - \alpha \chi(t, x)) e^{-\widehat{\Lambda}(T-t)} \leq 0. \quad (15.37)$$

Assume to the contrary we have $>$ instead of \leq in (15.37). Then by upper semi-continuity of ω^+ , the supremum is reached at some point $(t^*, x^*, i) \in \text{Int } \mathcal{E}$ on the left-hand side of (15.37), and

$$\omega^i(t^*, x^*)^+ \geq \omega^i(t^*, x^*)^+ - \alpha \chi(t^*, x^*) > 0. \quad (15.38)$$

Therefore on $[0, T] \times \mathbb{R}^d$ we have:

$$\begin{aligned} & (\omega^i(t, x) - \alpha \chi(t, x)) e^{-\widehat{\Lambda}(T-t)} \\ & \leq (\omega^i(t, x)^+ - \alpha \chi(t, x)) e^{-\widehat{\Lambda}(T-t)} \\ & \leq (\omega^i(t^*, x^*)^+ - \alpha \chi(t^*, x^*)) e^{-\widehat{\Lambda}(T-t^*)} = (\omega^i(t^*, x^*) - \alpha \chi(t^*, x^*)) e^{-\widehat{\Lambda}(T-t^*)}, \end{aligned}$$

and thus

$$\omega^i(t, x) - \alpha \chi(t, x) \leq (\omega^i(t^*, x^*) - \alpha \chi(t^*, x^*)) e^{-\widehat{\Lambda}(t-t^*)}.$$

In other words, (t^*, x^*) maximizes at zero $\omega^i - \varphi^i$ over $[0, T] \times \mathbb{R}^d$, with

$$\varphi^i(t, x) = \alpha \chi(t, x) + (\omega^i(t^*, x^*) - \alpha \chi(t^*, x^*)) e^{-\widehat{\Lambda}(t-t^*)}.$$

Whence, by part (i) (given that $\omega^i(t^*, x^*) > 0$, from (15.38)):

$$\begin{aligned} & -\tilde{\mathcal{G}}\varphi^i(t^*, x^*) \\ & - \widehat{\Lambda} \left(\max_{j \in I} \omega^j(t^*, x^*)^+ + |\partial \varphi^i(t^*, x^*) \sigma^i(t^*, x^*)| + \max_{l \in J} (\mathcal{I}^l \varphi^i(t^*, x^*))^+ \right) \leq 0. \end{aligned} \tag{15.39}$$

But the left-hand side in this inequality is nothing but

$$\begin{aligned} & -\alpha \tilde{\mathcal{G}}\chi(t^*, x^*) + \widehat{\Lambda}(\omega^i(t^*, x^*) - \alpha \chi(t^*, x^*)) \\ & - \widehat{\Lambda} \left(\omega^i(t^*, x^*) + \alpha |\partial \chi(t^*, x^*) \sigma^i(t^*, x^*)| + \alpha \max_{l \in J} (\mathcal{I}^l \chi^i(t^*, x^*))^+ \right) \\ & = -\alpha \tilde{\mathcal{G}}\chi(t^*, x^*) \\ & - \widehat{\Lambda} \left(\alpha \chi(t^*, x^*) + \alpha |\partial \chi(t^*, x^*) \sigma^i(t^*, x^*)| + \alpha \max_{l \in J} (\mathcal{I}^l \chi^i(t^*, x^*))^+ \right), \end{aligned}$$

which should be positive by (13.8) in (ii), in contradiction with (15.39).

15.2.4 Proof of Lemma 13.2.8

We only prove part (a), since the proof of part (b) is similar (cf. the comments preceding Lemma 13.2.4). Note that we only have $\bar{v} \leq u \leq \underline{v}$ at T in (b), and not necessarily $\bar{v} \leq u \leq \underline{v}$ on $\partial\mathcal{D}$; see the comments in (ii) below.

(i) We prove that \bar{u} is a viscosity subsolution to $(\mathcal{V}2)$ on $\text{Int } \mathcal{E}$. That \underline{u} is a viscosity supersolution to $(\mathcal{V}2)$ on $\text{Int } \mathcal{E}$ can be shown likewise. First note that $\bar{u} \leq h$, by (13.12) on $\text{Int } \mathcal{E} \cap \mathcal{E}_h$, the inequality $\Phi \leq h$ at T (cf. the assumption (M.2.ii) in Sect. 12.2.3) and the continuity of h and Φ . So let $(t^*, x^*, i) \in \text{Int } \mathcal{E}$ be such that $\bar{u}^i(t^*, x^*) > \ell^i(t^*, x^*)$ and such that $\bar{u}^i - \varphi^i$ is strictly maximum and null at (t^*, x^*) ,

for some function $\varphi \in \mathcal{C}^{1,2}(\mathcal{E})$. We need to show that (cf. (13.5)):

$$-\tilde{\mathcal{G}}\varphi^i(t^*, x^*) - g^i(t^*, x^*, \bar{u}(t^*, x^*), (\partial\varphi\sigma)^i(t^*, x^*), \mathcal{I}\varphi^i(t^*, x^*)) \leq 0. \quad (15.40)$$

By a classical argument in the theory of viscosity solutions (see e.g. Barles and Souganidis [23]), there exists, for every $h > 0$, a point (t, x) in $[0, T] \times \overline{B}_R$, where \overline{B}_R is a closed ball with large radius R around x^* , such that (we omit the dependence of t, x in h for notational simplicity):

$$u_h^i \leq \varphi^i + (u_h - \varphi)^i(t, x), \quad (15.41)$$

with equality at (t, x) and, as $h \rightarrow 0$, we have $\xi_h := (u_h - \varphi)^i(t, x)$ goes to 0 = $(\bar{u} - \varphi)^i(t^*, x^*)$. Hence $u_h^i(t, x)$ goes to $\bar{u}^i(t^*, x^*)$ (cf. an analogous statement and its justification in the second paragraph of (ii) below). Therefore $\bar{u}^i(t^*, x^*) > \ell^i(t^*, x^*)$ implies that $u_h^i(t, x) > \ell^i(t, x)$ for sufficiently small h , whence by (13.12):

$$-\tilde{\mathcal{G}}_h u_h^i(t, x) - g^i(t, x, u_h(t, x), (\partial_h u_h \sigma)^i(t, x), \mathcal{I}_h u_h^i(t, x)) \leq 0. \quad (15.42)$$

Given (15.41), by the monotonicity of the scheme and the assumption (A) on g , we have:

$$\begin{aligned} -\tilde{\mathcal{G}}_h(\varphi + \xi_h)^i(t, x) &\leq g^i(t, x, u_h(t, x), (\partial_h(\varphi + \xi_h)\sigma)^i(t, x), \mathcal{I}_h(\varphi + \xi_h)^i(t, x)) \\ &\leq g^i(t^*, x^*, \bar{u}(t^*, x^*), (\partial\varphi\sigma)^i(t^*, x^*), \mathcal{I}\varphi^i(t^*, x^*)) \\ &\quad + \eta(|t - t^*|) + \eta_R(|x - x^*|(1 + |(\partial\varphi\sigma)^i(t^*, x^*)|)) \\ &\quad + \widehat{\Lambda} \max_{j \in I} (u_h^j(t, x) - \bar{u}^j(t^*, x^*))^+ \\ &\quad + \Lambda |(\partial_h(\varphi + \xi_h)\sigma)^i(t, x) - (\partial\varphi\sigma)^i(t^*, x^*)| \\ &\quad + \widehat{\Lambda} \max_{l \in J} (\mathcal{I}_h^l(\varphi + \xi_h)^i(t, x) - \mathcal{I}^l\varphi^i(t^*, x^*))^+, \end{aligned}$$

where in the last inequality (cf. the proof of Lemma 13.2.4(i)):

- η is a modulus of continuity of g^i on a “large” compact set around

$$(t^*, x^*, \bar{u}(t^*, x^*), (\partial\varphi\sigma)^i(t^*, x^*), \mathcal{I}\varphi^i(t^*, x^*)),$$

$\widehat{\Lambda}$ stands for $(k + q)\Lambda$, and η_R is the modulus of continuity present in the assumption (U.2);

- the three last terms stem from the Lipschitz continuity and monotonicity properties of g .

The inequality (15.40) follows by letting h go to zero in the last inequality, using the consistency (13.11) of the scheme.

- (ii) Let us show further that \bar{u} and \underline{u} satisfy the boundary condition in the so-called weak viscosity sense at T , namely in the case of \bar{u} (the related statement

and proof are similar in the case of \underline{u} : the inequality (15.40) holds for every $(t^* = T, x^*, i)$ and $\varphi \in \mathcal{C}^{1,2}(\mathcal{E})$ such that

$$\bar{u}^i(t^*, x^*) > \Phi^i(t^*, x^*), \quad (15.43)$$

and $\bar{u}^i - \varphi^i$ is strictly maximum and null at (t^*, x^*) . As in part (i), there exists, for every $h > 0$, a point (t, x) in $[0, T] \times \overline{B}_R$, where \overline{B}_R is a closed ball with large radius R around x^* , such that inequality (15.41) holds with equality at (t, x) and we have that

$$\xi_h := (u_h - \varphi)^i(t, x), u_h^i(t, x) - \bar{u}^i(t^*, x^*)$$

both converge to zero as $h \rightarrow 0$. Therefore inequality (15.43) implies, for h sufficiently small, that $(t, x, i) \in \text{Int } \mathcal{E}$ and

$$u_h^i(t, x) > \ell^i(t, x),$$

whence, by (13.12),

$$-\tilde{\mathcal{G}}_h u_h^i(t, x) - g^i(t, x, u_h(t, x), (\partial_h u_h \sigma)^i(t, x), \mathcal{I}_h u_h^i(t, x)) \leq 0. \quad (15.44)$$

The inequality (15.40) follows as in part (i) above.

Now,⁶ by a classical argument in the theory of viscosity solutions (see the bottom of p. 303 of Alvarez and Tourin [3] or Amadori [4, 5]), any viscosity subsolution or supersolution to $(\mathcal{V}2)$ on $\text{Int } \mathcal{E}$ satisfying the boundary condition in the weak viscosity sense at T satisfies it pointwise at T . Thus, in our case, suppose to the contrary that

$$\bar{u}^i(T, x^*) > \varphi^i(T, x^*) \quad (15.45)$$

for some $x^* \in \mathbb{R}^d$ and introduce the functions

$$\varphi_\epsilon(x) = \frac{|x - x^*|^2}{\epsilon}, \quad \partial \varphi_\epsilon(x) = \frac{2(x - x^*)\sigma}{\epsilon} \quad (15.46)$$

$$\phi_\epsilon(t, x) = \varphi_\epsilon(x) + C_\epsilon(T - t),$$

in which

$$C_\epsilon > \sup_{(t, x) \in [T - \eta, T] \times \overline{B}_1} (\tilde{\mathcal{G}} \varphi_\epsilon^i(t, x) + g^i(t, x, \bar{u}(t, x), (\partial \varphi_\epsilon \sigma)^i(t, x), \mathcal{I} \varphi_\epsilon^i(t, x))) \quad (15.47)$$

goes to ∞ as $\epsilon \rightarrow 0$, where \overline{B}_1 denotes the closed unit ball centered at x^* in \mathbb{R}^d . There exists, for every $\epsilon > 0$, a point (t, x) in $[0, T] \times \overline{B}_R$, where \overline{B}_R is a closed ball with large radius R around x^* , such that:

⁶Note that the following argument only works at T and cannot be adapted to the case of problem $(\mathcal{V}1)$ on the whole of $\partial \mathcal{D}$; see the comment at the beginning of the proof.

- for every $\epsilon > 0$ the related point (t, x) maximizes $(\bar{u}^i - \phi_\epsilon)$ over $[0, T] \times \overline{B}_R$,
- $(t, x) \rightarrow (T, x^*)$ and $\bar{u}^i(t, x) \rightarrow \bar{u}^i(T, x^*)$ as $\epsilon \rightarrow 0$.

To justify the last point, note that by the maximizing property of (t, x) we have

$$\bar{u}^i(T, x^*) - \phi_\epsilon(T, x^*) \leq \bar{u}^i(t, x) - \phi_\epsilon(t, x),$$

whence (cf. (15.46))

$$0 \leq \frac{|x^* - x|^2}{\epsilon} + C_\epsilon(T - t) \leq \bar{u}^i(t, x) - \bar{u}^i(T, x^*), \quad (15.48)$$

so that

$$\bar{u}^i(T, x^*) \leq \bar{u}^i(t, x). \quad (15.49)$$

Since \bar{u} is locally bounded, (15.48) implies that $(t, x) \rightarrow (T, x^*)$ as $\epsilon \rightarrow 0$; this, along with the upper semi-continuity of \bar{u} and with (15.49), completes the proof that $\bar{u}^i(t, x)$ converges to $\bar{u}^i(T, x^*)$ as $\epsilon \rightarrow 0$.

Now, we have that $\ell \leq \varphi$ pointwise at T , so that this last convergence and (15.45) together imply that $\bar{u}^i(t, x) > \ell^i(t, x)$ for ϵ sufficiently small. By virtue of the results already established at this point of the proof, the function ϕ_ϵ must thus satisfy the related viscosity subsolution inequality at (t, x, i) , so that

$$C_\epsilon - \tilde{\mathcal{G}}\varphi_\epsilon^i(t, x)g^i(t, x, \bar{u}(t, x), (\partial\varphi_\epsilon\sigma)^i(t, x), \mathcal{I}\varphi_\epsilon^i(t, x)) \leq 0,$$

which contradicts (15.47) and completes the proof.

Chapter 16

Exercises

16.1 Discrete-Time Markov Chains

Recall our introductory example from Sect. 1.2.1.

1. Supposing that

$$\begin{aligned}\mathbb{P}(R_{n+2} = L \mid R_{n+1} = H, R_n = H) &= \mathbb{P}(R_{n+2} = H \mid R_{n+1} = L, R_n = L) = 1/8 \\ \mathbb{P}(R_{n+2} = L \mid R_{n+1} = L, R_n = H) &= \mathbb{P}(R_{n+2} = H \mid R_{n+1} = H, R_n = L) = 1/2,\end{aligned}$$

derive the transition matrix for the enlarged process $X_n = (R_n, R_{n+1})$.

2. Assuming further that $\mathbb{P}(R_0 = L) = 1/3$ and

$$\mathbb{P}(R_1 = L \mid R_0 = L) = \mathbb{P}(R_1 = H \mid R_0 = H) = 3/4,$$

what is the probability that the interest rates will be low for three consecutive days starting at day 0? starting at day 2?

3. Given that the initial interest rate is low, i.e. $R_0 = L$, what is the conditional probability that $R_4 = H$?
4. What is the probability that the interest rate will be high in the long run?

16.2 Discrete-Time Martingales

1. Show that, if X_n is a martingale, then it has uncorrelated increments, i.e.

$$\mathbb{E}[(X_m - X_n)(X_k - X_r)] = 0,$$

where $0 \leq r < k \leq n < m < +\infty$.

2. Let the ε_i be i.i.d. with mean μ . Let $S_0 = \bar{S}_0 = 0$ and for $n > 0$ let

$$S_n = \varepsilon_1 + \cdots + \varepsilon_n, \quad \bar{S}_n = S_n - n\mu.$$

Are all processes ε_n , S_n and \bar{S}_n Markov chains? Are all these processes martingales with respect to the filtration \mathcal{F}_n if $\mu \neq 0$? What is the answer to the latter question when $\mu = 0$?

- Are the processes ε_n , S_n and \bar{S}_n Markov chains?
 - Are the processes ε_n , S_n and \bar{S}_n martingales w.r.t. \mathcal{F}_n when $\mu \neq 0$?
 - Are the processes ε_n , S_n and \bar{S}_n martingales w.r.t. \mathcal{F}_n when $\mu = 0$?
3. Let $S_n = \sum_{i=1}^n \varepsilon_i$, $n \geq 0$, where ε_i , $i \geq 1$, is a sequence of i.i.d. random variables, and $\varepsilon_i \sim \mathcal{E}_\lambda$. Identify three different martingales associated with the process S_n . Represent these martingales as a function of S_n and parameter λ .
- [Hint: The first martingale $\bar{S}_n = S_n - f(n, \lambda)$, where $f(n, \lambda) =$ some function of n and λ ; the second martingale $M_n = \bar{S}_n^2 - g(n, \lambda)$, where $g(n, \lambda) =$ some function of n and λ ; the third martingale $Z_n = \frac{\exp(\theta S_n)}{h(n, \theta, \lambda)}$, where $h(n, \theta, \lambda) =$ some function of n and λ .]
- Let ε_i , $i \geq 1$, be i.i.d. random variables with $\mathbb{P}(\varepsilon_i = -1) = \mathbb{P}(\varepsilon_i = 1) = \frac{1}{2}$. Set $S_n = \sum_{i=1}^n \varepsilon_i$, $n = 0, 1, 2, \dots$, and let \mathcal{F}_n be the information contained in S_0, \dots, S_n (which is the same as information contained in $\varepsilon_1, \dots, \varepsilon_n$). Finally, let $D_n = e^{-S_n}$.
 - Verify whether the process D_n is a martingale, supermartingale, submartingale or none of these with respect to the filtration \mathcal{F}_n .
 - Find a numerical sequence z_n so that the process Z_n given as

$$Z_n = \frac{D_n}{z_n}$$

is a (Wald's) martingale.

- Exercise 1.3.11.
- Exercise 1.3.12.
- Let $S_n = 1 + \sum_{i=1}^n \varepsilon_i$, $n \geq 0$, be a random walk with $p = 1/4$ and $q = 3/4$, starting from $x = 1$. Suppose that

$$\nu = \min\{n \geq 0 : S_n = -2\}.$$

Use the Optional Stopping Theorem (assuming its conditions hold) to compute

- $\mathbb{E}\nu$
 - $\text{Var } \nu$
- Exercise 1.3.18.
 - [Polya's urn scheme, see Example 4, p. 109, and Example 3, p. 119, of Lawler [180].] An urn contains k red balls and m green balls at the initial time $n = 0$. One ball is chosen randomly from the urn. The ball is then put back into the urn together with another ball of the same color (hence, the number of total balls in the urn grows). Let X_n denote the proportion of green balls in the urn at time $n \geq 0$.

- a. Is X_n a Markov chain?
- b. Is X_n a martingale?
- c. Is it a UI martingale?

16.3 The Poisson Process and Continuous-Time Markov Chains

1. Let X_t be a Markov chain with state space $\{1, 2\}$ and intensities $\lambda(1, 2) = 1$, $\lambda(2, 1) = 4$. Find $Q(t)$.
2. Let N_t be a Poisson process with parameter $\lambda = 2$. Determine the following expectations:
 - a. $\mathbb{E}N_2$
 - b. $\mathbb{E}N_1^2$
 - c. $\mathbb{E}(N_1 N_2)$
 - d. $\mathbb{E}\tau_2, \mathbb{E}\tau_1$
 - e. $\mathbb{E}T_2$, where T_2 denotes the second jump time of N_t
 - f. $\mathbb{E}(\tau_1 \tau_2)$
3. Let N_t be a Poisson process with parameter λ . Is the process $Y_t = N_t^2 - \lambda t$ a martingale?

16.4 Brownian Motion

1. Suppose W_t is a standard Brownian motion and $B_t = \frac{1}{\sqrt{a}} W_{at}$ with $a > 0$. Show that B_t is a standard Brownian motion, known as a time rescaled Brownian motion.
2. Suppose W_t is a standard Brownian motion and $Z_t = t W_{1/t}$. Show that Z_t is a standard Brownian motion, known as a time reversed Brownian motion.
3. Let W_t be a standard Brownian motion. Compute the following conditional probability: $\mathbb{P}(W_2 > 0 | W_1 > 0)$. Are the events $\{W_1 > 0\}$ and $\{W_2 > 0\}$ independent?
4. Let W_t be SBM.
 - a. Compute $\mathbb{E}(W_6 | W_2, W_4)$.
 - b. Compute $\mathbb{E}(W_6 W_2 W_4)$.
 - c. Express the joint density function of (W_2, W_4, W_6) in terms of the transition density function of $W_t, t \geq 0$.
 - d. Compute the probability density function of W_4 conditional on $W_2 = 0$ and $W_6 = 0$.
5. Two independent Brownian motions X_t^1 and X_t^2 , with respective drift parameters $\mu_1 \leq \mu_2$ and the same variance parameter σ^2 , start out at respective positions $x_1 < x_2$. Calculate the probability that they will never meet.

[Hint: Consider the process $X_t = X_t^2 - X_t^1$ and note that it is also a Brownian motion. Let

$$\tau_0 = \min\{t \geq 0 : X_t = 0\}.$$

We want to compute

$$\mathbb{P}(X_t^1 \text{ and } X_t^2 \text{ will never meet}) = \mathbb{P}(\tau_0 = \infty) = 1 - \mathbb{P}(\tau_0 < +\infty). \quad (16.1)$$

Introduce an auxiliary stopping time

$$\tau_{0,b} = \tau = \min\{t \geq 0 : X_t = 0 \text{ or } X_t = b\}$$

and use the fact that

$$\mathbb{P}(\tau_0 < +\infty) = \lim_{b \rightarrow +\infty} \mathbb{P}(X_\tau = 0). \quad (16.2)$$

Then consider separately two cases: $\mu_1 = \mu_2$ and $\mu_1 < \mu_2$.]

16.5 Stochastic Integration

1. Do

- a. Exercise 3.1.3. Verify that $\mathbb{E}(Y_t^n | \mathcal{F}_s) = Y_s^n$ is true for $t_i \leq s \leq t_{i+1}$ and $t_k \leq t \leq t_{k+1}$, where $t_{i+1} \leq t_k$.
 - b. Exercise 3.1.4. Verify property (ii), i.e. verify that $\mathbb{E}Y_t^n = 0$ for every t .
 - c. Exercise 3.1.5.
 - d. Exercise 3.1.6.
 - e. Exercise 3.1.8.
2. Compute the Itô stochastic integral for the process $Z_t = 1$, $t \in [0, T]$.
3. Compute the Poisson stochastic integral I_T in (3.4) for the process $Z_t = 1$, $t \in [0, T]$.
4. Define $Z_t = N_{t-}$, $\forall t \in (0, T]$. Explain why process Z is predictable. Compute the stochastic integrals I_t and Y_t for the process Z .

16.6 Itô Formula

1. Verify the formulas presented in Examples 3.2.1–3.2.4 and (3.2.5)–(3.2.7).
2. Apply the Itô formula to $u(W_t)$.
 - a. $u(x) = x$,
 - b. $u(x) = x^3$.
3. Verify that process Y given as $Y_t = \int_0^t W_s dW_s$ is a martingale with respect to the natural filtration of process W .

4. a. Derive the Itô formula for $u(t, W_t, N_t)$, where W is a standard Brownian motion and N is a Poisson process with parameter λ .
b. Apply this formula to $u(t, w, n) = 3twn + 2t + w - n$. Is $u(t, W_t, N_t)$ a martingale in this case?
5. Let W_t be SBM. Define

$$Y_t = \int_0^t e^{W_s} dW_s, \quad t \geq 0.$$

Determine whether the process Y is a martingale.

16.7 Stochastic Differential Equations

1. Do
 - a. Exercise 3.3.4.
 - b. Exercise 3.3.6.
 - c. Exercise 3.3.11.
 - d. Exercise 3.3.14.
2. Let $W_t, t \geq 0$ be SBM.
 - a. Compute the expectation $\mathbb{E}X_T$ and the variance $\text{Var } X_T$, where X_t is the strong solution to the SDE

$$dX_t = X_t dt + X_t dW_t, \quad t \geq 0, \quad X_0 = 1.$$

Is the process X_t a martingale?

- b. Compute the covariance $\text{Cov}(Y_t, Y_s)$ for $0 \leq s \leq t$, where Y_t is the strong solution to the SDE

$$dY_t = -Y_t dt + dW_t, \quad t \geq 0, \quad Y_0 = 0$$

In addition, compute the mean and the variance of the limiting distribution of the process Y .

- [**Hint:** Process $e^t Y_t$ has independent increments.]
- c. What is the distribution of Y_t , where Y is the strong solution to the SDE

$$dY_t = -Y_t dt + dW_t, \quad t \geq 0, \quad Y_0 = \eta,$$

in which $\eta \sim \mathcal{N}(0, \frac{1}{4})$ is independent of the standard Brownian motion W ?

3. Let $N_t, t \geq 0$, be $\text{PP}(\lambda)$. Compute the covariance $\text{Cov}(Y_t, Y_s)$ for $0 \leq s \leq t$, where $Y_t = \ln Z_t$ and where Z is the strong solution to the SDE

$$dZ_t = Z_{t-}(dt + dN_t), \quad t \geq 0, \quad Z_0 = 1.$$

4. Let W_t be SBM and N_t be PP(λ). Let $X_t = 2^{N_t} e^{3t} + W_t^2$. Compute the differential dX_t .
5. Solve Exercises 3.4.1, 3.4.4, 3.4.5, 3.4.7, 3.4.8.
6. Let N_t be PP(λ) on a filtered probability space $(\Omega, \mathbb{F}, \mathbb{P})$. Let X be the strong solution to the following SDE on this space:

$$dX_t = X_{t-}(dt + dN_t), \quad t \geq 0, \quad X_0 = 1.$$

Define a new probability measure \mathbb{Q} on (Ω, \mathcal{F}) so that process X is a martingale under \mathbb{Q} .

Chapter 17

Corrected Problem Sets

The solutions to the problem sets of this chapter are available on <http://extras.springer.com>.

17.1 Exit of a Brownian Motion from a Corridor

EXERCISE Let N_t denote a Poisson process with intensity $\lambda > 0$, and let $M_t = N_t - \lambda t$ be the compensated martingale of N .

1. Verify that the process Y given as $Y_t = \int_0^t N_{s-} dM_s$ is a martingale:
 - a. First, by invocation of a general theorem of stochastic analysis;
 - b. Second, by a direct computation.
2. Compute $J_t = \int_0^t N_s dM_s$. Is process J_t a martingale?

PROBLEM Let $X_t = \mu t + \sigma W_t$, where W_t is SBM with respect to a filtration \mathbb{F} (with \mathcal{F}_0 trivial).

I Basic properties Let a function $u(t, x)$ be continuously differentiable in t and twice continuously differentiable in x .

1. State the Itô formula which is relevant for $u(t, X_t)$.
2. Justify that X is a Markov process.
3. Recalling the characterization $\mathbb{E}_t du(t, X_t) = (\partial_t + \mathcal{A})u(t, X_t) dt$ of the generator \mathcal{A} of a Markov process X , deduce from the above Itô formula the expression of the generator \mathcal{A} of the Brownian motion X , and rewrite the above Itô formula in terms of \mathcal{A} .

II Application Given reals $a < 0 < b$, let

$$\tau = \min\{t \geq 0 : X_t = a \text{ or } b\} \in [0, +\infty].$$

We want to compute $\mathbb{P}(X_\tau = b) = \mathbb{E} \mathbb{1}_{\{X_\tau = b\}}$ (with the convention that $\mathbb{1}_{\{X_\tau = b\}} = 0$ if τ is infinite).

1. Justify that on the random set $\{t \leq \tau\}$ we have $\mathbb{P}(X_\tau = b | \mathcal{F}_t) = u(t, X_t)$ for some function $u = u(t, x)$.
2. Show that $(\partial_t + \mathcal{A})u(t, X_t) dt$ is constant.
3. Derive a partial differential equation satisfied by u on the domain $[0, +\infty) \times [a, b]$, along with suitable boundary conditions at a and b .
4. Admitting uniqueness for a bounded solution to the equation derived in the question 3, show that for $b \neq 0$ we have

$$u(t, x) = v(x) := \frac{\exp(\theta x) - \exp(\theta a)}{\exp(\theta b) - \exp(\theta a)},$$

with $\theta = -2\mu/\sigma^2$.

5. Deduce the final formula for $\mathbb{P}(X_\tau = b)$.
6. Proceed as in questions 1–5 above to compute $\mathbb{P}(X_\tau = a)$.
7. Prove from the above that $\tau < +\infty$ with probability one.
8. What about the case $\mu = 0$?

17.2 Pricing with a Regime-Switching Volatility

EXERCISE We consider, in the risk-neutral Black–Scholes model $dS_t = \sigma S_t dW_t$, a forward start option with payoff at time T given by

$$\xi = (S_T - KS_\Theta)^+,$$

for a fixed time $0 < \Theta < T$. We write

$$d_{\pm}(t, S_t) = \frac{\ln(\frac{S_t}{KS_\Theta})}{\sigma\sqrt{T-t}} \pm \frac{1}{2}\sigma\sqrt{T-t},$$

for $t \in [0, T]$.

1. Show that the option's price at time Θ is given by

$$\Pi_\Theta = S_\Theta (\mathcal{N}(d_+(\Theta)) - K\mathcal{N}(d_-(\Theta))), \quad (17.1)$$

where \mathcal{N} is the standard Gaussian cumulative distribution function and

$$d_{\pm}(\Theta) = \frac{-\ln(K)}{\sigma\sqrt{T-\Theta}} \pm \frac{1}{2}\sigma\sqrt{T-\Theta}.$$

2. Show that the option's price at time t is given by

$$\Pi_t = \begin{cases} S_t (\mathcal{N}(d_+(\Theta)) - K\mathcal{N}(d_-(\Theta))) & \text{if } t < \Theta \\ S_t \mathcal{N}(d_+(t, S_t)) - KS_\Theta \mathcal{N}(d_-(t, S_t)) & \text{if } t \geq \Theta. \end{cases} \quad (17.2)$$

3. Show that one can replicate this option, using the underlying stock S and the riskless, constant asset as hedging instruments, so there exists a replication strategy Δ_t (number of stocks held at time t in the replication portfolio) such that $d\Pi_t = \Delta_t dS_t$.
4. Compute Δ_t .

PROBLEM We consider the following stochastic volatility model, under a risk-neutral probability measure \mathbb{P} :

$$dS_t = \sigma_t S_t dW_t, \quad (17.3)$$

where W is a standard Brownian motion, and the volatility σ follows a two-state Markov chain with constant intensity of transition $\lambda > 0$. In other words, in the time interval $(t, t + dt)$, the instantaneous volatility passes from the “old” value $\sigma_{t-} = \underline{\sigma}$ to the “new” value $\sigma_t = \bar{\sigma}$, and vice versa, with probability λdt , independent of W . Or, in an SDE formulation,

$$d\sigma_t = (\sigma'_{t-} - \sigma_{t-}) dN_t, \quad (17.4)$$

where:

- $\sigma' = \sigma'(\sigma)$ is a notation for $\underline{\sigma}$ if $\sigma = \bar{\sigma}$ and for $\bar{\sigma}$ if $\sigma = \underline{\sigma}$,
- N is a Poisson process with intensity $\lambda > 0$ and compensated martingale $M_t = N_t - \lambda t$.

Interest-rates are set to zero. The model filtration is $\mathbb{F} = \mathbb{F}^W \vee \mathbb{F}^N$. One considers a derivative with payoff $\phi(S_T)$ at time T , for a measurable and bounded payoff function ϕ .

I Derivation of the Pricing Equations

1. Comment on (17.3), (17.4).
2. Justify that the price process Π of this derivative can be represented as $\Pi_t = \mathbb{E}(\phi(S_T) | \mathcal{F}_t)$, for $t \in [0, T]$.
3. Justify that the process Π can be further represented as $\Pi_t = u(t, S_t, \sigma_t)$ for a pricing function $u = u(t, S, \sigma)$ over $[0, T] \times (0, +\infty) \times \{\underline{\sigma}, \bar{\sigma}\}$.
4. Show, under the assumption that the function u is “sufficiently regular” in (t, S) , that the dynamics of Π are given by:

$$\begin{aligned} du(t, S_t, \sigma_t) = & (\partial_t u + \mathcal{A}u)(t, S_t, \sigma_t) dt + \sigma_t S_t \partial_S u(t, S_t, \sigma_t) dW_t \\ & + \delta u(t, S_t, \sigma_{t-}) dM_t, \end{aligned} \quad (17.5)$$

where we have set

$$\begin{aligned} \delta u(t, S, \sigma) &= u(t, S, \sigma') - u(t, S, \sigma) \\ \mathcal{A}u(t, S, \sigma) &= \frac{\sigma^2 S^2}{2} \partial_{S^2}^2 u(t, S, \sigma) + \lambda \delta u(t, S, \sigma). \end{aligned}$$

5. Dede that the pricing function u satisfies the following system of equations:

$$\begin{cases} \left(\partial_t + \frac{\underline{\sigma}^2 S^2}{2} \partial_{S^2}^2 \right) u(t, S, \underline{\sigma}) + \lambda(u(t, S, \bar{\sigma}) - u(t, S, \underline{\sigma})) = 0 \\ \left(\partial_t + \frac{\bar{\sigma}^2 S^2}{2} \partial_{S^2}^2 \right) u(t, S, \bar{\sigma}) + \lambda(u(t, S, \underline{\sigma}) - u(t, S, \bar{\sigma})) = 0 \end{cases} \quad (17.6)$$

for $(t, S) \in [0, T) \times (0, +\infty)$, as well as a terminal condition to be specified at time T .

II Pricing by PDEs Given uniform time- and space-grids $(t_i)_{1 \leq i \leq n}$ and $(S^j)_{1 \leq j \leq m}$ of the respective steps h and k , with the S -grid centered around S_0 , let

$$\begin{aligned} \underline{\alpha}^j &= \frac{\underline{\sigma}^2 (S^j)^2}{2k^2}, & \underline{\beta}^j &= -\frac{\underline{\sigma}^2 (S^j)^2}{k^2}, & \underline{\gamma}^j &= \frac{\underline{\sigma}^2 (S^j)^2}{2k^2} \\ \bar{\alpha}^j &= \frac{\bar{\sigma}^2 (S^j)^2}{2k^2}, & \bar{\beta}^j &= -\frac{\bar{\sigma}^2 (S^j)^2}{k^2}, & \bar{\gamma}^j &= \frac{\bar{\sigma}^2 (S^j)^2}{2k^2}. \end{aligned}$$

1. Write in terms of the above coefficients an explicit finite difference scheme (or trinomial tree) to compute $\Pi_0 = u(0, S_0, \sigma_0)$ and $\Delta_0 = \partial_S u(0, S_0, \sigma_0)$ (in the (t, S) variables, without log-transformation). Derive the related stability condition.
2. Write, in terms of the above coefficients, a hybrid scheme explicit in the jump terms/implicit in the differential terms of (17.6), not subject to the previous stability condition, and discuss the implementation of this scheme.
3. What do you think about a fully implicit finite difference scheme for this problem?

III Pricing by Monte Carlo Now j refers to the j th simulated trajectory of S .

1. Write an exact simulation algorithm (without time-discretization error) of S_T in (17.3).
2. Describe the Monte Carlo standard pricing algorithm based on m simulated trajectories of S to compute Π_0 and Δ_0 . Give the expressions of the estimates $\hat{\Pi}_0$ and $\hat{\Delta}_0$ of Π_0 and Δ_0 , and of the corresponding standard errors, $\hat{\sigma}_0^{\Pi}$ and $\hat{\sigma}_0^{\Delta}$.
For the delta, proceed by derivation of the payoff along the trajectory of S , assuming the function ϕ almost everywhere differentiable.
3. Comment on the pros and cons of this Monte Carlo method and of the deterministic schemes of Part II for the computation of Π_0 and Δ_0 .

In particular, give the order of magnitude of the errors in both cases (consistency error for deterministic schemes, simulation error for Monte Carlo).

4. Describe qualitatively a Monte Carlo pricing scheme with variance reduction based on the control variate $\phi(S_T^{bs})$, where

$$dS_t^{bs} = \sigma_0 S_t^{bs} dW_t, \quad S_0^{bs} = S_0.$$

What are the properties of the method?

Describe the simulation scheme to be used for S_T^{bs} in this application.

Give the expressions of the resulting estimates $\tilde{\Pi}_0$ and $\tilde{\Delta}_0$ of Π_0 and Δ_0 , and of the corresponding standard errors, $\tilde{\sigma}_0^{\Pi}$ and $\tilde{\sigma}_0^{\Delta}$.

17.3 Hedging with a Regime-Switching Volatility

Still in the regime-switching volatility model of Sect. 17.2, note that we have

$$\mathbb{V}\text{ar}\left(\int_0^T \alpha_t dS_t + \int_0^T \beta_t dM_t\right) = \mathbb{E}\left(\int_0^T \alpha_t^2 \sigma_t^2 S_t^2 dt + \lambda \int_0^T \beta_t^2 dt\right) \quad (17.7)$$

for every progressive α_t and predictable β_t making both sides well defined in (17.7).

I Derivation of the Pricing Equations

1. Denoting $X_t = \ln(S_t)$, derive the SDE satisfied by X .
2. Justify that process Π_t can be represented as $\Pi_t = u(t, X_t, \sigma_t)$ for a pricing function $u = u(t, x, \sigma)$ over $[0, T] \times \mathbb{R} \times \{\underline{\sigma}, \bar{\sigma}\}$.
3. Show, under the assumption that the function u is “sufficiently regular” in (t, x) , that the dynamics of Π are given by:

$$\begin{aligned} d\Pi_t &= du(t, X_t, \sigma_t) = (\partial_t u + \mathcal{A}u)(t, X_t, \sigma_t) dt + \partial_x u(t, X_t, \sigma_t) \sigma_t dW_t \\ &\quad + \delta u(t, X_t, \sigma_{t-}) dM_t, \end{aligned} \quad (17.8)$$

where we have set

$$\begin{aligned} \delta u(t, x, \sigma) &= u(t, x, \sigma') - u(t, x, \sigma) \\ \mathcal{A}u(t, x, \sigma) &= \frac{\sigma^2}{2} (\partial_{x^2}^2 u - \partial_x u)(t, x, \sigma) + \lambda \delta u(t, x, \sigma). \end{aligned}$$

4. Deduce that the pricing function u satisfies the following system of equations:

$$\begin{cases} \left(\partial_t u + \frac{\sigma^2}{2} (\partial_{x^2}^2 u - \partial_x u) \right)(t, x, \underline{\sigma}) + \lambda(u(t, x, \bar{\sigma}) - u(t, x, \underline{\sigma})) = 0 \\ \left(\partial_t u + \frac{\bar{\sigma}^2}{2} (\partial_{x^2}^2 u - \partial_x u) \right)(t, x, \bar{\sigma}) + \lambda(u(t, x, \underline{\sigma}) - u(t, x, \bar{\sigma})) = 0 \end{cases} \quad (17.9)$$

for $(t, x) \in [0, T] \times \mathbb{R}$, as well as a terminal condition $u = \psi$ to be specified at time T .

II Pricing Let there be given uniform time- and space-grids $(t_i)_{1 \leq i \leq n}$ and $(x^j)_{1 \leq j \leq m}$ of respective steps h and k , with $t_n = nh = T$ and with the x -grid centered at X_0 . We write

$$\begin{aligned}\underline{\alpha} &= \frac{\sigma^2}{2k^2} + \frac{\sigma^2}{4k}, & \underline{\beta} &= -\frac{\sigma^2}{k^2}, & \underline{\gamma} &= \frac{\sigma^2}{2k^2} - \frac{\sigma^2}{4k} \\ \bar{\alpha} &= \frac{\sigma^2}{2k^2} + \frac{\sigma^2}{4k}, & \bar{\beta} &= -\frac{\sigma^2}{k^2}, & \bar{\gamma} &= \frac{\sigma^2}{2k^2} - \frac{\sigma^2}{4k}.\end{aligned}\quad (17.10)$$

- Denoting $\underline{u}_i^j \approx u(t_i, x^j, \underline{\sigma})$, $\bar{u}_i^j \approx u(t_i, x^j, \bar{\sigma})$, write an explicit finite difference scheme for u in the form of a suitable condition for \underline{u}_n , and then, for $i = n - 1, \dots, 0$, $j = 1, \dots, m$:

$$\begin{cases} \underline{u}_i^j = \underline{p}_- \underline{u}_{i+1}^{j-1} + \underline{p} \underline{u}_{i+1}^j + \underline{p}_+ \underline{u}_{i+1}^{j+1} + \underline{q} \bar{u}_{i+1}^j \\ \bar{u}_i^j = \bar{p}_- \bar{u}_{i+1}^{j-1} + \bar{p} \bar{u}_{i+1}^j + \bar{p}_+ \bar{u}_{i+1}^{j+1} + \bar{q} \underline{u}_{i+1}^j, \end{cases}$$

where the weights \underline{p}_- , \underline{p} , \underline{p}_+ , \underline{q} , \bar{p}_- , \bar{p} , \bar{p}_+ , \bar{q} are expressed in terms of the coefficients in (17.10).

- Give the approximations in the tree for $\Pi_0 = u(0, X_0, \sigma_0)$ and $\Delta_0 = S_0^{-1} \partial_x u(0, X_0, \sigma_0)$. Justify this expression for Δ_0 .
- Derive the stability condition of the scheme.
- Write a fully discrete Markov chain approximation $(\hat{X}_{t_i}, \hat{\sigma}_{t_i})$ to (X_t, σ_t) in the above pair of trees, in which \hat{X}_t evolves at the nodes of the trees and passes from one tree to the other with certain probabilities; more precisely, given that at time t_{i-1} the Markov chain is in the state $(\hat{X}_{t_{i-1}}, \hat{\sigma}_{t_{i-1}}) = (x, \underline{\sigma})$, write in terms of \underline{p}_- , \underline{p} , \underline{p}_+ , \underline{q} the probabilities that at time t_i we have $(\hat{X}_{t_i}, \hat{\sigma}_{t_i}) = (x - k, \underline{\sigma})$, $(x, \underline{\sigma})$, $(x + k, \underline{\sigma})$ or $(x, \bar{\sigma})$; express likewise the transition probabilities from $(\hat{X}_{t_{i-1}}, \hat{\sigma}_{t_{i-1}}) = (x, \bar{\sigma})$ to $(\hat{X}_{t_i}, \hat{\sigma}_{t_i}) = (x - k, \bar{\sigma})$, $(x, \bar{\sigma})$, $(x + k, \bar{\sigma})$ in terms of \bar{p}_- , \bar{p} , \bar{p}_+ , \bar{q} .
- Using the tree with transition probabilities as developed in the previous question, we simulate $m = 10^4$ different paths of $(\hat{X}, \hat{\sigma})$. Describe a Monte Carlo algorithm for Π_0 and Δ_0 based on the m trajectories of $(\hat{X}, \hat{\sigma})$ thus simulated. Give the expressions of the estimates $\hat{\Pi}_0$ and $\hat{\Delta}_0$ of Π_0 and Δ_0 , and of the corresponding standard errors, $\hat{\sigma}_0^\Pi$ and $\hat{\sigma}_0^\Delta$.

For the delta, proceed by derivation of the payoff function ϕ , assumed almost everywhere differentiable.

- Is this algorithm the best Monte Carlo solution for this model? Propose a better alternative, without time-discretization error.

III Hedging

- Now consider the issue of dynamically hedging, in continuous-time, one short option position by the underlying S and the riskless constant asset. Let a hedging

strategy $\zeta = (\zeta_t)_{t \in [0, T]}$ denote the number of units of stock held in the hedging portfolio at every point in time.

Justify that the tracking error (or profit-and-loss) process $e = e(\zeta)$ associated with the strategy ζ in S (and the quantity of riskless asset deduced from ζ by the self-financing condition) evolves according to

$$de_t = -d\Pi_t + \zeta_t dS_t$$

(with $e_0 = 0$).

2. One says that a strategy ζ replicates the payoff $\phi(S_T)$ if $e_T(\zeta) = 0$ almost surely under the historical probability measure \mathbb{P} .

Justify that a strategy ζ replicates $\phi(S_T)$ under $\widehat{\mathbb{P}}$ if and only if it replicates $\phi(S_T)$ under \mathbb{P} , i.e. $e_T(\zeta) = 0$ almost surely under \mathbb{P} .

3. Using (17.8) and (17.9), rewrite de_t as

$$de_t = \alpha_t dS_t + \beta_t dM_t$$

for integrands α_t and β_t to be determined.

4. Exhibit one elementary class of payoff functions ϕ for which there exists a replication strategy ζ ($e_T(\zeta) = 0$ almost surely) and describe this strategy.
5. Justify that, in general, there is no strategy ζ replicating the payoff $\phi(S_T)$.
6. Using (17.7), show that the strategy which minimizes the risk-neutral variance of e_T assumes the following form:

$$\zeta_t^{va} = S_t^{-1} \partial_x u(t, X_t, \sigma_t).$$

7. Does the strategy ζ^{va} minimize the variance under the historical probability measure?

Under which measure: risk-neutral or historical, are we interested in minimizing the variance of the hedging error in practice?

What do you think mathematically of the problem of minimizing the variance of the hedging error under the historical probability measure?

What do you think of the hedge ζ^{va} ?

8. One now considers the dynamic hedging in continuous-time of the option by its underlying S and an auxiliary option with payoff $\varphi(S_\Theta)$ at $\Theta > T$ (and the constant asset). The auxiliary option price process Φ is thus written $\Phi_t = v(t, X_t, \sigma_t)$, for some related pricing function $v = v(t, x, \sigma)$. The replication error $e = e(\zeta, \eta)$ associated with the strategy ζ in S and η in the auxiliary option (and the quantity of riskless asset deduced from the self-financing condition) evolves according to

$$de_t = -d\Pi_t + \zeta_t dS_t + \eta_t d\Phi_t$$

(with $e_0 = 0$).

Rewrite de_t as

$$de_t = \alpha_t dS_t + \beta_t dM_t$$

for integrands α_t and β_t to be determined.

Write the replication condition $e_T(\zeta, \eta) = 0$ in the form of a system of equations to be satisfied by the pair (ζ_t, η_t) for every $t < T$. Discuss the well-posedness of this system (write an explicit condition on the function v for it to have a unique solution) for every $t \in [0, T]$.

9. Describe a simulation algorithm in terms of the tree with transition probabilities as developed in the question II.4, allowing for numerical verification of the statistical properties of the hedging errors studied theoretically in the questions III.4–8.

17.4 Jump-to-Ruin

EXERCISE

1. Describe and justify the inverse method for simulating a real random variable X with a numerically invertible c.d.f. F .
2. Describe the rejection-acceptance method for simulating i.i.d. uniform points in the unit disk D .
3. Describe the Marsaglia method to simulate a pair of independent Gaussian random variables.
4. How to simulate a lognormal random variable X with mean m and standard deviation $s = 20\%m$?

PROBLEM We consider the following Merton Jump-to-Ruin (JR) model, under a risk-neutral probability measure \mathbb{P} :

$$dS_t = \lambda S_t dt + \sigma S_t dW_t - S_{t-} dN_t = \sigma S_t dW_t - S_{t-} dM_t, \quad (17.11)$$

where W is a standard Brownian motion, $\sigma > 0$ is a constant volatility parameter, and N is a Poisson process with intensity $\lambda > 0$ and compensated martingale $M_t = N_t - \lambda t$. The funding rate is supposed to be zero. The model filtration is $\mathbb{F} = \mathbb{F}^W \vee \mathbb{F}^N$. One considers a derivative with an \mathcal{F}_T -measurable payoff ξ at time T .

We denote by \mathcal{N} the Gaussian cumulative distribution function and, given the maturity T and strike K of an option, we let

$$d_{\pm} = \frac{\ln(\frac{S_0}{K}) + \lambda T}{\sigma \sqrt{T}} \pm \frac{1}{2} \sigma \sqrt{T}.$$

I Call Option In this part we consider the pricing of a vanilla call option, so that $\xi = \phi(S_T) = (S_T - K)^+$.

1. Prove that the JR price process C of this option can be represented, for $t \in [0, T]$, by:

$$C_t = \mathbb{E}(\phi(S_T) | \mathcal{F}_t) = u(t, S_t)$$

for a suitable pricing function $u = u(t, S)$ over $[0, T] \times \mathbb{R}_+$.

2. Show that we have for $t \in [0, T]$

$$dC_t = du(t, S_t) = (\partial_t u + \mathcal{A}u)(t, S_t) dt + \partial_S u(t, S_t) \sigma S_t dW_t + \delta u(t, S_{t-}) dM_t, \quad (17.12)$$

where

$$\delta u(t, S) = u(t, 0) - u(t, S)$$

$$\mathcal{A}u(t, S) = \lambda S \partial_S u(t, S) + \frac{\sigma^2 S^2}{2} \partial_{S^2}^2 u(t, S) + \lambda \delta u(t, S).$$

3. Prove that $u(t, 0) = 0$ and that the pricing function u satisfies the following pricing equation:

$$\begin{cases} u(T, S) = (S - K)^+, & S > 0 \\ \partial_t u(t, S) + \lambda S \partial_S u(t, S) + \frac{\sigma^2 S^2}{2} \partial_{S^2}^2 u(t, S) - \lambda u(t, S) = 0, & t < T, S > 0. \end{cases} \quad (17.13)$$

4. Deduce the expression of the price C_0 of the option at time 0.
 5. Introducing an auxiliary Black–Scholes (BS) model S_t^{bs} such that $S_0^{bs} = S_0$ and for $t \in [0, T]$:

$$dS_t^{bs} = \lambda S_t^{bs} dt + \sigma S_t^{bs} dW_t, \quad (17.14)$$

write an explicit representation of S_T in terms of S_T^{bs} and the first jump time τ of the driving Poisson process N .

Use this representation to compute C_0 directly, by probabilistic arguments.

II Put Option Now consider the pricing of a put option, in two forms: either a vanilla put with payoff $\xi = (K - S_T)^+$, or a vulnerable put with payoff $\xi^* = \mathbb{1}_{\{\tau > T\}}(K - S_T)^+$. A vulnerable put option thus only delivers its promised payoff $(K - S_T)^+$ if $\tau > T$. This corresponds to the case of a warrant issued by the firm with stock modeled by S . In this case, if the issuing firm is defaulted by T it will not be able¹ to deliver the promised payoff $(K - S_T)^+ = K$.

Note that for a call option, vulnerable or not makes no difference in JR, since in this model $S_T = (S_T - K)^+ = 0$ on $\tau < T$.

1. Following the lines of questions 1 and 2 in Part I, prove that the JR price process P of the vanilla put can be represented, for $t \in [0, T]$, as:

$$P_t = v(t, S_t)$$

¹Up to some recovery, assumed here to be zero for simplicity.

for a vanilla put pricing function $v = v(t, S)$ over $[0, T] \times \mathbb{R}_+$ such that

$$\begin{cases} v(T, S) = (K - S)^+, & S > 0 \\ \partial_t v(t, S) + \lambda S \partial_S v(t, S) + \frac{\sigma^2 S^2}{2} \partial_{S^2}^2 v(t, S) - \lambda v(t, S) \\ \quad + \lambda K = 0, & t < T, S > 0. \end{cases} \quad (17.15)$$

2. Recall and use the vanilla call-put parity relationship to show that

$$P_0 = Ke^{-\lambda T} \mathcal{N}(-d_-) - S_0 \mathcal{N}(-d_+) + K(1 - e^{-\lambda T}).$$

3. Prove that the JR price process P^* of the vulnerable put can be represented, for $t \in [0, T]$ as:

$$P_t^* = \mathbb{1}_{t < \tau} v^*(t, S_t)$$

for a pre-default vulnerable put pricing function $v^* = v^*(t, S)$ over $[0, T] \times \mathbb{R}_+$ such that

$$\begin{cases} v^*(T, S) = (K - S)^+, & S > 0 \\ \partial_t v^*(t, S) + \lambda S \partial_S v^*(t, S) + \frac{\sigma^2 S^2}{2} \partial_{S^2}^2 v^*(t, S) \\ \quad - \lambda v^*(t, S) = 0, & t < T, S > 0. \end{cases} \quad (17.16)$$

4. Prove directly by probabilistic arguments that

$$\begin{aligned} P_0^* &= C_0 - S_0 + Ke^{-\lambda T} \\ &= P_0 - (1 - e^{-\lambda T})K. \end{aligned} \quad (17.17)$$

III Barrier Option With a barrier option the right to exercise the payoff at maturity depends on the underlying having crossed a given barrier level on $[0, T]$. We will consider in this part an up-and-out call option with barrier level H corresponding to the following payoff:

$$\xi = \mathbb{1}_{\{\theta=T, S_T < H\}} (S_T - K)^+,$$

where

$$\theta = \inf\{t \geq 0; S_t \geq H\} \wedge T.$$

1. What is the interest of this barrier option as compared to the corresponding vanilla call? Which of the two should be cheaper?

2. Show that the price process B_t of this option is given over $[0, \theta]$ by $w(t, S_t)$, for a pricing function $w = w(t, S)$ such that

$$\begin{cases} w(T, S) = (S - K)^+, & 0 < S < H \\ w(t, H) = 0, & 0 \leq t \leq T \\ \partial_t w(t, S) + \lambda S \partial_S w(t, S) + \frac{\sigma^2 S^2}{2} w(t, S) \partial_{S^2}^2 w(t, S) \\ - \lambda w(t, S) = 0, & t < T, 0 < S < H. \end{cases} \quad (17.18)$$

3. Show that the time-0 JR and BS prices of the up-and-out call coincide.

IV Effective Volatility We recall the Dupire equation which is satisfied at time 0 by call prices C^{lo} in the (T, K) -variables, in a local volatility model with volatility function $\sigma(t, S)$ and risk-free interest rate r (assuming no dividends on S):

$$\begin{cases} C_0^{lo}(T = 0, K) = (S_0 - K)^+, & K > 0 \\ \partial_T C_0^{lo}(T, K) + r K \partial_K C_0^{lo}(T, K) \\ - \frac{1}{2} \sigma(T, K)^2 K^2 \partial_K^2 C_0^{lo}(T, K) = 0, & T, K > 0. \end{cases} \quad (17.19)$$

1. Recall what calibrating a model means.

What is calibration used for?

How is model calibration achieved in practice?

2. Using the Dupire equation (17.19), write the corresponding Dupire formula for $\sigma(T, K)$, and discuss the statement that “the class of local volatility models is calibratable to any reasonable market of vanilla call options”, where the meaning of “reasonable” will be specified.
3. Using the Dupire formulas in the JR and BS models, derive the following expression for the local volatility function $\sigma_0^{jr}(\cdot, \cdot)$ calibrated to the JR vanilla call prices at time 0 (the so-called JR-effective volatility function σ_0^{jr}):

$$\frac{\sigma_0^{jr}(T, K)^2}{2} = \frac{\sigma^2}{2} + \lambda \sigma \sqrt{2\pi T} \mathcal{N}(d_-) e^{\frac{d_-^2}{2}}. \quad (17.20)$$

4. Admitting that $\mathcal{N}(x) \sim_{x \rightarrow -\infty} \frac{-e^{-x^2}}{\sqrt{2\pi x}}$ and also using the fact that $\mathcal{N}(x) \rightarrow 1$ as $x \rightarrow +\infty$, compute the limit of $\sigma_0^{jr}(T, K)$ as $T \rightarrow 0+$ depending on whether $K < S_0, = S_0$ or $> S_0$. Interpret the result.

V Local Default Intensity In this part, we consider a driving point process N with a local intensity $\lambda(t, S_t)$, so that N is no longer a Poisson process (and N and W are no longer independent). All the above Markov and martingale analysis still holds true with λ replaced by $\lambda(t, S)$. But all explicit formulas collapse. So

the prices C_0 , P_0 , P_0^* and B_0 in Parts I, II, III, and the effective volatility function $\sigma_0(\cdot, \cdot)$ defined through the relevant Dupire formula (calibrated to the vanilla option prices in the jump-to-ruin model with local jump intensity model of this part) all need to be estimated numerically.

Given uniform time- and space-grids $(t_i)_{1 \leq i \leq n}$ and $(S^j)_{0 \leq j \leq m}$ of respective steps h and k , with the S -grid centered around S_0 , we denote

$$\alpha_i^j = \frac{\sigma^2(S^j)^2}{2k^2} - \frac{\lambda_i^j S^j}{2k}, \quad \beta_i^j = -\frac{\sigma^2(S^j)^2}{k^2}, \quad \gamma_i^j = \frac{\sigma^2(S^j)^2}{2k^2} + \frac{\lambda_i^j S^j}{2k}, \quad (17.21)$$

with $\lambda_i^j = \lambda(t_i, S^j)$.

1. In terms of the above coefficients, write an explicit finite difference scheme in the (t, S) variables for computing $C_0 = u(0, S_0)$. Derive the associated stability condition.
2. In terms of the above coefficients, write a hybrid scheme, implicit in the differential terms and explicit in the jump term, for computing $C_0 = u(0, S_0)$, not subject to the previous stability condition, and discuss the implementation of this scheme.
3. Discuss the changes required in the above schemes for computing P_0 , P_0^* and B_0 .
4. Propose a numerical method for calibrating the effective volatility function $\sigma_0(\cdot, \cdot)$ corresponding to the model of this part (jump-to-ruin model with local jump intensity).
5. Write, in terms of suitable coefficients, finite difference θ -schemes that can be used for pricing the above options in the local volatility model with volatility function $\sigma_0(\cdot, \cdot)$.

Discuss the particular issues related with explosion of $\sigma_0(t, S)$ (cf. in the situation of Part IV the function $\sigma_0^{jr}(\cdot, \cdot)$ at $t = 0+$, $S < S_0$).

For which of the options do we expect agreement (up to the numerical noise) between the prices in the jump-to-ruin model with local jump intensity of this part and the local volatility model with corresponding effective volatility $\sigma_0(\cdot, \cdot)$?

VI Hedging

1. Back at the situation of a constant intensity λ of N_t as in Parts I–IV, in this part we consider the issue of dynamically hedging, in continuous-time, one short vanilla call option position by the underlying S and the riskless constant asset. Let a hedging strategy $\zeta = (\zeta_t)_{t \in [0, T]}$ denote the number of units of stock which are held in the hedging portfolio at every point in time.

Justify that the tracking error (or profit-and-loss) process $e = e(\zeta)$ associated with the strategy ζ in S (and the quantity of riskless asset deduced from ζ by the self-financing condition) evolves following

$$de_t = -dC_t + \zeta_t dS_t$$

(with $e_0 = 0$).

2. A strategy ζ is said to replicate the payoff $\phi(S_T)$ if $e_T(\zeta) = 0$ almost surely under the historical probability measure $\widehat{\mathbb{P}}$.

Justify that a strategy ζ replicates $\phi(S_T)$ under $\widehat{\mathbb{P}}$ if and only if it replicates $\phi(S_T)$ under \mathbb{P} , i.e. $e_T(\zeta) = 0$ almost surely under \mathbb{P} .

3. Rewrite de_t as

$$de_t = \alpha_t dW_t + \beta_t dM_t$$

for integrands α_t and β_t to be specified.

4. Justify that there is no strategy ζ replicating the payoff $\phi(S_T)$.

5. Show that the strategy $\zeta_t^{va} = \frac{d\langle C, S \rangle_t}{d\langle S \rangle_t}$ that minimizes the risk-neutral variance of e_T assumes the following form:

$$\zeta_t^{va} = \frac{\sigma^2}{\sigma^2 + \lambda} \partial_S u(t, S_{t-}) - \frac{\lambda}{\sigma^2 + \lambda} \frac{\delta u(t, S_{t-})}{S_{t-}}.$$

6. Does the strategy ζ^{va} minimize the variance under the historical probability measure?

Under which measure: risk-neutral or historical, are we interested in minimizing the variance of the hedging error in practice?

What do you think mathematically of the problem of minimizing the variance of the hedging error under the historical probability measure?

What do you think of the hedge ζ^{va} ?

7. We now consider the dynamic hedging in continuous-time of the vanilla call option by its underlying S and an auxiliary put option with payoff $(K - S_\Theta)^+$ at $\Theta > T$ (and the constant funding asset). The auxiliary option price process P is thus written $P_t = v(t, S_t)$, for some related pricing function $v = v(t, S)$. The replication error $e = e(\zeta, \eta)$ associated with the strategy ζ in S and η in the auxiliary option (and the quantity of riskless asset deduced from the self-financing condition) evolves following

$$de_t = -dC_t + \zeta_t dS_t + \eta_t dP_t$$

(with $e_0 = 0$).

Write the replication condition $e_T(\zeta, \eta) = 0$ in the form of a system of equations to be satisfied by the pair (ζ_t, η_t) for every $t < T$.

Discuss the well-posedness of this system for every $t \in [0, T]$.

8. Describe simulation algorithms allowing for numerical verification of the hedging properties highlighted in the questions 4–7.
9. What changes are needed in this part if the option which is hedged is now a barrier up-and-out call as in Part III?

References

1. Achdou, Y., & Pironneau, O. (2005). *Computational methods for option pricing*. Philadelphia: SIAM.
2. Albanese, C., Campolieti, G., Carr, P., & Lipton, A. (2001). Black–Scholes goes hypergeometric. *Risk, December 2001*, 99–103.
3. Alvarez, O., & Tourin, A. (1996). Viscosity solutions of nonlinear integro-differential equations. *Annales de l’Institut Henri Poincaré (C) Analyse non linéaire*, 13(3), 293–317.
4. Amadori, A. L. (2003). Nonlinear integro-differential evolution problems arising in option pricing: a viscosity solutions approach. *Differential and Integral Equations*, 16(7), 787–811.
5. Amadori, A. L. (2007). The obstacle problem for nonlinear integro-differential operators arising in option pricing. *Ricerche Di Matematica*, 56(1), 1–17.
6. Andersen, L., & Brotherton-Ratcliffe, R. (1996). Exact exotics. *Risk, October 1996*, 85–89.
7. Andersen, L., & Sidenius, J. (2004). Extensions to the Gaussian copula: random recovery and random factor loadings. *The Journal of Credit Risk*, 1(1), 29–70.
8. Assefa, S., Bielecki, T., Crépey, S., & Jeanblanc, M. (2011). CVA computation for counterparty risk assessment in credit portfolios. In T. Bielecki, D. Brigo, & F. Patras (Eds.), *Credit risk frontiers* (pp. 397–436). New York: Wiley.
9. Avellaneda, M. (1999). Minimum-relative entropy calibration of asset-pricing models. *International Journal of Theoretical and Applied Finance*, 1(4), 447–472.
10. Avellaneda, M., Bu, R., Friedman, C., Grandchamp, N., Kruk, L., & Newman, J. (2001). Weighted Monte Carlo: a new technique for calibrating asset-pricing models. *International Journal of Theoretical and Applied Finance*, 4(1), 91–119.
11. Avellaneda, M., Friedman, C., Holmes, R., & Samperi, D. (1997). Calibrating volatility surfaces via relative-entropy minimization. *Applied Mathematical Finance*, 41, 37–64.
12. Avellaneda, M., & Laurence, P. (2000). *Quantitative modeling of derivative securities from theory to practice*. London: Chapman & Hall.
13. Avellaneda, M., Levy, A., & Paras, A. (1995). Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied Mathematical Finance*, 2(2), 73–88.
14. Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques de L’École Normale Supérieure*, 17, 21–86.
15. Balland, P. (2002). Deterministic implied volatility models. *Quantitative Finance*, 2(1), 31–44.
16. Bally, V., Caballero, E., Fernandez, B., & El-Karoui, N. (2002). *Reflected BSDE's PDE's and variational inequalities* (INRIA Technical Report No. 4455).
17. Bally, V., & Matoussi, A. (2001). Weak solutions for SPDEs and backward doubly stochastic differential equations. *Journal of Theoretical Probability*, 14(1), 125–164.
18. Bally, V., & Pagès, G. (2003). A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems. *Bernoulli*, 9(6), 1003–1049.

19. Bally, V., Pagès, G., & Printems, J. (2001). A stochastic quantization method for nonlinear problems. Monte Carlo and probabilistic methods for partial differential equations. *Monte Carlo Methods and Applications*, 7(1–2), 21–33.
20. Barles, G., Buckdahn, R., & Pardoux, E. (1997). Backward stochastic differential equations and integral-partial differential equations. *Stochastics & Stochastics Reports*, 60, 57–83.
21. Barles, G., & Imbert, C. (2008). Second-order elliptic integro-differential equations: viscosity solutions theory revisited. *Annales de L'IHP*, 25(3), 567–585.
22. Barles, G., & Lesigne, L. (1997). SDE, BSDE and PDE. In N. El Karoui & L. Mazliak (Eds.), *Pitman research notes in mathematics series: Vol. 364. Backward stochastic differential equations* (pp. 47–80).
23. Barles, G., & Souganidis, P. E. (1991). Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Analysis*, 4, 271–283.
24. Bates, D. (1996). Jumps and stochastic volatility: exchange rate processes implicit in deutsche mark options. *The Review of Financial Studies*, 9(1), 69–107.
25. Becherer, D., & Schweizer, M. (2005). Classical solutions to reaction–diffusion systems for hedging with interacting Itô and point processes. *The Annals of Applied Probability*, 15, 1111–1144.
26. Bellman, R. (1957). *Dynamic programming*. Princeton: Princeton University Press.
27. Bensoussan, A., & Lions, J.-L. (1982). *Applications of variational inequalities in stochastic control*. Amsterdam: North-Holland.
28. Bensoussan, A., & Lions, J.-L. (1984). *Impulse control and quasi-variational inequalities*. Paris: Gauthier-Villars.
29. Beveridge, C., & Joshi, M. (2011). Monte Carlo bounds for game options including convertible bonds. *Management Science*, 57(5), 960–974.
30. Bielecki, T., Brigo, D., & Crépey, S. (2013). Counterparty risk modeling—collateralization, funding and hedging. Taylor & Francis (in preparation).
31. Bielecki, T. R., Cousin, A., Crépey, S., & Herbertsson, A. (2012). Dynamic hedging of portfolio credit risk in a Markov copula model. *Journal of Optimization Theory and Applications* (doi:10.1007/s10957-013-0318-4, forthcoming).
32. Bielecki, T. R., Cousin, A., Crépey, S., & Herbertsson, A. (2012). A bottom-up dynamic model of portfolio credit risk with stochastic intensities and random recoveries. *Communications in Statistics—Theory and Methods* (in revision).
33. Bielecki, T., & Crépey, S. (2011). Dynamic hedging of counterparty exposure. In T. Zaphiropoulou, M. Rutkowski, & Y. Kabanov (Eds.), *The Musiela Festschrift*. Berlin: Springer (forthcoming).
34. Bielecki, T., Crépey, S., & Herbertsson, A. (2011). Markov chain models of portfolio credit risk. In A. Lipton & A. Rennie (Eds.), *Oxford handbooks in finance. Oxford handbook of credit derivatives* (pp. 327–382).
35. Bielecki, T. R., Crépey, S., & Jeanblanc, M. (2010). Up and down credit risk. *Quantitative Finance*, 10(10), 1137–1151.
36. Bielecki, T. R., Crépey, S., Jeanblanc, M., & Rutkowski, M. (2007). Valuation of basket credit derivatives in the credit migrations environment. In J. Birge & V. Linetsky (Eds.), *Handbook of financial engineering*. Amsterdam: Elsevier.
37. Bielecki, T. R., Crépey, S., Jeanblanc, M., & Rutkowski, M. (2008). Defaultable options in a Markovian intensity model of credit risk. *Mathematical Finance*, 18, 493–518 (updated version on www.defaultrisk.com).
38. Bielecki, T. R., Crépey, S., Jeanblanc, M., & Rutkowski, M. (2009). Valuation and hedging of defaultable game options in a hazard process model. *Journal of Applied Mathematics and Stochastic Analysis*. Article ID 695798 (long preprint version on www.defaultrisk.com).
39. Bielecki, T. R., Crépey, S., Jeanblanc, M., & Rutkowski, M. (2010). Arbitrage pricing of defaultable game options with applications to convertible bonds. *Quantitative Finance*, 8(8), 795–810.

40. Bielecki, T. R., Crépey, S., Jeanblanc, M., & Rutkowski, M. (2011). Convertible bonds in a defaultable diffusion model. In A. Kohatsu-Higa, N. Privault, & S. Sheu (Eds.), *Stochastic analysis with financial applications* (pp. 255–298). Basel: Birkhäuser.
41. Bielecki, T. R., Jakubowski, J., & Nieweglowskic, M. (2012). Study of dependence for some stochastic processes: symbolic Markov copulae. *Stochastic Processes and Their Applications*, 122(3), 930–951.
42. Bielecki, T. R., Jakubowski, J., Vidozzi, A., & Vidozzi, L. (2008). Study of dependence for some stochastic processes. *Stochastic Analysis and Applications*, 26(4), 903–924.
43. Bielecki, T. R., Jeanblanc, M., & Rutkowski, M. (2009). *CSFI lecture notes series: Vol. 2. Credit risk modeling*. Osaka University: Osaka University Press.
44. Bielecki, T. R., & Rutkowski, M. (2002). *Credit risk: modeling, valuation and hedging*. Berlin: Springer.
45. Bielecki, T. R., Vidozzi, A., & Vidozzi, L. (2008). A Markov copulae approach to pricing and hedging of credit index derivatives and ratings triggered step-up bonds. *The Journal of Credit Risk*, 4(1), 47–76.
46. Bismut, J. (1973). Conjugate convex functions in optimal stochastic control. *Journal of Mathematical Analysis and Applications*, 44, 384–404.
47. Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81, 635–654.
48. Boel, R., Varaiya, P., & Wong, E. (1975). Martingales on jump processes: I Representation results. *SIAM Journal on Control and Optimization*, 13(5), 999–1021.
49. Boel, R., Varaiya, P., & Wong, E. (1975). Martingales on jump processes: II Applications. *SIAM Journal on Control and Optimization*, 13(5), 1022–1061.
50. Bouchard, B., & Chassagneux, J.-F. (2008). Discrete time approximation for continuously and discretely reflected BSDE's. *Stochastic Processes and Their Applications*, 118(12), 2269–2293.
51. Bouchard, B., Ekeland, I., & Touzi, N. (2004). On the Malliavin approach to Monte Carlo approximations of conditional expectations. *Finance and Stochastics*, 8, 45–71.
52. Bouchard, B., & Menozzi, S. (2009). Strong approximations of BSDEs in a domain. *Bernoulli*, 15(4), 1117–1147.
53. Bouchard, B., & Warin, X. (2010). Monte Carlo valuation of American options: new algorithm to improve on existing methods. In R. Carmona, P. Del Moral, P. Hu, & N. Oudjane (Eds.) *Numerical methods in finance* (Chapter 7). Berlin: Springer.
54. Brace, A., Gatarek, D., & Musiela, M. (1997). The market model of interest rate dynamics. *Mathematical Finance*, 7(2), 127–154.
55. Breeden, D., & Litzenberger, R. (1978). Prices of state-contingent claims implicit in options prices. *Journal of Business*, 51, 621–651.
56. Brémaud, P. (1981). *Point processes and queues, martingale dynamics*. New York: Springer.
57. Briani, M., La Chioma, C., & Natalini, R. (2004). Convergence of numerical schemes for viscosity solutions to integro-differential degenerate parabolic problems arising in financial theory. *Numerische Mathematik*, 98(4), 607–646.
58. Brigo, D., & Mercurio, F. (2006). *Interest rate models—theory and practice. With smile, inflation and credit* (2nd ed.). Berlin: Springer.
59. Brigo, D., Morini, M., & Pallavicini, A. (2013). *Counterparty credit risk, collateral and funding with pricing cases for all asset classes*. New York: Wiley.
60. Broadie, M., & Glasserman, P. (1996). Estimating security price derivatives using simulation. *Management Science*, 42(2), 269–285.
61. Carmona, R., & Crépey, S. (2010). Importance sampling and interacting particle systems for the estimation of Markovian credit portfolios loss distribution. *International Journal of Theoretical and Applied Finance*, 13(4), 577–602.
62. Carr, P., & Madan, D. (1998). Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2, 61–73.
63. Cesari, G., Aquilina, J., Charpillon, N., Filipovic, Z., Lee, G., & Manda, I. (2010). *Modelling, pricing, and hedging counterparty credit exposure*. Berlin: Springer.

64. Chassagneux, J.-F. (2009). Discrete time approximation of doubly reflected BSDEs. *Advances in Applied Probability*, 41, 101–130.
65. Chassagneux, J.-F., & Crépey, S. (2012). Doubly reflected BSDEs with call protection and their approximation. *ESAIM: Probability and Statistics* (in revision).
66. Cherny, A., & Shiryaev, A. (2002). Vector stochastic integrals and the fundamental theorems of asset pricing. *Proceedings of the Steklov Institute of Mathematics*, 237, 6–49.
67. Coculescu, D., & Nikeghbali, A. (2012). Hazard processes and martingale hazard processes. *Mathematical Finance*, 22(3), 519–537.
68. Coleman, T., Li, Y., & Verma, A. (1999). Reconstructing the unknown volatility function. *Journal of Computational Finance*, 2(3), 77–102.
69. Cont, R., & Fournié, D. (2012). Functional Itô calculus and stochastic integral representation of martingales. *Annals of Probability*, 41(1), 109–133.
70. Cont, R., & Minca, A. (2011). Recovering portfolio default intensities implied by CDO quotes. *Mathematical Finance*, online first 22 June 2011.
71. Cont, R., & Tankov, P. (2003). *Financial modelling with jump processes*. London: Chapman & Hall.
72. Cont, R., & Voltchkova, K. (2005). A finite difference methods for option pricing in jump diffusion and exponential Lévy models. *SIAM Journal on Numerical Analysis*, 43(4), 1596–1626.
73. Cousin, A., Crépey, S., & Kan, Y. H. (2012). Delta-hedging correlation risk? *Review of Derivatives Research*, 15(1), 25–56.
74. Cox, J., Ross, S., & Rubinstein, M. (1979). Option pricing: a simplified approach. *Journal of Financial Economics*, 7(3), 229–263.
75. Crandall, M., Ishii, H., & Lions, P.-L. (1992). User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society*, 27, 1–67.
76. Crépey, S. (2001). *Contribution à des méthodes numériques appliquées à la finance et aux jeux différentiels*. PhD Thesis, Ecole Polytechnique/INRIA Sophia Antipolis.
77. Crépey, S. (2003). Calibration of the local volatility in a generalized Black–Scholes model using Tikhonov regularization. *SIAM Journal on Mathematical Analysis*, 34(5), 1183–1206.
78. Crépey, S. (2003). Calibration of the local volatility in a trinomial tree using Tikhonov regularization. *Inverse Problems*, 19, 91–127.
79. Crépey, S. (2004). Delta-hedging vega risk? *Quantitative Finance*, 4, 559–579.
80. Crépey, S. (2013). Bilateral counterparty risk under funding constraints—Part I: Pricing. *Mathematical Finance* (online December 2012, doi:[10.1111/mafi.12004](https://doi.org/10.1111/mafi.12004)).
81. Crépey, S. (2013). Bilateral counterparty risk under funding constraints—Part II: CVA. *Mathematical Finance* (online December 2012, doi:[10.1111/mafi.12005](https://doi.org/10.1111/mafi.12005)).
82. Crépey, S. (2013). Wrong way and gap risks modeling: a marked default time approach (in preparation).
83. Crépey, S., & Douady, R. (2012). LOIS: credit and liquidity. *Risk Magazine* (forthcoming in the June 2013 issue).
84. Crépey, S., Grbac, Z., & Nguyen, H. N. (2012). A multiple-curve HJM model of interbank risk. *Mathematics and Financial Economics*, 6(6), 155–190.
85. Crépey, S., Jeanblanc, M., & Wu, D. L. (2013). Informationally dynamized Gaussian copula. *International Journal of Theoretical and Applied Finance*, 16, 1350008. doi:[10.1142/S0219024913500088](https://doi.org/10.1142/S0219024913500088).
86. Crépey, S., Jeanblanc, M., & Zargari, B. (2010). Counterparty risk on a CDS in a Markov chain copula model with joint defaults. In M. Kijima, C. Hara, Y. Muromachi, & K. Tanaka (Eds.), *Recent advances in financial engineering 2009* (pp. 91–126). Singapore: World Scientific.
87. Crépey, S., & Matoussi, A. (2008). Reflected and doubly reflected BSDEs with jumps: a priori estimates and comparison principle. *The Annals of Applied Probability*, 18(5), 2041–2069.
88. Crépey, S., & Rahal, A. (2012). Pricing convertible bonds with call protection. *Journal of Computational Finance*, 15(2), 37–75.

89. Crépey, S., & Rahal, A. (2013). Simulation/regression pricing schemes for CVA computations on CDO tranches. *Communications in Statistics—Theory and Methods* (forthcoming).
90. Cvitanic, J., & Karatzas, I. (1996). Backward stochastic differential equations with reflection and Dynkin games. *Annals of Probability*, 24(4), 2024–2056.
91. Darling, R., & Pardoux, E. (1997). Backward SDE with random terminal time and applications to semilinear elliptic PDE. *Annals of Probability*, 25, 1135–1159.
92. Davis, M. H., & Varaiya, P. (1974). The multiplicity of an increasing family of σ -fields. *Annals of Probability*, 2, 958–963.
93. Delbaen, F., & Schachermayer, W. (1997). The fundamental theorem of asset pricing for unbounded stochastic processes. *Mathematische Annalen*, 312, 215–250.
94. Delbaen, F., & Schachermayer, W. (2005). *The mathematics of arbitrage*. Berlin: Springer.
95. Dellacherie, C., & Meyer, P.-A. (1982). *Probabilities and potential*. Amsterdam: North-Holland.
96. Delong, L. (2013). *Backward stochastic differential equations with jumps and their actuarial and financial applications*. Springer (forthcoming in the EAA Series).
97. Demeterfi, K., Derman, E., Kamal, M., & Zou, J. (1999). A guide to volatility and variance swaps. *The Journal of Derivatives*, 6(4), 9–32.
98. Derman, E., & Kani, I. (1994). Riding on a smile. *Risk, February 1994*, 139–145.
99. De Spiegeleer, J., & Schoutens, W. (2011). *Contingent convertible (CoCo) notes structure and pricing*. London: Euromoney Books.
100. De Spiegeleer, J., Schoutens, W., & Jabre, P. (2011). *The handbook of convertible bonds: pricing, strategies and risk management*. Chichester: Wiley.
101. 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.
102. Dubois, F., & Lelievre, T. (2005). Efficient pricing of Asian options by the PDE approach. *Journal of Computational Finance*, 8(2), 55–64.
103. Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6), 1343–1376.
104. Duffy, D. (2006). *Financial difference methods in financial engineering*. Chichester: Wiley.
105. Dupire, B. (1994). Pricing with a smile. *Risk, January 1994*, 18–20.
106. Dynkin, E. B. (1965). *Markov processes*. Berlin: Springer.
107. Dynkin, E. (1991). Branching particle systems and superprocesses. *Annals of Probability*, 19, 1157–1194.
108. Dynkin, E. (1993). Superprocesses and partial differential equations. *Annals of Probability*, 21, 1185–1262.
109. Egger, H., & Engl, H. W. (2005). Tikhonov regularization applied to the inverse problem of option pricing: convergence analysis and rates. *Inverse Problems*, 21, 1027–1045.
110. El Karoui, N., & Jiao, Y. (2009). Stein's method and zero bias transformation for CDOs tranche pricing. *Finance and Stochastics*, 13(2), 151–180.
111. El Karoui, N., Jiao, Y., & Kurtz, D. (2008). Gauss and Poisson approximation: applications to CDOs tranche pricing. *Journal of Computational Finance*, 12(2), 31–58.
112. El Karoui, N., Kapoudjian, E., Pardoux, C., Peng, S., & Quenez, M.-C. (1997). Reflected solutions of backward SDEs, and related obstacle problems for PDEs. *Annals of Probability*, 25, 702–737.
113. El Karoui, N., & Mazliak, L. (1997). *Backward stochastic differential equations*. Harlow: Longman.
114. El Karoui, N., Peng, S., & Quenez, M.-C. (1997). Backward stochastic differential equations in finance. *Mathematical Finance*, 7, 1–71.
115. El Karoui, N., & Quenez, M.-C. (1997). Nonlinear pricing theory and backward stochastic differential equations. In W. Runggaldier (Ed.), *Lecture notes in mathematics: Vol. 1656. Financial mathematics*, Bressanone, 1996 (pp. 191–246). Berlin: Springer.
116. Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems*. Dordrecht: Kluwer Academic.

117. Engl, H. W., Kunisch, K., & Neubauer, A. (1989). Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems. *Inverse Problems*, 5(4), 523–540.
118. Ern, A., Villeneuve, S., & Zanette, A. (2004). Adaptive finite element methods for local volatility European option pricing. *International Journal of Theoretical and Applied Finance*, 7(6), 659–684.
119. Ethier, H., & Kurtz, T. (1986). *Markov processes. Characterization and convergence*. New York: Wiley-Interscience.
120. Fermanian, J.-D., & Vigneron, O. (2010). Pricing and hedging basket credit derivatives in the Gaussian copula. *Risk, February 2010*, 92–96.
121. Filipović, D., & Trolle, A. (2011). The term structure of interbank risk. *Journal of Financial Economics* (forthcoming).
122. Fleming, W., & Soner, H. (2006). *Controlled Markov processes and viscosity solutions* (2nd ed.). New York: Springer.
123. Föllmer, H., & Sondermann, D. (1986). Hedging of non-redundant contingent claims. In A. Mas-Colell & W. Hildenbrand (Eds.), *Contributions to mathematical economics* (pp. 205–223). Amsterdam: North-Holland.
124. Fournié, E., Lasry, J.-M., Lebuchoux, J., & Lions, P.-L. (2001). Applications of Malliavin calculus to Monte Carlo methods in finance. II. *Finance and Stochastics*, 5, 201–236.
125. Fournié, E., Lasry, J.-M., Lebuchoux, J., Lions, P.-L., & Touzi, N. (1999). An application of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3, 391–412.
126. Freidlin, M. (1996). *Markov processes and differential equations: asymptotic problems*. Basel: Birkhäuser.
127. Frey, R., & Backhaus, J. (2010). Dynamic hedging of synthetic CDO tranches with spread and contagion risk. *Journal of Economic Dynamics & Control*, 34(4), 710–724.
128. Frey, R., & Backhaus, J. (2008). Pricing and hedging of portfolio credit derivatives with interacting default intensities. *International Journal of Theoretical and Applied Finance*, 11(6), 611–634.
129. Friedman, A. (1964). *Partial differential equations of parabolic type*. New York: Prentice Hall.
130. Fujita, H. (1966). On the blowing up of solutions of the Cauchy problem for $u_t = \nabla u + u^{1+\alpha}$. *Journal of Mathematical Sciences, The University of Tokyo*, 13, 109–124.
131. Gatheral, J. (2006). *The volatility surface: a practitioner's guide*. Hoboken: Wiley.
132. Giles, M., & Glasserman, P. (2006). Smoking adjoints: Fast Monte Carlo Greeks. *Risk, January 2006*, 88–92.
133. Glasserman, P. (2004). *Monte Carlo methods in financial engineering*. Berlin: Springer.
134. Gobet, E. (2004). Revisiting the Greeks for European and American options. In J. Akahori, S. Ogawa, & S. Watanabe (Eds.), *Proceedings of the international symposium on stochastic processes and mathematical finance at Ritsumeikan University*, Kusatsu, Japan, March 2003 (pp. 53–71). Singapore: World Scientific.
135. Guyon, J., & Henry-Labordère, P. (2012). *Nonlinear pricing methods in quantitative finance*. London: Chapman & Hall.
136. Györfi, L., Kohler, M., Krzyzak, A., & Walk, H. (2002). *A distribution-free theory of non-parametric regression*. New York: Springer.
137. Hagan, P., Kumar, D., Lesniewski, A., & Woodward, D. (2002). Managing smile risk. *Wilmott, July 2002*, 84–108.
138. Hamadène, S. (2006). Mixed zero-sum differential game and American game options. *SIAM Journal on Control and Optimization*, 45(2), 496–518.
139. Hamadène, S., & Hassani, M. (2006). BSDEs with two reflecting barriers driven by a Brownian motion and an independent Poisson noise and related Dynkin game. *Electronic Journal of Probability*, 11, 121–145.
140. Hamadène, S., & Ouknine, Y. (2003). Reflected backward stochastic differential equation with jumps and random obstacle. *Electronic Journal of Probability*, 8, 1–20.
141. Harrison, J. M., & Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications*, 11(3), 215–260.

142. Henry-Labordère, P. (2012). Cutting CVA's complexity. *Risk, July 2012*, 67–73.
143. Herbertsson, A., & Rootzén, H. (2008). Pricing kth-to-default swaps under default contagion: the matrix-analytic approach. *Journal of Computational Finance*, 12(1), 49–78.
144. Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2), 327–343.
145. Hout, K. J., & Welfert, B. D. (2007). Stability of ADI schemes applied to convection-diffusion equations with mixed derivative terms. *Applied Numerical Mathematics*, 57, 19–35.
146. Hull, J. (2011). *Options, futures and other derivatives* (8th ed.). New York: Prentice Hall.
147. Hull, J., & White, A. (2000). Forward rate volatilities, swap rate volatilities, and the implementation of the Libor market model. *The Journal of Fixed Income*, 10(2), 46–62.
148. Ikeda, N., Nagasawa, M., & Watanabe, S. (1968). Branching Markov processes I (II). *Journal of Mathematics of Kyoto University*, 8, 233–278 (365–410).
149. Ikeda, N., & Watanabe, S. (1989). *Stochastic differential equations and diffusion processes* (2nd ed.). Amsterdam: North-Holland.
150. Ishii, H., & Koike, S. (1991). Viscosity solutions for monotone systems of second-order elliptic PDEs. *Communications in Partial Differential Equations*, 16(6–7), 1095–1128.
151. Itô, K. (1944). Stochastic integral. *Proceedings of the Imperial Academy*, 20, 519–524.
152. Jacod, J., & Protter, P. (2004). *Probability essentials* (corrected 2nd ed.). Berlin: Springer.
153. Jacod, J., & Shiryaev, A. (2003). *Limit theorems for stochastic processes*. Berlin: Springer.
154. Jaille, P., Lamberton, D., & Lapeyre, B. (1990). Variational inequalities and the pricing of American options. *Acta Applicandae Mathematicae*, 21, 263–289.
155. Jakobsen, E. R., & Karlsen, K. H. (2005). Continuous dependence estimates for viscosity solutions of integro-PDEs. *Journal of Differential Equations*, 212(2), 278–318.
156. Jakobsen, E. R., & Karlsen, K. H. (2006). A “maximum principle for semi-continuous functions” applicable to integro-partial differential equations. *Nonlinear Differential Equations and Applications*, 13, 137–165.
157. Jakobsen, E. R., Karlsen, K. H., & La Chioma, C. (2008). Error estimates for approximate solutions to Bellman equations associated with controlled jump-diffusions. *Numerische Mathematik*, 110(2), 221–255.
158. Jarrow, R. A., & Turnbull, S. (1995). Pricing derivatives on financial securities subject to credit risk. *Journal of Finance*, 50(1), 53–85.
159. Jeanblanc, M., Yor, M., & Chesney, M. (2010). *Mathematical methods for financial markets*. Berlin: Springer.
160. Kahl, C., & Jackel, P. (2005). Not-so-complex logarithms in the Heston model. *Wilmott, September 2005*, 94–103.
161. Kallsen, J., & Kühn, C. (2005). Convertible bonds: financial derivatives of game type. In A. Kyprianou, W. Schoutens, & P. Wilmott (Eds.), *Exotic option pricing and advanced Lévy models* (pp. 277–288). Chichester: Wiley.
162. Kamrad, B., & Ritchken, P. (1991). Multinomial approximating models for options with k state variables. *Management Science*, 37, 1640–1652.
163. Karatzas, I. (1988). On the pricing of American options. *Applied Mathematics & Optimization*, 17, 37–60.
164. Karatzas, I., & Shreve, S. (1988). *Brownian motion and stochastic calculus*. New York: Springer.
165. Kemna, G. Z., & Vorst, A. C. (1990). A pricing method for options based on average asset values. *Journal of Banking & Finance*, 14(1), 113–129.
166. Kifer, Y. (2000). Game options. *Finance and Stochastics*, 4, 443–463.
167. Kunita, H. (2010). Itô's stochastic calculus: its surprising power for applications. *Stochastic Processes and Their Applications*, 120, 622–652.
168. Kunita, H., & Watanabe, T. (1963). Notes on transformations of Markov processes connected with multiplicative functionals. *Memoirs of the Faculty of Science. Kyushu University. Series A. Mathematics*, 17, 181–191.

169. Kushner, H. (1977). *Probability methods for approximations in stochastic control and for elliptic equations*. San Diego: Academic Press.
170. Kushner, H., & Dupuis, B. (1992). *Numerical methods for stochastic control problems in continuous time*. Berlin: Springer.
171. Kusuoka, S. (1999). A remark on default risk models. *Advances in Mathematical Economics*, 1, 69–82.
172. Kwok, Y. (1998). *Mathematical models of financial derivatives*. Berlin: Springer.
173. Lagnado, R., & Osher, S. (1997). A technique for calibrating derivative security pricing models: numerical solution of an inverse problem. *Journal of Computational Finance*, 1(1), 13–25.
174. Lamberton, D., & Lapeyre, B. (1996). *Introduction to stochastic calculus applied to finance*. London: Chapman & Hall.
175. Lando, D. (1994). *Three essays on contingent claims pricing*. PhD Thesis, Cornell University.
176. Lapeyre, B., Pardoux, E., & Sentis, R. (2003). *Introduction to Monte-Carlo methods for transport and diffusion equations*. London: Oxford University Press.
177. Lapeyre, B., & Temam, E. (2001). Competitive Monte Carlo methods for the pricing of Asian options. *Journal of Computational Finance*, 5(1), 39–57.
178. Laurent, J.-P., Cousin, A., & Fermanian, J.-D. (2011). Hedging default risks of CDOs in Markovian contagion models. *Quantitative Finance*, 11(12), 1773–1791.
179. Laurent, J.-P., & Gregory, J. (2005). Basket default swaps, CDOs and factor copulas. *The Journal of Risk*, 7(4), 103–122.
180. Lawler, G. (2004). *Introduction to stochastic processes* (2nd ed.). London: Chapman & Hall.
181. L'Ecuyer, P. (1994). Uniform random number generation. *Annals of Operations Research*, 53, 77–120.
182. L'Ecuyer, P. (1998). Random number generation. In J. Banks (Ed.), *The handbook of simulation* (Chapter 4, pp. 93–137). New York: Wiley.
183. Le Gall, J.-F. (1991). Brownian excursions, trees and measure-valued branching processes. *Annals of Probability*, 19, 1399–1439.
184. Lemor, J.-P. (2005). *Approximation par projections et simulations de Monte-Carlo des équations différentielles stochastiques rétrogrades*. PhD Thesis, Ecole Polytechnique.
185. Lemor, J.-P., Gobet, E., & Warin, X. (2006). Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli*, 12(5), 889–916.
186. Lepeltier, J.-P., & Maingueneau, M. (1984). Le jeu de Dynkin en théorie générale sans l'hypothèse de Mokobodski. *Stochastics*, 13, 25–44.
187. Lewis, A. (2001). *A simple option formula for general jump-diffusion and other exponential Levy processes* (working paper).
188. Li, D. (2000). On default correlation: a copula function approach. *The Journal of Fixed Income*, 9(4), 43–54.
189. Lions, P.-L., & Régnier, H. (2001). *Calcul du prix et des sensibilités d'une option américaine par une méthode de Monte Carlo* (working paper).
190. Lipton, A. (1999). Similarities via self-similarities. *Risk, September 1999*, 101–105.
191. Lipton, A. (2001). *Mathematical methods for foreign exchange—a financial engineer's approach*. Singapore: World Scientific.
192. Lipton, A. (2002). The vol smile problem. *Risk, February 2002*, 61–65.
193. Lipton, A. (2002). Assets with jumps. *Risk, September 2002*, 149–153.
194. Lipton, A., & Mc Ghee, W. (2002). Universal barriers. *Risk, May 2002*, 81–85.
195. Longstaff, F. A., & Schwartz, E. S. (2001). Valuing American options by simulations: a simple least-squares approach. *The Review of Financial Studies*, 14(1), 113–147.
196. Ma, J., & Cvitanic, J. (2001). Reflected forward–backward SDEs and obstacle problems with boundary conditions. *Journal of Applied Mathematics and Stochastic Analysis*, 14(2), 113–138.

197. Ma, J., & Yong, J. (1999). *Lecture notes in mathematics: Vol. 1702. Forward–backward stochastic differential equations and their applications*. Berlin: Springer.
198. McKean, H. (1975). Application of Brownian motion to the equation of Kolmogorov–Petrovskii–Piskunov. *Communications on Pure and Applied Mathematics*, *XXVIII*, 323–331.
199. Marshall, A., & Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, *2*, 84–98.
200. Matache, A.-M., Von Petersdorff, T., & Schwab, C. (2004). Fast deterministic pricing of options on Lévy driven assets. *Mathematical Modelling and Analysis*, *38*(1), 37–72.
201. Mercurio, F. (2010). *Interest rates and the credit crunch: new formulas and market models*. Bloomberg Portfolio Research Paper 2010-01-FRONTIERS.
202. Merton, R. C. (1973). The theory of rational option pricing. *The Bell Journal of Economics and Management Science*, *4*, 141–183.
203. Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, *3*, 125–144.
204. Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, *44*, 335–341.
205. Mikosch, T. (1998). *Elementary stochastic calculus with finance in view*. Singapore: World Scientific.
206. Morokoff, W., & Caflish, R. (1994). Quasi-random sequences and their discrepancies. *SIAM Journal on Scientific Computing*, *5*(6), 1251–1279.
207. Morton, K., & Mayers, D. (1994). *Numerical solution of partial differential equations*. Cambridge: Cambridge University Press.
208. Musiela, M., & Rutkowski, M. (2004). *Martingale methods in financial modelling* (2nd ed.). Berlin: Springer.
209. Niederreiter, H. (1987). Points sets ans sequences with small discrepancy. *Monatshefte für Mathematik*, *104*, 273–337.
210. Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods*. Philadelphia: SIAM.
211. Nikeghbali, A. (2006). An essay on the general theory of stochastic processes. *Probability Surveys*, *3*, 345–412.
212. Nikeghbali, A., & Yor, M. (2005). A definition and some characteristic properties of pseudo-stopping times. *Annals of Probability*, *33*, 1804–1824.
213. Nocedal, J., & Wright, S. J. (2006). *Numerical optimization* (2nd ed.). New York: Springer.
214. Nualart, D. (1995). *The Malliavin calculus and related topics*. New York: Springer.
215. O’Kane, D., & Livesey, M. (2004). Base correlation explained. *Lehman brothers fixed income quantitative credit research*. Nov 2004.
216. Pallavicini, A., Perini, D., & Brigo, D. (2011). Funding valuation adjustment: a consistent framework including CVA, DVA, collateral, netting rules and re-hypothecation (working paper).
217. Palmowski, Z., & Rolski, T. (2002). A technique for exponential change of measure for Markov processes. *Bernoulli*, *8*(6), 767–785.
218. Pardoux, E., & Peng, S. (1990). Adapted solution of a backward stochastic differential equation. *Systems & Control Letters*, *14*, 55–61.
219. Pardoux, E., Pradeilles, F., & Rao, Z. (1997). Probabilistic interpretation of systems of semi-linear PDEs. *Annales de l’Institut Henri Poincaré, série Probabilités–Statistiques*, *33*, 467–490.
220. Peaceman, D., & Rachford, H. (1955). The numerical solution of parabolic and elliptic differential equations. *SIAM Journal on Applied Mathematics*, *3*, 28–42.
221. Peng, S. (1999). Monotonic limit theorem of BSDE and nonlinear decomposition theorem of Doob–Meyer’s type. *Probability Theory and Related Fields*, *113*(4), 473–499.
222. Peng, S., & Xu, M. (2005). The smallest g-supermartingale and reflected BSDE with single and double L_2 obstacles. *Annales de L’I.H.P. Probabilités Et Statistiques*, *41*(3), 605–630.
223. Pham, H. (1998). Optimal stopping of controlled jump-diffusion processes: a viscosity solution approach. *Journal of Mathematical Systems Estimation and Control*, *8*, 1–27.

224. Pham, H. (2010). *Continuous-time stochastic control and optimization with financial applications*. Berlin: Springer.
225. Phillips, D. (1962). A technique for the numerical solution of certain integral equations of the first kind. *Journal of the Association for Computing Machinery*, 9, 84–97.
226. Platen, E., & Bruti-Liberati, N. (2010). *Numerical solution of stochastic differential equations with jumps in finance*. Berlin: Springer.
227. Protter, P. E. (2001). A partial introduction to financial asset pricing theory. *Stochastic Processes and Their Applications*, 91(2), 169–203.
228. Protter, P. E. (2004). *Stochastic integration and differential equations* (2nd ed.). Berlin: Springer.
229. Rahal, A. (2011). *Simulation/regression pricing schemes for credit risk applications*. PhD Thesis, Université d'Evry Val d'Essonne.
230. Rebonato, R. (2004). *Volatility and correlation: the perfect hedger and the fox* (2nd ed.). Chichester: Wiley.
231. Rebonato, R., McKay, K., & White, R. (2009). *The SABR/Libor market model: pricing, calibration and hedging for complex interest-rate derivatives*. New York: Wiley.
232. Reisinger, C. (2012). Analysis of linear difference schemes in the sparse grid combination technique. *IMA Journal of Numerical Analysis*, first published online September 5, 2012.
233. Revuz, D., & Yor, M. (2001). *Continuous martingales and Brownian motion* (3rd ed.). Berlin: Springer.
234. Ritchken, P. (1995). On pricing barrier options. *The Journal of Derivatives*, 3(2), 19–28.
235. Rogers, L. (2002). Monte Carlo valuation of American options. *Mathematical Finance*, 12, 271–286.
236. Rogers, L., & Shi, Z. (1995). The value of an Asian option. *Journal of Applied Probability*, 32(4), 1077–1088.
237. Rosen, D., & Sanders, D. (2009). Valuing CDOs of bespoke portfolios with implied multi-factor models. *The Journal of Credit Risk*, 5(3), 3–36.
238. Royer, M. (1996). Backward stochastic differential equations with jumps and related nonlinear expectations. *Stochastic Processes and Their Applications*, 13(3), 293–317.
239. Saad, Y., & Schultz, M. (1986). GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear system. *SIAM Journal on Scientific Computing*, 7, 856–869.
240. Samperi, D. (2002). Calibrating a diffusion pricing model with uncertain volatility: regularization and stability. *Mathematical Finance*, 12(1), 71–87.
241. Schönbucher, P. (2003). *Credit derivatives pricing models: model, pricing and implementation*. Chichester: Wiley.
242. Schoutens, W. (2012). In *The handbook of hybrid securities: strategies, pricing and risk management*. Wiley (e-book).
243. Schweizer, M. (1991). Option hedging for semimartingales. *Stochastic Processes and Their Applications*, 37, 339–363.
244. Schweizer, M. (2001). A guided tour through quadratic hedging approaches. In E. Jouini, J. Cvitanic, & M. Musiela (Eds.), *Option pricing, interest rates and risk management* (pp. 538–574). Cambridge: Cambridge University Press.
245. Shreve, S. (2005). *Stochastic calculus for finance II: continuous-time models*. Berlin: Springer.
246. Sobol, I. M. (1976). The distribution of points in a cube and the approximate evaluation of integrals. *U.S.S.R. Computational Mathematics and Mathematical Physics*, 16, 236–242.
247. Soner, M., Touzi, N., & Zhang, J. (2012). Wellposedness of second order backward SDEs. *Probability Theory and Related Fields*, 153, 149–190.
248. Tavella, D., & Randall, C. (2000). *Financial instruments: the finite difference method*. New York: Wiley.
249. Tikhonov, A. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics Doklady*, 4, 1035–1038. (English translation of *Doklady Akademii Nauk SSSR*, 151, 501–504.)

250. Titchmarsh, E. C. (1997). *The theory of functions* (2nd ed.). London: Oxford University Press.
251. Tsitsiklis, J. N., & Van Roy, B. (2001). Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12, 694–703.
252. Vasicek, O. (1991). *Limiting loan loss probability distribution*. San Francisco: KMV Corporation.
253. Villeneuve, S., & Zanette, A. (2002). Parabolic ADI methods for pricing American option on two stocks. *Mathematics of Operations Research*, 27(1), 121–149.
254. Wilmott, P. (1998). *Derivatives: the theory and practice of financial engineering*. New York: Wiley.
255. Wilmott, P. (2002). Cliquet options and volatility models. *Wilmott, December 2002*, 78–83.
256. Wilmott, P., Dewynne, J., & Howison, S. (1993). *Option pricing*. Oxford: Oxford Financial Press.
257. Windcliff, H., Forsyth, P., & Vetzal, K. (2006). Numerical methods for valuing cliquet options. *Applied Mathematical Finance*, 13, 353–386.
258. Windcliff, H., Forsyth, P., & Vetzal, K. (2006). Pricing methods and hedging strategies for volatility derivatives. *Journal of Banking & Finance*, 30, 409–431.
259. Yang, J., Hurd, T., & Zhang, X. (2006). Saddlepoint approximation method for pricing CDOs. *Journal of Computational Finance*, 10(1), 1–20.
260. Zhu, Y., Wu, X., & Chern, I. (2004). *Derivative securities and difference methods*. Berlin: Springer.
261. Zvan, R., Forsyth, P. A., & Vetzal, K. R. (1998). Robust numerical methods for PDE models of Asian option. *Journal of Computational Finance*, 1(2), 39–78.

Index

A

- (A), 366
- A priori bound and error estimates, 331, 375
- Absolutely continuous, 71, 173
- ADI, 230, 231
- Admissible, 88
- Affine jump-diffusions, 128, 150
- American and game options, 84, 91, 92, 261, 272
- American Monte Carlo, 80, 195, 269
- Antithetic variables, 171
- Arbitrage, 84, 86, 88, 108
- Arbitrage price, 93, 101
- Asian options, 193, 235
- Azema supermartingale, 114

B

- Backward simulation, 183
- Backward stochastic differential equations (BSDEs), vii, x, 77, 83, 109, 321, 328
- Barles–Souganidis theorem, 218
- Barrier option, 192, 234, 436
- Base implied correlation, 146
- Basis functions, 320
- Bayes formula, 4, 10, 25, 75
- Bermudan options, 209
- BGM model, 133
- Bid-ask spread, 121, 246
- Bivariate Black–Scholes model, 179, 210, 230
- Black model, 131
- Black–Scholes formulas, 125
- Black–Scholes model, 43, 104, 123
- Black–Scholes equation, 124
- Bottom-up models, 297
- Bound and error estimates, 327
- Boundary condition, 218

Boundary condition in the weak viscosity sense, 418

- Box–Müller transformation, 167
- Branching Markov processes, 159
- Branching particles, 80, 194, 385, 387
- Breeden–Litzenberger, 248
- Brownian bridge, 69
- Brownian motion (BM), 33, 63, 423
- BSDEs stopped at a random time, 329
- BSDEs with random terminal time, 329
- Burkholder–Davis–Gundy inequality, 327, 331

C

- Càdlàg, 52, 86
- Calibration, viii, 121, 243, 307, 437
- Call protection, 93, 262, 343, 378
- Call/put, 91
- Cap, 134
- Caplet, 133
- Carr–Madan, 153
- Carré du champ, 60
- Cash flows, 87
- Cauchy cascade, 284, 371, 383
- Cauchy cascade of variational inequalities, 285
- CDO, 91, 138, 293, 305
- Central limit theorem, 170
- Change of measure, 323
- Change of variables formula for densities, 167
- Characteristic function, 127, 150, 151, 204
- Cholesky decomposition, 138, 169, 181, 182
- Classical solution, 214, 362
- Clean price, 369
- Cliquet options, 237, 239, 373
- Collateral, 310, 315
- Collateralized Debt Obligation (CDO), 139
- Common factor, 141
- Common random numbers, 177

- Comparison, 375
 Comparison principle, 214, 363
 Comparison theorem, 340
 Compensated jump-to-default process, 114
 Compensated measure, 324, 336
 Compensated sum of jumps, 59
 Compensator, 324
 Completeness, 90
 Compound implied correlation, 146
 Conditional expectation, 3, 115, 268
 Conditionally independent, 141
 Confidence interval, 161, 170, 175, 195
 Consistency, 217, 223, 224
 Constant elasticity of variance, 65
 Contingent convertible bonds (“CoCo”), 262
 Continuous Cauchy cascade, 383
 Continuous Euler scheme, 185
 Continuous-time Markov chains, 24, 104, 337
 Control variates, 172
 Convergence in law, 199, 207
 Convergence rate, 218
 Convertible bonds, 91, 92, 113, 261
 Correlation, 232
 Correlation matrix, 135
 Cost functions, 76, 339
 Cost process, 95, 106
 Counterparty risk, 92, 104, 111, 305, 338
 Covariance matrix, 36
 Cox–Ross–Rubinstein tree, 201
 Crank–Nicholson scheme, 223
 Credit default swap (CDS), 91, 139
 Credit derivatives, 92, 111, 138
 Credit ratings, 338
 Credit risk, 263
 Credit risk adjusted discount factor, 112, 114
 Credit Valuation Adjustment (CVA), 310
 Credit-risk-adjusted-interest-rate, 114
 Crisis, 107, 132, 138
 Cumulative dividend value process, 87
 Cumulative price, 87, 94
 Cure period, 311
 Curse of dimensionality, 80, 158, 194, 197,
 259, 287, 295
 CVA, vii, viii, 293
- D**
 Decoupled forward backward stochastic differential equation (decoupled FBSDE), 340
 Default contagion, 338
 Default intensity, 112
 Default protection, 91
 Defaultable derivatives, 111
 Defaultable game option, 112
- Delta, 223
 Delta-hedging, 146
 Density, 71
 Derivative, 90
 Deterministic jumps, 371
 Deterministic pricing equations, 83, 265, 274,
 281
 Deterministic schemes, 291
 Diagonal-dominant, 225
 Differentiation of the payoff, 176
 Differentiation of the transition probability density, 176
 Diffusion coefficient, 63
 Diffusions, 63, 104
 Dirichlet boundary condition, 221
 Discontinuous viscosity solutions comparison principle, 359
 Discount factor, 87, 108
 Discrepancy principle, 247
 Discrete coupons, 265
 Discrete dividends, 321, 369, 371
 Discrete FFT, 228
 Discrete Fourier transform, 155, 228
 Discrete path-dependence, 321, 369, 373
 Discrete time pricing models, 199
 Discretely monitored Asian options, 373
 Discretely monitored call protections, 120
 Discretely path-dependent options, 237
 Discretely sampled Asian options, 237, 240
 Dividend, 87
 Dividend process, 90
 Doob decomposition, 22
 Doob-martingale, 75
 Doob-Meyer decomposition, 52, 99, 331
 Doubling strategy, 16
 Doubly reflected backward stochastic differential equations (R2BSDEs), 115,
 262, 328
 Doubly reflected BSDE with an intermittent upper barrier, 330
 Down-and-out, 208
 Drift, 72, 74
 Drift coefficient, 63
 Duality, 256
 Dupire equation, 249
 Dupire formula, 250, 437
 Dynamic hedging, 433
 Dynamic programming, viii, 79, 195, 274
 Dynkin games, 94, 116, 266, 330, 359, 375,
 384
- E**
 (E.1), 360
 (E.2), 360

- (E.3), 360
 (\mathcal{E}) , 266
 \mathcal{E}_λ , 1
 Early exercise, viii, 84, 90, 213
 Effective dimension, 175
 Effective upper barrier, 330
 Effective volatility, 247, 250, 437
 Enlarged state space, 373
 Enlargement of the state space, 7
 Entropic regularization, 252
 Entropy, 255
 Equity tranche, 294
 Equity-to-credit, 113
 Equivalent, 71
 Essential supremum, 79
 Essinf, 94
 Essup, 94
 Estimation, 243
 Euler scheme, 184, 270
 European, 91
 European claims, 92
 Exercise boundary, 80
 Exercise region, 195
 Exit time, 40
 Exotic options, 121
 Explicit finite difference, 207
 Explicit scheme, 223
 Exponential distribution, 1, 30
 Exponential random variable, 166
- F**
 Factor process, 102, 109, 213, 291, 323
 Fast Fourier transform (FFT), 142, 153
 Feynman-Kac formula, 78
 Feynman-Kac representation, 78, 273
 Filtration, 85
 Financial derivative, 95
 Finite differences, 176, 213, 430
 Finite elements, 218
 First Fundamental Theorem of Asset Pricing, 90
 First passage time, 42
 Fokker-Planck equation, 249
 Forward simulation, 183
 Forward start option, 428
 Forward-neutral, 130
 Fourier transform, 128, 150
 Fully implicit finite difference schemes, 276
 Fully implicit scheme, 223
 Functional Itô calculus, 233, 373
 Fundamental Theorem of Asset Pricing, 89
 Funding, 76, 86, 307
- G**
 Gain, 97
 Galtchouk-Kunita-Watanabe decomposition, 107
 Game claim, 92
 Gamma, 223
 Gauss factorization, 224
 Gauss-Poisson approximations, 144
 Gaussian distribution, 1
 Gaussian pair, 167
 Gaussian process, 68
 Generator, 57, 60, 70, 103, 199, 221, 308, 336, 427
 Geometric Brownian motion, 43, 65
 Girsanov, 189
 Girsanov theorem, 72, 74, 135
 Girsanov transformations, 24, 58, 71, 173, 255
 Global regression, 277
 Gradient descent, 219, 247
 Greeks, 176, 213
 Gronwall lemma, 327, 331
- H**
 $(H.0)$, 328
 $(H.1)$, 328
 $(H.1)'$, 332
 $(H.2)$, 328
 $(H.2)'$, 334
 (\mathcal{H}) -hypothesis, 117
 Hazard intensity, 114
 Heat equation, 230
 Hedge, 100
 Hedging, 84, 95, 121
 Heston model, 104, 126, 186
 Historical probability measure, 85, 343, 433
 Homogeneous groups model, 296
 Homogeneous models, 153
 Hybrid scheme, 438
- I**
 Immersion, 114
 Immersion hypothesis, 114
 Implied correlation, 145, 243
 Implied volatility, 145, 157, 243
 Implied volatility surface, 250
 Importance sampling, 173
 Incompleteness, 102
 Infinite activity, 59, 104
 Integer-valued random measure, 324
 Integration by parts formulas, 62
 Intermittent call protection, 119, 265
 Inverse Gaussian cumulative distribution function, 166
 Inverse Gaussian density, 39

- Inverse problems, 243
 Inverse simulation method, 166
 Isometry property, 48, 50, 68
 Iterated logarithm, 165
 Iteration on the policies, 195, 270
 Iteration on the values, 195, 270
 Iterative scheme, 225
 Itô formula, 54, 75, 105, 110, 119, 249, 425
 Itô formula for a Poisson process, 57
 Itô formula for an Itô process, 58
 Itô process, 58, 69
 Itô stochastic integral, 424
- J**
 Jensen inequality, 13, 327
 Jensen-Ishii Lemma, 410
 Joint defaults, 315
 Jordan components of a finite variation process, 98
 Jump condition, 239, 285
 Jump intensity measure, 103
 Jump measures, 325
 Jump-diffusions with regimes, 337, 338
 Jump-diffusions, 103, 194, 201, 337
 Jump-to-Ruin model (JR), 434
- K**
 Kamrad–Ritchken tree, 206
 Koksma–Hlawka inequality, 175
 Kolmogorov equations, 11, 26, 32, 37, 295
 Kolmogorov ODEs, 319
 Kushner’s theorem, 200
- L**
 Lagrange multipliers, 256
 Laplace transform, 42, 142, 144
 Largest viscosity subsolution, 284, 286
 Latin Hypercube, 319
 Law of large numbers, 170
 Lax equivalence theorem, 217
 Lax–Milgram theorem, 219
 Lebesgue–Stieltjes, 324
 Lebesgue–Stieltjes integral, 51
 Lévy characterization theorem, 67
 Lévy measures, 337
 Lévy-driven models, 104
 Libor, 132
 Libor Market Model, 133
 Libor Swap Model, 136
 Linear system, 224, 273
 Liquidity, 86
 Local consistency conditions, 200
 Local intensity, 297
 Local martingale cost process, 84
 Local martingales, 75, 89, 111, 391
 Local volatility, 104, 247, 297, 437
 Localization, 216, 220
 Locally bounded, 87
 Lognormal density, 168
 Longstaff and Schwartz, 195, 270
 Lookback option, 190, 210, 233
 “*l* out of the last *d*” call protection, 262
 Low-discrepancy sequences, 162, 164
- M**
 (M), 360
 (M)', 372
 (M.2)', 372
 (M.3), 352
 (M.3)', 372
 Malliavin calculus, 105, 110, 119, 176, 196, 358
 Marked jump-diffusions, 377, 382
 Market imperfections, 343
 Markov, 36, 104
 Markov chain approximation, 432
 Markov chains, 7, 103, 199, 211, 273, 294, 307, 422
 Markov copula property, 308
 Markov property, 7, 10, 25, 37, 70, 75, 79, 124, 294, 358, 384
 Markov setup, 90, 213
 Markovian BSDE, 351
 Markovian change of probability measure, 348
 Markovian consistency conditions, 341
 Markovian copula common shock model, 307
 Markovian FBSDE, 359
 Markovian forward SDE, 358
 Markovian model, 307
 Markovian reflected backward SDEs, 358
 Markovian SDE, 69
 Markovian solution, 343, 359, 371
 Markovian stability estimates, 352
 Marsaglia method, 168
 Marshall–Olkin, 309, 311
 Martingale, 13, 24, 104, 421
 Martingale predictable representation property, 347, 350, 353
 Martingale problem, 337, 347, 350
 Martingale representation, 77
 Martingale representation property, 333
 Matrix exponentiation schemes, 295
 Matrix generator, 26, 159
 Maximum principle, 216
 MC backward estimate, 270, 276
 MC forward estimate, 270, 276, 280
 Mean-reversion, 126
 Mean-reverting square-root process, 65

- Mean-self-financing, 96
 Measure-stochastic integration, 59
 Merton model, 104, 127, 179, 226
 Method of cells, 196, 269, 277, 288, 302
 Milstein Scheme, 185
 Min-variance hedging, 298
 Minimality conditions, 79, 266, 376
 Minimax theorem, 256
 Mokobodski condition, 99, 334, 352, 353
 Moment generating function, 142
 Monotone, stable and consistent approximation scheme, 217
 Monotone operator, 214, 217
 Monte Carlo, viii, 161
 Monte Carlo loop, 171
 Moro's numerical approximation, 166
 Multilinear regression, 258, 316
 Multinomial tree, 207, 273
 Multiplicity, 159
- N**
 $\mathcal{N}(\mu, \Gamma)$, 1
 Neumann boundary condition, 222
 No Free Lunch with Vanishing Risk (NFLVR), 88
 Nominal dimension, 176
 Nonlinear, 78, 122
 Nonlinear Feynman-Kac formula, 78
 Nonlinear least squares problem, 244, 247
 Nonlinear minimization, 247, 252
 Nonlinear Monte Carlo, 196
 Nonlinear regression, 196, 268, 275, 286, 295
 Nonlinear regression in time, 319
 Nonlinear simulation pricing schemes, 194
 Nonlinear simulation/regression schemes, 159
 Nonparametric models, 247
 Numéraire, 84, 104, 108, 130, 236
 Numerical computation of conditional expectations, 196
 Numerical differentiation, 250
- O**
 Objective probability measure, 121
 Obstacle problems, 105, 194, 233, 360
 ODE schemes, 295
 ODEs, 105
 One time-step BSDE, 202
 One time-step reflected BSDE, 203
 One-factor Gaussian copula model, 138
 Optimal stopping, 330, 359
 Option, 95
 Optional sampling theorem, 19
 Optional stopping theorem (OST), 19, 422
 Order of consistency, 218
- Ornstein-Uhlenbeck process, 65, 68
- P**
 \mathbb{P} -price, 101
 \mathbb{P} -solution, 362, 365
 \mathbb{P} -viscosity solution, 362, 370
 \mathbb{P} -viscosity subsolution, 362, 370
 \mathbb{P} -viscosity supersolution, 362, 370
 \mathcal{P}_γ , 1
 Parallel computing, 159
 Partial differential equations (PDEs), vii
 Partial integro-differential equations (PIDEs), 75, 98, 213, 226, 321
 Partial integro-differential variational inequalities, 359
 Particle simulation schemes, 159, 320
 Path dependence, 103, 188, 262, 283
 Path-dependent call protection, 291
 Path-dependent options, 233
 Picard iteration, 84, 225
 PIDE cascade, 118
 PIDEs, 105
 Poisson distribution, 1
 Poisson process, 27, 187, 423
 Poisson stochastic integral, 424
 Portfolio credit risk, 293
 $PP(\lambda)$, 27, 74
 Pre-default, 118, 263
 Pre-default cumulative price, 117
 Pre-default values, 114
 Predictable, 88, 112, 324
 Predictable integrand, 51
 Predictable process, 23
 Predictable random function, 59
 Price process, 87
 Pricing and hedging, 83
 Pricing equations, 81
 Pricing function, 102, 120, 213, 267, 281, 358, 431
 Pricing measure, 121
 Primary strategy, 100
 Principal Component Analysis, 169
 Probability measure, 71
 Probit regression, 318
 Profit-and-loss, 97, 182
 Protection pricing function, 118
 Pseudo Monte Carlo, 161
 Pseudo-random generator, 162
 Pseudo-stopping time, 114
 Pure jump models, 293
- Q**
 Quadratic covariation, 61
 Quadratic variation, 61

- Quadrature, 144
 Quantization, 197
 Quasi Monte Carlo, 161, 175
 Quasi-left continuity, 334
 Quasi-random generator, 162, 184
 Quasi-random numbers, 164, 191
 Quasimartingales, 99, 331
- R**
 Radial limit, 237
 Radon-Nikodym density, 111, 135, 349
 Random walk, 8, 35, 206
 Random measure, 59
 Rao decomposition, 331
 RDBSDE, 329, 369
 Real world (also called objective, or statistical, or historical, or physical, or actuarial) probability measure, 85
 Realized covariance and variance, 61
 Recombining tree, 207
 Reduced-form, 111
 Reduced-form approach, 114
 Reference filtration, 114, 263
 Reflected BSDEs (RBSDEs), 79, 97, 194, 273, 328
 Reflecting process, 84, 100, 267
 Regime switching, 321, 323
 Regression, 196, 302
 Regression basis, 276
 Regularization, 244
 Regularization parameter, 246
 Rejection-acceptance method, 163
 Replication, 125, 299, 434, 439
 Representation property, 159, 297
 RIBSDE, 262, 330
 RIBSDEs, 369, 374
 Riccati equation, 129
 Risk-neutral, 73, 84
 Risk-neutral measure, 85, 86, 97
 Risk-neutral variance, 107
 Running supremum of the Brownian motion, 188
- S**
 SABR model, 157
 Saddle-point, 330
 Saddle-point methods, 144
 Savings account, 84, 86, 108
 SDE uniqueness, 347
 Second Fundamental Theorem of Asset Pricing, 90
 Self-financing, 88
 Self-financing property, 108
 Semigroup and Markov properties, 337
 Semigroup properties, 382
 Semilinear PIDE, 78
 Semimartingale cost processes, 97
 Semimartingale integration theory, 51
 Semimartingale Itô formula, 233
 Semimartingales, 52, 62, 86, 108
 Senior tranche, 294
 Sharp bracket, 60
 Sigma martingales, 87
 Simple process, 46
 Simulation error, 188
 Simulation regression scheme, 291
 Simulation schemes, 385, 387
 Simulation/regression, 196, 295
 Simulation/regression numerical schemes, 259
 Simulation/regression pricing schemes, 293
 Single-name credit risk, 114
 Skew, 145
 Sklar's theorem, 141
 Smile, 121, 145
 Smirk, 145
 Snell envelope, 94, 272
 Sobol sequence, 165
 Special semimartingale, 52, 99
 Special semimartingale form of the Itô formula, 60
 Special semimartingale Itô formula for a jump-diffusion, 70
 Spectral decomposition, 169
 Splines, 252
 Splitting, 229, 233
 Spread risk, 147
 Square brackets, 61
 Square integrable, 99
 Square integrable martingale, 107, 334
 Stability, 217, 432
 Stability, convergence and convergence rates, 252
 Stable, 223, 224
 Stable, monotone and consistent approximation schemes, 321, 359, 365, 368
 Standard Brownian motion (SBM), 34
 Static hedging, 258
 Sticky delta, 149
 Stochastic differential equations (SDEs), 63, 67, 326, 345
 Stochastic exponential, 68
 Stochastic flows, 176
 Stochastic grid, 285
 Stochastic integrals, 45

- Stochastic integrals with respect to a Poisson process, 51
 Stochastic pricing equations, 83, 99, 106, 265, 274
 Stochastic processes, 1, 85
 Stochastic volatility, 104, 126, 338, 429
 Stopped RBSDE, 329
 Stopping time, 17, 24
 Strong solution, 64
 Submartingale, 13, 24
 Successive over-relaxation scheme, 225
 Superhedge, 105, 122
 Superhedging, 203, 267
 Supermartingale, 13, 24, 203
 Survival indicator process, 112
 Swap market model, 136
 Swaptions, 134
- T**
 θ -schemes, 222
 Taylor expansions, 53, 206, 221, 273
 Terminal cash flows, 90
 Terminal measure, 134
 Thomas algorithm, 225, 232
 Tikhonov regularization, 245
 Time homogeneous Markov chain, 25
 Time-discretization, 184
 Time-discretization error, 188
 Top models, 297
 Totally inaccessible, 114
 Tower rule, 5
 Tracking error, 97, 433, 438
 Trading strategy, 88
 Transaction costs, 86, 102, 122
 Transition density function, 37
 Transition matrix, 10, 421
 Transition probabilities, 12
 Tree pricing schemes, 199
 Tridiagonal matrices, 224
 Tsitsiklis and VanRoy, 195, 270
- U**
 $(U.1)$, 363
 $(U.2)$, 363
 $(U.3)$, 363
 $\mathcal{U}_{\mathcal{D}}$, 1
 Uncertain volatility model, 158
 Uncertain Volatility Model, 237
 Uniform distribution, 1
 Uniform ellipticity, 356
 Uniformly elliptic, 381
 Uniqueness of a risk-neutral measure, 90
 Upwind discretization, 221
 Usual conditions, 52, 86
- V**
 $(V1)$, 360
 $(V2)$, 360
 Value function, 80, 359, 362, 365
 Value process, 330
 Variance and volatility swaps, 240
 Variance reduction, 171, 319
 Variational inequalities, 273, 283, 321, 359, 360, 384
 Verification principle, 266, 330, 342, 375
 Viscosity solution, 105, 109, 119, 215, 229, 273, 359, 361, 370
 Viscosity subsolution, 361, 370
 Viscosity supersolution, 361, 370
 Volatility and variance swaps, 237, 373
 Volatility of the volatility, 126
 Vulnerable put, 435
- W**
 Wald's martingale, 14, 40, 41
 Watanabe characterization of a Poisson process, 345
 Weak solution, 64, 67
 Wealth process, 88
 Weighted Monte Carlo, 254
 Wrong-way risk, 315