LECTURE NOTES IN COMPUTATIONAL SCIENCE AND ENGINEERING 102

Stephan Dahlke · Wolfgang Dahmen
Michael Griebel · Wolfgang Hackbusch
Klaus Ritter · Reinhold Schneider
Christoph Schwab · Harry Yserentant *Editors*

Extraction of Quantifiable Information from Complex Systems

Springer

Lecture Notes in Computational Science and Engineering

102

Editors:

More information about this series at
http://www.springer.com/series/3527

Stephan Dahlke • Wolfgang Dahmen •
Michael Griebel • Wolfgang Hackbusch •
Klaus Ritter • Reinhold Schneider •
Christoph Schwab • Harry Yserentant
Editors

Extraction of Quantifiable Information from Complex Systems

Springer

Editors
Stephan Dahlke
Fachbereich Mathematik und Informatik
Philipps-Universität Marburg
Marburg, Germany

Wolfgang Dahmen
Institut für Geometrie und Praktische Mathematik
RWTH Aachen
Aachen, Germany

Michael Griebel
Institut für Numerische Simulation
Universität Bonn
Bonn, Germany

Wolfgang Hackbusch
Max-Planck-Institut für Mathematik in den Naturwissenschaften
Leipzig, Germany

Klaus Ritter
Fachbereich Mathematik
Technische Universität Kaiserslautern
Kaiserslautern, Germany

Reinhold Schneider
Institut für Mathematik
Technische Universität Berlin
Berlin, Germany

Christoph Schwab
Seminar für Angewandte Mathematik
ETH Zürich
Zürich, Switzerland

Harry Yserentant
Institut für Mathematik
Technische Universität Berlin
Berlin, Germany

ISSN 1439-7358 ISSN 2197-7100 (electronic)
ISBN 978-3-319-08158-8 ISBN 978-3-319-08159-5 (eBook)
DOI 10.1007/978-3-319-08159-5
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014954584

Mathematics Subject Classification (2010): 35-XX, 41-XX, 60-XX, 65-XX

Printed on acid-free paper

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

Foreword

In April 2007, the Deutsche Forschungsgemeinschaft (DFG) approved the Priority Program 1324 "Mathematical Methods for Extracting Quantifiable Information from Complex Systems". The objective of this volume is to offer a comprehensive overview of the scientific highlights obtained in the course of this priority program.

Mathematical models of complex systems are gaining rapidly increasing importance in driving fundamental developments in various fields such as science and engineering at large but also in new areas such as computational finance. Ever-increasing hardware capacities and computing power encourage and foster the development of more and more realistic models. On the other hand, the necessarily growing complexity of such models keeps posing serious and even bigger challenges to their numerical treatment.

Principal obstructions such as the *curse of dimensionality* suggest that a proper response to these challenges cannot be based solely on further increasing computing power. Instead, recent developments in mathematical sciences indicate that significant progress can only be achieved by contriving novel and much more powerful numerical solution strategies by systematically exploiting synergies and conceptual interconnections between the various relevant research areas. Needless to stress that this requires a deeper understanding of the mathematical foundations as well as exploring new and efficient algorithmic concepts. Fostering such well-balanced developments has been a central objective of this priority program.

The understanding and numerical treatment of spatially high-dimensional systems is clearly one of the most challenging tasks in applied mathematics. The problem of spatial high dimensionality is encountered in numerous application contexts such as machine learning, design of experiments, parameter-dependent models and their optimization, mathematical finance, PDEs in high-dimensional phase space, to name only a few, which already reflect the conceptual breadth. It is this seeming variability that makes a substantial impact of better exploiting conceptual and methodological synergies conceivable and in fact likely. It seems that to be really successful, theoretical research and practical applications should go hand in hand. In fact, this volume reflects an attempt to realize a proper balance between research with a primary methodological focus and challenging concrete application areas, although these two regimes can, of course, not be strictly separated. To that end, it has appeared to be necessary to combine different fields of mathematics such as numerical analysis and computational stochastics. On the other hand, to keep the whole programme sufficiently focused, it seemed advisable to concentrate on specific but related fields of application that share some common characteristics that allow one to benefit from conceptual similarities.

On the methodological side, several important new numerical approximation methods have been developed and/or further investigated in the course of the priority program. First of all, as one of the central techniques, let us mention tensor approximations. New tensor formats have been developed, and efficient tensor approximation schemes for various applications, e.g. in quantum dynamics and computational finance, have been studied; see Chaps. 2, 10, 12, 16 and 19. Adaptive strategies with all their facets have been employed in most of the projects; see, e.g., Chaps. 2, 4, 5, 9, 10, 14 and 16. Closely related with adaptivity is of course the concept of sparsity/compressed sensing; see Chaps. 14 and 18. As further techniques, sparse grids (Chap. 9), ANOVA decompositions (Chap. 11) and Fourier methods (Chap. 17) have been investigated. As a quite new technique, the reduced basis methods also came into play (see Chap. 2), in particular in the second period of SPP 1324. Of course, tensor methods as well as model order reduction concepts such as the reduced basis method address spatially high-dimensional problems. Both paradigms use the separation of variables as the central means to reduce computational complexity. Moreover, they can be viewed as trying to exploit sparsity by determining specific problem- and solution-dependent dictionaries that are able to approximate the searched object by possibly few terms. Moreover, Chaps. 1, 6 and 20 are concerned with Monte Carlo and Multilevel Monte Carlo methods in the context of stochastic applications.

One of the major themes within SPP 1324 has been high-dimensional problems in physics. Chapter 21 is concerned with the regularity of the solution to the electronic Schrödinger equation. Chapter 19 studies problems in quantum dynamics, the chemical master equation is one of the topics in Chap. 15, and Chap. 11 is concerned with electronic structure problems. Another very important issue within SPP 1324 has been differential equations with random or parameter-dependent coefficients and their various applications. The theory and numerical treatment of these problems are discussed in Chaps. 2 and 7. Closely related with this topic are stochastic differential equations and stochastic partial differential equations. The adaptive numerical treatment of SPDEs is studied in Chap. 5. SDEs with their various applications such as stochastic filtering are discussed in Chaps. 1, 6 and 8. Additional fields of application have been computational finance (see Chap. 16) and inverse problems (see Chaps. 3 and 18).

Overall, the network of SPP 1324 comprised more than 60 scientists, and 20 projects were funded in two periods. Up to now, more than 170 papers have been published by the participants of SPP 1324. The aim of this volume is of course not to give a complete presentation of all these results but rather to collect the scientific highlights in order to demonstrate the impact of SPP 1324 on further researches. The editors and authors hope that this volume will arouse interest in the reader in the various new mathematical concepts and numerical algorithms that have been developed in the priority program. For further information concerning SPP 1324, please visit http://www.dfg-spp1324.de/.

Marburg, Germany	Stephan Dahlke
Aachen, Germany	Wolfgang Dahmen
Bonn, Germany	Michael Griebel
Leipzig, Germany	Wolfgang Hackbusch
Kaiserslautern, Germany	Klaus Ritter
Berlin, Germany	Reinhold Schneider
Zürich, Switzerland	Christoph Schwab
Berlin, Germany	Harry Yserentant
June 2014	

Acknowledgements

First of all, the authors and the editors of this book thank the Deutsche Forschungsgemeinschaft (DFG) for the support within the DFG Priority Program 1324 “Extraction of Quantifiable Information from Complex Systems”. Moreover, we thank the referees Folkmar Bornemann, Joachim Buhmann, Hans Georg Feichtinger, Ursula Gather, Markus Hegland, Des Higham, George Karniadakis, Claudia Klüppelberg, Stig Larsson, Claude Le Bris, Gilles Pagès, Otmar Scherzer, Ian H. Sloan and Endre Süli for their hard work and conscientious refereeing. We also feel very grateful to our SPP fellows Kyeong-Hun Kim, Mihály Kovács, Stig Larsson, Kijung Lee, Raul Tempone and Henryk Woźniakowski, who contributed a lot to the success of SPP 1324. Special thanks are devoted to the DFG representatives Frank Kiefer and Carsten Balleier for the very productive cooperation. Last but not least, we thank Frank Eckhardt for assistance and support during the production of this book.

Contents

1 Solving Stochastic Dynamic Programs by Convex Optimization and Simulation 1
Denis Belomestny, Christian Bender, Fabian Dickmann, and Nikolaus Schweizer
1.1 Introduction 1
1.2 The Primal-Dual Approach to Convex Dynamic Programs 3
1.3 Construction of Lower Bounds via Martingale Basis Functions ... 8
1.4 Construction of Upper Bounds: Multilevel Monte Carlo and Sieve Optimization 11
1.5 Numerical Experiments 17
References 22

2 Efficient Resolution of Anisotropic Structures 25
Wolfgang Dahmen, Chunyan Huang, Gitta Kutyniok, Wang-Q Lim, Christoph Schwab, and Gerrit Welper
2.1 Introduction 26
2.2 Anisotropic Approximations 28
2.3 Well-Conditioned Stable Variational Formulations 30
2.4 Reduced Basis Methods 41
2.5 Sparse Tensor Approximation for Radiative Transfer 46
References 50

3 Regularity of the Parameter-to-State Map of a Parabolic Partial Differential Equation 53
Rudolf Ressel, Patrick Dülk, Stephan Dahlke, Kamil S. Kazimierski, and Peter Maass
3.1 Introduction 53
3.2 Function Spaces 55

3.3 The Model PDE as an Evolution Equation 56
3.4 The Parameter-to-State Map 61
References 67

4 Piecewise Tensor Product Wavelet Bases by Extensions and Approximation Rates 69
Nabi G. Chegini, Stephan Dahlke, Ulrich Friedrich, and Rob Stevenson
4.1 Introduction 69
4.2 Approximation by Tensor Product Wavelets on the Hypercube ... 71
4.3 Construction of Riesz Bases by Extension 73
4.4 Approximation by – Piecewise – Tensor Product Wavelets 76
4.5 Numerical Results 79
References 80

5 Adaptive Wavelet Methods for SPDEs 83
Petru A. Cioica, Stephan Dahlke, Nicolas Döhring, Stefan Kinzel, Felix Lindner, Thorsten Raasch, Klaus Ritter, and René L. Schilling
5.1 Introduction 84
5.2 Preliminaries 85
5.3 Regularity Analysis in Besov Spaces 91
5.4 Nonlinear Approximation for Elliptic Equations 97
References 106

6 Constructive Quantization and Multilevel Algorithms for Quadrature of Stochastic Differential Equations 109
Martin Altmayer, Steffen Dereich, Sangmeng Li, Thomas Müller-Gronbach, Andreas Neuenkirch, Klaus Ritter, and Larisa Yaroslavtseva
6.1 Introduction 110
6.2 Constructive Quantization of Systems of SDEs 111
6.3 Multilevel Methods for Discontinuous Payoffs in the Generalized Heston Model 118
6.4 Multilevel Methods for Lévy-Driven SDEs 124
References 130

7 Bayesian Inverse Problems and Kalman Filters 133
Oliver G. Ernst, Björn Sprungk, and Hans-Jörg Starkloff
7.1 Introduction 133
7.2 Bayesian Approach to Inverse Problems 135
7.3 Analysis of Kalman Filters for Bayesian Inverse Problems 146
7.4 Numerical Example: 1D Elliptic Boundary Value Problem 152
7.5 Conclusions 156
References 157

8 Robustness in Stochastic Filtering and Maximum Likelihood Estimation for SDEs 161
Joscha Diehl, Peter K. Friz, Hilmar Mai, Harald Oberhauser, Sebastian Riedel, and Wilhelm Stannat
8.1 Introduction 162
8.2 Robustness of the Stochastic Filter 165
8.3 Maximum Likelihood Estimation for SDEs 175
8.4 Practical Implications 177
References 177

9 Adaptive Sparse Grids in Reinforcement Learning 179
Jochen Garcke and Irene Klompmaker
9.1 Introduction 179
9.2 Reinforcement Learning 181
9.3 Planning with Sparse Grids 184
9.4 Sparse Grid Based Scheme for Reinforcement Learning 190
9.5 Experiments 191
9.6 Conclusion 192
References 193

10 A Review on Adaptive Low-Rank Approximation Techniques in the Hierarchical Tensor Format 195
Jonas Ballani, Lars Grasedyck, and Melanie Kluge
10.1 Introduction 195
10.2 Low-Rank Tensor Representations 197
10.3 Adaptive Tensor Sampling 202
10.4 Non-adaptive Tensor Sampling 206
10.5 Tensor Completion 207
References 209

11 A Bond Order Dissection ANOVA Approach for Efficient Electronic Structure Calculations 211
Michael Griebel, Jan Hamaekers, and Frederik Heber
11.1 Introduction 212
11.2 Schrödinger Equation in the Born-Oppenheimer Approximation 215
11.3 ANOVA Decomposition Scheme 216
11.4 Numerical Results 225
Concluding Remarks 232
References 233

12 Tensor Spaces and Hierarchical Tensor Representations 237
Wolfgang Hackbusch and Reinhold Schneider
12.1 Introduction 237
12.2 Subspace Approximation and Tucker Format 240
12.3 Hierarchical Tensor Representations 243
12.4 Hierarchical Tensors as Differentiable Manifolds 248

12.5 Numerical Methods 251
12.6 Tensorisation 256
References 259

13 Nonlinear Eigenproblems in Data Analysis: Balanced Graph Cuts and the RatioDCA-Prox 263
Leonardo Jost, Simon Setzer, and Matthias Hein
13.1 Introduction 263
13.2 Exact Relaxation of Balanced Graph Cuts 264
13.3 Minimization of Ratios of Non-negative Differences of Convex Functions via the RatioDCA-Prox 265
13.4 The RatioDCA-Prox for Ratios of Lovasz Extensions: Application to Balanced Graph Cuts 273
13.5 Experiments 277
References 278

14 Adaptive Approximation Algorithms for Sparse Data Representation 281
Mijail Guillemard, Dennis Heinen, Armin Iske, Sara Krause-Solberg, and Gerlind Plonka
14.1 Introduction 281
14.2 The Easy Path Wavelet Transform 282
14.3 Dimensionality Reduction on High-Dimensional Signal Data 292
14.4 Audio Signal Separation and Signal Detection 295
References 301

15 Error Bound for Hybrid Models of Two-Scaled Stochastic Reaction Systems 303
Tobias Jahnke and Vikram Sunkara
15.1 Introduction 303
15.2 The Chemical Master Equation of Two-scale Reaction Systems 304
15.3 Model Reduction Based on Conditional Expectations 307
15.4 Error Analysis for the Hybrid Model 309
References 318

16 Valuation of Structured Financial Products by Adaptive Multiwavelet Methods in High Dimensions 321
Rüdiger Kiesel, Andreas Rupp, and Karsten Urban
16.1 Introduction 321
16.2 Variational Formulation 324
16.3 Multiwavelets 326
16.4 Discretization 328
16.5 The Hierarchical Tucker Format (HTF) 330
16.6 Numerical Experiments 335
16.7 The Kronecker Product 342
References 343

17 Computational Methods for the Fourier Analysis of Sparse High-Dimensional Functions ... 347
Lutz Kämmerer, Stefan Kunis, Ines Melzer, Daniel Potts, and Toni Volkmer
17.1 Introduction ... 347
17.2 Evaluation of Multivariate Trigonometric Polynomials ... 348
17.3 Reconstruction Using Multivariate Trigonometric Polynomials ... 352
References ... 360

18 Sparsity and Compressed Sensing in Inverse Problems ... 365
Evelyn Herrholz, Dirk Lorenz, Gerd Teschke, and Dennis Trede
18.1 Introduction ... 365
18.2 Exact Recovery for Ill-Posed Problems ... 367
18.3 Compressive Sensing Principles for Ill-Posed Problems ... 372
References ... 377

19 Low-Rank Dynamics ... 381
Christian Lubich
19.1 Introduction ... 381
19.2 Projecting onto the Tangent Space: The Dirac–Frenkel Time-Dependent Variational Approximation Principle ... 382
19.3 Dynamical Low-Rank Approximation of Matrices ... 384
19.4 A Projector-Splitting Integrator for Dynamical Low-Rank Approximation ... 387
19.5 Dynamical Low-Rank Approximation of Tensors ... 389
19.6 The MCTDH Method for Quantum Dynamics ... 392
19.7 Low-Rank Differential Equations for Structured Matrix Nearness Problems ... 394
References ... 395

20 Computation of Expectations by Markov Chain Monte Carlo Methods ... 397
Erich Novak and Daniel Rudolf
20.1 Introduction ... 397
20.2 Approximation of Expectations by MCMC ... 398
20.3 Application of the Error Bound and Limitations of MCMC ... 404
20.4 Open Problems and Related Comments ... 410
References ... 410

21 Regularity, Complexity, and Approximability of Electronic Wavefunctions ... 413
Harry Yserentant
21.1 Introduction ... 413
21.2 The Variational Form of the Equation ... 415
21.3 The Mixed Regularity of the Wavefunctions ... 416
21.4 The Transcorrelated Formulation and the Regularity Proof ... 418

21.5 The Radial-Angular Decomposition 420
21.6 Sparse Grids, Hyperbolic Cross Spaces, and Antisymmetry 421
21.7 Eigenfunction and Wavelet Expansions 427
References 428

Index 429

List of Contributors

Martin Altmayer University of Mannheim, Germany

Jonas Ballani RWTH Aachen, Germany

Denis Belomestny University of Duisburg-Essen, Germany

Christian Bender University of Saarland, Saarbrücken, Germany

Nabi Godarzvand Chegini University of Amsterdam, The Netherlands

Petru A. Cioica Philipps-University of Marburg, Germany

Stephan Dahlke Philipps-University of Marburg, Germany

Wolfgang Dahmen RWTH Aachen, Germany

Steffen Dereich WWU Münster, Germany

Fabian Dickmann University of Duisburg-Essen, Germany

Joscha Diehl Technical University of Berlin, Germany

Nicolas Döhring Technical University of Kaiserslautern, Germany

Patrick Dülk University of Bremen, Germany

Oliver G. Ernst Technical University of Chemnitz, Germany

Ulrich Friedrich Philipps-University of Marburg, Germany

Peter K. Friz Weierstrass Institute, Technical University of Berlin, Germany

Jochen Garcke University of Bonn, Germany

Lars Grasedyck RWTH Aachen, Germany

Michael Griebel University of Bonn, Germany

Mijail Guillemard Technical University of Berlin, Germany

Wolfgang Hackbusch MPI Mathematik in den Naturwiss., Leipzig, Germany

Jan Hamaekers University of Bonn, Germany

Frederik Heber University of Bonn, Germany

Matthias Hein University of Saarland, Saarbrücken, Germany

Dennis Heinen University of Göttingen, Germany

Evelyn Herrholz Neubrandenburg University of Applied Sciences, Neubrandenburg, Germany

Chunyan Huang School of Applied Mathematics, Central University of Finance and Economics, Beijing, P.R.China

Armin Iske University of Hamburg, Germany

Tobias Jahnke Karlsruhe Institute of Technology, Karlsruhe, Germany

Leonardo Jost University of Saarland, Saarbrücken, Germany

Lutz Kämmerer Technical University of Chemnitz, Germany

Kamil S. Kazimierski University of Graz, Austria

Rüdiger Kiesel University of Duisburg-Essen, Germany

Stefan Kinzel Philipps-University of Marburg, Germany

Irene Klompmaker Technical University of Berlin, Germany

Melanie Kluge RWTH Aachen, Germany

Sara Krause-Solberg University of Hamburg, Germany

Stefan Kunis University of Osnabrück, Germany

Gitta Kutyniok Technical University of Berlin, Germany

Sangmeng Li WWU Münster, Germany

Wang-Q Lim Technical University of Berlin, Germany

Felix Lindner Technical University of Kaiserslautern, Germany

Dirk Lorenz Technical University of Braunschweig, Germany

Christian Lubich University of Tübingen, Germany

Peter Maass University of Bremen, Germany

Hilmar Mai Weierstrass Institute, Berlin, Germany

Ines Melzer University of Osnabrück, Germany

Thomas Müller-Gronbach University of Passau, Germany

Andreas Neuenkirch University of Mannheim, Germany

Erich Novak Friedrich Schiller University of Jena, Germany

Harald Oberhauser University of Oxford, UK

Gerlind Plonka-Hoch University of Göttingen, Germany

Daniel Potts Technical University of Chemnitz, Germany

Thorsten Raasch University of Mainz, Germany

Rudolf Ressel DLR Oberpfaffenhofen, EOC, Wessling, Germany

Sebastian Riedel Technical University of Berlin, Germany

Klaus Ritter Technical University of Kaiserslautern, Germany

Daniel Rudolf Friedrich Schiller University of Jena, Germany

Andreas Rupp University of Ulm, Germany

René L. Schilling Technical University of Dresden, Germany

Reinhold Schneider Technical University of Berlin, Germany

Christoph Schwab ETH Zürich, Switzerland

Nikolaus Schweizer University of Saarland, Saarbrücken, Germany

Simon Setzer University of Saarland, Saarbrücken, Germany

Björn Sprungk Technical University of Chemnitz, Germany

Wilhelm Stannat Technical University of Berlin, Germany

Hans-Jörg Starkloff University of Applied Sciences Zwickau, Germany

Rob Stevenson University of Amsterdam, The Netherlands

Vikram Sunkara Karlsruhe Institute of Technology, Karlsruhe, Germany

Gerd Teschke Neubrandenburg University of Applied Sciences, Neubrandenburg, Germany

Dennis Trede University of Bremen, Germany

Karsten Urban University of Ulm, Germany

Toni Volkmer Technical University of Chemnitz, Germany

Gerrit Welper Texas A & M University, College Station, TX, USA

Larisa Yaroslavtseva University of Passau, Germany

Harry Yserentant Technical University of Berlin, Germany

Chapter 1
Solving Stochastic Dynamic Programs by Convex Optimization and Simulation

Denis Belomestny, Christian Bender, Fabian Dickmann, and Nikolaus Schweizer

Abstract In this chapter we review some recent progress on Monte Carlo methods for a class of stochastic dynamic programming equations, which accommodates optimal stopping problems and time discretization schemes for backward stochastic differential equations with convex generators. We first provide a primal maximization problem and a dual minimization problem, based on which confidence intervals for the value of the dynamic program can be constructed by Monte Carlo simulation. For the computation of the lower confidence bounds we apply martingale basis functions within a least-squares Monte Carlo implementation. For the upper confidence bounds we suggest a multilevel simulation within a nested Monte Carlo approach and, alternatively, a generic sieve optimization approach with a variance penalty term.

1.1 Introduction

In this chapter we review some recent progress on Monte Carlo methods for dynamic programming equations of the form

$$Y_j^* = F_j(E_j[\beta_{j+1} Y_{j+1}^*]), \ j = 0, \ldots, J-1, \quad Y_J^* = F_J(0) \tag{1.1}$$

on a complete filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_j)_{j=0,\ldots J}, P)$ in discrete time. In this equation an adapted $\mathbb{R}^{D+1}$-valued process β and the adapted random field $F : \{0, \ldots, J\} \times \Omega \times \mathbb{R}^{D+1} \to \mathbb{R}$ are given. Moreover, $E_j[\cdot]$ denotes the conditional expectation given $\mathscr{F}_j$. Assumptions on β and F will be specified later on.

D. Belomestny • F. Dickmann
Universität Duisburg-Essen, Thea-Leymann-Str. 9, 45127 Essen, Germany
e-mail: denis.belomestny@uni-due.de; fabian.dickmann@uni-due.de

C. Bender (✉) • N. Schweizer
Universität des Saarlandes, Postfach 151150, 66041 Saarbrücken, Germany
e-mail: bender@math.uni-sb.de; schweizer@math.uni-sb.de

S. Dahlke et al. (eds.), *Extraction of Quantifiable Information from Complex Systems*, Lecture Notes in Computational Science and Engineering 102, DOI 10.1007/978-3-319-08159-5_1

Several time discretization schemes for backward stochastic differential equations (BSDEs) with or without reflection and for fully nonlinear second order parabolic PDEs lead to dynamic programs of the form (1.1), see [10,11,15,20,28]. In financial engineering, equations of the form (1.1) appear (after a time discretization is performed) in many nonlinear option pricing problems. These include:

- *Bermudan option pricing*: Here $\beta \equiv 1$ and $F_j(y) = \max\{S_j, y\}$, where the adapted process S_j denotes the discounted payoff of the Bermudan option, when called at the jth exercise time. Then, Y_0^* is the price of the Bermudan option (in discounted units), see e.g. [25].
- *Credit value adjustment*: Here $\beta \equiv 1$, $F_j(y) = (1 - r\Delta)y - (1 - R)\lambda\Delta(y)_+$ for $j < J$, where $r \geq 0$ is the risk-free interest rate, $\lambda > 0$ is the default intensity of the counterparty, $R \in [0, 1)$ is the recovery rate in case of default, and $(\cdot)_+$ denotes the positive part. The random variable $F_J(0)$ represents the payoff of the option at maturity T, if there is no default prior to maturity, and the interval $[0, T]$ is divided into J equidistant subintervals of length Δ. Then, Y_j^* is the price of the option at time $j\Delta$ including credit value adjustment (in a reduced form approach), provided that default did not occur prior to $j\Delta$. See e.g. [13,14] for BSDE approaches to pricing under credit risk.
- *Funding costs:* We now assume that funding costs are incorporated in the valuation mechanism, when at time $j\Delta$ the hedging costs for the delta hedge in the risky stocks $X_j^1, \ldots, X_j^D$ exceeds the price of the option with payoff $F_J(0)$ (at maturity T). In this case $F_j(y_0, \ldots, y_D) = (1-r\Delta)y_0 - R(\sum_{d=1}^D y_d - y_0)_+\Delta$ for $j < J$, where r is the interest rate, at which money can be lent, and $(R+r)$ is the rate, at which money can be borrowed. This is a classical example of nonlinear option pricing by BSDEs, for which we refer to the survey paper [19]. The variable y_0 represents the price of the option and the variables y_d, $d = 1, \ldots, D$, describe the amount of money required for the delta hedge in the dth stock. Correspondingly one chooses $\beta_j^0 = 1$ and β_j^d as (a suitable approximation of) $X_{j-1}^d(X_j^d - E_{j-1}[X_j^d])/E_{j-1}[(X_j^d - E_{j-1}[X_j^d])^2]$.

The main difficulty when solving equations of the form (1.1) numerically is that, going backwards in time, in each time step a conditional expectation must be approximated which depends on the numerical approximation of Y^* one time step ahead. Therefore one needs to apply an approximate operator for the conditional expectation which can be nested without exploding costs. In particular, when the generator F depends on ω through a high-dimensional Markovian process, Monte Carlo methods are usually applied to estimate the conditional expectations. In this respect, the least-squares Monte Carlo method, which was suggested for Bermudan option pricing by [21,26] and for BSDEs by [20], is certainly among the most popular choices. For the Bermudan option pricing problem this approximate dynamic programming approach (i.e. solving the dynamic program with the conditional expectation replaced by an approximate operator) is often complemented with the primal-dual methodology of [1]. In a nutshell, the solution of the approximate dynamic program is taken as an input in order to construct confidence intervals for the price Y_0^* of the option. This approach crucially relies on the dual representation of [18, 23] for Bermudan option pricing.

In Sect. 1.2 we first provide a review of this primal-dual approach for Bermudan option pricing. Following the lines of [7] we then generalize the theory behind this approach to dynamic programs of the form (1.1) under the assumptions that the driver F is convex and that a discrete comparison principle holds. The remaining sections are devoted to making this general primal-dual approach practical by designing and analyzing algorithms, which improve on the existing literature in various aspects. In Sect. 1.3 we suggest to run the least-squares Monte Carlo method for the approximate dynamic program with a set of basis functions which satisfy a martingale property. While this corresponds to the 'regression later' approach of [17] for the Bermudan option problem, it was recently observed by [8] that the use of martingale basis functions can significantly reduce the propagation of the projection error over time and the variance in the context of time discretization schemes for BSDEs.

Given the corresponding approximate solution of the dynamic program (1.1), the construction of a lower confidence bound for Y_0^* is usually a straightforward application of the primal-dual methodology. Contrarily, the construction of the upper bound requires a martingale as input, which should be close to the Doob martingale of βY^*. In the context of Bermudan option pricing, Andersen and Broadie [1] suggested a method to approximate this martingale starting from the solution of the approximate dynamic program and applying one layer of nested simulation in order to compute the Doob decomposition numerically. Based on [5] we present in Sect. 1.4.1 a multilevel variant of this algorithm, where varying numbers of paths are applied for the two layers of simulations at different levels. This multilevel variant can be shown to reduce the complexity of the Andersen-Broadie algorithm from ε^{-4} (generic nested Monte Carlo) to $\varepsilon^{-2}\log^2(\varepsilon)$, which up to the logarithmic factor is the same complexity as a plain non-nested Monte Carlo implementation. As an alternative to the Andersen-Broadie type algorithms we also present a completely generic approach to the approximation of the Doob martingale of βY^* via sieve optimization combined with a variance penalty term in Sect. 1.4.2. Convergence of this algorithm was analyzed in [3] for the Bermudan option pricing problem as the number of martingales in the sieve and the number of simulated samples converges to infinity. Finally, we illustrate the proposed algorithms by numerical experiments in the context of nonlinear expectations under model uncertainty and of option pricing under credit value adjustment.

1.2 The Primal-Dual Approach to Convex Dynamic Programs

In this section we first recall how the primal-dual approach works for the Bermudan option pricing problem. Then we present a generalization to dynamic programs of the form (1.1) with convex generator.

As stated in the introduction, the Bermudan option pricing problem leads to a dynamic program of the form

$$Y_j^* = \max\{S_j, E_j[Y_{j+1}^*]\}, \quad Y_J^* = S_J \tag{1.2}$$

for some adapted and integrable process S with $S_J \geq 0$. The starting point of the primal-dual approach is the well-known observation that this dynamic program is the one associated to the *optimal stopping problem* (primal problem), i.e.

$$Y_0^* = \sup_{\tau \in \mathscr{S}} E[S_\tau], \tag{1.3}$$

where $\mathscr{S}$ is the set of stopping times with values greater than or equal to j, and the (smallest) optimal stopping time τ^* can be expressed as

$$\tau^* = \inf\{i \geq 0;\ S_i \geq E_i[Y_{i+1}^*]\}.$$

Hence, for any stopping time τ, $Y_0^{low} := E[S_\tau]$ yields a lower bound for the Bermudan option price Y_0^*. In practice, a 'close-to-optimal' stopping time τ is often constructed as follows: One first rephrases the dynamic program in terms of the continuation value $Z_j^* := E_j[Y_{j+1}^*]$ as

$$Z_j^* = E_j[\max\{S_{j+1}, Z_{j+1}^*\}], \quad Z_J^* = 0.$$

Then, one solves this dynamic program numerically, replacing the conditional expectation by some approximate operator, which leads to an approximation Z of Z^*. Finally, based on Z one constructs the lower bound Y_0^{low} via the stopping time $\tau = \inf\{i \geq 0;\ S_i \geq Z_i\}$. The primal lower bound is then complemented by a dual upper bound. Indeed, Rogers [23] and Haugh and Kogan [18] showed independently that Y_0^* can be expressed via the dual minimization problem

$$Y_0^* = \inf_{M \in \mathscr{M}_1} E[\max_{j=0,\ldots,J}(S_j - M_j)], \tag{1.4}$$

where $\mathscr{M}_{D+1}$ denotes the set of $\mathbb{R}^{D+1}$-valued martingales with $M_0 = 0$, and that the Doob martingale of Y^* is optimal. Hence, the construction of a tight upper bound requires the numerical approximation of the Doob decomposition of Y^*. The nested Monte Carlo algorithm by [1] is popular to perform such numerical Doob decompositions, but in Sect. 1.4 we present algorithms that can produce tight upper bounds at the cost of a non-nested Monte Carlo implementation.

Following the approach of [7], which is detailed there for the case of discrete time reflected BSDEs, we now generalize the construction of a primal maximization problem and a dual minimization problem to dynamic programs of the form (1.1). The following assumptions are in force:

(R) $(\beta_j)_j = (\beta_{0,j}, \ldots, \beta_{D,j})_j$ is a bounded, adapted $D+1$-dimensional process with $\beta_{0,j} \equiv 1$ for all j. The adapted random field $F : \{0, \ldots, J\} \times \Omega \times \mathbb{R}^{D+1} \to \mathbb{R}$ is Lipschitz continuous in $z \in \mathbb{R}^{D+1}$ uniformly in (j, ω) and satisfies $E[|F_j(0)|^2] < \infty$ for every $j = 0, \ldots, J$.

(Comp) For every j and any two $\mathscr{F}_{j+1}$-measurable, integrable real-valued random variables $Y, \tilde{Y}$ such that $Y \geq \tilde{Y}$ a.s., it holds that

$$F_j(E_j[\beta_{j+1}Y]) \geq F_j(E_j[\beta_{j+1}\tilde{Y}]).$$

(Conv) The map $z \mapsto F_j(\omega, z)$ is convex for every j and almost every ω.

We briefly comment on the first two assumptions. The regularity condition (R) makes sure that the dynamic program (1.1) recursively defines square-integrable random variables Y_j^*, $j = J, \ldots, 0$. Condition (Comp) entails a *comparison principle* for the dynamic program (1.1). Indeed, if Y is a *subsolution* of (1.1), i.e.

$$Y_j \leq F_j(E_j[\beta_{j+1}Y_{j+1}]), \; j = 0, \ldots, J-1, \quad Y_J \leq F_J(0),$$

then one can easily show by backward induction that, thanks to (Comp),

$$Y_j \leq Y_j^*, \quad j = 0, \ldots, J. \tag{1.5}$$

Of course, the analogous statement holds for supersolutions.

Primal lower bounds. The construction of the primal maximization problem relies on a linearization of F in terms of its convex conjugate and is analogous to Proposition 3.4 in [19] for BSDEs in continuous time. Recall that the convex conjugate $F_j^\#$ of F_j is defined by

$$F_j^\#(\rho) = \sup_{z \in \mathbb{R}^{D+1}} \rho^\top z - F_j(z) \tag{1.6}$$

and lives on the (bounded uniformly in ω) domain $D_{F^\#}^{j,\omega} \subseteq \mathbb{R}^{D+1}$ where the supremum in (1.6) is finite. We denote by $\mathscr{A}$ the set of adapted, $\mathbb{R}^{D+1}$-valued processes ρ such that ρ_j takes values in $D_{F^\#}^{j,\omega}$ and satisfies $E[F_j^\#(\rho_j)] < \infty$ for $j = 0, \ldots, J-1$. For a fixed $\rho \in \mathscr{A}$, we define recursively the typically non-adapted process $\theta^{low} := \theta^{low}(\rho)$ via $\theta_J^{low} := F_J(0)$ and

$$\theta_j^{low} := \rho_j^\top \beta_{j+1} \theta_{j+1}^{low} - F_j^\#(\rho_j) = F_J(0) \prod_{k=j}^{J-1} \rho_k^\top \beta_{k+1} - \sum_{i=j}^{J-1} F_i^\#(\rho_i) \prod_{k=j}^{i-1} \rho_k^\top \beta_{k+1}. \tag{1.7}$$

Then, the adapted process defined by $Y_j^{low} := Y_j^{low}(\rho) := E_j[\theta_j^{low}]$ satisfies

$$Y_j^{low} = \rho_j^\top E_j[\beta_{j+1} Y_{j+1}^{low}] - F_j^\#(\rho_j) \leq \sup_{\rho \in D_{F^\#}^{j,\omega}} (\rho^\top E_j[\beta_{j+1} Y_{j+1}^{low}] - F_j^\#(\rho))$$

$$= F_j(E_j[\beta_{j+1} Y_{j+1}^{low}]), \quad j = 0, \dots, J-1. \tag{1.8}$$

where the final step uses that $F_j = F_j^{\#\#}$ by convexity. As $Y_J^{low} = F_J(0) = Y_J^*$, we observe that Y^{low} is a subsolution, and, hence, (1.5) yields $Y_j^{low}(\rho) \leq Y_j^*$ for every $j = 0, \dots, J$. Finally, by the Lipschitz assumption there exists an adapted process ρ^* such that

$$\rho_j^{*\top} E_j[\beta_{j+1} Y_{j+1}^*] - F_j^\#(\rho_j^*) = F_j(E_j[\beta_{j+1} Y_{j+1}^*]). \tag{1.9}$$

One can now show by induction that $Y_j^* = Y_j^{low}(\rho^*)$ for every $j = 0, \dots, J$.

We can summarize these considerations in the following theorem.

Theorem 1.1 (Primal problem). *Under assumptions (R), (Comp), and (Conv), Y_0^* can be represented as value of the maximization problem*

$$Y_0^* = \sup_{\rho \in \mathscr{A}} E[\theta_0^{low}(\rho)] = \sup_{\rho \in \mathscr{A}} E\left[F_J(0) \prod_{k=0}^{J-1} \rho_k^\top \beta_{k+1} - \sum_{i=0}^{J-1} F_i^\#(\rho_i) \prod_{k=0}^{i-1} \rho_k^\top \beta_{k+1} \right].$$

Moreover, any process $\rho^* \in \mathscr{A}$, which satisfies (1.9), is optimal.

Dual upper bounds. For the construction of the dual minimization problem we apply a pathwise dynamic programming approach, i.e. the conditional expectations are dropped in (1.1), but some martingale increments are added to the equation instead. To this end we first fix an $\mathbb{R}^{D+1}$-valued martingale, i.e. an integrable and adapted process M with $E_j[M_{j+1} - M_j] = 0$ and $M_0 = 0$. Define recursively the typically non-adapted process $\theta^{up} := \theta^{up}(M)$ via $\theta_J^{up} := F_J(0)$ and

$$\theta_j^{up} = F_j(\beta_{j+1}\theta_{j+1}^{up} - (M_{j+1} - M_j)).$$

Taking conditional expectations and applying Jensen's inequality shows that the adapted process $Y_j^{up} = E_j[\theta_j^{up}]$ satisfies

$$Y_j^{up} \geq F_j(E_j[\beta_{j+1}\theta_{j+1}^{up}]) = F_j(E_j[\beta_{j+1} Y_{j+1}^{up}]), \quad j = 0, \dots, J-1. \tag{1.10}$$

As $Y_J^* = F_J(0) = Y_J^{up}$, Y^{up} is a supersolution of (1.1), and hence the comparison principle implies that $Y_j^{up} \geq Y_j^*$ for all j. Finally, choosing M^* as the the Doob martingale of βY^*, i.e., $M_j^* - M_{j-1}^* = \beta_j Y_j^* - E_{j-1}[\beta_j Y_j^*]$ for all j, one can check inductively that $\theta^{up}(M^*)$ is adapted and that $\theta^{up}(M^*) = Y^*$. We, thus, arrive at the following result.

Theorem 1.2 (Dual problem). *Under assumptions (R), (Comp), and (Conv), Y_0^* can be represented as value of the minimization problem*

$$Y_0^* = \inf_{M \in \mathscr{M}_{D+1}} E[\theta_0^{up}(M)].$$

Moreover, the Doob martingale of βY^* is optimal even in the sense of pathwise control, i.e. $\theta_0^{up}(M^*) = Y_0^*$

Remark 1.1. (i) As explained in Remark 3.5 of [7], the above minimization problem can be re-interpreted as the dual problem to the maximization problem in Theorem 1.1 in the sense of information relaxation. For the general theory of information relaxation duals for discrete time stochastic control problems we refer to [12].

(ii) The results in [7] also cover constructions of minimization and maximization problems with value given by Y_0^* for implicit dynamic programs of the form

$$Y_j^* = F_j(Y_j,\ E_j[\beta_{j+1} Y_{j+1}^*]),\ j = 0, \ldots, J-1, \quad Y_J^* = F_J(0),$$

even without imposing the convexity assumption on F.

(iii) The primal-dual methodology can also be applied for problems with a multi-dimensional value process Y^* such as multiple stopping problems, see [6, 24].

Examples. (i) We first revisit the Bermudan option problem, which is governed by the dynamic programming equation (1.2). As $D = 0$, $\beta \equiv 1$ and $F_j(z) = \max\{S_j, z\}$, the standing assumptions are satisfied. One easily computes $F_j^{\#}(\rho) = (\rho - 1)S_j$ with domain $D_{F^{\#}}^{j,\omega} = [0, 1]$. The primal problem of Theorem 1.1 then reads

$$Y_0^* = \sup_{\rho} E\left[S_J \prod_{k=0}^{J-1} \rho_k + \sum_{i=0}^{J-1} S_i(1-\rho_i) \prod_{k=0}^{i-1} \rho_k, \right]$$

where ρ runs over the set of adapted process with values in $[0, 1]$. By the optimality condition (1.9), it obviously suffices to take the supremum over the set of adapted processes ρ with values in $\{0, 1\}$. The primal problem is then seen to be a reformulation of the optimal stopping problem (1.3), if one maps ρ on the stopping time $\inf\{i \geq 0;\ \rho_i = 0\}$. Concerning the dual minimization problem, one can check inductively that in this case $\theta_j^{up}(M) = \max_{i \in \{j,\ldots,J\}}(S_i - (M_i - M_j))$. Hence, the dual minimization problem in Theorem 1.2 collapses to the dual formulation in (1.4) due to [18, 23].

(ii) The second example is concerned with an Euler type time discretization scheme for *backward stochastic differential equations* (BSDEs) driven by a D-dimensional Brownian motion W. For a BSDE of the form

$$d\mathscr{Y}_t = -f(t, \mathscr{Y}_t, \mathscr{Z}_t)dt + \mathscr{Z}_t^\top dW_t, \quad \mathscr{Y}_T = h$$

we consider Y^* as discretization over the time grid $\{t_0, \ldots, t_J\}$, where:

$$Y_j^* = E_j[Y_{j+1}^*] + (t_{j+1} - t_j) f\left(t_j, E_j[Y_{j+1}], E_j\left[Y_{j+1}^* \frac{W_{t_{j+1}} - W_{t_j}}{t_{j+1} - t_j}\right]\right) \tag{1.11}$$

with terminal condition $Y_J^* = h$. The generator f is an adapted, square-integrable, convex and (uniformly in (t, ω)) Lipschitz continuous random field and h is a square-integrable $\mathscr{F}_J$-measurable random variable. This is a slight variant of the schemes studied by [11, 28] and coincides with the one suggested by [15] in the more general context of second order BSDEs. As filtration in discrete time we can choose the one generated by the Brownian motion up to the jth point in the time grid. By defining $\beta_1, \ldots, \beta_D$ as suitably normalized and truncated increments of the Brownian motion, this recursion is of the form $F_j(z) = z_0 + (t_{j+1} - t_j) f(t_j, z)$. Assumptions (R) and (Conv) are then certainly fulfilled. The truncation of β depends on the time grid and the Lipschitz constants of f in an appropriate way and is necessary to ensure that (Comp) is satisfied, see [7] for details.

1.3 Construction of Lower Bounds via Martingale Basis Functions

This section reviews the popular least-squares Monte Carlo approach for the approximate solution of a dynamic program of the form (1.1) via empirical regression on a set of basis functions, see e.g. [20,21,26]. A special emphasis will be on the particular situation where the basis functions form a set of martingales. This case was studied by [17] for optimal stopping problems and by [8] for the BSDE case.

In view of the optimality condition (1.9) for the primal maximization problem we first rewrite the dynamic program in terms of $Z_j^* := E_j[\beta_{j+1} Y_{j+1}^*]$ as

$$Z_j^* = E_j[\beta_{j+1} F_{j+1}(Z_{j+1}^*)], \quad Z_J^* = 0, \tag{1.12}$$

and note that the solution of the dynamic program (1.1) can be recovered from Z^* as $Y_j^* = F_j(Z_j^*)$. The basic idea of the least-squares Monte Carlo approach is to replace the conditional expectations in (1.12) by an orthogonal projection on a linear subspace of $L^2(\mathscr{F}_j)$, which is spanned by a set of basis functions. The orthogonal projection is then calculated numerically via Monte Carlo simulation by replacing the expectations in the definition of the orthogonal projection by empirical means. More precisely, denote by $\eta_{d,j}$ a row vector of Λ $\mathscr{F}_j$-measurable random variables for every time index j and every $d = 0, \ldots, D+1$. We then define an approximation Z_j of Z_j^* by

$$Z_{d,j} = \eta_{d,j} \alpha_{d,j}, \quad d = 0, \ldots, D,$$

where the coefficients $\alpha_{d,j}$ are computed as follows: Assume we have N independent copies ('regression paths') of

$$\left\{ (F_j^{(n)}, \beta_j^{(n)}, \eta_j^{(n)}),\ j = 0, \ldots, J,\ n = 1, \ldots, N \right\}$$

at hand. We now define $\alpha_{d,J} = 0$ for every $d = 0, \ldots, D$ and

$$\alpha_{d,j} = \arg \min_{\alpha \in \mathbb{R}^{\Lambda}} \frac{1}{N} \sum_{n=1}^{N} \left| \beta_{d,j+1}^{(n)} F_{j+1}^{(n)}(\eta_{0,j+1}^{(n)} \alpha_{0,j+1}, \ldots, \eta_{D,j+1}^{(n)} \alpha_{D,j+1}) - \eta_{d,j}^{(n)} \alpha \right|^2. \tag{1.13}$$

Given these coefficients we can compute on the one hand an approximation of Y^* by $Y_j = F_j(\eta_{0,j}\alpha_{0,j}, \ldots, \eta_{D,j}\alpha_{D,j})$, and on the other hand we can (approximatively) solve for the optimality criterion (1.9) with Z^* replaced by Z in order to get an approximation ρ of the optimizer ρ^* of the primal problem, i.e. ρ_j satisfies

$$(\eta_{0,j}\alpha_{0,j}, \ldots, \eta_{D,j}\alpha_{D,j})\rho_j - F_j^{\#}(\rho_j) \approx F_j((\eta_{0,j}\alpha_{0,j}, \ldots, \eta_{D,j}\alpha_{D,j})).$$

Then,

$$E\left[F_J(0) \prod_{k=0}^{J-1} \rho_k^\top \beta_{k+1} - \sum_{i=0}^{J-1} F_i^{\#}(\rho_i) \prod_{k=0}^{i-1} \rho_k^\top \beta_{k+1} \right]$$

is a lower bound for Y_0^* which is expected to be good, if the basis functions are well-chosen and the number of simulated sample paths is sufficiently large. For a detailed analysis of the projection error due to the choice of the basis and of the simulation error for least-squares Monte Carlo algorithms we refer to [27] and [2] for the Bermudan option pricing problems and to [20] for the BSDE case. Lower confidence bounds for Y_0^* can finally be calculated by replacing the expectation by a sample mean over a new set of independent samples ('outer paths') of $\{F, \beta, \eta\}$ (which are independent of the regression paths). We note that the complexity of this type of algorithm can be reduced by a multilevel approach, which balances the cost between the effort for approximating the conditional expectations and the number of outer paths at different levels, see [4] for the Bermudan option problem. We do not dwell on the details here, but present a similar idea for the computation of upper bounds in Sect. 1.4.1.

In order to illustrate the above least-squares Monte Carlo scheme, let us denote the simulation based projection on the dth set of basis functions at time j by $\mathscr{P}_{d,j}$.

Then the algorithm can be written (informally) as

$$Z_j = \mathscr{P}_{d,j}\left(\beta_{j+1} F_{j+1}(Z_{j+1})\right),$$

i.e. the conditional expectations of the dynamic program are replaced by the empirical projections. We now modify this algorithm by adding an additional projection. Precisely we replace the above Z_j by

$$\tilde{Z}_j = \mathscr{P}_{d,j}\left(\beta_{j+1} \mathscr{P}_{0,j+1}(F_{j+1}(\tilde{Z}_{j+1}))\right).$$

A-priori this does not look like a good idea, because each additional empirical projection is expected to increase the numerical error. However, this scheme can be simplified, if the basis satisfies the following additional martingale property:

(MB) The basis functions $\eta_{0,j}$ form a system of martingales, i.e. $E_j[\eta_{0,j+1}] = \eta_{0,j}$ for $j = 0, \dots, J-1$ and, for $d = 1, \dots, D$, the basis functions are defined via $\eta_{d,j} := E_i[\beta_{d,j+1}\eta_{0,j+1}]$ (which entails that these conditional expectations are available in closed form).

Under this martingale basis assumption one chooses one set of basis functions $\eta_{0,J}$ at terminal time only, and all the other basis functions are computed from this set. The main advantage of assumption (MB) is that conditional expectations of linear combinations of the basis functions (even if multiplied by the β-weights) are at hand in closed form. Hence, the outer empirical projections in the definition of $\tilde{Z}$ need not be performed, but should rather be replaced by the true conditional expectations. These considerations lead to the the *martingale basis algorithm*

$$\tilde{Z}_j^{(MB)} = E_j\left[\beta_{j+1} \mathscr{P}_{0,j+1}(F_{j+1}(\tilde{Z}_{j+1}^{(MB)}))\right].$$

More precisely, one modifies the construction of the coefficients $\alpha_{d,i}$ compared to the standard least-square Monte Carlo scheme as follows. Define $\alpha_{d,i} = \alpha_i$ for all $d = 0, \dots, D$, where $\alpha_J = 0$ and

$$\alpha_j = \arg\min_{\alpha \in \mathbb{R}^\Lambda} \frac{1}{N} \sum_{n=1}^{N} \left| F_{j+1}^{(n)}(\eta_{0,j+1}^{(n)}\alpha_{j+1}, \dots, \eta_{D,j+1}^{(n)}\alpha_{j+1}) - \eta_{0,j+1}^{(n)}\alpha \right|^2 \tag{1.14}$$

Once the coefficients are computed, one constructs the approximations of Y^*, Z^*, ρ^* and the lower bound for Y_0^* in exactly the same way as described above. An obvious advantage of this martingale basis algorithm for $D \geq 1$ is, that only one empirical regression is performed at each time step, while the original least-squares Monte Carlo algorithm requires $(D+1)$ empirical regressions per time step.

In the setting of discrete time approximations of BSDEs one has $F_j(z) = z_0 + (t_{j+1} - t_j) f(t_j, z)$. Hence, (with a slight abuse of notation), the martingale basis