


Pricing and Hedging under High-Dimensional Jump-Diffusion Models using Partial Differential Equations

Peter Thomas Hepperger

Department of Mathematics
Technische Universität München
85748 Garching, Germany

2011

TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Mathematische Statistik

Pricing and Hedging under High-Dimensional Jump-Diffusion Models using Partial Differential Equations

Peter Thomas Hepperger

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Martin Brokate
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Claudia Klüppelberg
	2.	Univ.-Prof. Dr. Rüdiger Kiesel, Universität Duisburg-Essen
	3.	Prof. Dr. Fred Espen Benth, University of Oslo, Norwegen (schriftliche Beurteilung)

Die Dissertation wurde am 02.02.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 02.05.2011 angenommen.

Abstract

This thesis is concerned with high- and infinite-dimensional jump-diffusion models. Two applications serve as motivating examples: stock basket options and options on electricity swaps. A general Hilbert space valued model is introduced, which can be applied to both settings. We discuss the stochastic properties of the model in detail.

The focus of the thesis then lies on numerical techniques for the solution of pricing and hedging problems. The "curse of dimensionality" prevents us from applying well-known methods, like partial integro-differential equations (PIDEs) or Fourier methods, directly to the high-dimensional model. Instead, we use a projection approach called proper orthogonal decomposition (POD). This dimension reduction method is based on a similar idea as principal component analysis in statistics. It makes use of the correlation structure of the assets and involves the solution of an eigenvalue problem for the covariance operator of the driving stochastic process. We apply POD to various pricing problems and option types and establish corresponding convergence results.

First, we consider European option prices. We derive the corresponding PIDE and employ POD to find a low-dimensional, approximating equation. Based on the variational formulation, we show existence, uniqueness, and regularity of the PIDE solution. Next, we study the hedging problem for European options in the context of electricity markets. The basic challenge when hedging electricity is to hedge an infinite-dimensional forward curve with a finite set of traded assets. Thus, the electricity market is incomplete. We derive the stochastic dynamics of hedging portfolios and solve the quadratic hedging problem. We show that the optimal hedging strategy at each point in time is the solution of a linear equation system. Similar to a classical delta hedge, the strategy depends on partial derivatives of the option price, which can be obtained from a PIDE.

The dimension reduction can also be applied to Asian options. In this case, we obtain an additional variable in the PIDE, the arithmetic average. We examine the properties of the average in the Hilbert space valued context and extend the convergence theorems for the POD method accordingly. Finally, we price high-dimensional Bermudan options. We improve the convergence speed of the well-established Longstaff–Schwarz Monte Carlo (MC) approach through a variance reduction scheme. To this end, we use the POD approximation as a control variable. The variance of the least-squares MC method, and thus the overall computational effort, is reduced significantly. The expectation of the approximated price process can be computed with any algorithm suitable for pricing low-dimensional Bermudan options, in particular Fourier or PIDE methods.

In various numerical experiments, we evaluate the performance of the POD method. The numerical solution of the PIDEs is based on a sparse grid combination method for spatial discretization and a Discontinuous Galerkin time stepping method. We study the influence of correlation and basket size on the efficiency of the pricing algorithm, the impact of different hedging portfolios on the hedging error, and the accuracy of direct versus duality based Bermudan pricing methods. We find that POD is an efficient means to perform a dimension reduction if the underlying asset price processes feature sufficiently strong correlation.

Zusammenfassung

Diese Arbeit befasst sich mit hoch- und unendlichdimensionalen Sprung-Diffusions-Modellen. Zwei Anwendungsbeispiele dienen dabei als Motivation: Optionen auf Aktienkörbe und Elektrizitäts-Swaps. Ein allgemeines, Hilbertraum-wertiges Modell, das auf beide Produkte angewandt werden kann, wird eingeführt. Wir diskutieren die stochastischen Eigenschaften des Modells im Detail.

Der Schwerpunkt der Arbeit liegt dann auf numerischen Verfahren für die Lösung von Preis- und Hedge-Problemen. Der „Fluch der hohen Dimension“ hindert uns daran bekannte Methoden, etwa partielle Integro-Differentialgleichungen (PIDEs) oder Fourier Methoden, direkt auf das hochdimensionale Modell anzuwenden. Wir verwenden stattdessen einen Projektionsansatz namens Proper Orthogonal Decomposition (POD). Diese Dimensionsreduktion beruht auf einer ähnlichen Idee wie die Hauptkomponentenanalyse in der Statistik. Sie macht sich die Korrelationsstruktur der Wertpapiere zunutze und basiert auf der Lösung des Eigenwertproblems für den Kovarianzoperator des stochastischen Prozesses. Wir wenden die Projektion auf verschiedene Optionstypen an und beweisen entsprechende Konvergenzresultate.

Zunächst untersuchen wir europäische Optionen. Wir leiten die zugehörige PIDE her und setzen die POD ein, um eine niederdimensionale, approximierende Gleichung zu erhalten. Mit Hilfe der variationellen Form zeigen wir Existenz, Eindeutigkeit und Regularität der Lösung. Danach untersuchen wir Hedging-Probleme für europäische Optionen auf Elektrizitätsmärkten. Die grundlegende Herausforderung besteht hierbei darin, die unendlichdimensionale Forwardkurve mit einer kleinen Auswahl handelbarer Kontrakte zu hedgen. Der Elektrizitätsmarkt ist daher unvollständig. Wir leiten die stochastische Dynamik des Hedge-Portfolios her und lösen das quadratische Hedge-Problem. Wir zeigen, dass die optimale Handelsstrategie als Lösung eines linearen Gleichungssystems dargestellt werden kann. Genau wie ein klassischer Delta-Hedge hängt diese Strategie von den partiellen Ableitungen des Optionspreises ab, die wir aus der PIDE erhalten.

Die Dimensionsreduktion kann auch auf asiatische Optionen angewandt werden. In diesem Fall erhalten wir eine zusätzliche Variable in der PIDE, das arithmetische Mittel. Wir untersuchen die Eigenschaften dieser Variable und erweitern die Konvergenzsätze für die POD-Methode entsprechend. Darüber hinaus bewerten wir bermudische Optionen. Dabei verbessern die Konvergenzgeschwindigkeit des bekannten Longstaff–Schwarz Monte Carlo (MC) Algorithmus durch Varianzreduktion. Hierzu setzen wir die POD-Näherung als Kontrollvariable ein. Die Varianz des MC Verfahrens, und damit der Gesamtrechenaufwand, kann dadurch signifikant reduziert werden. Der Erwartungswert des approximierten Preisprozesses besipielsweise mit Fourier- oder PIDE-Methoden berechnet werden.

Wir evaluieren die Leistungsfähigkeit der POD-Methode in zahlreichen numerischen Experimenten. Für die numerische Lösung der PIDEs wird die Kombinationstechnik für dünne Gitter in der räumlichen Diskretisierung, sowie ein Discontinuous-Galerkin Zeitschrittverfahren eingesetzt. Wir untersuchen den Einfluss der Korrelation und der Anzahl der Wertpapiere auf die Effizienz des Algorithmus, die Auswirkungen verschiedener Hedge-Portfolios auf den Hedge-Fehler und die Genauigkeit von direkten und dualen Preismethoden für amerikanische Optionen. Es zeigt sich, dass die POD-Zerlegung eine effiziente Methode zur Dimensionsreduktion ist, falls die zugrunde liegenden Preisprozesse hinreichend starke Korrelation aufweisen.

Acknowledgments

Foremost, it is a pleasure to thank my supervisor Prof. Claudia Klüppelberg. She introduced me to the stochastic side of applied mathematics, which was a whole new and fascinating area of research for me during my PhD studies. I am grateful for her constant encouragement, helpful advice, and for the excellent working conditions at her chair. She provided the perfect balance of steady support and the freedom to develop my own ideas.

It is an honor for me to express my thanks to Prof. Fred Espen Benth. Without his guidance and experience, this thesis would not have been possible. Not only did he initiate my project on electricity pricing and hedging, he also provided great ideas, patient explanations, and encouraging feedback.

Many thanks to my colleagues in the departments of stochastics as well as numerics at the TU Munich. Their knowledge and professional advice gave me a better insight to various aspects of my thesis. Their good company made my PhD studies an enjoyable time.

I owe my gratitude to the International Graduate School of Science and Engineering (IGSSE) for financial support and for the opportunity to look beyond my own subject.

Lastly, and most importantly, I am grateful to my family for their loving support and steady encouragement. To them I dedicate this thesis.

Contents

1 Introduction and Motivating Examples **1**
- 1.1 Example 1: Stock Basket Options . . . 2
- 1.2 Example 2: Electricity Swaptions . . . 3
- 1.3 Outline and Objectives of the Thesis . . . 5

2 Hilbert Space Valued Jump-Diffusion **9**
- 2.1 Model Description . . . 9
- 2.2 Properties of the Model . . . 11
- 2.3 Corresponding Time-Homogeneous Model . . . 16

3 Dimension Reduction **19**
- 3.1 Proper Orthogonal Decomposition (POD) . . . 19
- 3.2 Numerical Computation of POD Components . . . 22

4 European Options **25**
- 4.1 Partial Integro-Differential Equation (PIDE) . . . 25
- 4.2 Pricing European Options with POD . . . 29
  - 4.2.1 Finite Dimensional PIDE . . . 29
  - 4.2.2 Variational Formulation and Uniqueness . . . 33
  - 4.2.3 Convergence of Finite-Dimensional Approximations . . . 38

5 Hedging **41**
- 5.1 Dynamics of Swap Rates and Swaption Prices . . . 41
  - 5.1.1 Swap Rate Derivatives . . . 42
  - 5.1.2 Applying Itô's Formula . . . 45
  - 5.1.3 Time-Inhomogeneous Swaption Price Dynamics . . . 47
- 5.2 Quadratic Hedging of Electricity . . . 49
- 5.3 Hedging with POD . . . 54

6 Asian Options **61**
- 6.1 Value of an Asian Option . . . 61
- 6.2 Pricing Asian Options with POD . . . 65

7 Bermudan Options **71**
- 7.1 Pricing Bermudan Options with POD . . . 72
- 7.2 Variance-Reduced Least-Squares Monte Carlo . . . 76
- 7.3 Dual Method . . . 79

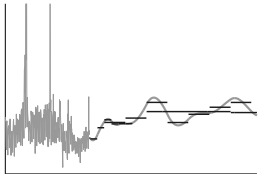
8 Numerical Experiments **81**
- 8.1 Numerical PIDE Solution . . . 81
- 8.2 Test Problems . . . 86
- 8.3 Experiments for European Options . . . 89
- 8.4 Experiments for Hedging . . . 92
- 8.5 Experiments for Bermudan Options . . . 101

1 Introduction and Motivating Examples

Ever since Black and Scholes published their seminal work on option pricing [15], the evaluation of financial derivatives has been closely related to the solution of partial differential equations (PDEs). In fact, the link between stochastic processes and deterministic PDEs had already been established in the Feynman–Kac formula [50] decades earlier. While the Black–Scholes model relies on a geometric Brownian motion to model an asset price, Feynman–Kac type results can be extended to much more sophisticated settings [39]. It is, e.g., possible to introduce jumps to the stochastic process. A jump-diffusion model combines a Wiener process with a compound Poisson process. The name "diffusion" stems from the fact that there is a diffusive (elliptic) term in the corresponding differential equation, similar to the heat equation. Moreover, one can add time-dependence to the coefficients in the model. This yields additive processes, also known as time-inhomogeneous Lévy processes. Further possible extensions, which we will not discuss in the present thesis, include stochastic volatility and infinite jump activity. For an overview of jump models in finance, see, e.g., [24, 74].

Elaborate models allow for a much better fit of financial data than a plain geometric Brownian motion. They come, however, at a price. While the famous Black–Scholes formula gives an explicit solution for the price and hedge of a plain-vanilla European option, it is generally impossible to find analytically tractable solutions for these problems under more realistic assumptions. Therefore, efficient numerical algorithms are essential for the application of such models in practice. The Feynman–Kac representation enables the use of PDE methods for pricing and hedging. These methods include the PDE discretization with finite differences or finite elements [25, 36]. Their convergence and performance is well-studied [18, 72, 77]. Jumps in the stochastic model pose an additional difficulty. They result in nonlocal integral terms in the equation, yielding partial integro-differential equations (PIDEs) [26]. These require specialized discretization schemes [49, 61]. An alternative approach relies on Fourier methods. The characteristic function of additive processes is usually known explicitly. The general idea is to consider the Fourier transform of an option payoff (or an integrable modification of the payoff) and to factorize it in such a way that the characteristic function of the process occurs in the resulting expression. Finally, an inverse transform yields the value of the option. Since Fourier transforms can be approximated with discrete transforms, efficient algorithms can be obtained using fast Fourier transforms (FFTs) [22, 23, 58].

Pricing and hedging problems become much more elaborate when the derivative depends on more than one asset. In principle, both PIDE and Fourier methods can be extended to the multivariate case. The main challenge here is the exponential increase of computational effort for higher-dimensional problems. This *curse of dimensionality* makes it virtually impossible to use regular, full discretizations for options depending on more than three assets. Sparse grids are an effective means to counter the curse. PIDE methods have recently been extended to such multidimensional settings using sparse grids [66, 81]. Computationally feasible dimensions n, however, are still moderate, usually $n \leq 10$. On the other hand, traded products often imply much higher dimensions. Derivatives (e.g., swaptions) may even depend on a continuum of "assets" and thus be infinite-dimensional. The following two examples illustrate the need for high-dimensional models. The first one is concerned with basket options on the stock



Flip-book: electricity spot, swaps, and forward curve; explanation on page 4

market. The second example is more elaborate and involves electricity contracts. Both examples motivate the use of the subsequently presented dimension reduction techniques. We will refer to them frequently throughout the thesis. In particular, all of the numerical experiments are based on either of them.

1.1 Example 1: Stock Basket Options

We consider an option depending on n stocks. Let T be the time horizon (terminal date of maturity) of the option. The multivariate asset price process is denoted

$$S_t := \big(S_t(1), \dots, S_t(n)\big) := \big(S_0(1)\, e^{X_t(1)}, \dots, S_0(n)\, e^{X_t(n)}\big) \in \mathbb{R}^n,\ t \in [0, T]. \tag{1.1}$$

For each $i = 1, \dots, n$, the price $S(i)$ of the ith asset is modeled as the product of its initial value $S_0(i) > 0$ and the ordinary exponential of a time-inhomogeneous jump-diffusion process $(X_t(i))_{t \geq 0}$. We will specify the model in more detail in Section 2. As an example for market data, Figure 1 displays the daily prices for the top eight S&P 500 stocks over two years. Although some stocks (namely IBM) display rather singular paths, one can easily identify clusters of stocks, which are highly correlated. We will make use of these correlations during the dimension reduction to obtain low-dimensional approximations.

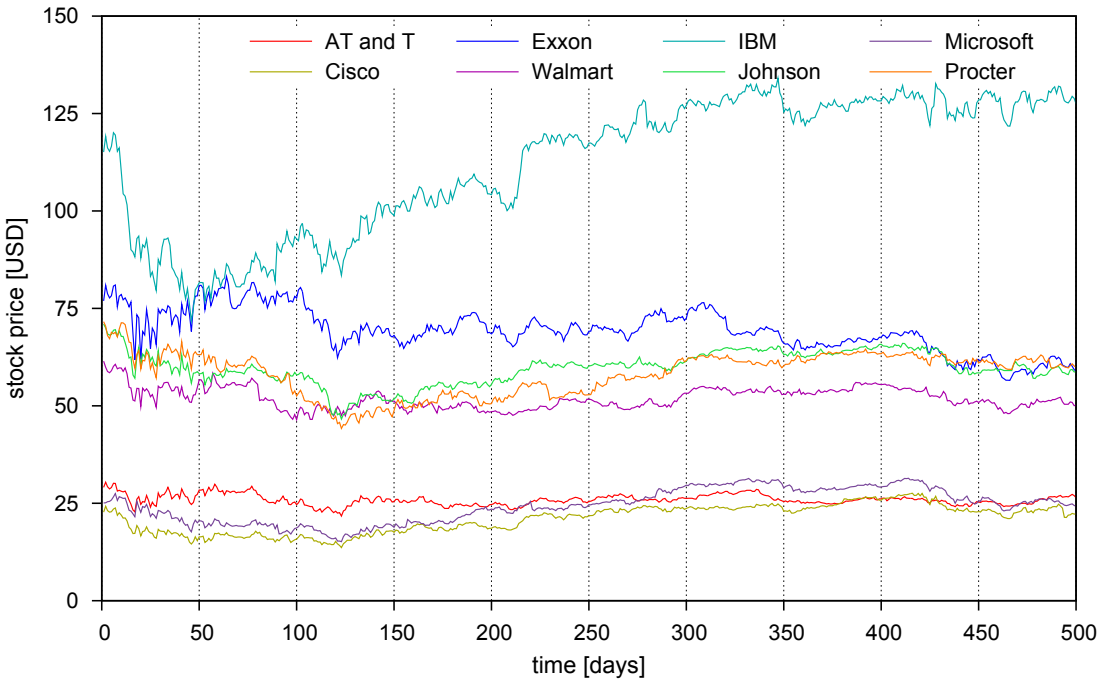


Figure 1: Prices for eight S&P 500 stocks on the trading days from 18 Sept. 2008 to 20 Oct. 2010.

There are various types of options depending on such a basket of assets. Index options written on a weighted average are a common example. An index put has the terminal

value

$$P_{\text{index}}(T, S_T) = \left(K - \sum_{i=1}^{n} w_i S_T(i) \right)^+, \tag{1.2}$$

where $K > 0$ is the strike and $w_i \in \mathbb{R}$ are constant weights. Other examples include maximum or minimum options. Maximum or minimum basket put options correspond to the payoff

$$P_{\max}(T, S_T) = \left(K - \max_{i=1,\dots,n} S_T(i) \right)^+ \quad \text{or} \quad P_{\min}(T, S_T) = \left(K - \min_{i=1,\dots,n} S_T(i) \right)^+,$$

respectively. Basket call options can be defined analogously.

1.2 Example 2: Electricity Swaptions

During the last two decades, energy markets all over the world have been liberalized. Electricity is now traded liquidly on exchanges like the Scandinavian Nordpool and the German Energy Exchange (EEX). These relatively young markets are open to producers, consumers, and speculating investors. Traded products include spot, futures, forwards, and options on these. A compendium of mathematical finance for electricity markets can be found in [13]. There are substantial differences between stock and electricity markets. The electricity spot price exhibits several unique stylized features, including seasonality, large jumps (many times higher than the average price), and mean reversion. In addition, while stocks are sold at a single point in time, electricity contracts always imply the delivery over a certain period of time. Therefore, electricity forwards and derivatives are written on a delivery period (a week, a month, or even a year). The most liquidly traded products on energy exchanges like EEX or Nordpool are contracts of futures type. These are agreements on a constant delivery of 1 MW of electricity over a certain future period of time $[T_1, T_2]$. In return, a rate $F(t; T_1, T_2)$, fixed at time $t \le T_1$, is paid during this delivery period. Since a payment of a fixed rate is made in exchange for the (unknown) future spot price, these contracts are also known as *electricity swaps*. The concept is illustrated in Figure 2.

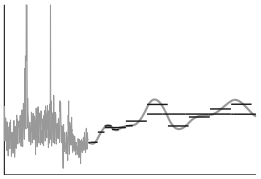
The relation of spot and forward prices is not clearly defined for electricity because of its non-storability [9, 12]. This difficulty can be avoided by directly modeling the forward curve under a risk neutral (with respect to swap rates) measure [3, 10, 51]. For every maturity $u \in [T_1, T_2]$, let

$$f_t(u) := \lim_{v \to u} F(t; u, v) \tag{1.3}$$

be the corresponding value of the forward curve at time $t \le u$. Due to no-arbitrage considerations, the following equality must hold for every $t \le T_1$.

$$\int_{T_1}^{T_2} e^{-r(u-t)} F(t; T_1, T_2)\, du = \int_{T_1}^{T_2} e^{-r(u-t)} f_t(u)\, du,$$

where r is the constant risk free interest rate. Thus, the swap rate F can be written as



the weighted integral

$$F(t;T_1,T_2) = \int_{T_1}^{T_2} \omega(u;T_1,T_2) f_t(u)\, du,$$

with the nonnegative discounting factor

$$\omega(u;T_1,T_2) := \frac{e^{-ru}}{\int_{T_1}^{T_2} e^{-rv}\, dv}. \tag{1.4}$$

Since no initial payment is needed to enter a swap contract, the swap rate $F(t;T_1,T_2)$ is a martingale under the risk neutral measure.

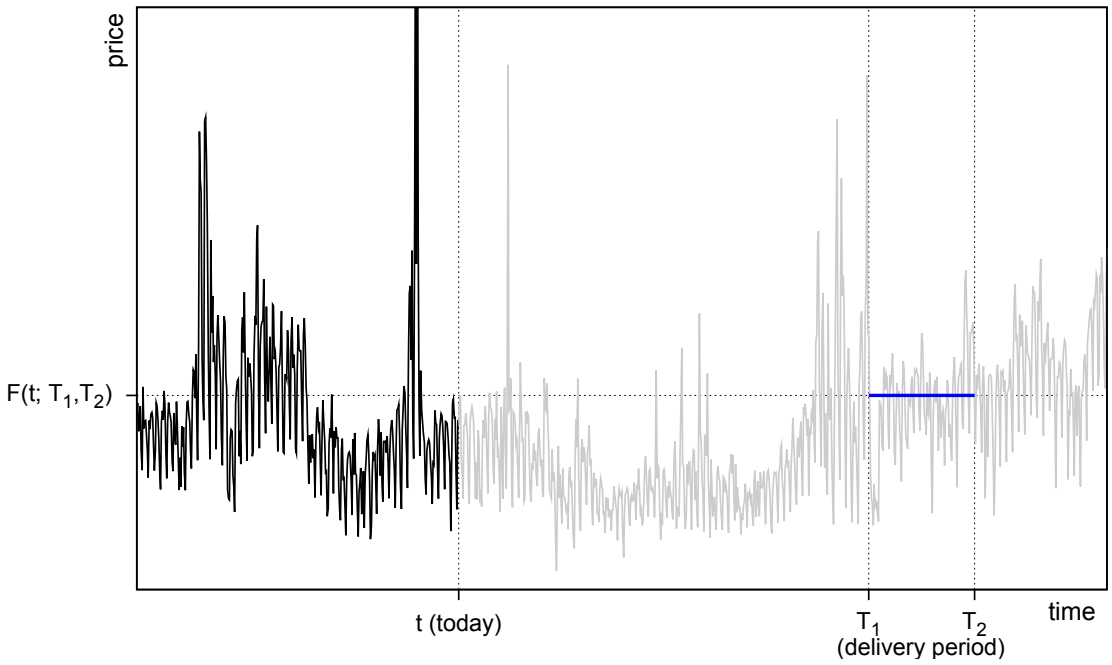


Figure 2: Electricity spot price, swap delivery period $[T_1, T_2]$, and swap rate $F(t;T_1,T_2)$.

Figure 3 displays the different prices and concepts from the energy market and their relation. One year worth of daily spot prices are taken from actual EEX data. The seasonality function is a truncated Fourier series fitted to the spot. Each traded swap contract is represented by a single horizontal line; these are market data as well. The longest lines correspond to contracts with a delivery period of one year, shorter lines represent quarterly and monthly products. Finally, the forward curve is obtained by smooth interpolation of the swap data, taking also the seasonality into account. For an overview of the fitting methods confer, e.g., [11, 52]. In the lower right corner of this thesis, a **flip-book animation** of the graph in Figure 3 is displayed. It illustrates the dynamics of the spot and forward prices on the EEX. In order to watch it, place the thesis on a table and flip the pages from back to front (the last page contains the first image).

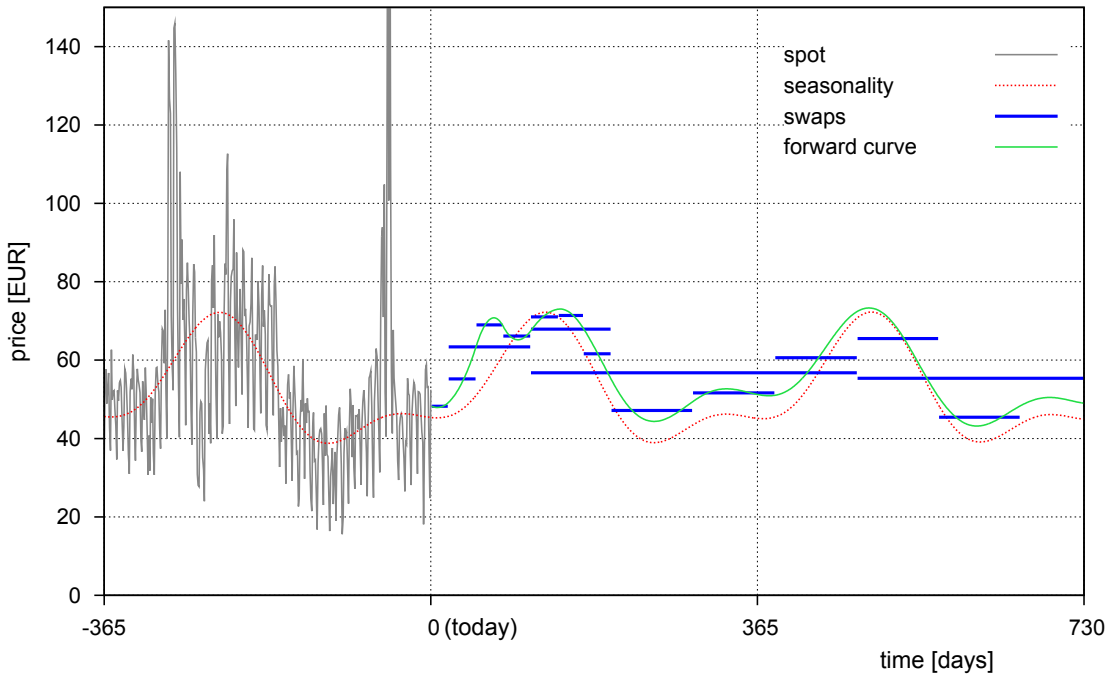


Figure 3: EEX spot price data, fitted seasonality, traded swaps, and interpolated forward curve.

The derivatives we are primarily interested in are options on electricity swaps with a given delivery period $[T_1, T_2]$. The associated payoff of a European put option with strike rate $K > 0$ and maturity $T \leq T_1$ is

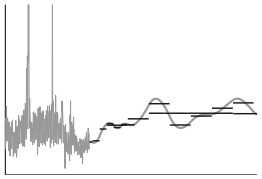
$$\begin{aligned} P_{\text{swaption}}\big(T, F(T;T_1,T_2)\big) &= \left(\int_{T_1}^{T_2} e^{-r(u-T)} K \, du - \int_{T_1}^{T_2} e^{-r(u-T)} F(T;T_1,T_2) \, du \right)^+ \\ &= \kappa \left(K - \int_{T_1}^{T_2} \omega(u;T_1,T_2) f_t(u) \, du \right)^+, \end{aligned} \tag{1.5}$$

where

$$\kappa := \kappa(T;T_1,T_2) := \int_{T_1}^{T_2} e^{-r(u-T)} \, du. \tag{1.6}$$

1.3 Outline and Objectives of the Thesis

The payoff for an index option (1.2) and the payoff for an electricity swaption (1.5) are very similar in their structure. The main difference is that for the swaption, the finite weighted sum of assets has been replaced by a weighted integral. We can interpret the latter as an infinite-dimensional "index" on a continuum of forward prices. With this intuition in mind, we employ a common theoretical framework for both the stock baskets and the electricity forward curve. We use a Hilbert space valued stochastic process to model the assets. In the case of an n-dimensional basket of stocks, the Hilbert space is n-dimensional. For the electricity market, on the other hand, we choose a function space



containing the possible forward curves. In *Section 2*, we discuss the driving stochastic process in detail. In particular, we discuss how to construct an exponential jump-diffusion taking values in a Hilbert space. We also examine how the drift has to be chosen in order to make such an exponential a martingale. Furthermore, we state the characteristic function and examine further basic properties of the process.

The main objective of this thesis is to develop efficient numerical algorithms for the solution of basic pricing and hedging problems in this high-dimensional setting. The main technique used to break the curse of dimensionality is a dimension reduction method called proper orthogonal decomposition (POD), which we present in *Section 3*. We discuss the construction of low-dimensional projections, which can be used to approximate the Hilbert space valued process, and derive error estimates for the approximation. Each of the subsequent sections is concerned with the application of the POD method to different types of pricing and hedging problems. They are based on separate papers by this author. In each of these sections, convergence results for the dimension reduction are established. They rely on the general theorems in Sections 2 and 3.

Section 4 deals with European options and is based on [44]. We derive the Hilbert space valued PIDE describing the option price. The equation is projected with POD, yielding a finite-dimensional PIDE. Assuming Lipschitz continuity of the payoff, regularity of the projected price process is shown. We also obtain existence and uniqueness results for the PIDE. To this end, we derive the variational formulation of the PIDE, which is also the starting point for finite element methods. Finally, we establish an upper bound for the approximation error.

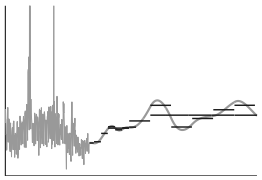
The quadratic hedging problem for European options is solved in *Section 5*. We restrict ourselves to the case of electricity swaptions. This market is incomplete and features certain intrinsic difficulties. The challenge here is to hedge an option depending on an infinite-dimensional object (the forward curve) with a small set of traded contracts (swaps with various delivery periods). We may, e.g., want to hedge a monthly swaption with several weekly and one monthly swap. It is inherent to the problem that no perfect hedge is possible, even in a pure diffusion setting. There is a so called *basis risk*, which cannot be avoided or hedged with the given underlyings. Presenting the results from [45, 47], we derive a representation of the variance-minimal hedging strategy. Similar to a classical delta hedge, the optimal hedging strategy depends on partial derivatives of the option price. These derivatives can be approximated numerically using POD. We obtain convergence results for the associated hedging error.

In *Section 6*, we show how the approach can be extended to Asian options. It contains the results from [46]. We introduce an additional variable to our model, the arithmetic time average. After studying the basic properties of the average process, we describe the POD method for Asian options in detail. Then, we generalize the low-dimensional PIDE satisfied by the approximated price process to the case of Asian options. We show convergence of the PIDE solution to the true value of the Asian option.

Finally, we consider Bermudan options in *Section 7*. Based on [48], we apply the POD method for a finite number of exercise dates. An accurate calculation of Bermudan option prices with the PIDE approach is hardly possible, even after the dimension reduction, because sparse grids are not directly applicable here. This is due to insufficient regularity of the price process. Instead, we choose the dimension of the POD approximation rather low (less or equal to three) and operate on full grids. We use the

results as a control variable to achieve a significant improvement of the well-established least-squares Monte Carlo method proposed by Longstaff and Schwartz [60]. The expectation of the control variable can be computed with any numerical algorithm suitable for low-dimensional Bermudan option pricing. In particular, PIDE solvers and Fourier methods are applicable. The variance reduction allows for a considerable decrease in computational effort. As an alternative to the variance reduction, the POD approximation can also be used as a candidate for the minimizing martingale in the dual pricing approach suggested by Rogers [67]. We evaluate the performance of both approaches in our numerical tests. Choosing the number of exercise possibilities sufficiently large, the same techniques can be used to approximate American options.

The last section of the thesis, *Section 8*, is concerned with the implementation of the presented methods. The numerical solution of the PIDEs using a discontinuous Galerkin discretization and, in the case of European options, the sparse grid combination technique is discussed. We then introduce a set of test problems. Numerical experiments are performed to demonstrate the effectiveness of the dimension reduction method. In particular, we present results for European option pricing and hedging as well as Bermudan option pricing.



2 Hilbert Space Valued Jump-Diffusion

Several authors propose exponential additive processes (also known as exponential time-inhomogeneous Lévy processes) of diffusion or jump-diffusion type to model electricity forward curves [13, 51]. Generalizing their approach, we now state the Hilbert space valued model used throughout this thesis. We define the driving time-inhomogeneous jump-diffusion process, declare assumptions concerning the coefficients, and state relevant properties of the model. In particular, we discuss moments and martingale conditions. The model is assumed to be stated under the risk neutral pricing measure. Finally, we derive a time-homogeneous process which is equivalent to the additive model in the sense that it has the same terminal distribution. It thus yields the same results for European option prices, but it allows for a reduced computational complexity due to time-constant coefficients. This is of course only applicable to derivatives which are not path-dependent.

For a definition of stochastic processes and integration in Hilbert spaces with respect to Brownian motion see, e.g., [27, 56]. An overview of Poisson random measures in Hilbert spaces can be found in [42], the case of Lévy processes is treated in [62]. Infinite-dimensional stochastic analysis, its applications to interest-rate theory, and Heath–Jarrow–Morton models are presented in [21].

2.1 Model Description

Let $D \subset \mathbb{R}^m$ be an arbitrary subset, and let μ_D be a measure on D. Then

$$H := L^2(D, \mu_D)$$

is a separable Hilbert space. For every $h \in H$, we denote the corresponding norm by

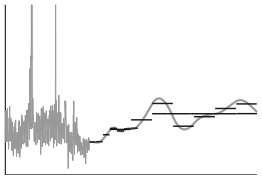
$$\|h\|_H := \sqrt{\int_D \big[h(u)\big]^2 \, \mu_D(du)}.$$

For electricity swaptions, we choose $D = [T_1, T_2]$ and $\mu_D = \lambda_D$ (the Lebesgue measure on D). Then H can be interpreted as the space of forward curves on the delivery period D. For basket options, we choose D to be a finite (though possibly large) index set with n entries, each index corresponding to one asset. The measure μ_D is then simply the counting measure, and the norm $\|\cdot\|_H$ is the Euclidean norm on $\mathbb{R}^n$.

The basic driving stochastic process for our model is the H-valued additive process

$$X_t := \int_0^t \gamma_s \, ds + \int_0^t \sigma_s \, dW_s + \int_0^t \int_E \eta_s(y) \, \widetilde{M}(dy, ds), \; t \geq 0. \tag{2.1}$$

The diffusion part is driven by a U-valued Wiener process W, where $(U, \|\cdot\|_U)$ is a separable Hilbert space. The covariance operator of W is a symmetric, nonnegative definite trace class operator Q. The mark space $(E, \|\cdot\|_E)$ of the Poisson part of the process is a Banach space. The jumps are characterized by $\widetilde{M}$, the compensated random



measure of an E-valued compound Poisson process

$$J_t = \sum_{i=1}^{N_t} Y_i, \qquad t \geq 0,$$

which is independent of W. Here, N denotes a Poisson process with intensity λ and $Y_i \sim P^Y$ $(i = 1, 2, \ldots)$ are iid on E (and independent of N). The corresponding Lévy measure is denoted by $\nu = \lambda P^Y$. We denote by $\mathrm{L}(U, H)$ and $\mathrm{L}(E, H)$ the spaces of all bounded linear operators mapping U and E to H, respectively. We assume the drift $\gamma : [0, T] \to H$, the volatility $\sigma : [0, T] \to \mathrm{L}(U, H)$, and the jump dampening factor $\eta : [0, T] \to \mathrm{L}(E, H)$ to be deterministic functions. The time-dependent coefficients enable us to reproduce the Samuelson effect of increasing volatilities close to maturity, which can be observed in electricity prices. Subsequently, we will write $f_t(u) := f(t)(u) \in \mathbb{R}$ for every $f : [0, T] \to H$, $u \in D$, and similarly $g_t(h) := g(t)(h) \in H$ for every $g : [0, T] \to \mathrm{L}(H, H)$, $h \in H$. We make the following assumption, which is similar to the finite-dimensional moment conditions in [71, Sec. 25]. For an introduction to time-dependent Bochner spaces, such as $\mathrm{L}^2(0, T; H)$, see [33, Ch. 5.9].

Assumption 2.1. *The second exponential moment of the jump distribution Y exists:*

$$E\Big[e^{2\|Y\|_E}\Big] = \int_E e^{2\|y\|_E} P^Y(dy) < \infty.$$

We assume further $\gamma \in \mathrm{L}^2(0, T; H)$, $\sigma \in \mathrm{L}^2(0, T; \mathrm{L}(U, H))$, and

$$\|\eta_t\|_{\mathrm{L}(E,H)} \leq 1 \quad \textit{for every } t \in [0, T].$$

Under Assumption 2.1, $(X_t)_{t \geq 0}$ is an additive process with finite activity and finite expectation. This simplifies notation, since no truncation of large jumps is needed in the characteristic function.

We would now like to model the vector of asset prices S_t and the forward curve f_t, introduced in (1.1) and (1.3), respectively, as the exponential of the driving process X in some sense. For a finite set of stocks, we have $H = \mathbb{R}^n$. Since in this case $X_t = \big(X_t(1), \ldots, X_t(n)\big)$, we will obviously stick to (1.1) and set

$$S_t := \big(S_t(1), \ldots, S_t(n)\big) := \big(S_0(1)\, e^{X_t(1)}, \ldots, S_0(n)\, e^{X_t(n)}\big) \in \mathbb{R}^n, \quad t \in [0, T]. \tag{2.2}$$

In the context of electricity swaps, we have an infinite-dimensional, continuous setting. There, we could take the point-wise exponential $f_t(u) = f_0(u) \exp\big(X_t(u)\big)$, $u \in D = [T_1, T_2]$. While this would be possible, several technical assumptions would have to be made to ensure that f_t is again square integrable (and thus an element of the Hilbert space H). Since we are interested in swap rates, and not in pointwise evaluations of forward curves, a more natural way to define the exponential is the following: Choose an orthonormal basis $\{e_k\}_{k=1}^{\dim(H)}$ of H and set

$$f_t := \sum_{k=1}^{\dim(H)} \langle f_0, e_k \rangle_H \; e^{\langle X_t, e_k \rangle_H} \, e_k \tag{2.3}$$

for $t \geq 0$, with $f_0 \in H$. Note that the choice of the basis $\{e_k\}_{k\in\mathbb{N}}$ is part of the modeling process, similar to the choice of jump distributions and correlation structures. This allows us to solve the pricing and hedging problem for various modeling paradigms with the same, unified theoretical framework. One may use, e.g., eigenfunctions obtained from principal component analysis of the forward market data. For this choice, (2.3) is nothing more than a factor model describing the dynamics of each principal component. Since electricity is usually traded on an hourly basis, another reasonable approach is to use piecewise constant indicator functions on hourly intervals. The actually considered Hilbert space then is a high- but finite-dimensional subspace of H. Thus, (2.3) describes a *family* of models corresponding to different ways of modeling the forward curves.

It is worth mentioning that for a finite-dimensional Hilbert space, the process f is equivalent to the point-wise exponential if $\{e_k\}$ are canonical unit vectors. In particular, we can recover the setting for the multivariate stock basket (2.2) if $H = \mathbb{R}^n$. Consequently, we will use definition (2.3) for both the stock market and the electricity market, except for those sections of the thesis which are applicable exclusively to a finite number of assets.

2.2 Properties of the Model

The way we have defined the exponential in (2.3) makes it possible to show the existence of moments and to derive sufficient conditions for f to be an H-valued martingale. We start with a proposition concerning the properties of the additive process X defined in (2.1). As before, let $T > 0$ be the finite time horizon of the option. We denote the trace of a nuclear operator by $\mathrm{tr}(\cdot)$. In all subsequent proofs, C denotes a generic positive constant with no fixed value.

Proposition 2.2. *The process $(X_t)_{t\geq 0}$ is square-integrable:*

$$\sup_{0\leq t\leq T} E\,\|X_t\|_H^2 < \infty. \tag{2.4}$$

Proof. The definition of the process $(X_t)_{t\geq 0}$ yields

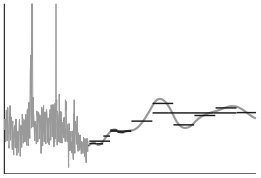
$$E\,\|X_t\|_H^2 \leq 3E\left[\left\|\int_0^t \gamma_s\,ds\right\|_H^2 + \left\|\int_0^t \sigma_s\,dW_s\right\|_H^2 + \left\|\int_0^t\int_E \eta_s(y)\,\widetilde{M}(dy,ds)\right\|_H^2\right]$$

We apply three different results to the three integrals on the right-hand side. For the first one, we use the basic properties of Bochner integrals and Jensen's inequality to obtain

$$\left\|\int_0^t \gamma_s\,ds\right\|_H^2 \leq \int_0^t \|\gamma_s\|_H^2\,ds \leq \|\gamma\|^2_{\mathrm{L}^2(0,T;H)}\,.$$

By definition of the integral with respect to H-valued Gaussian processes, and the results from [27, eqs. (4.8) and (4.10)], we have

$$E\left\|\int_0^t \sigma_s\,dW_s\right\|_H^2 \leq \mathrm{tr}(Q)\;E\int_0^t \|\sigma_s\|^2_{\mathrm{L}(U,H)}\,ds \leq \mathrm{tr}(Q)\;\|\sigma\|^2_{\mathrm{L}^2(0,T;\mathrm{L}(U,H))}\,,$$



where $\operatorname{tr}(Q)$ denotes the trace of the covariance operator of W. Finally, from Young's inequality and [41, Prop. 3.3] we get

$$\begin{aligned} E\Big\| \int_0^t \int_E \eta_s(y)\, \widetilde{M}(dy, ds) \Big\|_H^2 &\le C \int_0^t \int_E \|\eta_s(y)\|_H^2 \; \nu(dy)\, ds \\ &\le C\lambda T \int_E \|y\|_E^2 \; P^Y(dy). \end{aligned}$$

Combining the above estimates and employing Assumption 2.1 yields (2.4), since the right-hand side in each estimate is independent of t. □

Proposition 2.3. *The characteristic function of $X(t)$ is given by*

$$\begin{aligned} E\Big[e^{i\langle X_t, h\rangle_H}\Big] = \exp\Bigg[& i\Big\langle \int_0^t \gamma_s\, ds, h \Big\rangle_H - \frac{1}{2}\Big\langle \Big[\int_0^t \sigma_s Q \sigma_s^*\, ds\Big](h), h \Big\rangle_H \\ & + \int_0^t \int_E \Big(e^{i\langle \eta_s(y), h\rangle_H} - 1 - i\,\langle \eta_s(y), h\rangle_H \Big)\, \nu(dy)\, ds \Bigg] \end{aligned} \tag{2.5}$$

for every $h \in H$, where σ_s^* is the adjoint operator of σ_s.

Proof. The drift γ is deterministic, and so the first term on the right-hand side of (2.5) is trivial. Since we have finite second moments by Proposition 2.2, we may apply [56, Thm. 4] to obtain the characteristic function of the diffusion and jump parts. It remains to verify the expression for the covariance operator of the diffusion. Applying [27, Prop. 4.13] yields

$$E\Big[\Big\langle \int_0^t \sigma_s dW_s, h_1 \Big\rangle_H \Big\langle \int_0^t \sigma_s dW_s, h_2 \Big\rangle_H \Big] = \Big\langle \Big[\int_0^t \sigma_s Q \sigma_s^*\, ds \Big](h_1), h_2 \Big\rangle_H$$

for every $h_1, h_2 \in H$. The integral on the right-hand side is a Bochner integral with values in $\mathrm{L}(H, H)$. □

Moreover, we can show the existence of certain Laplace transforms of X_t. This is similar to the properties of additive processes in the finite-dimensional case presented, e.g., in [71].

Proposition 2.4. *There are constants $C_1, C_2 > 0$ such that for every $h \in H$ with $\|h\|_H \le 2$ and a.e. $t \in [0, T]$, we have*

$$\begin{aligned} E\Big[e^{\langle X_t, h\rangle_H}\Big] = \exp\Bigg[& \Big\langle \int_0^t \gamma_s\, ds, h \Big\rangle_H + \frac{1}{2}\Big\langle \Big[\int_0^t \sigma_s Q \sigma_s^*\, ds\Big](h), h \Big\rangle_H \\ & + \int_0^t \int_E \Big(e^{\langle \eta_s(y), h\rangle_H} - 1 - \langle \eta_s(y), h\rangle_H \Big)\, \nu(dy)\, ds \Bigg] \\ \le\ & C_1\, e^{C_2 T}. \end{aligned} \tag{2.6}$$

Proof. Using the Cauchy–Schwarz inequality and Assumption 2.1, we obtain

$$\int_E e^{\langle \eta_t(y), h\rangle_H}\, \nu(dy) \le \int_E e^{\|y\|_E \|h\|_H}\, \nu(dy) \le \lambda \int_E e^{2\|y\|_E}\, P^Y(dy) < \infty.$$

By [62, Th. 4.30], this is sufficient for the equality in (2.6). A theorem for interchanging linear operators and Bochner integrals [33, App. E, Th. 8] yields

$$\left\langle \int_0^t \gamma_s \, ds, h \right\rangle_H = \int_0^t \langle \gamma_s, h \rangle_H \, ds,$$
$$\left\langle \left[\int_0^t \sigma_s Q \sigma_s^* ds \right](h), h \right\rangle_H = \int_0^t \left\langle \left[\sigma_s Q \sigma_s^*\right](h), h \right\rangle_H \, ds.$$

Hence, we have the estimate

$$\begin{aligned}
&\exp\left[\left\langle \int_0^t \gamma_s \, ds, h \right\rangle_H + \frac{1}{2} \left\langle \left[\int_0^t \sigma_s Q \sigma_s^* ds \right](h), h \right\rangle_H \right. \\
&\qquad \left. + \int_0^t \int_E \left(e^{\langle \eta_s(y), h \rangle_H} - 1 - \langle \eta_s(y), h \rangle_H \right) \nu(dy) \, ds \right] \\
\leq &\exp\left[2 \int_0^T \|\gamma_s\|_H \, ds + \frac{4}{2} \int_0^T \|Q\| \, \|\sigma_s\|^2_{\mathrm{L}(U,H)} \, ds \right. \\
&\qquad \left. + T\lambda \left(\int_E e^{2\|y\|_E} \, P^Y(dy) + 1 + 2 \int_E \|y\|_E \, P^Y(dy) \right) \right] \\
\leq &\exp\left[2C \, \|\gamma\|_{\mathrm{L}^2(0,T;H)} + 2 \, \|Q\| \, \|\sigma\|^2_{\mathrm{L}^2(0,T;\mathrm{L}(U,H))} \right. \\
&\qquad \left. + T\lambda \left(\int_E e^{2\|y\|_E} \, P^Y(dy) + 1 + 2 \int_E \|y\|_E \, P^Y(dy) \right) \right]
\end{aligned}$$

This implies the statement of the proposition, again by Assumption 2.1. □

The next important step is to show that the process $(f_t)_{0 \leq t \leq T}$ takes indeed values in the Hilbert space H. Let $I = \{1, 2, \ldots, \dim(H)\}$, if H has finite dimension, or $I = \mathbb{N}$ otherwise.

Proposition 2.5. *The process $(f_t)_{0 \leq t \leq T}$, which is defined as an exponential of X_t by (2.3), satisfies $\|f_t\|_H < \infty$ almost surely. Moreover, there are constants $C_1, C_2 > 0$ such that*

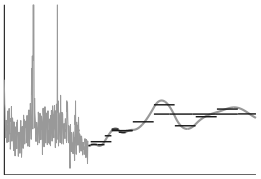
$$E \, \|f_t\|_H^2 \leq C_1 \, e^{C_2 T} \, \|f_0\|_H^2 \qquad \text{for a.e. } t \in [0, T]. \tag{2.7}$$

Proof. It is enough to show (2.7), since this implies $\|f_t\|_H < \infty$ almost surely. To this end, we use monotone convergence and calculate

$$E \, \langle f_t, f_t \rangle_H = E\left[\sum_{k \in I} \langle f_0, e_k \rangle_H^2 \; e^{2\langle X_t, e_k \rangle_H} \right] = \sum_{k \in I} \langle f_0, e_k \rangle_H^2 \, E\left[e^{\langle X_t, 2e_k \rangle_H} \right].$$

Applying Proposition 2.4 with $h = 2e_k$ yields (2.7). □

Finally, we can calculate the unique drift $\gamma \in \mathrm{L}^2(0, T; H)$ which makes all stock prices or swap rates, i.e., all bounded linear functionals of f, martingales. We define martingales on Hilbert spaces in the sense of Kunita [56]: f is considered a Hilbert space valued



martingale if and only if

$$\big(\,\langle f_t, h\rangle_H\,\big)_{t\geq 0}.$$

is a real-valued martingale for every $h \in H$.

Proposition 2.6. *The process $\big(f_t\big)_{0\leq t\leq T}$ is an H-martingale in the sense of Kunita, if and only if*

$$\gamma_t = \sum_{k\in I}\left[-\frac{1}{2}\Big\langle\big[\sigma_t Q\sigma_t^*\big](e_k), e_k\Big\rangle_H - \int_E\Big(e^{\langle\eta_t(y),e_k\rangle_H} - 1 - \langle\eta_t(y), e_k\rangle_H\Big)\nu(dy)\right]e_k \tag{2.8}$$

for a.e. $t \in [0,T]$.

Proof. By definition, f is an H-martingale if and only if

$$\big(\,\langle f_t, h\rangle_H\,\big)_{t\geq 0} = \left(\sum_{k\in I}\langle f_0, e_k\rangle_H\, e^{\langle X_t, e_k\rangle_H}\,\langle h, e_k\rangle_H\right)_{t\geq 0}$$

is a martingale for every $h \in H$. By Proposition 2.4 and the Cauchy–Schwarz inequality, we obtain

$$E\left[\sum_{k\in I}|\langle f_0, e_k\rangle_H|\; e^{\langle X_t, e_k\rangle_H}\;|\langle h, e_k\rangle_H|\right] \leq C_1\, e^{C_2 T}\,\|f_0\|_H\,\|h\|_H\,.$$

Hence, we may use dominated convergence to calculate

$$\begin{aligned}
&E\,\langle f_t, h\rangle_H \\
&\quad= \sum_{k\in I}\langle f_0, e_k\rangle_H\,\langle h, e_k\rangle_H\, E\Big[e^{\langle X_t, e_k\rangle_H}\Big] \\
&\quad= \sum_{k\in I}\langle f_0, e_k\rangle_H\,\langle h, e_k\rangle_H \exp\left[\int_0^t\big\langle\gamma_s, e_k\big\rangle_H\,ds + \frac{1}{2}\int_0^t\Big\langle\big[\sigma_s Q\sigma_s^*\big](e_k), e_k\Big\rangle_H\,ds\right. \\
&\qquad\qquad\left. + \int_0^t\int_E\Big(e^{\langle\eta_s(y),e_k\rangle_H} - 1 - \langle\eta_s(y), e_k\rangle_H\Big)\,\nu(dy)\,ds\right].
\end{aligned}$$

Consequently, the drift γ given by (2.8) makes f an H-martingale, since

$$\sum_{k\in I}\langle f_0, e_k\rangle_H\,\langle h, e_k\rangle_H = \langle f_0, h\rangle_H\,.$$

On the other hand, setting $h = e_k$ ($k \in I$) in the calculation above shows that this is indeed the only possible choice for γ. □

As we are working under the risk neutral measure, we will subsequently assume that γ is defined by (2.8), possibly modified with an additional deterministic drift due to the interest rate.

Due to the existence of second moments, the bounded linear covariance operator

$$\mathcal{C}_{X_T} : \begin{cases} H \to H' \cong H \\ h \;\mapsto E\big[\,\langle X_T - E[X_T], h\rangle_H \,\langle X_T - E[X_T], \cdot\rangle_H\,\big] \end{cases} \tag{2.9}$$

is well-defined. The following theorem gives a summary of its properties. For notational convenience, we define the centered process

$$Z_t := X_t - E[X_t] = X_t - \int_0^t \gamma_s \, ds \quad t \in [0,T]. \tag{2.10}$$

Proposition 2.7. *The operator $\mathcal{C}_{X_T}$ defined in* (2.9) *is a symmetric and nonnegative definite trace class operator (and thus compact).*

Proof. Since $\mathcal{C}_{X_T}$ is a covariance operator, it is symmetric and nonnegative definite by construction. It remains to prove that the trace of $\mathcal{C}_{X_T}$ is finite. By dominated convergence,

$$\begin{aligned} \operatorname{tr}\left(\mathcal{C}_{X_T}\right) &= \sum_{k\in I} \big\langle \mathcal{C}_{X_T} e_k, e_k \big\rangle_H = \sum_{k\in I} E\big\langle Z_T, e_k\big\rangle_H^2 \\ &= E\left[\sum_{k\in I} \big\langle Z_T, e_k\big\rangle_H^2\right] = E\big\|Z_T\big\|_H^2 \\ &< \infty \end{aligned}$$

holds. We use [27, Prop. C.3] to conclude that $\mathcal{C}_X$ is compact and hence a trace class operator. □

The following assumption states a non-vanishing volatility, which is usually assumed when pricing with PIDEs. We generalize this hypothesis to the Hilbert space valued setting. To this end, let $E_0(\mathcal{C}_{X_T})$ be the eigenspace of the covariance operator $\mathcal{C}_{X_T}$ corresponding to eigenvalue 0 (the kernel). Its orthogonal complement is $E_0(\mathcal{C}_{X_T})^\perp$. This is the subspace of H to which the centered process $\big(Z_t\big)_{0\le t\le T}$ is restricted almost surely, since

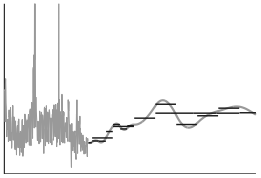
$$E\,\langle Z_t, h\rangle_H^2 = 0 \quad \text{for every } h \in E_0(\mathcal{C}_{X_T}) \text{ and a.e. } t \ge 0.$$

As before, let Q be the covariance operator of W.

Assumption 2.8. *For every $t \in [0,T]$, the restriction of the operator $\sigma_t Q \sigma_t^*$ to the subspace $E_0(\mathcal{C}_X)^\perp \subset H$ is positive definite, i.e.,*

$$\langle \sigma_t Q \sigma_t^* h, h\rangle_H > 0 \quad \textit{for every } h \in E_0(\mathcal{C}_{X_T})^\perp \backslash \{0\}.$$

This means that X has a non-vanishing Brownian component for all directions in H, which are *not* almost surely orthogonal to the trajectory of the process. This is necessary for the derivation and the regularity of the PIDE, which we are going to study in Section 4.



2.3 Corresponding Time-Homogeneous Model

For European options, the payoff depends only on the value of f (and thus X) at the time of maturity T. The characteristic function of X_T completely determines the price of European derivatives. Therefore, it is possible to construct a time-homogeneous process $\big(X_L(t)\big)_{t\geq 0}$ (the index L is for *Lévy*) which produces the same terminal distribution and thus the same price (cf. [71, Ch. 11]). In this section, we state the characteristics of this process. Since it involves time-constant coefficients, it can be used to simplify analytical calculations and, most importantly, to speed up numerical computations whenever we are not interested in path-dependent properties.

We denote the Borel sets on H by $\mathcal{B}(H)$ and the indicator function of a set $B \in \mathcal{B}(H)$ by χ_B. The Lévy–Khinchin triplet of the time-homogeneous process is given by

$$
\begin{aligned}
A_L(h_1,h_2) &= \frac{1}{T}\left\langle \left[\int_0^T \sigma_s Q \sigma_s^*\, ds\right](h_1), h_2\right\rangle_H, \\
\gamma_L &= \frac{1}{T}\int_0^T \gamma(t)dt, \\
\nu_L(B) &= \frac{1}{T}\int_0^T \int_E \chi_B\big(\eta_t(y)\big)\nu(dy)\,dt \quad \text{for } B\in\mathcal{B}(H),
\end{aligned}
\tag{2.11}
$$

where $A_L : H \times H \to \mathbb{R}$ is a bilinear covariance operator, $\gamma_L \in H$, and ν_L is a finite activity Lévy measure on H. Note that the resulting characteristic function of X_L at time $t = T$ is identical to (2.5).

In order to obtain a more explicit representation for $X_L(T)$, we define

$$
Q_L : \begin{cases} H \to H, \\ h \;\mapsto \frac{1}{T}\Big[\int_0^T \sigma_s Q \sigma_s^*\, ds\Big](h). \end{cases}
$$

This is a symmetric nonnegative definite operator with finite trace, since by construction, and by the proof of Proposition 2.2, we have

$$
\begin{aligned}
\sum_{k\in I} \langle Q_L e_k, e_k\rangle_H &= \sum_{k\in I} \frac{1}{T} E\left[\left\langle \int_0^t \sigma_s dW_s, e_k\right\rangle_H \left\langle \int_0^t \sigma_s dW_s, e_k\right\rangle_H\right] \\
&= \frac{1}{T}\, E\left\| \int_0^t \sigma_s dW_s\right\|_H^2 \leq (\operatorname{tr} Q)\, \|\sigma\|^2_{\mathrm{L}^2(0,T;\mathrm{L}(U,H))}\,.
\end{aligned}
$$

Consequently, the operator Q_L is compact by [27, Prop. C.3] and, in particular, a trace class operator. By [28, Thm. 1.2.1], there is a unique Gaussian probability measure with mean 0 and covariance operator Q_L. Moreover, there is a corresponding Q_L-Wiener process by [27, Prop. 4.2], which we denote W_L. Let further $Y_{L,i} \sim P^{Y_L}$ be iid random variables with values in H and

$$
P^{Y_L}(Y_{L,i}\in B) = \begin{cases} \frac{1}{\lambda T}\int_0^T \int_H \chi_B\big(\eta_t(y)\big)\nu(dy)\,dt & \text{if } \lambda > 0, \\ \chi_B(0) & \text{if } \lambda = 0, \end{cases}
$$

for every Borel set $B \in \mathcal{B}(H)$. We introduce the H-valued compound Poisson process

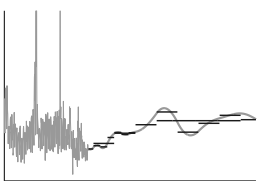
$$J_L(t) = \sum_{i=1}^{N_t} Y_{L,i}.$$

with intensity λ. The random measure corresponding to J_L is denoted by M_L, and its compensated version by $\widetilde{M}_L$. This produces exactly the same average number and height of jumps on $[0,T]$ as $\widetilde{M}$. Combined, we obtain the time-homogeneous jump-diffusion process

$$X_L(t) = \gamma_L t + W_L(t) + \int_0^t \int_H \xi \widetilde{M}_L(d\xi, ds),$$

with the same distribution as X at $t = T$. The last two summands are H-martingales. We denote the corresponding centered process by

$$Z_L(t) = W_L(t) + \int_0^t \int_H \xi \widetilde{M}_L(d\xi, ds).$$



3 Dimension Reduction

The main goal of this section is to introduce a low-dimensional approximation of the H-valued process X. To this end, we use *proper orthogonal decomposition* (POD), which is very similar to a Karhunen–Loève expansion. For an overview of POD methods in the context of deterministic differential equations, see [55]. An introduction to Karhunen–Loève expansions of stochastic processes can be found in [59, Ch. 37]. Numerical aspects, and most of the projection theory we need here, are presented in [75]. POD is also closely related to principal component analysis (which is commonly used for data analysis) and factor analysis (which uses additional error terms in the decomposition). All of these methods are based on the construction of a small set of orthogonal basis elements which can be used to approximate X in L^2. The POD components can be used to derive low-dimensional versions of PIDEs for approximate pricing and hedging.

3.1 Proper Orthogonal Decomposition (POD)

While principal component analysis and factor analysis are usually employed to analyze empirical data, we apply the POD method directly to our model. The dimension reduction takes place in the state space of the previously H-valued process. Let us now give a mathematically precise formulation of what is meant by "approximating X". Recall that $I = \{1, 2, \ldots, \dim(H)\}$, if H has finite dimension, or $I = \mathbb{N}$ otherwise.

Definition 3.1. *A sequence of orthonormal elements $(p_l)_{l\in I} \subset H$ is called a* POD basis *for X_T, if it solves the minimization problem*

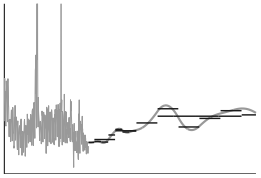
$$\min_{\langle p_i, p_j\rangle_H = \delta_{ij}} E\left\| X_T - \left(E[X_T] + \sum_{l=1}^{d} p_l \langle Z_T, p_l\rangle_H \right)\right\|_H^2$$

for every $d \in I$.

In other words, a POD basis is a set of deterministic orthonormal functions such that we expect the projection of the random vector $Z_T = X_T - E[X_T] \in H$ onto the first d elements of this basis to be a good approximation. The number d of components in the projection will later be the dimension of the approximating problem. Projection to a POD basis is equivalent to using the partial sum of the first d elements of a Karhunen–Loève expansion, which itself is closely connected to the eigenvector problem of the covariance operator $\mathcal{C}_{X_T}$ defined in (2.9). It is worth mentioning that $\mathcal{C}_{X_T}$ is also the covariance operator of the centered random variable Z_T. The following theorem shows that the eigenvectors of $\mathcal{C}_{X_T}$ are indeed the POD basis we are looking for.

Theorem 3.2. *A sequence of orthonormal eigenvectors $(p_l)_{l\in I}$ of the operator $\mathcal{C}_{X_T}$, ordered by the size of the corresponding eigenvalues $\mu_1 \geq \mu_2 \geq \ldots \geq 0$, solves the maximization problem*

$$\max_{\langle p_i, p_j\rangle_H = \delta_{ij}} \sum_{l=1}^{d} \langle \mathcal{C}_{X_T} p_l, p_l\rangle_H$$



for every $d \in I$. The maximum value is

$$\sum_{l=1}^{d} \langle \mathcal{C}_{X_T} p_l, p_l \rangle_H = \sum_{l=1}^{d} \mu_l.$$

Moreover, the eigenvectors are a POD basis in the sense of definition 3.1. The expectation of the projection error satisfies

$$E\left\| Z_T - \sum_{l=1}^{d} p_l \langle Z_T, p_l \rangle_H \right\|_H^2 = \sum_{l=d+1}^{\dim(H)} \mu_l. \tag{3.1}$$

Proof. This is an application of [75, Thm. 2.7 and Prop. 2.8]. □

An expression similar to the right-hand side of (3.1) will occur in almost every convergence theorem throughout this thesis. Subsequently, let $(p_l)_{l\in I}$ and $(\mu_l)_{l\in I}$ denote the POD basis and eigenvalues from Theorem 3.2. Further, let

$$U_d := \operatorname{span}\{p_1, \ldots, p_d\} \subset H$$

be the d-dimensional subspace spanned by the POD components corresponding to the d largest eigenvalues. We will assume that $d \leq \dim(E_0(\mathcal{C}_{X_T})^\perp)$, i.e., $\mu_1 \geq \ldots \geq \mu_d > 0$, as there is no need to include eigenvectors of the covariance operator corresponding to eigenvalue 0. Indeed, the projection of Z on $E_0(\mathcal{C}_{X_T})$ is almost surely 0. We define the projection operator

$$\mathcal{P}_d : \begin{cases} H \to U_d \cong \mathbb{R}^d, \\ z \mapsto x := \sum_{l=1}^{d} \langle z, p_l \rangle_H p_l. \end{cases} \tag{3.2}$$

Wherever appropriate, we identify U_d with $\mathbb{R}^d$ via the isometry

$$\iota : \begin{cases} \left(U_d, \|\cdot\|_H\right) \to \left(\mathbb{R}^d, \|\cdot\|\right), \\ x \mapsto \left(\langle x, p_l \rangle_H\right)_{l=1}^d. \end{cases} \tag{3.3}$$

In particular, we identify $\mathcal{P}_d z$ with the sequence $(\langle z, p_l \rangle_H)_{l=1}^d$.

So far, we have approximated the value of Z only at time T. It turns out, however, that this is indeed sufficient to obtain small projection errors for arbitrary $t \in [0, T]$.

Corollary 3.3. *The following holds:*

$$\sup_{t\in[0,T]} E \|Z_t - \mathcal{P}_d Z_t\|_H^2 \leq \sum_{l=d+1}^{\dim(H)} \mu_l. \tag{3.4}$$

Proof. This is a direct consequence of the independent increments of Z. Using the

Pythagorean theorem, we obtain

$$\begin{aligned} E\,\|Z_T - \mathcal{P}_d Z_T\|_H^2 &= E\,\|Z_t - \mathcal{P}_d Z_t + (Z_T - Z_t) - \mathcal{P}_d(Z_T - Z_t)\|_H^2 \\ &= E\,\|Z_t - \mathcal{P}_d Z_t\|_H^2 + E\,\|(Z_T - Z_t) - \mathcal{P}_d(Z_T - Z_t)\|_H^2 \\ &\geq E\,\|Z_t - \mathcal{P}_d Z_t\|_H^2 . \end{aligned}$$

Applying Proposition 3.2 yields (3.4). □

Consequently, it is not necessary to change Definition 3.1 in order to approximate every Z_t, $t \in [0,T]$. This is due to the fact that by approximating Z_T, we capture also the events up to time T. In the time-homogeneous case, we even obtain the following, t-dependent equality.

Corollary 3.4. *Suppose that Z is a time-homogeneous jump-diffusion process, i.e., σ and η in (2.1) do not depend on t. For every $t \in [0,T]$, we then have*

$$E\,\|Z_t - \mathcal{P}_d Z_t\|_H^2 = \frac{t}{T} \sum_{l=d+1}^{\dim(H)} \mu_l. \tag{3.5}$$

Proof. Due to iid increments, the covariance operator of $Z(t)$ is given by

$$\mathcal{C}_{X_t} = \frac{t}{T} \mathcal{C}_{X_T}.$$

Hence, the eigenpairs of $\mathcal{C}_{X_t}$ are given by $(\frac{t}{T}\mu_l, p_l)$, $l \in \mathbb{N}$. Applying Theorem 3.2 (setting $T = t$) yields (3.5). □

Depending on the properties of the covariance operator $\mathcal{C}_{X_T}$, bounds for the decay of the eigenvalues μ_l can be found. To this end, we define the integral kernel

$$K : \begin{cases} D \times D \to \mathbb{R}, \\ (u,v) \quad \mapsto E\big[Z_T(u) Z_T(v)\big], \end{cases} \tag{3.6}$$

where as before $D \subset \mathbb{R}^m$ and $H = \mathrm{L}^2(D, \mu_D)$. Then K is indeed the integral kernel of the covariance operator $\mathcal{C}_{X_T}$, since by Fubini's theorem

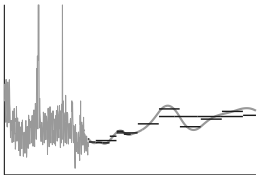
$$\int_D \int_D K(u,v) h_1(u)\, \mu_D(du)\; h_2(v)\, \mu_D(dv) = E\big[\, \langle Z_T, h_1 \rangle_H \langle Z_T, h_2 \rangle_H \,\big] = \langle \mathcal{C}_{X_T} h_1, h_2 \rangle_H$$

for every $h_1, h_2 \in H$. Consequently,

$$\mathcal{C}_{X_T} h_1(\cdot) = \int_D K(\cdot, v) h_1(v)\, \mu_D(dv).$$

Moreover, K is an element of $\mathrm{L}^2(D \times D)$, since

$$\begin{aligned} \int_D \int_D K^2(u,v)\, \mu_D(du)\, \mu_D(du) &\leq \int_D \int_D E\big[Z_T(u)^2\big] E\big[Z_T(v)^2\big]\, \mu_D(du)\, \mu_D(du) \\ &= \big[E\,\|Z_T\|_H^2\big]^2 < \infty \end{aligned}$$



by Theorem 2.2. We now give a result for the eigenvalue decay, depending on the properties of K.

Proposition 3.5. *Let $D \subset \mathbb{R}^m$ be a bounded Borel set and μ_D be the Lebesgue measure. If the kernel K is piecewise $\mathrm{H}^k \otimes \mathrm{L}^2$ on $D \times D$ for a $k \in \mathbb{N}$, then there exists a constant C such that*

$$\mu_l \leq C\, l^{-\frac{k}{m}} \quad \text{for every } l \in I.$$

Moreover, if the kernel K is piecewise analytic, then there are constants C_1, C_2 such that

$$\mu_l \leq C_1 e^{-C_2 l^{\frac{1}{m}}} \quad \text{for every } l \in I.$$

Proof. These results are shown in [75, Props. 2.18 and 2.21]. □

Remark 3.6. *A precise definition of "piecewise $\mathrm{H}^k \otimes \mathrm{L}^2$" as used in the hypothesis of Proposition 3.5 is given in [78, Def. 3.1]. The hypothesis of the proposition is fulfilled for the jump-diffusion model if the drift γ, the volatility σ, the Brownian covariance operator Q, and the jump-dampening factors η satisfy corresponding piecewise smoothness criteria.*

3.2 Numerical Computation of POD Components

As we have seen in Theorem 3.2, the construction of a POD basis for X_T can be reduced to the eigenvalue problem

$$\mathcal{C}_{X_T} p_l = \mu_l\, p_l, \quad l \in I,$$

with the covariance operator $\mathcal{C}_{X_T} : H \to H' \cong H = L^2(D, \mu_D)$ defined in (2.9). In general, the eigenvectors p_l are not known analytically. However, it is possible to compute good approximations numerically. For finite index sets D (basket options), the solution of the symmetric eigenvalue problem can be obtained with standard methods (e.g., the QR-algorithm). For all subsequent results, we will therefore assume the more complicated setting that $D \subset \mathbb{R}^m$ is a bounded Borel set and μ_D is the Lebesgue measure. This holds, e.g., for electricity swaptions (with $m = 1$). In this case, we employ a finite element discretization and solve a projected eigenvalue problem. Convergence results for this technique can be shown under certain regularity conditions on the covariance kernel K defined in (3.6). The general theory of Galerkin approximations of Karhunen–Loève expansions is discussed in [75].

Let $U_{\Delta x} \subset H = \mathrm{L}^2(D, \mu_D)$ be a finite element subspace with discretization parameter Δx. For $l = 1, \ldots, \dim U_{\Delta x}$, the Galerkin approximations $(\mu_l^{\Delta x}, p_l^{\Delta x}) \subset \mathbb{R} \times U_{\Delta x}$ of the eigenpairs $(\mu_l, p_l) \subset \mathbb{R} \times H$ are, by definition, solutions of the following problem:

$$\begin{aligned} \forall \varphi \in U_{\Delta x} : \left\langle \mathcal{C}_{X_T} p_l^{\Delta x}, \varphi \right\rangle_H &= \int_D \left(\int_D K(u,v) p_l^{\Delta x}(v)\, \mu_D(dv) \right) \varphi(u)\, \mu_D(du) \\ &\overset{!}{=} \mu_l^{\Delta x} \left\langle p_l^{\Delta x}, \varphi \right\rangle_H . \end{aligned}$$

This is equivalent to the eigenvalue problem

$$\mathcal{P}_{\Delta x}\mathcal{C}_{X_T}\mathcal{P}_{\Delta x}\, p_l^{\Delta x} = \mu_l^{\Delta x} p_l^{\Delta x}, \tag{3.7}$$

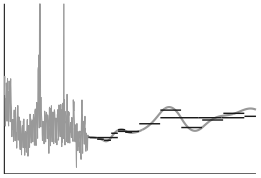
where $\mathcal{P}_{\Delta x} : H \to U_{\Delta x}$ is the projection operator onto the finite element subspace. The operator $\mathcal{P}_{\Delta x}\mathcal{C}_{X_T}\mathcal{P}_{\Delta x}$ is self-adjoint and compact due to the properties of the projection and Proposition 2.7. The following proposition gives an error bound for the approximation of Z_T obtained with this Galerkin method.

Proposition 3.7. *Let the covariance kernel K defined in* (3.6) *be a piecewise smooth function. Further, let $U_{\Delta x} \subset H$ be a finite element space of piecewise polynomials of degree $q \in \mathbb{N}$, where Δx denotes the mesh width of the regular triangulation. Finally, let $(\mu_l)_{l\in I}$ be the true eigenvalues of the covariance operator $\mathcal{C}_{X_T}$, and let $(p_l^{\Delta x})_{l\in I}$ be orthonormal solutions of the projected eigenvalue problem* (3.7)*. Then there exists a constant C such that*

$$E\left\| Z_T - \sum_{l=1}^{d} \left\langle Z_T, p_l^{\Delta x}\right\rangle_H p_l^{\Delta x} \right\|_H^2 \leq C(\Delta x)^{2q+1} + \sum_{l=d+1}^{\dim(H)} \mu_l$$

holds for every $d \leq \dim U_{\Delta x}$.

Proof. The estimate is taken from [75, Prop. 3.3]. The necessary assumption [75, Ass. 3.1] is satisfied due to the piecewise smoothness of the kernel and [78, Thm. 1.5] □



4 European Options

In this section, we derive the PIDE describing the price process of a European option. We then apply the POD method introduced in the previous section to project the PIDE. We show regularity of the low-dimensional, approximating equation, and obtain existence and uniqueness results. Bounds for the approximation error are established. We also derive the variational formulation of the projected PIDE, which is useful for theoretical considerations as well as numerical finite element schemes.

4.1 Partial Integro-Differential Equation (PIDE)

We consider a European option depending on the value of the process f, defined in (2.3), at maturity T. The terminal value f_T is a deterministic function of X_T, and X_T is distributed like its time-homogeneous analogue $X_L(T) = \gamma_L T + Z_L(T)$. Thus, we may equivalently price a European option whose payoff G is a function of $Z_L(T)$. If $z \in H$ is the value of Z_L at time t, the value V of the option at time $t \leq T$ is

$$V(t,z) := E\big[G\big(z + Z_L(T-t)\big)\big].$$

We will subsequently discount the value to time 0 and work with

$$\widehat{V}(t,z) := e^{-rT} E\big[G\big(z + Z_L(T-t)\big)\big]. \tag{4.1}$$

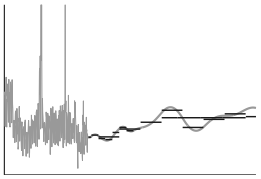
Note that $\widehat{V}$ is a martingale under the risk neutral measure. In this section, an Itô formula for Hilbert space valued random variables and the martingale property of $\widehat{V}$ are employed to derive a PIDE for $\widehat{V}$. We make the following assumption concerning the payoff of the option.

Assumption 4.1. *Suppose that the payoff function G is Lipschitz continuous on H with Lipschitz constant L_G.*

Remark 4.2. *Assumption 4.1 is not necessarily satisfied for payoffs depending on the exponential of $Z_L(T)$, e.g., a plain call option depending on f_T. However, this can be remedied easily. In the specific case of a call, we can apply a put-call parity. More generally, every payoff can be truncated to a bounded domain (e.g., by multiplying with a smooth cutoff function). A payoff function has finite expectation; hence the error introduced by truncation is arbitrarily small. Since we have to localize the computational domain for any numerical calculation anyway (compare Section 8), Assumption 4.1 is no substantial restriction.*

Let us first recall the definition of derivatives on a Hilbert space (see, e.g., [31, Ch. VIII]). We denote the first and second Fréchet derivative of $\widehat{V}$ at $(t,z) \in [0,T] \times H$ by $D_z\widehat{V}(t,z) \in L(H,\mathbb{R})$ and $D_z^2\widehat{V}(t,z) \in L(H,H)$ respectively. These are continuous linear operators such that

$$\widehat{V}(t,z+\zeta) = \widehat{V}(t,z) + D_z\widehat{V}(t,z)\,\zeta + \frac{1}{2}\left\langle D_z^2\widehat{V}(t,z)\,\zeta,\zeta\right\rangle_H + o(\|\zeta\|_H^2)$$



for every $\zeta \in H$. The partial derivative with respect to time is denoted $\partial_t \widehat{V}(t,z)$. It is often convenient to identify $D_z^2 \widehat{V}(t,z)$ with a bilinear form on $H \times H$, setting

$$D_z^2 \widehat{V}(t,z)(\zeta_1,\zeta_2) := \left\langle D_z^2 \widehat{V}(t,z)\,\zeta_1, \zeta_2 \right\rangle_H .$$

If $\widehat{V}$ is Fréchet differentiable, then the Gâteaux derivatives

$$\partial_\zeta \widehat{V}(t,z) := \lim_{\varepsilon \to 0} \frac{\widehat{V}(t,z+\varepsilon\zeta) - \widehat{V}(t,z)}{\varepsilon}$$

are also well-defined for every $\zeta \in H$. They satisfy

$$\partial_\zeta \widehat{V}(t,z) = D_z \widehat{V}(t,z)\,\zeta.$$

If, on the other hand, $\widehat{V}$ has linear and continuous Gâteaux derivatives, and the mapping $z \mapsto \partial_\cdot \widehat{V}(t,z) \in \mathrm{L}(H,\mathbb{R})$ is continuous, then $\widehat{V}(t,z)$ is continuously Fréchet differentiable in z.

We assume the following regularity condition for $\widehat{V}$, which is, in particular, a prerequisite for Itô's formula. In the context of pricing with PIDEs, it is common to assume that $\widehat{V}$ is twice continuously differentiable, see, e.g., [24, 49, 61]. Note, however, that this hypothesis is not necessary for the convergence results in Section 4.2.

Assumption 4.3. *Suppose that $\widehat{V} \in C^{1,2}((0,T)\times H,\mathbb{R}) \cap C([0,T]\times H,\mathbb{R})$; i.e., $\widehat{V}$ is continuously differentiable with respect to t and twice continuously differentiable with respect to z. Moreover, assume that the operator norms $\left\|D_z^2 \widehat{V}(t,z)\right\|$, $\left\|D_z \widehat{V}(t,z)\right\|$, and $\left\|\partial_t \widehat{V}(t,z)\right\|$ are bounded.*

As a direct consequence of this assumption, V satisfies the Lipschitz condition

$$\left|\widehat{V}(t,z) - \widehat{V}(t,z+\zeta)\right| \le L_V \left\|\zeta\right\|_H \quad \text{for every } \zeta \in H$$

with the constant $L_V := \sup_{(s,\widetilde{z})\in[0,T]\times H} \left\|D_z \widehat{V}(s,\widetilde{z})\right\|$. We are now able to calculate the stochastic dynamics of $\widehat{V}$ using Itô's formula.

Theorem 4.4. *The discounted price $\widehat{V}$ of a European option, given by* (4.1)*, satisfies*

$$\begin{aligned} d\widehat{V}\big(t, Z_L(t)\big) = {} & \\ & \partial_t \widehat{V}\big(t, Z_L(t-)\big)\,dt + \frac{1}{2}\operatorname{tr}\Big(D_z^2 \widehat{V}\big(t, Z_L(t-)\big)\,Q_L\Big)\,dt \\ & + \int_H \Big[\widehat{V}\big(t, Z_L(t-)+\zeta\big) - \widehat{V}\big(t, Z_L(t-)\big) - D_z \widehat{V}\big(t, Z_L(t-)\big)\,\zeta\Big]\,\nu_L(d\zeta)\,dt \\ & + D_z \widehat{V}\big(t, Z_L(t-)\big)\,dW_L(t) + \int_H \Big[\widehat{V}\big(t, Z_L(t-)+\zeta\big) - \widehat{V}\big(t, Z_L(t-)\big)\Big]\,\widetilde{M}_L(d\zeta, dt). \end{aligned} \tag{4.2}$$

Proof. By Itô's formula [56, Thm. 3], we obtain

$$\begin{aligned}\widehat{V}\big(t, Z_L(t)\big) &= \widehat{V}\big(0, Z_L(0)\big) + \int_0^t \partial_t \widehat{V}\big(s, Z_L(s-)\big) ds \\ &\quad + \frac{1}{2}\Big\langle\!\!\Big\langle \int_0^{\cdot} D_z^2 \widehat{V}\big(s, Z_L(s-)\big)\, dW_L(s)\,;\, W_L(\cdot)\Big\rangle\!\!\Big\rangle_t \\ &\quad + \int_0^t D_z \widehat{V}\big(s, Z_L(s-)\big)\, dZ_L(s) \\ &\quad + \sum_{0\le s\le t} \Big[\widehat{V}\big(s, Z_L(s-) + \Delta Z_L(s)\big) - \widehat{V}\big(s, Z_L(s-)\big) \\ &\qquad\qquad - D_z \widehat{V}\big(s, Z_L(s-)\big)\, \Delta Z_L(s)\Big],\end{aligned} \tag{4.3}$$

where $\langle\!\langle X\,;\,Y\rangle\!\rangle$ denotes the predictable quadratic covariation of $\langle X, Y\rangle_H$, and $\Delta Z_L(s) = Z_L(s) - Z_L(s-)$ is the jump height of Z_L at time s.

We first calculate the covariation. From [27, Cor. 4.14], we know that

$$\begin{aligned}E\Big\langle \int_{t_1}^{t_2} D_z^2 \widehat{V}\big(s, Z_L(s-)\big) dW_L(s), W_L(t_2) - W_L(t_1)\Big\rangle_H \\ = \int_{t_1}^{t_2} \operatorname{tr}\Big(D_z^2 \widehat{V}\big(s, Z_L(s-)\big)\, Q_L\Big) ds\end{aligned}$$

for every $0 \le t_1 \le t_2$. Consequently, by independence of the increments of W_L, we find

$$\Big\langle\!\!\Big\langle \int_0^{\cdot} D_z^2 \widehat{V}\big(s, Z_L(s-)\big)\, dW_L(s)\,;\, W_L(\cdot)\Big\rangle\!\!\Big\rangle_t = \int_0^t \operatorname{tr}\Big(D_z^2 \widehat{V}\big(s, Z_L(s-)\big)\, Q_L\Big) ds.$$

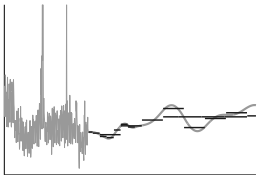
For the next term in (4.3), we use the dynamics of Z_L to obtain

$$\begin{aligned}&\int_0^t D_z \widehat{V}\big(s, Z_L(s-)\big) dZ_L(s) \\ &\qquad = \int_0^t D_z \widehat{V}\big(s, Z_L(s-)\big) d\Big[W_L(s) + \int_0^s \int_H y \widetilde{M}_L(dy, ds_2)\Big] \\ &\qquad = \int_0^t D_z \widehat{V}\big(s, Z_L(s-)\big)\, dW_L(s) \\ &\qquad\quad + \int_0^t D_z \widehat{V}\big(s, Z_L(s-)\big)\, d\Big[\sum_{0\le s_2\le s} \Delta Z_L(s_2) - \int_0^s \int_H y\, \nu_L(d\zeta)\, ds_2\Big].\end{aligned}$$

A theorem for interchanging linear operators and Bochner integrals [33, App. E, Thm. 8] yields

$$\int_0^t D_z \widehat{V}\big(s, Z_L(s-)\big)\, d\Big[\int_0^s \int_H y\, \nu_L(dy)\, ds_2\Big] = \int_0^t \int_H D_z \widehat{V}\big(s, Z_L(s-)\big)\, y\, \nu_L(dy)\, ds.$$

It remains to reorganize the jump terms in (4.3). Due to the bounded derivatives of the



price function $\widehat{V}$ (Assumption 4.3), the following holds:

$$\int_0^t D_z\widehat{V}\big(s, Z_L(s-)\big) \int_H z\, \widetilde{M}_L(dz, ds)$$
$$= \sum_{0\le s\le t} D_z\widehat{V}\big(s, Z_L(s-)\big)\big(Z_L(s) - Z_L(s-)\big) - \int_0^t \int_H D_z\widehat{V}\big(s, Z_L(s-)\big)\, z\, \nu_L(dz)ds.$$

Moreover, we have

$$\sum_{0\le s\le t} \Big[\widehat{V}\big(s, Z_L(s)\big) - \widehat{V}\big(s, Z_L(s-)\big)\Big]$$
$$= \int_0^t \int_H \Big[\widehat{V}\big(s, Z_L(s-) + z\big) - \widehat{V}\big(s, Z_L(s-)\big)\Big]\, \widetilde{M}_L(dz, ds)$$
$$+ \int_0^t \int_H \Big[\widehat{V}\big(s, Z_L(s-) + z\big) - \widehat{V}\big(s, Z_L(s-)\big)\Big]\, \nu_L(dz)\, ds.$$

Combining all of the results with the Itô dynamics (4.3) completes the proof. □

In order to get a slightly more explicit form of the trace expression in (4.2), we insert the orthonormal basis $(e_k)_{k\in I}$ of H into the definition of the trace. This yields

$$\operatorname{tr}\Big(D_z^2\widehat{V}\big(t, Z_L(t-)\big)\, Q_L\Big) = \sum_{k\in I} D_z^2\widehat{V}\big(t, Z_L(t-)\big)\,\big(Q_L p_k, p_k\big).$$

Theorem 4.5. *The discounted price $\widehat{V}$ of a European option with payoff $G(Z_L(T))$ at maturity T satisfies the PIDE*

$$\begin{aligned} -\partial_t\widehat{V}(t,z) &= \frac{1}{2}\sum_{k\in I} D_z^2\widehat{V}(t,z)\,(Q_L p_k, p_k) \\ &\quad + \int_H \Big[\widehat{V}(t, z+\zeta) - \widehat{V}(t,z) - D_z\widehat{V}(t,z)\,\zeta\Big]\nu_L(d\zeta), \end{aligned} \tag{4.4}$$

with terminal condition

$$\widehat{V}(T,z) = e^{-rT}G(z),$$

for every $t\in(0,T)$, $z\in E_0(\mathcal{C}_{X_T})^\perp$.

Proof. We employ Theorem 4.4. The penultimate term in (4.2),

$$\int_0^t D_z\widehat{V}\big(s, Z_L(s-)\big)\, dW_L(s),$$

is a martingale by [27, Thm. 4.12]. In order to show that the integral with respect to the compensated Poisson measure is a martingale too, we apply [69, Thm. 3.11]. The prerequisite for this theorem is a strong integrability condition, which is satisfied due to

[69, Thm. 3.12], since

$$\int_0^t \int_H E\left|\widehat{V}\big(s, Z_L(s-)+\zeta\big) - \widehat{V}\big(s, Z_L(s-)\big)\right| \nu_L(d\zeta)ds \leq t \int_H L_V \left\|\zeta\right\|_H \nu_L(d\zeta) < \infty.$$

The remaining integral terms in (4.2) are continuous in t and of finite variation. Consequently, the martingale property of $\widehat{V}$, together with the fact that continuous martingales of finite variation are almost surely constant [65, Th. 27], yields the PIDE along almost every trajectory of Z_L.

It remains to show that the PIDE indeed holds for arbitrary $t \in (0,T)$ and $z \in E_0(\mathcal{C}_{X_T})^\perp$. We denote by $B_\varepsilon(z)$ the ball with radius ε around z. Due to the non-vanishing diffusion (Assumption 2.8), the probability for $Z_L(t) \in B_\varepsilon(z)$ is non-zero for every $\varepsilon > 0$. Thus, for every $\varepsilon > 0$, we can find a $z_\varepsilon \in B_\varepsilon(z)$ such that the PIDE holds in (t, z_ε). Due to the regularity of $\widehat{V}$ (Assumption 4.3), we can conclude that the PIDE holds for every $(t,z) \in (0,T) \times E_0(\mathcal{C}_{X_T})^\perp$. □

4.2 Pricing European Options with POD

We use the POD projection operator $\mathcal{P}_d$ defined in (3.2) to construct the finite-dimensional approximation

$$\widehat{V}_d(t,x) := e^{-rT} E\big[G\big(x + \mathcal{P}_d Z_L(T-t)\big)\big] \tag{4.5}$$

for $x \in U_d \cong \mathbb{R}^d$. In this section, we show that $\widehat{V}_d$ is a good approximation of $\widehat{V}$. Moreover, we derive the finite-dimensional PIDE solved by $\widehat{V}_d$ and verify that this solution is unique.

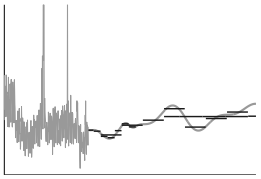
4.2.1 Finite Dimensional PIDE

Before we can derive the finite-dimensional PIDE for $\widehat{V}_d$, we have to deal with the question of differentiability. The following theorem shows that Assumption 4.1 is actually enough to recover regularity of V_d.

Theorem 4.6. *The finite-dimensional approximation $\widehat{V}_d : [0,T] \times U_d \to \mathbb{R}$ defined in (4.5) satisfies $\widehat{V}_d \in C^{1,2}((0,T) \times U_d, \mathbb{R}) \cap C([0,T] \times U_d, \mathbb{R})$. Moreover, the partial derivatives $\partial_{x_i}\widehat{V}_d(t,x)$, $\partial_{x_i}\partial_{x_j}\widehat{V}_d(t,x)$, and $\partial_t \widehat{V}_d(t,x)$ are functions of at most linear growth in $\|x\|$ for $i,j = 1,\dots,d$.*

Proof. The first step of the proof is to show the existence of a smooth density for the random variable $\mathcal{P}_d Z_L(t)$ for arbitrary $t \geq 0$. To achieve this, a fast decay condition for its characteristic function

$$\widehat{\mu}_t(x) := E\Big(e^{i\langle x, \mathcal{P}_d Z_L(t)\rangle_{U_d}}\Big) \in \mathbb{C}, \quad x \in U_d,$$



is useful. We have

$$E\Big[\big\langle \mathcal{P}_d W_L(t), x_1\big\rangle_{U_d}\big\langle \mathcal{P}_d W_L(t), x_2\big\rangle_{U_d}\Big] = \sum_{k,l=1}^{d} \langle x_1, p_k\rangle_{U_d}\langle x_2, p_l\rangle_{U_d} E\Big[\big\langle W_L(t), p_k\big\rangle_H \big\langle W_L(t), p_l\big\rangle_H\Big] = \langle Q_L x_1, x_2\rangle_H$$

for every $x_1, x_2 \in U_d$. Thus, the covariance operator for the diffusion part of $\mathcal{P}_d Z_L$ is given by $\mathcal{P}_d Q_L \mathcal{P}_d$. The same arguments as in the proof of Proposition 2.3 yield

$$\widehat{\mu}_t(x) = \exp\bigg(-\frac{1}{2} t \big\langle \mathcal{P}_d Q_L x, x\big\rangle_{U_d} + \int_0^t \int_H \Big(e^{i\langle \mathcal{P}_d \xi, x\rangle_{U_d}} - 1 - i\,\langle \mathcal{P}_d \xi, x\rangle_{U_d}\Big)\,\nu_L(d\xi) ds\bigg)$$

for every $x \in U_d$. Using Assumptions 2.1 and 2.8, this implies

$$\begin{aligned} |\widehat{\mu}_t(x)| &\le \exp\bigg(-\frac{1}{2} t\,\langle \mathcal{P}_d Q_L x, x\rangle_H \\ &\qquad + \int_0^t \int_H \big| e^{i\langle \mathcal{P}_d \xi, x\rangle_H} - 1 - i\,\langle \mathcal{P}_d \xi, x\rangle_H\big|\,\nu_L(d\xi) ds\bigg) \\ &\le \exp\bigg[t\left(-\frac{1}{2} C_1 \|x\|^2 + C_2 \|x\| + C_3\right)\bigg], \end{aligned}$$

with positive constants C_1, C_2, and C_3 depending on d. In particular, we have

$$\lim_{\|x\|\to\infty} \|x\|^n\, \widehat{\mu}_t(x) = 0 \quad \text{for every } n \in \mathbb{N}.$$

Similarly, we obtain

$$|\partial_x^\alpha \widehat{\mu}_t(x)| \le p_\alpha(t, \|x\|)\,|\widehat{\mu}_t(x)| \quad \text{and} \quad |\partial_x^\alpha\, \partial_t \widehat{\mu}_t(x)| \le q_\alpha(t, \|x\|)\,|\widehat{\mu}_t(x)|$$

for every multiindex $\alpha \in \mathbb{N}_0^d$, where p_α and q_α are polynomials. Consequently, for every $t \in (0,T)$, $\widehat{\mu}_t$ and $\partial_t \widehat{\mu}_t$ are elements of the Schwartz space

$$\mathcal{S}(\mathbb{R}^d) = \Big\{ f \in C^\infty(\mathbb{R}^d) : \lim_{\|x\|\to\infty} \|x\|^n\, \partial_x^\alpha f(x) = 0 \text{ for every } \alpha \in \mathbb{N}_0^d, n \in \mathbb{N}_0\Big\}.$$

From [71, Prop. 28.1], we know that $\mathcal{P}_d Z_L(t)$ has a density $g_t \in C^\infty(\mathbb{R}^d)$, given by

$$g_t(y) := (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle y, x\rangle} \widehat{\mu}_t(x)\, dx.$$

Moreover, by the properties of $\widehat{\mu}_t$ and [80, Thm. V.2.8], we obtain $g_t \in \mathcal{S}(\mathbb{R}^d)$ and $\partial_t g_t \in \mathcal{S}(\mathbb{R}^d)$. Finally, note that $\widehat{V}_d$ can be written as a convolution of the payoff and the density:

$$\widehat{V}_d(t,x) = e^{-rT} \int_{\mathbb{R}^d} G(x+y) g_{T-t}(y) dy = e^{-rT} \int_{\mathbb{R}^d} G(y) g_{T-t}(y-x) dy.$$

Due to Assumption 4.1, we have $|G(y)| \leq |G(0)| + L_G \|y\|$. Hence, for $x \in U_d$, $t \in (0,T)$, we may compute

$$\partial_x^\alpha \widehat{V}_d(t,x) = -e^{-rT} \int_{\mathbb{R}^d} G(x+y)\, \partial_y^\alpha g_{T-t}(y)\, dy$$

and

$$\partial_t \widehat{V}_d(t,x) = -e^{-rT} \int_{\mathbb{R}^d} G(x+y)\, \partial_t g_{T-t}(y)\, dy$$

for every $\alpha \in \mathbb{N}_0^d$. This proves continuity of the derivatives. In addition, we obtain

$$\begin{aligned}\left|\partial_x^\alpha \widehat{V}_d(t,x)\right| &\leq e^{-rT} \int_{\mathbb{R}^d} |G(y+x)|\, |\partial_x^\alpha g_{T-t}(y)|\, dy \\ &\leq \int_{\mathbb{R}^d} \big(\, |G(0)| + L_G \|x\| + L_G \|y\| \,\big)\, |\partial_x^\alpha g_{T-t}(y)|\, dy\end{aligned}$$

for every $\alpha \in \mathbb{N}_0^d$. Similarly,

$$\begin{aligned}\left|\partial_t \widehat{V}_d(t,x)\right| &= e^{-rT} \left| \int_{\mathbb{R}^d} G(y) \partial_t g_{T-t}(y-x) dy \right| \\ &\leq \int_{\mathbb{R}^d} \big(\, |G(0)| + L_G \|x\| + L_G \|y\| \,\big)\, |\partial_t g_{T-t}(y)|\, dy.\end{aligned}$$

Thus, the growth condition is shown. It remains to prove that $\widehat{V}_d$ is also continuous for $t \to T$. This is, however, a direct consequence of the fact that

$$\lim_{t \to 0} E \|Z_L(t)\|_H = 0,$$

and thus

$$\begin{aligned}\left|\widehat{V}_d(t,x) - \widehat{V}_d(T,x)\right| &\leq e^{-rT} E \left|G\big(x + \mathcal{P}_d Z_L(T-t)\big) - G(x)\right| \\ &\leq C\, E \|Z_L(T-t)\|_H \to 0 \quad \text{for } t \to T.\end{aligned}$$

□

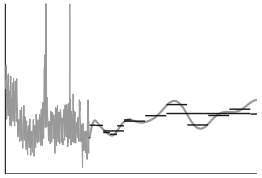
Theorem 4.7. *The function $\widehat{V}_d$ defined in (4.5) is a classical solution of the finite-dimensional PIDE*

$$\begin{aligned}-\partial_t \widehat{V}_d(t,x) &= \frac{1}{2} \sum_{l=1}^{d} D_x^2 \widehat{V}_d(t,x)\, (\mathcal{P}_d Q_L p_l, \mathcal{P}_d p_l) \\ &\quad + \int_H \Big[\widehat{V}_d(t, x + \mathcal{P}_d \zeta) - \widehat{V}_d(t,x) - D_x \widehat{V}_d(t,x)\, \mathcal{P}_d \zeta \Big] \nu_L(d\zeta),\end{aligned} \tag{4.6}$$

with terminal condition

$$\widehat{V}_d(T,x) = \widehat{V}(T,x) = e^{-rT} G(x),$$

for $t \in (0,T)$, $x \in U_d$.



Proof. The stochastic dynamics of $\mathcal{P}_d Z_L(t)$ are given by

$$d\big(\mathcal{P}_d Z_L(t)\big) = d\big(\mathcal{P}_d W_L(t)\big) + \int_H \mathcal{P}_d\zeta\, \widetilde{M}_L(d\zeta, dt).$$

The process $\mathcal{P}_d W_L(t) = \sum_{i=1}^d \langle W_L(t), p_l\rangle_H p_l$ is a d-dimensional Wiener process with correlation operator $\mathcal{P}_d Q_L \mathcal{P}_d$. The integral with respect to $\widetilde{M}_L$ can easily be rewritten as an integral over U_d, since the integrand depends only on the projection $\mathcal{P}_d\zeta \in U_d$.

We apply the finite-dimensional version of Itô's formula (cf., e.g., [24, Thm. 8.18]) to $\widehat{V}_d\big(t, \mathcal{P}_d Z_L(t)\big)$. In contrast to the Hilbert space valued case, bounded derivatives are not needed here. The properties of $\widehat{V}_d$ shown in Theorem 4.6 are sufficient. By the same arguments as in the proof of Theorem 4.4, we obtain the following:

$$\begin{aligned}
&d\widehat{V}_d\big(t, \mathcal{P}_d Z_L(t)\big) = \\
&\quad \partial_t \widehat{V}_d\, dt + \frac{1}{2}\operatorname{tr}\Big(D_x^2 \widehat{V}_d\, \mathcal{P}_d Q_L \mathcal{P}_d\Big)\, dt \\
&\quad + \int_H \Big[\widehat{V}_d\big(t, \mathcal{P}_d Z_L(t-) + \mathcal{P}_d\zeta\big) - \widehat{V}_d\big(t, \mathcal{P}_d Z_L(t-)\big) - D_x\widehat{V}_d\, \mathcal{P}_d\zeta\Big]\, \nu_L(d\zeta)\, dt \\
&\quad + D_x\widehat{V}_d\, d\big(\mathcal{P}_d W_L(t)\big) + \int_H \Big[\widehat{V}_d\big(t, \mathcal{P}_d Z_L(t-) + \mathcal{P}_d\zeta\big) - \widehat{V}_d\big(t, \mathcal{P}_d Z_L(t-)\big)\Big]\, \widetilde{M}_L(d\zeta, dt).
\end{aligned}$$

Proceeding exactly as in the proof of Theorem 4.5, we obtain (4.6). □

The PIDE in Theorem 4.7 is of course nothing more than a projected version of the PIDE (4.4) for $\widehat{V}$. The derivatives $D_x\widehat{V}_d(t,x) \in \mathrm{L}(\mathbb{R}^d, \mathbb{R})$ and $D_x^2\widehat{V}_d(t,x) \in \mathrm{L}(\mathbb{R}^d, \mathbb{R}^d)$ can be interpreted as a vector and a matrix, respectively. In particular, we have

$$\begin{aligned}
\sum_{l=1}^d D_x^2\widehat{V}_d(t,x)\,(\mathcal{P}_d Q_L p_l, \mathcal{P}_d p_l) &= \sum_{l=1}^d \sum_{i,j=1}^d \partial_{x_i}\partial_{x_j}\widehat{V}_d(t,x)\,\langle \mathcal{P}_d Q_L p_l, p_j\rangle_H\,\langle \mathcal{P}_d p_l, p_i\rangle_H \\
&= \sum_{i,j=1}^d \langle Q_L p_i, p_j\rangle_H\, \partial_{x_i}\partial_{x_j}\widehat{V}_d(t,x).
\end{aligned}$$

To simplify notation, we define coefficients

$$a_{ij} := \frac{1}{2}\langle Q_L p_i, p_j\rangle_H, \quad i,j = 1,\dots,d.$$

The PIDE can then be written as

$$\begin{aligned}
-\partial_t\widehat{V}_d(t,x) &= \sum_{i,j=1}^d a_{ij}\,\partial_{x_i}\partial_{x_j}\widehat{V}_d(t,x) \\
&\quad + \int_H \Big[\widehat{V}_d(t, x+\mathcal{P}_d\zeta) - \widehat{V}_d(t,x) - \sum_{i=1}^d \langle \zeta, p_i\rangle_H\, \partial_{x_i}\widehat{V}_d(t,x)\Big]\nu_L(d\zeta).
\end{aligned} \tag{4.7}$$

Moreover, we have the following ellipticity property.

Proposition 4.8. *The matrix $(a_{ij})_{i,j=1}^d$ is symmetric positive definite.*

Proof. The symmetry is obvious, since Q_L is self-adjoint. For arbitrary $y \in \mathbb{R}^d$, we find

$$\sum_{i,j=1}^d y_i\, a_{ij}\, y_j = \frac{1}{2}\left\langle Q_L \sum_{i=1}^d y_i p_i, \sum_{i=1}^d y_i p_i \right\rangle_H = \frac{1}{2}\frac{1}{T}\int_0^T \left\langle \sigma_s Q \sigma_s^* \sum_{i=1}^d y_i p_i, \sum_{i=1}^d y_i p_i \right\rangle_H ds.$$

Since $(p_l)_{l=0}^d$ are eigenvectors of $\mathcal{C}_{X_T}$ corresponding to strictly positive eigenvalues, we have $\sum_{i=1}^d y_i p_i \in E_0(\mathcal{C}_{X_T})^\perp$. In addition, $\left\|\sum_{i=1}^d y_i p_i\right\|_H = \|y\| > 0$ due to the isometry (3.3). Thus, we can conclude that $(a_{ij})_{i,j=1}^d$ is positive definite by Assumption 2.8. □

4.2.2 Variational Formulation and Uniqueness

We have already shown that the approximation $\widehat{V}_d(t,x)$ is a classical solution of the finite-dimensional PIDE (4.7). In this section, we introduce the corresponding variational formulation in appropriate Hilbert spaces and show uniqueness of the weak solution. Since the payoff is not necessarily bounded, and thus not an element of $\mathrm{L}^2(\mathbb{R}^d)$, we use weighted Sobolev spaces instead. Let ρ_θ be the weight function with exponential decay defined by

$$\rho_\theta : \begin{cases} \mathbb{R}^d \to \mathbb{R}, \\ x \;\mapsto e^{-\theta\sqrt{1+\|x\|^2}}, \end{cases}$$

with a parameter $\theta > 0$. We define the scalar products

$$\langle \psi, \varphi \rangle_{\mathrm{L}^{2,\theta}} := \int_{\mathbb{R}^d} \psi(x)\varphi(x)\rho_\theta(x)\, dx$$

and

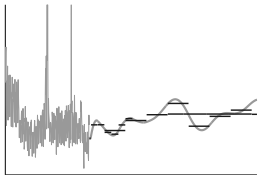
$$\langle \psi, \varphi \rangle_{\mathrm{H}^{k,\theta}} := \sum_{\alpha \in \mathbb{N}_0^d, |\alpha| \le k} \langle \partial^\alpha \psi, \partial^\alpha \varphi \rangle_{\mathrm{L}^{2,\theta}}$$

for functions $\psi, \varphi : \mathbb{R}^d \to \mathbb{R}$. The corresponding Hilbert spaces are denoted by $\mathrm{L}^{2,\theta}(\mathbb{R}^d)$ and $\mathrm{H}^{k,\theta}(\mathbb{R}^d)$ (cf., e.g., [8, Chap. 3.1]). In particular, we consider the Gelfand triple

$$\mathrm{H}^{1,\theta}(\mathbb{R}^d) \hookrightarrow \mathrm{L}^{2,\theta}(\mathbb{R}^d) \hookrightarrow \left(\mathrm{H}^{1,\theta}(\mathbb{R}^d)\right)'.$$

Finally, we define the bilinear form

$$\begin{aligned} a(\psi,\varphi) := &\int_{\mathbb{R}^d} \sum_{i,j=1}^d a_{ij}\, \partial_{x_i}\psi(x)\, \partial_{x_j}\varphi(x)\, \rho_\theta(x)\, dx + \int_{\mathbb{R}^d} \sum_{i=1}^d b_i(x)\, \partial_{x_i}\psi(x)\, \varphi(x)\, \rho_\theta(x)\, dx \\ &- \int_H \int_{\mathbb{R}^d} \left[\psi(x + \mathcal{P}_d\zeta) - \psi(x) - \sum_{i=1}^d \langle \zeta, p_i \rangle_H\, \partial_{x_i}\psi(x)\right] \varphi(x)\, \rho_\theta(x)\, dx\, \nu_L(d\zeta) \end{aligned} \tag{4.8}$$



for $\psi, \varphi \in \mathrm{H}^{1,\theta}(\mathbb{R}^d)$, where

$$b_i(x) := -\frac{\theta \sum_{j=1}^d a_{ij} x_j}{\sqrt{1 + \|x\|^2}} \quad \text{for } i = 1, \dots, d.$$

The coefficients $b_i(x)$ satisfy

$$\big|b_i(x)\big| \le \theta\, d \max_{j=1,\dots,d} |a_{ij}|$$

and are therefore bounded on $\mathbb{R}^d$. We can now state a variational form of (4.7).

Proposition 4.9. *The function $\widehat{V}_d$ defined in* (4.5) *satisfies*

$$-\Big\langle \partial_t \widehat{V}_d(t,\cdot), \varphi \Big\rangle_{\mathrm{L}^{2,\theta}} + a(\widehat{V}_d(t,\cdot), \varphi) = 0 \tag{4.9}$$

for every $\varphi \in \mathrm{H}^{1,\theta}(\mathbb{R}^d)$ and every $t \in (0,T)$, with terminal condition

$$\widehat{V}_d(T,x) = G(x). \tag{4.10}$$

Proof. We first note that $\widehat{V}_d(t,\cdot) \in \mathrm{H}^{2,\theta}(\mathbb{R}^d)$ and $\partial_t \widehat{V}_d(t,\cdot) \in \mathrm{L}^{2,\theta}(\mathbb{R}^d)$ hold for every $t \in (0,T)$ due to Theorem 4.6. Starting with (4.7), partial integration yields

$$\begin{aligned}
&-\Big\langle \partial_t \widehat{V}_d(t,x), \varphi \Big\rangle_{\mathrm{L}^{2,\theta}} \\
&\quad = -\sum_{i,j=1}^d a_{ij} \int_{\mathbb{R}^d} \partial_{x_i} \widehat{V}_d(t,x)\, \partial_{x_j}(\varphi \rho_\theta)(x)\, dx \\
&\qquad + \int_H \int_{\mathbb{R}^d} \Big[\widehat{V}_d(t, x + \mathcal{P}_d \zeta) - \widehat{V}_d(t,x) - \sum_{i=1}^d \langle \zeta, p_i \rangle_H\, \partial_{x_i} \widehat{V}_d(t,x) \Big] \varphi(x) \rho_\theta(x)\, dx\, \nu_L(d\zeta).
\end{aligned}$$

Using the product rule, we obtain

$$\sum_{j=1}^d a_{ij} \partial_{x_j}(\varphi \rho_\theta)(x) = \sum_{j=1}^d a_{ij}\, \partial_{x_j} \varphi(x)\, \rho_\theta(x) + \varphi(x)\, b_i(x)\, \rho_\theta(x).$$

□

Proposition 4.10. *The bilinear form a defined in* (4.8) *is continuous and satisfies Gårding's inequality. More precisely, there are constants $C > 0$, $c_1 \ge 0$, and $c_2 > 0$ (possibly depending on d) such that*

$$|a(\psi, \varphi)| \le C \, \|\psi\|_{\mathrm{H}^{1,\theta}} \, \|\varphi\|_{\mathrm{H}^{1,\theta}} \tag{4.11}$$

and

$$a(\psi,\psi) + c_1 \, \|\psi\|^2_{\mathrm{L}^{2,\theta}} \ge c_2 \, \|\psi\|^2_{\mathrm{H}^{1,\theta}} \tag{4.12}$$

hold for every $\psi, \varphi \in \mathrm{H}^{1,\theta}(\mathbb{R}^d)$.

Proof. First, we show continuity. From the definition of a, we obtain

$$\begin{aligned}
&|a(\psi,\varphi)| \\
&\quad\le \sum_{i,j=1}^{d} |a_{ij}| \int_{\mathbb{R}^d} \big|\partial_{x_i}\psi(x)\big|\big|\partial_{x_j}\varphi(x)\big|\rho_\theta(x)\,dx + \sum_{i=1}^{d} \int_{\mathbb{R}^d} \big|b_i(x)\big|\big|\partial_{x_i}\psi(x)\big|\big|\varphi(x)\big|\rho_\theta(x)\,dx \\
&\qquad + \int_H \int_{\mathbb{R}^d} \big|\psi(x+\mathcal{P}_d\zeta)\big|\big|\varphi(x)\big|\rho_\theta(x)\,dx\,\nu_L(d\zeta) + \int_H \int_{\mathbb{R}^d} \big|\psi(x)\big|\big|\varphi(x)\big|\rho_\theta(x)\,dx\,\nu_L(d\zeta) \\
&\qquad + \sum_{i=1}^{d} \int_H \int_{\mathbb{R}^d} \big|\langle\zeta,p_i\rangle_H\big|\big|\partial_{x_i}\psi(x)\big|\big|\varphi(x)\big|\rho_\theta(x)\,dx\,\nu_L(d\zeta).
\end{aligned}$$

The Cauchy–Schwarz inequality yields

$$\begin{aligned}
|a(\psi,\varphi)| &\le \max_{i,j=1,\dots,d} |a_{ij}| \sum_{i,j=1}^{d} \big\|\partial_{x_i}\psi\big\|_{\mathrm{L}^{2,\theta}} \big\|\partial_{x_j}\varphi\big\|_{\mathrm{L}^{2,\theta}} \\
&\quad + \max_{i=1,\dots,d} \big\|b_i\big\|_{\mathrm{L}^\infty} \sum_{i=1}^{d} \big\|\partial_{x_i}\psi\big\|_{\mathrm{L}^{2,\theta}} \big\|\varphi\big\|_{\mathrm{L}^{2,\theta}} + 2\int_H \big\|\psi\big\|_{\mathrm{L}^{2,\theta}} \big\|\varphi\big\|_{\mathrm{L}^{2,\theta}}\,\nu_L(d\zeta) \\
&\quad + \int_H \sum_{i=1}^{d} \big|\langle\zeta,p_i\rangle_H\big|\,\big\|\partial_{x_i}\psi\big\|_{\mathrm{L}^{2,\theta}}\,\big\|\varphi\big\|_{\mathrm{L}^{2,\theta}}\,\nu_L(d\zeta).
\end{aligned}$$

Due to

$$\int_H \nu_L(d\zeta) < \infty \quad \text{and} \quad \int_H \|\zeta\|_H\,\nu_L(d\zeta) < \infty,$$

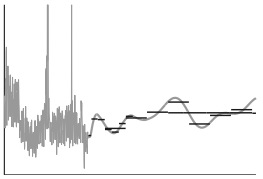
this proves (4.11).

For the proof of Gårding's inequality, we start with the ellipticity property from Proposition 4.8. For every $\zeta \in \mathbb{R}^d$,

$$\sum_{i,j=1}^{d} a_{ij}\,\zeta_i\,\zeta_j \ge c \sum_{i=1}^{d} \zeta_i^2$$

holds, with a constant $c > 0$. Hence

$$\begin{aligned}
c\int_{\mathbb{R}^d} \sum_{i=1}^{d} \big|\partial_{x_i}\psi(x)\big|^2\,\rho_\theta(x)\,dx &\le \int_{\mathbb{R}^d} \sum_{i,j=1}^{d} a_{ij}\,\partial_{x_i}\psi(x)\partial_{x_j}\psi(x)\,\rho_\theta(x)\,dx \\
&= a(\psi,\psi) - \int_{\mathbb{R}^d} \sum_{i=1}^{d} b_i(x)\,\partial_{x_i}\psi(x)\,\psi(x)\,\rho_\theta(x)\,dx \\
&\quad + \int_H \int_{\mathbb{R}^d} \Big[\psi(x+\mathcal{P}_d\zeta) - \psi(x) - \sum_{i=1}^{d} \langle\zeta,p_i\rangle_H\;\partial_{x_i}\psi(x)\Big]\psi(x)\,\rho_\theta(x)\,dx\,\nu_L(d\zeta).
\end{aligned}$$



The same calculations as in the proof of continuity above yield

$$\begin{aligned} c\sum_{i=1}^{d}\big\|\partial_{x_i}\psi\big\|^2_{\mathrm{L}^{2,\theta}} &\le a(\psi,\psi) + C_1\sum_{i=1}^{d}\big\|\partial_{x_i}\psi\big\|_{\mathrm{L}^{2,\theta}}\big\|\psi\big\|_{\mathrm{L}^{2,\theta}} + C_2\big\|\psi\big\|^2_{\mathrm{L}^{2,\theta}} \\ &\le a(\psi,\psi) + C_1\left(\frac{\varepsilon}{2}\sum_{i=1}^{d}\big\|\partial_{x_i}\psi\big\|^2_{\mathrm{L}^{2,\theta}} + \frac{d}{2\varepsilon}\big\|\psi\big\|^2_{\mathrm{L}^{2,\theta}}\right) + C_2\big\|\psi\big\|^2_{\mathrm{L}^{2,\theta}}, \end{aligned}$$

where we have used Young's inequality in the last estimate. Choosing ε so small that

$$C_1\frac{\varepsilon}{2} \le \frac{1}{2}c$$

and setting

$$c_1 = \frac{C_1 d}{2\varepsilon} + C_2 \quad \text{and} \quad c_2 = \frac{1}{2}c$$

yields (4.12). □

With the following two theorems, we show that $\widehat{V}_d$ is indeed the unique solution of the PIDE. We start with a lemma requiring stronger regularity hypotheses for G. Afterwards, we give a result for arbitrary Lipschitz continuous payoffs.

Lemma 4.11. *Suppose that $G \in \mathrm{H}^{2,\theta}(\mathbb{R}^d)$ has bounded first and second derivatives. Then $\widehat{V}_d$ is the unique solution of* (4.9)*, with terminal condition* (4.10)*, in the space $\mathcal{W}(0,T)$ defined by*

$$\mathcal{W}(0,T) := \Big\{ f : \mathbb{R}^d \to \mathbb{R} : f \in \mathrm{L}^2\big(0,T;\mathrm{H}^{1,\theta}\big),\, \partial_t f \in \mathrm{L}^2\Big(0,T;\big(\mathrm{H}^{1,\theta}\big)'\Big)\Big\}.$$

Proof. The bilinear form a is continuous and satisfies Gårding's inequality by Proposition 4.10. Therefore, the PIDE has a unique solution in the space $\mathcal{W}(0,T)$ by [82, Thm. 26.1]. On the other hand, we know from Theorem 4.7 that $\widehat{V}_d$ satisfies (4.9). It remains to prove that $\widehat{V}_d \in \mathcal{W}(0,T)$.

Using the same notation as in the proof of Theorem 4.6, we have

$$\widehat{V}_d(t,x) = e^{-rT}\int_{\mathbb{R}^d} G(x+y)g_{T-t}(y)dy,$$

where $g_{T-t} \in \mathcal{S}(\mathbb{R}^d)$ is the density of $\mathcal{P}_d Z_L(T-t)$. Consequently,

$$\int_{\mathbb{R}^d}\widehat{V}_d^2(t,x)\rho_\theta(x)\,dx \le \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\big(C_1 + C_2\,\|x\| + C_3\,\|y\|\,\big)g_{T-t}(y)\,dy\right)^2\rho_\theta(x)\,dx.$$

Since

$$\int_{\mathbb{R}^d} g_{T-t}(y)\,dy = 1, \quad \int_{\mathbb{R}^d}\|y\|\,g_{T-t}(y)\,dy = E\,\|\mathcal{P}_d Z_L(T-t)\| < \infty,$$

and

$$\int_{\mathbb{R}^d} \|x\|^i \, \rho_\theta(x)\, dx < \infty \;\; (i \in \{1,2\})$$

hold for every $t \in (0,T)$, this implies $\widehat{V}_d \in \mathrm{L}^2\big(0,T;\mathrm{L}^{2,\theta}(\mathbb{R}^d)\big)$. Moreover, we have

$$\partial_x^\alpha \widehat{V}_d(t,x) = e^{-rT} \int_{\mathbb{R}^d} \partial^\alpha G(x+y)\, g_{T-t}(y)\, dy$$

for every $\alpha \in \mathbb{N}_0^d$, $|\alpha| \in \{1,2\}$. Due to the boundedness of the derivatives of G, the following holds:

$$\int_{\mathbb{R}^d} \big(\partial_x^\alpha \widehat{V}_d(t,x)\big)^2 \rho_\theta(x)\, dx \le C \int_{\mathbb{R}^d} \bigg(\int_{\mathbb{R}^d} g_{T-t}(y)\, dy \bigg)^2 \rho_\theta(x)\, dx = C \int_{\mathbb{R}^d} \rho_\theta(x)\, dx.$$

Consequently, $\widehat{V}_d \in L^2\big(0,T;\mathrm{H}^{2,\theta}(\mathbb{R}^d)\big)$. Since $\widehat{V}_d$ satisfies the PIDE (4.7), its time derivative $\partial_t \widehat{V}_d$ can be expressed in terms of its first and second spatial derivatives. Thus, we also have $\partial_t \widehat{V}_d \in L^2\big(0,T;\mathrm{L}^{2,\theta}(\mathbb{R}^d)\big)$, and the proof is complete. □

Theorem 4.12. *Let G satisfy Assumption 4.1. Then $\widehat{V}_d$ is an element of the space $\mathcal{W}(0,T)$ defined in Lemma 4.11 and the unique solution of* (4.9) *with terminal condition* (4.10).

Proof. The key of the proof is to approximate G with a sequence of smooth functions to which Lemma 4.11 can be applied. To this end, let $(\psi_n)_{n\in\mathbb{N}} \subset \mathrm{C}_0^\infty(\mathbb{R}^d)$ be a sequence of standard mollifiers with compact support. Define the convolution

$$G_n(x) := \int_{\mathbb{R}^d} G(x-y)\psi_n(y)dy \in \mathrm{C}^\infty(\mathbb{R}^d), \quad x \in \mathbb{R}^d.$$

Since G is by assumption Lipschitz, and thus uniformly continuous, the following uniform approximation property holds by [33, Thm. C.6]:

$$\forall \varepsilon > 0 \; \exists\, N \in \mathbb{N} : |G_n(x) - G(x)| \le \varepsilon \quad \text{for every } n \ge N, x \in \mathbb{R}^d.$$

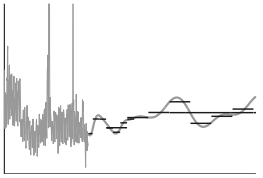
Moreover, for every $\xi \in \mathbb{R}^d$ and every multiindex $\alpha \in \mathbb{N}_0^d$, we have

$$\begin{aligned} |\partial^\alpha G_n(x+\xi) - \partial^\alpha G_n(x)| &\le \int_{\mathbb{R}^d} \big|G(x+\xi-y) - G(x-y)\big| \; |\partial^\alpha \psi_n(y)|\, dy \\ &\le L_G \, \|\xi\| \int_{\mathbb{R}^d} |\partial^\alpha \psi_n(y)|\, dy. \end{aligned}$$

Thus, in particular, the first and second derivatives of G_n are bounded for every $n \in \mathbb{N}$. Now let

$$\widehat{V}_d^n(t,x) := e^{-rT} E\big[G_n\big(x + \mathcal{P}_d Z_L(T-t)\big)\big]$$

be the price function associated with payoff G_n. By Lemma 4.11, $\widehat{V}_d^n \in \mathcal{W}(0,T)$ is the unique solution of (4.7) with terminal condition $\widehat{V}_d^n(T,x) = G_n(x)$. Moreover, the PIDE



with terminal value $G(x)$ also has a unique solution, which we denote by $\widetilde{V}_d \in \mathcal{W}(0,T)$. From [82, Thm. 26.1], we obtain

$$\left\| \widehat{V}_d^n - \widetilde{V}_d \right\|_{\mathrm{L}^2(0,T;\mathrm{L}^{2,\theta})} \leq C \left\| G_n - G \right\|_{\mathrm{L}^{2,\theta}} \to 0 \quad \text{for } n \to \infty. \tag{4.13}$$

On the other hand, by the proof of Lemma 4.11, we have $\widehat{V}_d \in \mathrm{L}^2(0,T;\mathrm{L}^{2,\theta})$ for every payoff G satisfying Assumption 4.1. Thus, using the notation from the proof of Theorem 4.6, we get

$$\begin{aligned} &\left\| \widehat{V}_d^n - \widehat{V}_d \right\|^2_{\mathrm{L}^2(0,T;\mathrm{L}^{2,\theta})} \\ &\quad = e^{-rT} \int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \big(G(x+y) - G_n(x+y) \big) g_{T-t}(y) dy \right)^2 \rho_\theta(x)\, dx\, dt \\ &\quad \leq e^{-rT} \int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \big| G(x+y) - G_n(x+y) \big| g_{T-t}(y) dy \right)^2 \rho_\theta(x)\, dx\, dt \\ &\quad \to 0 \quad \text{for } n \to \infty. \end{aligned} \tag{4.14}$$

Combining (4.13) and (4.14), we obtain $\widehat{V}_d = \widetilde{V}_d$. □

4.2.3 Convergence of Finite-Dimensional Approximations

We are interested in the convergence of $\widehat{V}_d$ to the true price function $\widehat{V}$. The fair price of the option is given by $\widehat{V}(0,0)$, since we assume the time-homogeneous process Z_L starts in 0. In particular, we are looking for a pointwise convergence result.

Theorem 4.13. *Let $\mu_1 \geq \mu_2 \geq \ldots \geq 0$ be the eigenvalues of the covariance operator $\mathcal{C}_{X_T}$ defined in* (2.9)*. Then there exists a constant $C > 0$ such that*

$$\left| \widehat{V}_d(0,0) - \widehat{V}(0,0) \right| \leq C \sqrt{\sum_{l=d+1}^{\dim(H)} \mu_l}.$$

Proof. By definition of $\widehat{V}$ and $\widehat{V}_d$, and Assumption 4.1, we have

$$\begin{aligned} \left| \widehat{V}_d(0,0) - \widehat{V}(0,0) \right| &= e^{-rT} \left| E\big[G\big(Z_L(T)\big) - G\big(\mathcal{P}_d Z_L(T)\big) \big] \right| \\ &\leq e^{-rT} E\big[L_G \left\| Z_L(T) - \mathcal{P}_d Z_L(T) \right\|_H \big]. \end{aligned}$$

Since $\|\cdot\|_{\mathrm{L}^1} \leq C \|\cdot\|_{\mathrm{L}^2}$ for finite measure spaces, we may apply Theorem 3.2 to obtain

$$\left| \widehat{V}_d(0,0) - \widehat{V}(0,0) \right| \leq C \sqrt{E\big[\left\| Z_L(T) - \mathcal{P}_d Z_L(T) \right\|_H^2 \big]} = C \sqrt{\sum_{l=d+1}^{\dim(H)} \mu_l}.$$

□

Note that all the results concerning regularity, variational formulation, and unique solvability of the finite-dimensional PIDE are still valid when we use the Galerkin approxi-

mation $(p_l^{\Delta x})_{l=1}^d$ from Section 3.2 instead of $(p_l)_{l=1}^d$, if we replace Assumption 2.8 with the following hypothesis:

$$\langle Q_L h, h\rangle_H > 0 \quad \text{for every } h \in \operatorname{span}\{p_l^{\Delta x}, l = 1, \dots, d\}\backslash\{0\}. \tag{4.15}$$

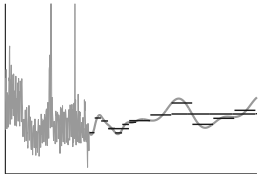
In this case, we denote the corresponding finite-dimensional price process (projected to $\operatorname{span}\{p_l^{\Delta x}, l = 1, \dots, d\}$) by $\widehat{V}_d^{\Delta x}$. We obtain the following slightly modified convergence theorem.

Corollary 4.14. *Let the hypotheses of Proposition 3.7 and* (4.15) *hold. Then there exists a constant $C > 0$ such that*

$$\left|\widehat{V}_d^{\Delta x}(0,0) - V(0,0)\right| \leq C\sqrt{(\Delta x)^{2q+1} + \sum_{l=d+1}^{\dim(H)} \mu_l},$$

where q is the degree of the finite-element polynomials.

Proof. The estimate follows from Proposition 3.7 by exactly the same arguments as in the proof of Theorem 4.13. □



5 Hedging

The main objective of this section is to solve the quadratic hedging problem for European electricity swaptions with the dimension reduction method presented in Section 3. While this type of contracts bears some similarity to derivatives on interest rate markets, there are major differences concerning the hedging problem. Contracts similar to bonds, which imply delivery at a single point in time, are not available. Moreover, the driving factors of the forward curve are not necessarily related to the short rate and thus the discounting factors. In particular, we will assume a constant interest rate for simplicity. The challenge here is that we can only trade in various averages of the forward curve, the swaps. With these finitely many assets, we cannot expect to hedge every possible movement of the forward curve, which has infinitely many degrees of freedom. This discrepancy is clearly visible in Figure 3. Consequently, the market is inherently incomplete, even under a pure diffusion model.

Before we state a solution to the hedging problem, we study the properties of swap rates in detail. We show differentiability and calculate their stochastic dynamics. Moreover, we derive the time-dependent PIDE satisfied by the swaption price (without relying on the time-homogeneous model used for option pricing in Section 4). We then find a representation of the (not necessarily unique) optimal hedging strategy as a solution of a linear equation system. The result is in fact a generalization of the hedging formulas in one-dimensional jump-diffusion models. Similar to a classical delta hedge, the optimal strategy depends on partial derivatives of the option price. These derivatives can be approximated numerically with the POD approach. Finally, we show convergence of this approximation.

5.1 Dynamics of Swap Rates and Swaption Prices

We consider a portfolio of n swap contracts available for trading, whose delivery periods are given by $D_i := [T_1^i, T_2^i]$, $i = 1, \dots, n$. We may, for example, hedge a quarterly swaption by trading the quarterly swap itself as well as three monthly swaps. The swap rates corresponding to the swaps in our portfolio are given by

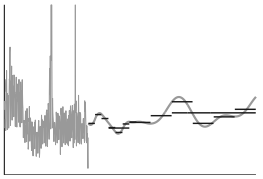
$$F(t; T_1^i, T_2^i) = \int_{T_1^i}^{T_2^i} \omega_i(u) f(t,u)\, du,$$

where

$$\omega_i(u) := \omega(u; T_1^i, T_2^i) = \frac{e^{-ru}}{\int_{T_1^i}^{T_2^i} e^{-ru}\, du}$$

is the discounting factor defined in (1.4). We consider a European option with maturity T written on the swap with delivery period $D = [T_1, T_2]$. Since we cannot hedge with swaps whose delivery periods start before maturity of the option, we will assume $T \le T_1^i$ for every $i = 1, \dots, n$.

For the computation of an optimal hedging strategy, the swap rates $F(t; T_1^i, T_2^i)$, $i = 1, \dots, n$, play a central role. Each rate $F(t; T_1^i, T_2^i)$ is a real-valued, deterministic function



of the forward curve f. More precisely,

$$\big(F(t;T_1^i,T_2^i)\big)_{0\le t\le T} = \big(\,\langle \omega_i, f_t\rangle_H\,\big)_{0\le t\le T}$$

is a real-valued martingale, since f is an H-martingale by Proposition 2.6. By (2.3), the forward curve is in turn a deterministic function of the driving jump-diffusion X defined in (2.1). We may thus introduce

$$F_i : \begin{cases} H \to \mathbb{R} \\ x \;\mapsto \big\langle \omega_i,\, \sum_{k\in I} \langle f_0, e_k\rangle_H \; e^{\langle x, e_k\rangle_H}\, e_k \big\rangle_H \end{cases}$$

and obtain

$$F_i(X_t) = F(t;T_1^i,T_2^i), \qquad t\in[0,T].$$

Similarly, we write

$$F(X_t) = \left\langle \omega, \sum_{k\in I} \langle f_0, e_k\rangle_H \; e^{\langle X_t, e_k\rangle_H}\, e_k \right\rangle_H = F(t;T_1,T_2), \qquad t\in[0,T], \tag{5.1}$$

for the swap rate on which the option is written.

The following three subsections are devoted to the calculation of the stochastic dynamics of both swap rates and swaption prices. They are of a rather technical nature, since the unbounded derivatives of the swap rates prevent us from applying the same Itô formula which we have used in Section 4. Instead, we will show that the properties of the derivatives satisfy the premises of a different version of Itô' s formula. The main results are presented in Theorems 5.5 and 5.8.

5.1.1 Swap Rate Derivatives

The first step towards calculating the stochastic dynamics of swap rates is the calculation of their derivatives as functions of the driving process X. We denote the Fréchet derivative of F_i at $x\in H$ by $D_xF_i(x)\in L(H,\mathbb{R})$. The following theorem shows that the swap rates are indeed twice differentiable. Note that all subsequent theorems on differentiability hold for every F_i, $i=1,\dots,n$, and also for F as defined in (5.1). To simplify notation, we will thus drop the index i and show the results for the generic swap rate F.

Proposition 5.1. *The swap rate function F defined in* (5.1) *is of class C^2, i.e., it is twice continuously Fréchet differentiable. For every $x\in H$ and arbitrary $\xi,\xi_1,\xi_2\in H$, the derivatives satisfy*

$$\begin{aligned} D_xF(x)\,\xi &= \sum_{k\in I} \langle \omega, e_k\rangle_H \,\langle f_0, e_k\rangle_H \, e^{\langle x, e_k\rangle_H}\, \langle \xi, e_k\rangle_H \qquad \textit{and} \\ D_x^2F(x)\,(\xi_1,\xi_2) &= \sum_{k\in I} \langle \omega, e_k\rangle_H \,\langle f_0, e_k\rangle_H \, e^{\langle x, e_k\rangle_H}\, \langle \xi_1, e_k\rangle_H \,\langle \xi_2, e_k\rangle_H \,. \end{aligned} \tag{5.2}$$

Proof. We start by computing Gâteaux derivatives of F. By definition, we have

$$F(x+\varepsilon\xi) = \sum_{k\in I} \langle \omega, e_k\rangle_H \langle f_0, e_k\rangle_H \, e^{\langle x+\varepsilon\xi, e_k\rangle_H} \tag{5.3}$$

for every $\varepsilon > 0$. We define $c_k := \langle \omega, e_k\rangle_H \langle f_0, e_k\rangle_H$ and find that

$$\sum_{k\in I} |c_k| \le \|\omega\|_H \, \|f_0\|_H < \infty.$$

Using the chain rule, we obtain

$$\frac{\partial}{\partial\varepsilon}\Big(c_k e^{\langle x+\varepsilon\xi, e_k\rangle_H}\Big) = c_k e^{\langle x+\varepsilon\xi, e_k\rangle_H} \langle \xi, e_k\rangle_H .$$

Moreover, the partial sums of these derivatives converge uniformly in ε for $|\varepsilon| < 1$, since

$$\left| \sum_{k=m}^{\dim(H)} c_k e^{\langle x+\varepsilon\xi, e_k\rangle_H} \langle \xi, e_k\rangle_H \right| \le e^{\|x\|_H} e^{\|\xi\|_H} \|\xi\|_H \sum_{k=m}^{\dim(H)} |c_k| \to 0 \qquad \text{for } m \to \infty.$$

Thus, we may differentiate (5.3) term by term. This yields

$$\frac{\partial}{\partial\xi} F(x) = \frac{\partial}{\partial\varepsilon} F(x+\varepsilon\xi)\Big|_{\varepsilon=0} = \sum_{k\in I} c_k e^{\langle x, e_k\rangle_H} \langle \xi, e_k\rangle_H .$$

These derivatives are obviously continuous in x, since $\left|e^{\langle x+\varepsilon\xi, e_k\rangle_H} - e^{\langle x, e_k\rangle_H}\right| \to 0$ for $\varepsilon \to 0$, uniformly in k. Since the Gâteaux derivatives of F are continuous, F is continuously Fréchet differentiable and

$$D_x F(x)\,\xi = \frac{\partial}{\partial\xi} F(x) = \sum_{k\in I} c_k e^{\langle x, e_k\rangle_H} \langle \xi, e_k\rangle_H .$$

Due to the isometric isomorphism $\mathrm{L}(H, \mathbb{R}) \cong H$, we may identify $D_x F(x)$ with an element in H and write

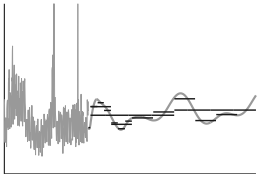
$$D_x F(x) = \sum_{k\in I} c_k e^{\langle x, e_k\rangle_H} \, e_k.$$

By the very same arguments as for the first derivative, we obtain

$$\frac{\partial}{\partial\xi} D_x F(x) = \sum_{k\in I} c_k e^{\langle x, e_k\rangle_H} \langle \xi, e_k\rangle_H \, e_k.$$

This implies the second equation in (5.2). □

In order to apply an Itô formula to F, we need one additional property for its derivatives, which we state in the following lemma. We denote by $\mathrm{L}_{\mathrm{HS}}(H, H) \subset \mathrm{L}(H, H)$ the space of Hilbert-Schmidt operators defined on H.



Lemma 5.2. *The values of the second Fréchet derivative of the function $F : H \to \mathbb{R}$ defined in* (5.1) *are Hilbert-Schmidt operators. The mapping*

$$D_x^2 F : \begin{cases} H \to \mathrm{L_{HS}}(H,H) \\ x \mapsto D_x^2 F(x) \end{cases}$$

is uniformly continuous on bounded subsets.

Proof. The Hilbert-Schmidt norm of $D_x^2 F(x)$ is given by

$$\begin{aligned} \left\| D_x^2 F(x) \right\|_{\mathrm{L_{HS}}(H,H)}^2 &= \sum_{k \in I} \left\langle D_x^2 F(x)\, e_k, D_x^2 F(x)\, e_k \right\rangle_H \\ &= \sum_{k \in I} \langle f_0, e_k \rangle_H^2 \, \langle \omega, e_k \rangle_H^2 \, e^{2\langle x, e_k \rangle_H} \\ &\leq \|f_0\|_H^2 \, \|\omega\|_H^2 \, e^{2\|x\|_H}. \end{aligned}$$

A similar calculation shows

$$\begin{aligned} &\left\| D_x^2 F(x_1) - D_x^2 F(x_2) \right\|_{\mathrm{L_{HS}}(H,H)}^2 \\ &\quad = \sum_{k \in I} \langle f_0, e_k \rangle_H^2 \, \langle \omega, e_k \rangle_H^2 \left(e^{\langle x_1, e_k \rangle_H} - e^{\langle x_2, e_k \rangle_H} \right)^2 \\ &\quad \leq \|f_0\|_H^2 \, \|\omega\|_H^2 \, e^{2 \max\{\|x_1\|_H, \|x_2\|_H\}} \, \|x_1 - x_2\|_H^2 \end{aligned}$$

for every $x_1, x_2 \in H$. This implies the uniform continuity on bounded subsets. □

The statement of the following lemma is a prerequisite for applying [62, Th. 8.23]. This will be useful for splitting the result of Itô's formula in a martingale and a finite variation part.

Lemma 5.3. *The integrals*

$$\int_0^T \int_E E \left| F\big(X_{t-} + \eta_t(y)\big) - F\big(X_{t-}\big) \right|^2 \, \nu(dy) dt$$

and

$$\int_0^T \int_E E \left| D_x F(X_{t-}) \eta_t(y) \right|^2 \, \nu(dy) dt$$

are both finite.

Proof. For the first integral, we apply Young's inequality to obtain

$$\left| F\big(X_{t-} + \eta_t(y)\big) - F(X_{t-}) \right|^2 \leq 2 \left| F\big(X_{t-} + \eta_t(y)\big) \right|^2 + 2 \left| F(X_{t-}) \right|^2.$$

We deal with the two terms separately. Using the definition of F we calculate

$$\begin{aligned} \int_0^T \int_E E \left| F(X_{t-}) \right|^2 \, \nu(dy) dt &= \lambda \int_0^T E \left| \langle \omega, f_{t-} \rangle_H \right|^2 \, dt \\ &\leq \lambda \, \|\omega\|_H^2 \int_0^T E \, \|f_{t-}\|_H^2 \, dt. \end{aligned}$$

This expression is finite by Proposition 2.5. Similarly,

$$\begin{aligned}
&\int_0^T \int_E E\left|F\big(X_{t-}+\eta_t(y)\big)\right|^2 \,\nu(dy)\,dt \\
&\quad= \int_0^T \int_E E\left|\sum_{k\in I} \langle\omega, e_k\rangle_H \langle f_{t-}, e_k\rangle_H \, e^{\langle \eta_t(y), e_k\rangle_H}\right|^2 \nu(dy)\,dt \\
&\quad\le \|\omega\|_H^2 \int_0^T E\,\|f_{t-}\|_H^2\,dt \int_E e^{2\|y\|_E}\,\nu(dy).
\end{aligned}$$

This is finite by Proposition 2.5 and Assumption 2.1.

In order to show that the second integral in the statement of the lemma is finite, we insert the derivative of F calculated in Proposition 5.1. This yields

$$\begin{aligned}
&\int_0^T \int_E E\left|D_xF(X_{t-})\eta_t(y)\right|^2 \,\nu(dy)\,dt \\
&\quad= \int_0^T \int_E E\left|\sum_{k\in I} \langle\omega, e_k\rangle_H \langle f_{t-}, e_k\rangle_H \langle \eta(y), e_k\rangle_H\right|^2 \nu(dy)dt \\
&\quad\le C\,\|\omega\|_H^2 \int_0^T E\,\|f_{t-}\|_H^2\,dt \int_E \|y\|_E^2\,\nu(dy).
\end{aligned}$$

We proceed with Proposition 2.5 and Assumption 2.1 as above to complete the proof. □

5.1.2 Applying Itô's Formula

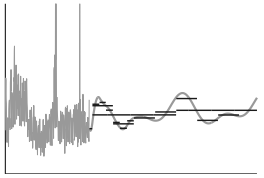
The second step towards calculating the dynamics of the swap rate F is the application of a Hilbert space valued version of Itô's formula.

Lemma 5.4. *The function $F : H \to \mathbb{R}$ defined in* (5.1) *satisfies*

$$\begin{aligned}
dF(X_t) = &\frac{1}{2}\operatorname{tr}\left(D_x^2F(X_{t-})\sigma_{t-}Q\sigma_{t-}^*\right)dt \\
&+ \int_E \Big[F\big(X_{t-}+\eta_t(y)\big) - F(X_{t-}) - D_xF(X_{t-})\,\eta_t(y)\Big]\,\nu(dy)\,dt \\
&+ D_xF(X_{t-})\,\gamma_t\,dt + D_xF(X_{t-})\,\sigma_t\,dW_t \\
&+ \int_E \Big[F\big(X_{t-}+\eta_t(y)\big) - F(X_{t-})\Big]\,\widetilde{M}(dy,dt).
\end{aligned} \tag{5.4}$$

Proof. By Proposition 5.1 and Lemma 5.2, the premises of Itô's formula [62, Th. D.2] are satisfied. The theorem yields

$$\begin{aligned}
F(X_t) = F(X_0) &+ \int_0^t D_xF(X_{s-})\,dX_s + \frac{1}{2}\int_0^t D_x^2F(X_{s-})\,d[X,X]_s^c \\
&+ \sum_{0\le s\le t}\Big[F(X_s) - F(X_{s-}) - D_xF(X_{s-})\,(X_s - X_{s-})\Big],
\end{aligned} \tag{5.5}$$



where $[X,X]^c$ denotes the continuous part of the predictable quadratic covariation as defined in [62]. By definition, we have

$$[X,X]^c_t = \sum_{i,j\in I} e_i \otimes e_j \Big([X_i, X_j]^c_t \Big),$$

where $e_i \otimes e_j$ denotes the tensor product of the two basis elements and $X_i(t) := \langle X(t), e_i \rangle_H$ for $i \in I$. Let $\mathcal{P}_i$ denote the operator represented by the basis element e_i ($i \in I$), and $\mathcal{P}_i^*$ be its adjoint operator:

$$\mathcal{P}_i : \begin{cases} H \to \mathbb{R} \\ h \ \mapsto \langle h, e_i \rangle_H \end{cases} \quad \text{and} \quad \mathcal{P}_i^* : \begin{cases} \mathbb{R} \to H \\ a \ \mapsto a \cdot e_i. \end{cases}$$

By the properties of quadratic variations for real-valued processes and [27, Cor. 4.14], we obtain

$$\begin{aligned} [X_i, X_j]^c_t = [X_i^c, X_j^c]_t &= \left[\left\langle \int_0^\cdot \sigma_s \, dW_s, e_i \right\rangle_H, \left\langle \int_0^\cdot \sigma_s \, dW_s, e_j \right\rangle_H \right]_t \\ &= \int_0^t \mathcal{P}_i \, \sigma_s Q \sigma_s^* \, \mathcal{P}_j^* \, ds = \int_0^t \langle \sigma_s Q \sigma_s^* e_j, e_i \rangle_H \, ds. \end{aligned}$$

Thus, we have

$$\int_0^t D_x^2 F(X_{s-}) d[X,X]^c_s = \int_0^t \sum_{i,j\in I} D_x^2 F(X_{s-}) \, (e_i, e_j) \, \langle \sigma_s Q \sigma_s^* e_j, e_i \rangle_H \, ds$$

and hence

$$\begin{aligned} \int_0^t D_x^2 F(X_{s-}) d[X,X]^c_s &= \int_0^t \sum_{j\in I} D_x^2 F(X_{s-}) \big(\sigma_s Q \sigma_s^* e_j, e_j \big) \, ds \\ &= \int_0^t \operatorname{tr} \Big(D_x^2 F(X_{s-}) \, \sigma_s Q \sigma_s^* \Big) \, ds. \end{aligned} \tag{5.6}$$

It remains to reorganize the jump terms in (5.5). Due to Lemma 5.3, the following holds:

$$\begin{aligned} &\int_0^t D_x F(X_{s-}) \int_E \eta_s(y) \, \widetilde{M}(dy, ds) \\ &\qquad = \sum_{0\le s\le t} D_x F(X_{s-}) \, (X_s - X_{s-}) - \int_0^t \int_E D_x F(X_{s-}) \, \eta_s(y) \, \nu(dy) ds. \end{aligned}$$

Moreover, we have

$$\begin{aligned} \sum_{0\le s\le t} \Big[F(X_s) - F(X_{s-}) \Big] &= \int_0^t \int_E \Big[F\big(X_{s-} + \eta_s(y)\big) - F(X_{s-}) \Big] \, \widetilde{M}(dy, ds) \\ &\quad + \int_0^t \int_E \Big[F\big(X_{s-} + \eta_s(y)\big) - F(X_{s-}) \Big] \, \nu(dy) \, ds. \end{aligned}$$

Combined with (5.5) and (5.6), this implies (5.4). □

As a direct consequence of this lemma, we can state the following, shorter version of the stochastic dynamics.

Theorem 5.5. *The function $F : H \to \mathbb{R}$, defined in* (5.1)*, satisfies*

$$dF(X_t) = D_x F(X_{t-})\,\sigma_t\,dW_t + \int_E \Big[F\big(X_{t-} + \eta_t(y)\big) - F(X_{t-})\Big]\,\widetilde{M}(dy, dt).$$

Proof. By construction, $F(X)$ is a real-valued martingale. The last two integrals in (5.4) are local martingales by definition of the stochastic integral [62, Ths. 8.7, 8.23]. Since continuous local martingales of finite variation are a.s. constant [65, Th. 27], the sum of the remaining integral terms in (5.4) must equal 0. □

5.1.3 Time-Inhomogeneous Swaption Price Dynamics

The remainder of the section is concerned with the stochastic dynamics of the swaption price $\widehat{V}$. Since hedging is path dependent, we cannot rely on an equivalent, time-homogeneous model. In contrast to Section 4.1, we will hence derive a PIDE with time dependent coefficients. Moreover, we will replace Assumption 4.3 with the following regularity assumption. It is slightly less restrictive than Assumption 4.3 and does not require bounded second derivatives. Consequently, we will also use a different version of Itô's formula than for the pricing problem.

Assumption 5.6. *Suppose that $\widehat{V} \in C^{1,2}((0,T) \times H, \mathbb{R}) \cap C([0,T] \times H, \mathbb{R})$, i.e., $\widehat{V}$ is continuously differentiable with respect to t and twice continuously Fréchet differentiable with respect to z. Moreover, assume that the second derivative satisfies $D_z^2\widehat{V}(t,z) \in \mathrm{L}_{\mathrm{HS}}(H,H)$ for every $(t,z) \in [0,T] \times H$ and the mapping $D_z^2\widehat{V} : (t,z) \to \mathrm{L}_{\mathrm{HS}}(H,H)$ is uniformly continuous on bounded subsets.*

Similar to Lemma 5.3, we need a technical lemma in order to be able to rearrange the jump terms in the dynamics of $\widehat{V}$.

Lemma 5.7. *The integrals*

$$\int_0^T \int_E E\left|\widehat{V}\big(t, Z_{t-} + \eta_t(y)\big) - \widehat{V}(t, Z_{t-})\right|^2 \nu(dy)\,dt$$

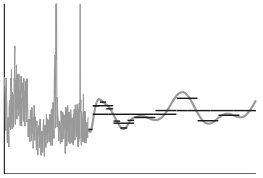
and

$$\int_0^T \int_E E\left|D_z\widehat{V}(t, Z_{t-})\,\eta_t(y)\right|^2 \nu(dy)\,dt$$

are both finite.

Proof. By definition of $\widehat{V}$ and the Lipschitz continuity of the payoff (Assumption 4.1) we obtain

$$\begin{aligned}\left|\widehat{V}(t, Z_{t-} + z) - \widehat{V}(t, Z_{t-})\right| &\le e^{-rT} E\big[\,|G(Z_T + z) - G(Z_T)|\,\big|\mathcal{F}_t\big] \\ &\le e^{-rT} L_G\,\|z\|_H\end{aligned} \tag{5.7}$$



for every $z \in H$. Hence, the first integral satisfies

$$\int_0^T \int_E E\left|\widehat{V}\big(t, Z_{t-} + \eta_t(y)\big) - \widehat{V}(t, Z_{t-})\right|^2 \nu(dy)\, dt \le e^{-2rT} L_G^2 T \int_E \|y\|_E^2 \, \nu(dy) < \infty.$$

For the second integral, note that by (5.7) we have

$$\left\| D_z \widehat{V}(t, Z_{t-}) \right\|_{\mathrm{L}(H,\mathbb{R})} \le e^{-rT} L_G.$$

This implies

$$\int_0^T \int_E E\left|D_z \widehat{V}(t, Z_{t-})\, \eta_t(y)\right|^2 \nu(dy)\, dt \le e^{-2rT} L_G^2 T \int_E \|y\|_E^2 \, \nu(dy) < \infty.$$

□

The regularity assumption made for $\widehat{V}$ is almost identical to the properties of the swap rate F shown in Lemma 5.2. Hence, it is not surprising that we can derive very similar stochastic dynamics for $\widehat{V}$, using once again Itô's formula on Hilbert spaces. As before, we denote the kernel of the covariance operator $\mathcal{C}_{X_T}$ by $E_0(\mathcal{C}_{X_T})$ and its orthogonal complement by $E_0(\mathcal{C}_{X_T})^\perp$. Note that $\mathcal{C}_{X_T}$ is, by construction, also the covariance operator of the centered process Z. The following theorem is the analog of Theorem 4.5 with time-dependent coefficients.

Theorem 5.8. *For every $t \in [0,T]$, the discounted price $\widehat{V}$ of a European swaption, defined in* (4.1)*, satisfies*

$$d\widehat{V}(t, Z_t) = D_z \widehat{V}(t, Z_{t-}) \sigma_t \, dW_t + \int_E \Big[\widehat{V}\big(t, Z_{t-} + \eta_t(y)\big) - \widehat{V}(t, Z_{t-})\Big] \widetilde{M}(dy, dt).$$

Moreover, it is a classical solution of the PIDE

$$\begin{aligned} -\partial_t \widehat{V}(t,z) = {} & \frac{1}{2} \operatorname{tr}\left(D_z^2 \widehat{V}(t,z) \sigma_t Q \sigma_t^*\right) \\ & + \int_E \Big[\widehat{V}\big(t, z + \eta_t(y)\big) - \widehat{V}(t,z) - D_z \widehat{V}(t,z)\, \eta_t(y)\Big] \nu(dy), \end{aligned}$$

with terminal condition

$$\widehat{V}(T,z) = e^{-rT} G(z),$$

for every $t \in (0,T)$, $z \in E_0(\mathcal{C}_{X_T})^\perp$.

Proof. The proof is mainly a collection of arguments we have already used before. By Assumption 5.6, we may apply Itô's formula [62, Th. D.2] to obtain

$$\begin{aligned} \widehat{V}(t, Z_t) = {} & \widehat{V}(0, Z_0) + \int_0^t \partial_t \widehat{V}(s, Z_{s-})\, ds + \int_0^t D_z \widehat{V}(s, Z_{s-})\, dZ_s \\ & + \frac{1}{2} \int_0^t D_z^2 \widehat{V}(s, Z_{s-})\, d[Z,Z]_s^c \\ & + \sum_{0 \le s \le t} \Big[\widehat{V}(s, Z_s) - \widehat{V}(s, Z_{s-}) - D_z \widehat{V}(s, Z_{s-})\,(Z_s - Z_{s-})\Big]. \end{aligned}$$

Since Z and X differ only with respect to a deterministic drift of finite variation, we have $[Z,Z]^c_s = [X,X]^c_s$. Consequently, the calculations in the proof of Lemma 5.4 yield

$$\begin{aligned} d\widehat{V}(t,Z_t) &= \partial_t \widehat{V}(t,Z_{t-})\,dt + \frac{1}{2}\operatorname{tr}\left(D_z^2\widehat{V}(t,Z_{t-})\,\sigma_{t-}Q\sigma_{t-}^*\right)dt \\ &\quad + \int_E \left[\widehat{V}\big(t,Z_{t-}+\eta_t(y)\big) - \widehat{V}(t,Z_{t-}) - D_z\widehat{V}(t,Z_{t-})\,\eta_t(y)\right]\nu(dy)\,dt \\ &\quad + D_z\widehat{V}(t,Z_{t-})\,\sigma_t\,dW_t \\ &\quad + \int_E \left[\widehat{V}\big(t,Z_{t-}+\eta_t(y)\big) - \widehat{V}(t,Z_{t-})\right]\widetilde{M}(dy,dt). \end{aligned}$$

Using the same arguments as in the proof of Theorem 5.5, this yields the stochastic dynamics (and the PIDE) for $\widehat{V}$ along almost every trajectory of Z. Finally, the last paragraph of the proof for Theorem 4.5 is applied without any change to show that the PIDE indeed holds for every $(t,z) \in (0,T)\times E_0(\mathcal{C}_{X(T)})^\perp$. □

5.2 Quadratic Hedging of Electricity

As stated before, options on electricity swaps cannot be perfectly replicated with the products available on the market. Quadratic hedging therefore seems to be a reasonable approach. For an introduction to quadratic hedging in the Brownian case see, e.g., [32, 57, 76]. Hedging with more general driving processes is discussed in [16, 64].

In this section, we derive the optimal hedging strategy for quadratic hedging with a portfolio of swaps. Before we can compute the hedge, we need to discuss the set of admissible strategies and the corresponding value of the portfolio. A trading strategy is given by $\big(\theta_0(t),\theta(t)\big)$, $0 \le t \le T$, where $\theta_0 \in \mathbb{R}$ is the risk free investment and $\theta(t) = \big(\theta_1(t),\dots,\theta_n(t)\big) \in \mathbb{R}^n$ describes the investment in each of the n swaps at time t. The value of the portfolio at time t is denoted by $V^\theta(t)$. The value S_0 of the risk free asset solves the differential equation

$$dS_0(t) = rS_0(t)dt.$$

Since a swap has no inherent value (you can enter the contract without paying anything), we have

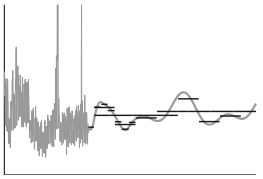
$$V^\theta(t) = \theta_0(t)S_0(t).$$

Nevertheless, changes of the swap rates affect the wealth of the investor. In order to be self-financing, the discounted value $\widehat{V}^\theta$ of the portfolio must satisfy the following equation:

$$d\widehat{V}^\theta(t) = \sum_{i=1}^n \theta_i(t)e^{-rt}\kappa_i\,dF_i(t), \tag{5.8}$$

where

$$\kappa_i := \kappa(T;T_1^i,T_2^i) = \int_{T_1^i}^{T_2^i} e^{-r(u-T)}\,du.$$



The discounting factor κ has been introduced before in (1.6). A strategy $\big(\theta_0(t), \theta(t)\big)$ is admissible, if it is predictable, càglàd, and satisfies

$$E\left|\int_0^T \sum_{i=1}^n \theta_i(t) e^{-rt} \kappa_i \, dF_i(t)\right|^2 < \infty.$$

Quadratic hedging is equivalent to minimizing the expected global quadratic hedging error

$$J(\theta) := E\left|\widehat{V}^\theta(T) - \widehat{V}(T)\right|^2. \tag{5.9}$$

In order to simplify and shorten notation, we define abbreviations for the jumps of swap rates and option prices:

$$\begin{aligned}\Delta F_i(t,y) &:= F_i\big(X_{t-} + \eta_t(y)\big) - F_i(X_{t-}), \quad i = 1,\dots,n,\\ \Delta\widehat{V}(t,y) &:= \widehat{V}\big(t, Z_{t-} + \eta_t(y)\big) - \widehat{V}(t, Z_{t-}), \quad \text{for } y \in E.\end{aligned}$$

Moreover, we will omit some of the more obvious function arguments and write e.g. $D_x F_i$ for $D_x F_i(X_{t-})$ and $D_z\widehat{V}$ for $D_z\widehat{V}(t, Z_{t-})$. The following matrix valued process A is essential for all our calculations. It describes the sensitivity of the traded swaps to changes of the driving stochastic processes.

$$\begin{aligned}a_{ij}(t) &:= e^{-2rt}\kappa_i\kappa_j\left(D_x F_i \, \sigma_t Q \sigma_t^* \, D_x F_j + \int_E \Delta F_i \Delta F_j \, \nu(dy)\right),\ i,j = 1,\dots,n,\\ A(t) &:= (a_{ij}(t))_{i,j=1}^n \in \mathbb{R}^{n\times n}.\end{aligned} \tag{5.10}$$

Note that A is symmetric nonnegative definite by construction. It is worth mentioning that we do not assume A to be strictly positive definite. Consequently, we allow for swaps in the portfolio, which are redundant or irrelevant to the hedging strategy. In particular, we cannot expect a unique optimal strategy under these weak assumptions. In practice, we could then introduce a second optimization criterion, e.g., minimizing the norm of θ.
The following proposition states a representation of the hedging error.

Theorem 5.9. *Let $A(t) \in \mathbb{R}^{n\times n}$ be the matrix valued process defined in* (5.10)*. Define further*

$$b_i(t) := e^{-rt}\kappa_i\left(D_x F_i \, \sigma_t Q \sigma_t^* \, D_z\widehat{V} + \int_E \Delta F_i \, \Delta\widehat{V} \, \nu(dy)\right),\ i = 1,\dots,n,$$

and

$$c(t) := D_z\widehat{V} \, \sigma_t Q \sigma_t^* \, D_z\widehat{V} + \int_E \big(\Delta\widehat{V}\big)^2 \, \nu(dy).$$

Then the quadratic hedging error with strategy θ can be written as

$$J(\theta) = E\int_0^T \big[\theta(t)^T A(t)\theta(t) - 2b(t)^T\theta(t) + c(t)\big]\, dt. \tag{5.11}$$

Proof. Inserting the dynamics of F_i and $\widehat{V}$, calculated in Theorems 5.5 and 5.8, into the definition (5.9) of J yields

$$\begin{aligned} J(\theta) = E\Bigg[& \int_0^T \sum_{i=1}^n \theta_i(t) e^{-rt} \kappa_i \, D_x F_i \sigma_t \, dW_t \\ & + \int_0^T \sum_{i=1}^n \theta_i(t) e^{-rt} \kappa_i \int_E \Delta F_i(t,y) \, \widetilde{M}(dy,dt) \\ & - \int_0^T D_z \widehat{V} \sigma_t \, dW_t - \int_0^T \int_E \Delta \widehat{V}(t,y) \, \widetilde{M}(dy,dt) \Bigg]^2 . \end{aligned}$$

By independence of W and $\widetilde{M}$ we hence have

$$\begin{aligned} J(\theta) = {} & E\Bigg[\int_0^T \sum_{i=1}^n \theta_i(t) e^{-rt} \kappa_i \, D_x F_i \sigma_t \, dW_t - \int_0^T D_z \widehat{V} \sigma_t \, dW_t \Bigg]^2 \\ & + E\Bigg[\int_0^T \sum_{i=1}^n \theta_i(t) e^{-rt} \kappa_i \int_E \Delta F_i(t,y) \, \widetilde{M}(dy,dt) \\ & - \int_0^T \int_E \Delta \widehat{V}(t,y) \, \widetilde{M}(dy,dt) \Bigg]^2 . \end{aligned} \tag{5.12}$$

We denote the two expectations on the right-hand side of (5.12) with J_1 and J_2. We apply [27, Cor. 4.14] to the Brownian term J_1 and obtain

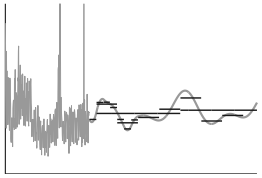
$$\begin{aligned} J_1 = E \int_0^T \operatorname{tr} \Bigg\{ & \left[e^{-rt} \sum_{i=1}^n \theta_i(t) \kappa_i \, D_x F_i - D_z \widehat{V} \right] \sigma_t Q \sigma_t^* \\ & \times \left[e^{-rt} \sum_{j=1}^n \theta_j(t) \kappa_j \, D_x F_j - D_z \widehat{V} \right]^* \Bigg\} \, dt. \end{aligned}$$

Note that the argument of the trace operator in this equation is a function mapping $\mathbb{R}$ to $\mathbb{R}$. Consequently, its "trace" is in fact the application of this function to 1. Moreover, the operators $D_x F_i$ and $D_z \widehat{V}$ are elements of $\mathrm{L}(H, \mathbb{R})$, which we can identify with elements of H. Hence, we have

$$\big[D_x F\big]^*(1) = D_x F \quad \text{and} \quad \big[D_z \widehat{V}\big]^*(1) = D_z \widehat{V}.$$

Combined, we obtain

$$\begin{aligned} J_1 = E \int_0^T \Bigg[& e^{-2rt} \sum_{i=1}^n \sum_{j=1}^n \theta_i(t) \theta_j(t) \kappa_i \kappa_j \, D_x F_i \, \sigma_t Q \sigma_t^* \, D_x F_j \\ & - 2 e^{-rt} \sum_{i=1}^n \theta_i(t) \kappa_i \, D_x F_i \, \sigma_t Q \sigma_t^* \, D_z \widehat{V} + D_z \widehat{V} \, \sigma_t Q \sigma_t^* \, D_z \widehat{V} \Bigg] \, dt. \end{aligned}$$



We use Theorem [62, Th. 23] to deal with the jump term J_2 in (5.12). This yields

$$J_2 = E \int_0^T \int_E \left[e^{-rt} \sum_{i=1}^n \theta_i(t) \kappa_i \Delta F_i(t,y) - \Delta \widehat{V}(t,y) \right]^2 \nu(dy) dt.$$

Adding the expressions for J_1 and J_2, we obtain (5.11) by definition of A, b, and c. □

The expression (5.11) for the hedging error in the previous theorem involves a quadratic form with respect to θ. The following lemma states an important property of this quadratic form, which we will use to show existence of an optimal hedging strategy.

Lemma 5.10. *The vector $b(t)$ defined in Theorem 5.9 satisfies*

$$\forall y \in \mathbb{R}^n : \big(y^T A(t) = 0 \Rightarrow y^T b(t) = 0 \big)$$

for every $t \in [0,T]$.

Proof. Let $y^T A(t) = 0$. By definition of A, we have

$$\begin{aligned} 0 = y^T A(t) y = e^{-2rt} \Bigg[& \left(\sum_{i=1}^n y_i \kappa_i D_x F_i \right) \sigma_t Q \sigma_t^* \left(\sum_{j=1}^n y_j \kappa_j D_x F_j \right) \\ & + \int_E \left(\sum_{i=1}^n y_i \kappa_i \Delta F_i \right) \left(\sum_{j=1}^n y_j \kappa_j \Delta F_j \right) \nu(dy) \Bigg] \end{aligned}$$

Due to the positive semi-definiteness of Q, this yields

$$\left(\sum_{i=1}^n y_i \kappa_i D_x F_i \right) \sigma_t Q \sigma_t^* \left(\sum_{i=1}^n y_i \kappa_i D_x F_i \right) = 0$$

and

$$\int_E \left(\sum_{i=1}^n y_i \kappa_i \Delta F_i \right)^2 \nu(dy) = 0.$$

Using the Cauchy–Schwarz inequality, we obtain the estimate

$$\begin{aligned} \left| y^T b(t) \right| &= \left| e^{-rt} \sum_{i=1}^n y_i \kappa_i \left(D_x F_i \, \sigma_t Q \sigma_t^* \, D_z \widehat{V} + \int_E \Delta F_i \, \Delta \widehat{V} \, \nu(dy) \right) \right| \\ &\le e^{-rt} \left[\left(\sum_{i=1}^n y_i \kappa_i D_x F_i \right) \sigma_t Q \sigma_t^* \left(\sum_{i=1}^n y_i \kappa_i D_x F_i \right) \right]^{\frac{1}{2}} \left[D_z \widehat{V} \, \sigma_t Q \sigma_t^* \, D_z \widehat{V} \right]^{\frac{1}{2}} \\ &\quad + e^{-rt} \left[\int_E \left(\sum_{i=1}^n y_i \kappa_i \Delta F_i \right)^2 \nu(dy) \right]^{\frac{1}{2}} \left[\int_E (\Delta \widehat{V})^2 \nu(dy) \right]^{\frac{1}{2}} \\ &= 0. \end{aligned}$$

□

We are now able to derive the main result of this section, a representation of the optimal hedging strategy for portfolios containing an arbitrary number of swaps.

Theorem 5.11. *An investment strategy $\overline{\theta}$ minimizes the hedging error if and only if it solves*

$$A(t)\overline{\theta}(t) = b(t) \qquad \text{for a.e. } t \in [0,T]. \tag{5.13}$$

There is at least one solution to this equation. It is unique if and only if $A(t)$ is strictly positive definite.

Proof. The minimal hedging error is achieved when the integrand

$$h : \begin{cases} \mathbb{R}^n \ \to \mathbb{R} \\ \theta(t) \mapsto \theta(t)^T A(t)\theta(t) - 2b(t)^T\theta(t) + c(t) \end{cases}$$

in (5.11) is minimized point-wise for a.e. $t \in [0,T]$. The matrix $A(t)$ is nonnegative definite by construction. Consequently, h is a convex function (though not strictly convex). For convex functions, the necessary optimality condition of first order is already sufficient for a global minimum. Thus, every solution $\overline{\theta}(t)$ of (5.13) is an optimal hedging strategy. It is a direct consequence of Lemma 5.10 that $b(t)$ is an element of the range of $A(t)$. Hence, there is at least one such solution. The uniqueness property is then obvious. □

Comparison to One-Dimensional Hedging The hedging portfolio computed in Theorem 5.11 is in fact a generalization of the optimal hedge in a one-dimensional jump-diffusion model. For the special case of a portfolio containing only a single swap ($n = 1$), with the same delivery period $D = [T_1, T_2]$ as the hedged swap itself, we obtain the following strategy.

Corollary 5.12. *The optimal investment for quadratic hedging with a single swap is given by*

$$\overline{\theta}(t) := \frac{D_x F\, \sigma_t Q \sigma_t^*\, D_z \widehat{V} + \int_E \Delta F(t,y) \Delta \widehat{V}(t,y)\, \nu(dy)}{e^{-rt}\kappa \Big[D_x F\, \sigma_t Q \sigma_t^*\, D_x F + \int_E \big(\Delta F(t,y)\big)^2\, \nu(dy) \Big]}. \tag{5.14}$$

We will now briefly show how this result relates to the well-known hedging strategy for a single stock. To this end, we set all the Hilbert and Banach spaces in our model to $H = U = E = \mathbb{R}$. Since W is then a one-dimensional Brownian motion, we set $Q = \mathrm{Id}\,|_{\mathbb{R}}$, $\eta \equiv \mathrm{Id}\,|_{\mathbb{R}}$ and $\kappa = 1$. Furthermore, in this case the stock price is modeled by

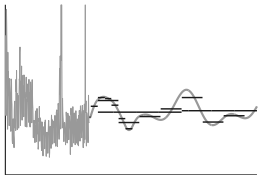
$$S_t = F(t, X_t) = S_0 \exp\left(\int_0^t \gamma(s) ds + Z_t \right) \in \mathbb{R},$$

with an appropriate drift term γ. The option price can be written as a function of S_{t-}:

$$\widetilde{V}(t, S_{t-}) := e^{rt} \widehat{V}(t, Z_{t-}).$$

Hence, the following holds for the derivative of the price with respect to S:

$$D_z \widehat{V}(t, Z_{t-}) = e^{-rt} D_S \widetilde{V}(t, S_{t-})\, S_{t-}.$$



Finally, we calculate

$$D_x F(t, X_{t-}) = F(t, X_{t-}) = S_{t-}, \quad \Delta F(t, y) = (e^y - 1)S_t,$$

and

$$\Delta \widehat{V}(t, y) = \widehat{V}(t, Z_{t-} + y) - \widehat{V}(t, Z_{t-}) = e^{-rt}\big[\widetilde{V}(t, S_{t-}e^y) - \widetilde{V}(t, S_{t-})\big].$$

Putting everything together, we can do a change of variable in (5.14) and obtain

$$\overline{\theta}(t) = \frac{\sigma_t^2 D_S \widetilde{V}(t, S_{t-}) + \frac{1}{S_{t-}} \int_{\mathbb{R}} (e^y - 1)\big[\widetilde{V}(t, S_{t-}e^y) - \widetilde{V}(t, S_{t-})\big]\, \nu(dy)}{\sigma_t^2 + \int_{\mathbb{R}} (e^y - 1)^2\, \nu(dy)}.$$

Note that this is exactly the formula for the optimal quadratic hedge in a stock market, calculated, e.g., in [24, Rem. 10.3].

5.3 Hedging with POD

Theorem 5.11 describes the optimal quadratic hedge for a portfolio of swaps. The matrix $A(t)$ in the linear equation system (5.13) is determined by the swap rate functions F_i, $i = 1, \dots, n$, and their derivatives, which are known in closed form. Hence, they can be calculated directly, using a (possibly high-dimensional) discretization of the Hilbert space H. In contrast, the right-hand side $b(t)$ depends on the unknown option price $\widehat{V}$ and its derivatives. Since it is not feasible to solve the corresponding PIDE with a high-dimensional discretization, we will apply POD.

The goal of this section is to compute the optimal hedge with equation (5.13) by replacing all expressions depending on $\widehat{V}$ with similar expressions depending on $\widehat{V}_d$ defined in (4.5). Note that applying the chain rule [31, Th. 8.2.1] yields

$$D_z \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) = \frac{\partial}{\partial Z_{t-}} \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) = D_x \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) \circ \mathcal{P}_d \in \mathrm{L}(H, \mathbb{R}).$$

The derivative $D_x \widehat{V}_d$ is a finite-dimensional object, which we can approximate numerically by solving the PIDE for $\widehat{V}_d$. Similarly, the jump term

$$\Delta \widehat{V}_d(t, \mathcal{P}_d \zeta) := \widehat{V}_d\big(t, \mathcal{P}_d(Z_{t-} + \zeta)\big) - \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) \tag{5.15}$$

can be evaluated numerically for every $\zeta \in H$.

To simplify subsequent notation, we further introduce

$$\langle h_1, h_2 \rangle_{\sigma_t Q \sigma_t^*} := h_1\, \sigma_t Q \sigma_t^*\, h_2 \quad \text{and} \quad \|h_1\|_{\sigma_t Q \sigma_t^*} := \sqrt{\langle h_1, h_1 \rangle_{\sigma_t Q \sigma_t^*}}$$

for every $h_1, h_2 \in H$. Due to the fact that the Cauchy–Schwarz inequality holds not only for scalar products, but also for symmetric nonnegative definite bilinear forms (see, e.g., [31, Thm. 6.2.1]), $\|\cdot\|_{\sigma_t Q \sigma_t^*}$ defines a semi-norm on H.

In order to approximate b, we define

$$\widetilde{b}_i(t) := e^{-rt}\kappa_i\Big(\big\langle D_x F_i, D_z \widehat{V}_d(t, \mathcal{P}_d Z_{t-})\big\rangle_{\sigma_t Q \sigma_t^*} + \int_E \Delta F_i \, \Delta \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) \, \nu(d\zeta)\Big), \tag{5.16}$$

for $i = 1, \ldots, n$, where $\widehat{V}_d$ and $\Delta \widehat{V}_d$ are the approximations defined in (4.5) and (5.15), respectively. For $t \in [0, T]$, we consider a (not necessarily unique) solution $\widetilde{\theta}$ of

$$A(t)\widetilde{\theta}(t) = \widetilde{b}(t), \tag{5.17}$$

which is an approximation for an optimal strategy $\overline{\theta}(t)$. Similar to (5.13), the linear equation for $\widetilde{\theta}(t)$ has at least one solution, although A might be singular.

Lemma 5.13. *The vector $\widetilde{b}(t)$ defined in* (5.16) *satisfies*

$$\forall y \in \mathbb{R}^n : \big(y^T A(t) = 0 \Rightarrow y^T \widetilde{b}(t) = 0\big)$$

for every $t \in [0, T]$. In particular, $\widetilde{b}(t)$ is an element of the range of $A(t)$.

Proof. The proof is identical to the proof of Lemma 5.10. Compare also the proof of Lemma 5.14 below for a very similar argument. □

In contrast to $\overline{\theta}$, the equation for $\widetilde{\theta}$ contains only expressions which are either analytically known or can be approximated numerically. We will show that the *additional* hedging error introduced by using $\widetilde{\theta}$ converges to 0 for increasing dimension d of the approximating problem. (The unhedgeable part of the hedging risk remains unchanged by definition.)

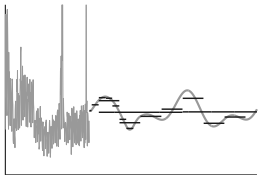
Before we can show the convergence result, we analyze the relation of $b(t)$ and $\widetilde{b}(t)$ in greater detail. For all subsequent computations, let $t \in [0, T]$ be arbitrary, but fixed. Since the matrix $A(t)$ is positive semi-definite, we can find an orthonormal basis of eigenvectors $Q_A = (q_1 | \ldots | q_n) \in \mathbb{R}^{n \times n}$ with corresponding eigenvalues λ_i, $i = 1, \ldots, n$, such that

$$A(t) = Q_A \operatorname{diag}(\lambda_1, \ldots, \lambda_n) Q_A^T,$$

with $\operatorname{diag}(\ldots)$ denoting a diagonal matrix. The following lemma states an upper bound for the projection of the difference $b(t) - \widetilde{b}(t)$ onto each of the eigenvectors q_i of $A(t)$.

Lemma 5.14. *For every $l \in \{1, 2, \ldots, n\}$, the following estimate holds:*

$$\begin{aligned}\Big[\big(\widetilde{b}(t) - b(t)\big)^T q_l\Big]^2 \\ \leq 2\lambda_l \Big[\big\|D_z \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) - D_z \widehat{V}\big\|^2_{\sigma_t Q \sigma_t^*} + \int_H \big(\Delta \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) - \Delta \widehat{V}\big)^2 \, \nu(d\zeta)\Big].\end{aligned} \tag{5.18}$$



Proof. By definition of $b(t)$ and $\widetilde{b}(t)$, we have

$$\left[\left(\widetilde{b}(t)-b(t)\right)^T q_l\right]^2 = e^{-2rt}\left[\left\langle \sum_{i=1}^n q_l(i)\kappa_i\, D_x F_i\,, D_z\widehat{V}_d(t,\mathcal{P}_d Z_{t-}) - D_z\widehat{V}\right\rangle_{\sigma_t Q\sigma_t^*} + \int_H \left(\sum_{i=1}^n q_l(i)\kappa_i\,\Delta F_i\right)\left(\Delta\widehat{V}_d(t,\mathcal{P}_d Z_{t-}) - \Delta\widehat{V}\right)\nu(d\zeta)\right]^2$$

Applying the inequalities by Young and Cauchy–Schwarz yields

$$\begin{aligned}&\left[\left(\widetilde{b}(t)-b(t)\right)^T q_l\right]^2\\ &\quad\le 2e^{-2rt}\left\|\sum_{i=1}^n q_l(i)\kappa_i D_x F_i\right\|^2_{\sigma_t Q\sigma_t^*}\left\|D_z\widehat{V}_d(t,\mathcal{P}_d Z_{t-}) - D_z\widehat{V}\right\|^2_{\sigma_t Q\sigma_t^*}\\ &\qquad + 2e^{-2rt}\left(\int_H\left(\sum_{i=1}^n q_l(i)\kappa_i\Delta F_i\right)^2\nu(d\zeta)\right)\left(\int_H\left(\Delta\widehat{V}_d(t,\mathcal{P}_d Z_{t-}) - \Delta\widehat{V}\right)^2\nu(d\zeta)\right)\end{aligned}$$

On the other hand, the eigenvalues of $A(t)$ satisfy

$$\lambda_l = q_l^T A(t) q_l = e^{-2rt}\left\|\sum_{i=1}^n q_l(i)\kappa_i D_x F_i\right\|^2_{\sigma_t Q\sigma_t^*} + e^{-2rt}\int_H\left(\sum_{i=1}^n q_l(i)\kappa_i\Delta F_i\right)^2\nu(d\zeta).$$

Combined, we obtain (5.18). □

We can now prove the main convergence result for the quadratic hedging with POD. It gives a bound on the additional hedging error caused by the approximation of the optimal strategy.

Theorem 5.15. *There is a constant $C>0$ such that the additional hedging error when using $\widetilde{\theta}$, defined in* (5.17)*, instead of an optimal hedging strategy $\overline{\theta}$, defined in* (5.13)*, satisfies*

$$0 \le J(\widetilde{\theta}) - J(\overline{\theta}) \le Ce^{-2rT}\sum_{l=d+1}^{\dim(H)}\mu_l.$$

As before, μ_l are the eigenvalues of the covariance operator $\mathcal{C}_{X_T}$.

Proof. Since $\overline{\theta}$ is an optimal strategy, we have $0 \le J(\widetilde{\theta}) - J(\overline{\theta})$. Using Theorem 5.9, we obtain

$$J(\widetilde{\theta}) - J(\overline{\theta}) = E\int_0^T\left[\widetilde{\theta}^T A\widetilde{\theta}\,(t) - 2b^T\widetilde{\theta}\,(t) - \overline{\theta}^T A\overline{\theta}\,(t) + 2b^T\overline{\theta}\,(t)\right]dt. \tag{5.19}$$

By construction, $\widetilde{\theta}$ and $\overline{\theta}$ are solutions of

$$A(t)\widetilde{\theta}(t) = \widetilde{b}(t) \quad\text{and}\quad A(t)\overline{\theta}(t) = b(t), \tag{5.20}$$

respectively. Since $A(t)$ may be singular, we cannot invert it. Instead, we introduce the Moore-Penrose pseudo-inverse $A^+(t)$ of $A(t)$. Suppose without loss of generality that the eigenvalues of $A(t)$ satisfy $0 \le \lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$. Suppose further that $\lambda_m > 0$ is the lowest non-zero eigenvalue. Then $M^+(t)$ is given by

$$A^+(t) = Q_A \operatorname{diag}\Big(0, \ldots, 0, \frac{1}{\lambda_m}, \ldots, \frac{1}{\lambda_n}\Big) Q_A^T.$$

Every pair of solutions $\big(\widetilde{\theta}(t), \overline{\theta}(t)\big)$ to (5.20) yields the same additional hedging error. Thus, we may replace $\widetilde{\theta}(t)$ with the specific solution $A^+(t)\widetilde{b}(t)$ and $\overline{\theta}(t)$ with $A^+(t)b(t)$ in (5.19). Since the pseudo-inverse satisfies $A^+AA^+(t) = A^+(t)$, we obtain

$$\begin{aligned} J(\widetilde{\theta}) - J(\overline{\theta}) &= E \int_0^T \Big[\widetilde{b}^T(t) A^+(t) \widetilde{b}(t) - 2b^T(t) A^+ \widetilde{b}(t) + b^T(t) A^+(t) b(t)\Big]\, dt \\ &= E \int_0^T \big(\widetilde{b}(t) - b(t)\big)^T A^+(t) \big(\widetilde{b}(t) - b(t)\big)\, dt. \end{aligned}$$

Plugging in the definition of $A^+(t)$ yields

$$J(\widetilde{\theta}) - J(\overline{\theta}) = E \int_0^T \sum_{i=m}^{n} \frac{1}{\lambda_i} \Big[\big(\widetilde{b}(t) - b(t)\big)^T q_i\Big]^2 dt.$$

Applying Lemma 5.14, we find

$$\begin{aligned} J(\widetilde{\theta}) - J(\overline{\theta}) \le 2(n-m) E \int_0^T \Big[&\big\| D_z \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) - D_z \widehat{V} \big\|^2_{\sigma_t Q \sigma_t^*} \\ &+ \int_H \big(\Delta \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) - \Delta \widehat{V}\big)^2 \nu(d\zeta) \Big]\, dt \end{aligned}$$

and hence

$$\begin{aligned} J(\widetilde{\theta}) - J(\overline{\theta}) = 2(n-m) E \Big| &\int_0^T \big(D_z \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) - D_z \widehat{V}\big) \sigma_t\, dW_t \\ &+ \int_0^T \int_H \big(\Delta \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) - \Delta \widehat{V}\big)\, \widetilde{M}(d\zeta, dt) \Big|^2, \end{aligned}$$

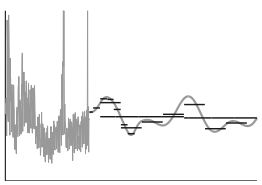
where, for the last equality, we have used [27, Cor. 4.14] for the Brownian part, [62, Thm. 8.23] for the jump part, and in addition the independence of W and $\widetilde{M}$.

The very same arguments as in the proof of Theorem 5.8 yield

$$d\widehat{V}_d(t, \mathcal{P}_d Z_t) = D_z \widehat{V}_d(t, \mathcal{P}_d Z_{t-}) \sigma_t\, dW_t + \int_H \Delta \widehat{V}_d(t, \mathcal{P}_d \zeta)\, \widetilde{M}(d\zeta, dt).$$

Hence, we have

$$\begin{aligned} J(\widetilde{\theta}) - J(\overline{\theta}) &\le 2nE \left| \int_0^T d\widehat{V}_d(t, \mathcal{P}_d Z_t)\, dt - \int_0^T d\widehat{V}(t, Z_t)\, dt \right|^2 \\ &= 2nE \Big| \big[\widehat{V}_d(T, \mathcal{P}_d Z_T) - \widehat{V}_d(0,0)\big] - \big[\widehat{V}(T, Z_T) - \widehat{V}(0,0)\big] \Big|^2. \end{aligned}$$



Using Young's inequality and the definition of $\widehat{V}$ and $\widehat{V}_d$, we see that

$$J(\widetilde{\theta}) - J(\overline{\theta}) \leq 4n \left|\widehat{V}(0,0) - \widehat{V}_d(0,0)\right|^2 + 4ne^{-2rT} E\left|G(Z_T) - G(\mathcal{P}_d Z_T)\right|^2.$$

From Assumption 4.1 and Theorem 3.2 we finally obtain

$$J(\widetilde{\theta}) - J(\overline{\theta}) \leq 8ne^{-2rT} L_G^2 \, E\left\|Z_T - \mathcal{P}_d Z_T\right\|_H^2 = 8ne^{-2rT} L_G^2 \sum_{l=d+1}^{\dim(H)} \mu_l.$$

□

Theorem 5.15 shows that every solution of (5.17) is a good approximation to an optimal hedging strategy, if the dimension d is chosen appropriately. Note that although hedging errors are path-dependent, choosing a POD basis which minimizes the projection error for the terminal value Z_T is sufficient for the computation of $\widetilde{\theta}$. We do *not* need to approximate the whole path of the process explicitly.

Semi-Definite Quadratic Optimization Due to the results in the previous section, the numerical hedging algorithm will be based on solving the linear equation

$$A(t)\widetilde{\theta}(t) = \widetilde{b}(t). \tag{5.21}$$

By Lemma 5.13, there is always at least one solution. In practice, however, finding such a solution yields numerical issues, since the matrix $A(t)$ is not *strictly* positive definite. Even small discretization or rounding errors in the computation of $A(t)$ and $\widetilde{b}(t)$ may result in an unsolvable equation, where $\widetilde{b}(t)$ is no longer an element of the range of $A(t)$. For the remainder of the section, let once again $t \in [0,T]$ be arbitrary but fixed.

Depending on the eigenvalues of $A(t)$, there are two possible remedies for this situation. Suppose as before that the eigenvalues satisfy $0 \leq \lambda_1 \leq \ldots \leq \lambda_n$, with $\lambda_m > 0$ being the lowest non-zero eigenvalue. The corresponding matrix of orthonormal eigenvectors is $Q_A = (q_1|\ldots|q_n)$. We first consider the case when the ratio $\frac{\lambda_n}{\lambda_m}$ is small, i.e. λ_m is well separated from 0. Instead of looking for an arbitrary solution of (5.21), we look for the one with the smallest norm $\left\|\widetilde{\theta}(t)\right\|_{\mathbb{R}^n}$. Equation (5.21) is equivalent to

$$\operatorname{diag}(0,\ldots,0,\lambda_m,\ldots,\lambda_n) Q_A^T \widetilde{\theta}(t) = Q_A^T \widetilde{b}(t).$$

Since

$$\left\|\widetilde{\theta}(t)\right\|_{\mathbb{R}^n} = \sum_{i=1}^{m-1} \left\langle \widetilde{\theta}(t), q_i \right\rangle_{\mathbb{R}^n}^2 + \sum_{i=m}^{n} \left\langle \widetilde{\theta}(t), q_i \right\rangle_{\mathbb{R}^n}^2,$$

minimizing the norm of $\widetilde{\theta}(t)$ is equivalent to $\widetilde{\theta}$ being orthogonal to every q_i, $i = 1,\ldots m-1$. Consequently, we project (5.21) to the orthogonal complement of the kernel of $A(t)$. Hence, we solve

$$\operatorname{diag}(\lambda_m,\ldots,\lambda_n)\varphi(t) = (q_m|\ldots|q_n)^T \widetilde{b}(t) \tag{5.22}$$

for $\varphi \in \mathbb{R}^{n-m+1}$, and set

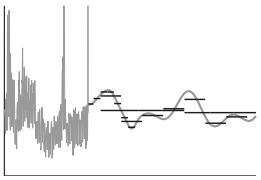
$$\widetilde{\theta}(t) = (q_m | \dots | q_n)\varphi(t).$$

Note that the condition number of the linear system (5.22) is given by $\frac{\lambda_n}{\lambda_m}$, which is small by our hypothesis. Thus, the unique solution φ can be computed in a numerically stable way.

We consider now the case $\frac{\lambda_n}{\lambda_m} \gg 1$. The projected equation (5.22) is then ill-conditioned. Moreover, it is difficult to identify the eigenvalues of $A(t)$ which are equal to 0 numerically. In this case, we can either change our portfolio and use swaps which are less correlated, or we can apply regularization to the equation. Tikhonov regularization proved to be very effective in numerical experiments. We replace (5.13) with the minimization problem

$$\overline{\theta}(t) = \operatorname{argmin}_{\theta(t)} \theta(t)^T \big(A(t) + \delta \, \|A\| \, I_n\big)\theta(t) - 2b(t)^T\theta(t), \qquad \text{for a.e. } t \in [0,T],$$

where $I_n \in \mathbb{R}^{n\times n}$ denotes the unit matrix, and $\delta \ll 1$ is a small regularization parameter. Just like the projection method discussed above, the regularization approach also gives preference to solutions with smaller norms. For convergence results concerning the Tikhonov regularization, we refer to [37, Ch. 6.4.3].



6 Asian Options

In this section, we study arithmetic average Asian options. It is well known [68] that the price of an Asian option on a single asset, modeled by a geometric Brownian motion, is the solution of a PDE in two variables: the value of the underlying and its average up to the current time. Using a special parametrization, it is also possible to obtain a PIDE with only one space variable [79], but there is no obvious way to extend this parametrization to the multi- or infinite-dimensional case. Introducing the average as additional variable in our Hilbert space valued model, the option price can be written as a function of time t, the average value A_t, and the current state of the assets Z_t. Hence, we arrive at a PIDE similar to (4.4), with one extra term accounting for the average. We derive the low-dimensional PIDE satisfied by the dimension-reduced price process. Convergence of the finite-dimensional PIDE solution to the true value of the Asian option is shown. In general, the PIDEs corresponding to Asian options cannot be solved analytically. They are, however, the basis for numerical pricing methods. Using appropriate algorithms, the PIDEs can be solved in a numerically stable way, see [84, 30] and the references therein. For an overview of methods for pricing Asian options see, e.g., [73].

6.1 Value of an Asian Option

Before we can define the value of an arithmetic average Asian option, we need to clarify what exactly *average* means in our Hilbert space valued setting. Consider an index option on a basket of stocks as defined in (1.2). An Asian option could, e.g., depend on the time-average of the index, which in turn is a weighted sum of the individual stock values. The weight factors are nothing more than a linear mapping working on the vector of asset prices. More generally, we consider an arbitrary bounded linear mapping $w : H \to \mathbb{R}$, which we identify with $w \in H$ by the Fréchet–Riesz representation theorem. The arithmetic average up to time $t > 0$ is then given by

$$A_t := \frac{1}{t} \int_0^t \langle w, S_s \rangle_H \, ds \in \mathbb{R}.$$

Note that the very same formula can be applied to electricity swap rates, replacing S_s with f_s:

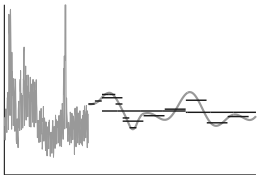
$$A_t := \frac{1}{t} \int_0^t F(s; T_1, T_2) \, ds = \frac{1}{t} \int_0^t \langle w, f_s \rangle_H \, ds. \tag{6.1}$$

We will subsequently use the latter notation.

Using the Jensen inequality, the Cauchy–Schwarz inequality, and Fubini's theorem, we obtain

$$E\left[A_t^2\right] = \frac{1}{t^2} E\left[\left(\int_0^t \langle w, f_s \rangle_H \, ds\right)^2\right] \leq \frac{1}{t^2} \|w\|_H^2 \int_0^t E\,\|f_s\|_H^2 \, ds.$$

This expression is finite by Proposition 2.5. Hence, the average is a well defined random



variable in $\mathrm{L}^2(\Omega)$ for $t > 0$ (where $(\Omega, \mathcal{F})$ denotes the measurable space on which the risk neutral measure is defined). The defining equation (6.1) is, however, not valid for $t = 0$. Intuitively,

$$A_0 := \langle w, f_0 \rangle_H \tag{6.2}$$

is the obvious continuation for A. The following theorem shows that this is indeed the correct choice.

Theorem 6.1. *The following convergence holds almost surely:*

$$\lim_{t \to 0} A_t = \langle w, f_0 \rangle_H .$$

Proof. Using the definition of A, we find

$$|A_t - \langle w, f_0 \rangle_H| \leq \frac{1}{t} \int_0^t |\langle w, f_s - f_0 \rangle_H| \, ds. \tag{6.3}$$

In order to find a bound for $\langle w, f_u - f_0 \rangle_H$, we consider the driving process X. From the proof of Proposition 2.2, we know that

$$E \, \|X_t\|_H^2 \leq \int_0^t \left(\|\gamma_s\|_H^2 + \operatorname{tr}(Q) \, \|\sigma_s\|_{\mathrm{L}(U,H)}^2 + C \int_E \|\eta_s(y)\|_H^2 \, \nu(dy) \right) ds.$$

Thus, $\lim_{t \to 0} \|X_t\|_H = 0$ in $\mathrm{L}^2(\Omega)$. Consequently, there is a sequence $\{t_n\}_{n \in \mathbb{N}} \subset \mathbb{R}^+$ satisfying $\lim_{n \to \infty} t_n = 0$ such that almost surely

$$\lim_{n \to \infty} \|X_{t_n}\|_H = 0.$$

Moreover, almost surely there exists $\delta > 0$ such that the path of X is continuous in $[0, \delta)$. Consequently, we have almost surely

$$\lim_{t \to 0} \|X_t\|_H = 0.$$

Due to the Cauchy–Schwarz inequality, this yields almost surely $\lim_{t \to 0} \langle X_t, e_k \rangle_H = 0$ and thus

$$\lim_{t \to 0} e^{\langle X_t, e_k \rangle_H} = 1$$

uniformly in k. Hence, we have almost surely

$$|\langle w, f_t - f_0 \rangle_H| = \sum_{k \in I} \langle f_0, e_k \rangle_H \, \langle w, e_k \rangle_H \left(e^{\langle X_t, e_k \rangle_H} - 1 \right) \to 0 \quad \text{for } t \to 0.$$

We apply this limit to (6.3) and the proof is complete. □

Let $T > 0$ be the maturity of the Asian option. By definition, the value of the option depends on A_T. In addition, it may depend on the state f_T of the underlyings at maturity, e.g., in the case of a floating strike. The state f_T in turn is a function of the

centered driving process Z_T defined in (2.10). As before, we can write f_t as the function

$$f_t : \begin{cases} H \to H, \\ z \;\mapsto \sum_{k\in I} \langle f_0, e_k\rangle_H \; e^{\left\langle \int_0^t \gamma_s \, ds + z,\, e_k \right\rangle_H} e_k \end{cases} \tag{6.4}$$

for $t \in [0,T]$ (compare (2.3)).

We denote the value of the option at time $t \in [0,T]$, discounted to time 0, by

$$\widehat{V}(t,z,a) := e^{-rT} E\big[G(Z_T, A_T)\big| Z_t = z, A_t = a\big] \quad \text{for every } z \in H, a \in \mathbb{R}. \tag{6.5}$$

This is the conditional expectation of the payoff $G : H \times \mathbb{R} \to \mathbb{R}$ at maturity T given the current state $z \in H$ of the underlying assets and the average $a \in \mathbb{R}$. Generalizing Assumption 4.1, we make the following assumption concerning the payoff.

Assumption 6.2. *We assume that there are constants L_z^G and L_a^G such that the payoff function G satisfies the Lipschitz conditions*

$$\begin{aligned} |G(z_1,a) - G(z_2,a)| &\le L_z^G \, \|z_1 - z_2\|_H \quad \text{for every } z_1, z_2 \in H,\ a \in \mathbb{R}, \\ |G(z,a_1) - G(z,a_2)| &\le L_a^G \, |a_1 - a_2| \qquad \text{for every } z \in H,\ a_1, a_2 \in \mathbb{R}. \end{aligned}$$

Note that this assumption is satisfied, e.g., for Asian call and put options on A_T with fixed or floating strike. The one-dimensional partial derivative of $\widehat{V}$ with respect to the average is denoted by $\partial_a \widehat{V}$. We can now state the Hilbert space valued PIDE for the Asian option price $\widehat{V}$.

Theorem 6.3. *Suppose that the discounted price $\widehat{V}$ defined in* (6.5) *is continuously differentiable with respect to t and twice continuously differentiable with respect to z and a. Moreover, assume that the second derivative with respect to z restricted to an arbitrary bounded subset of H is a uniformly continuous mapping to the Hilbert–Schmidt space $\mathrm{L_{HS}}(H,H)$. Then $\widehat{V}$ is a classical solution of the PIDE*

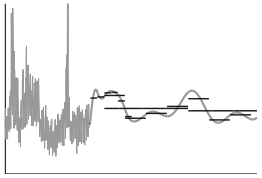
$$\begin{aligned} -\partial_t \widehat{V}(t,z,a) = {} & \frac{1}{2} \operatorname{tr}\Big(D_z^2 \widehat{V}(t,z,a)\, \sigma_t Q \sigma_t^* \Big) + \frac{1}{t}\big(\langle w, f_t(z)\rangle_H - a \big)\, \partial_a \widehat{V}(t,z,a) \\ & + \int_E \Big[\widehat{V}\big(t, z + \eta_t(\zeta), a\big) - \widehat{V}(t,z,a) - D_z \widehat{V}(t,z,a)\, \eta_t(\zeta) \Big] \nu(d\zeta) \end{aligned} \tag{6.6}$$

with terminal condition

$$\widehat{V}(T,z,a) = e^{-rT} G(z,a)$$

for every $t \in (0,T)$, $z \in E_0(\mathcal{C}_{X_T})^\perp$, and $a \in \mathbb{R}$.

Proof. The proof is very similar to that of Theorem 5.8 for the European hedging PIDE. Applying Itô's formula for Hilbert space valued processes [62, Thm. D.2] to $\widehat{V}(t, Z_t, A_t)$,



$t > 0$, yields

$$
\begin{aligned}
&\widehat{V}(t, Z_t, A_t) = \\
&\quad \widehat{V}(0, Z_0, A_0) + \int_0^t \partial_t \widehat{V}(s-, Z_{s-}, A_s)\, ds + \int_0^t D_z \widehat{V}(s-, Z_{s-}, A_s)\, dZ_s \\
&\quad + \int_0^t \partial_a \widehat{V}(s-, Z_{s-}, A_s)\, dA_s + \frac{1}{2} \int_0^t D_z^2 \widehat{V}(s-, Z_{s-}, A_s)\, d[Z, Z]_s^c \\
&\quad + \sum_{0 \le s \le t} \Big[\widehat{V}(s, Z_s, A_s) - \widehat{V}(s-, Z_{s-}, A_s) - D_z \widehat{V}(s-, Z_{s-}, A_s)\, (Z_s - Z_{s-}) \Big],
\end{aligned}
\tag{6.7}
$$

where $[Z, Z]^c$ denotes the continuous part of the square bracket process as defined in [62]. Note that the average process A is continuous and of finite variation. Hence, the jump part of the equation does not contain the partial derivative $\partial_a \widehat{V}$. For the same reason, the square bracket processes $[A, A]^c$ and $[A, Z]^c$ do not occur in the equation. We first simplify the covariation term. From the proof of Lemma 5.4, we know that

$$
[Z, Z]_t^c = \sum_{i,j \in I} e_i \otimes e_j \bigg(\int_0^t \langle \sigma_s Q \sigma_s^*\, e_j, e_i \rangle_H \; ds \bigg),
$$

where $e_i \otimes e_j$ denotes the tensor product of the two basis elements. Thus, we get

$$
\int_0^t D_z^2 \widehat{V}(s-, Z_{s-}, A_s)\, d[Z, Z]_s^c = \int_0^t \operatorname{tr} \Big(D_z^2 \widehat{V}(s-, Z_{s-}, A_s)\, \sigma_s Q \sigma_s^* \Big)\, ds.
$$

Next we calculate dA_s. By definition (6.1) of A we have

$$
\langle w, f_s \rangle_H \, ds = d(sA_s) = A_s\, ds + s\, dA_s.
$$

Hence, we obtain

$$
dA_s = \frac{1}{s} \left(\langle w, f_s(Z_s) \rangle_H - A_s \right) ds.
$$

Finally, we reorganize the jump terms in (6.7). The result is

$$
\begin{aligned}
&d\widehat{V}(t, Z_t, A_t) = \\
&\quad \partial_t \widehat{V}(t-, Z_{t-}, A_t) dt + \frac{1}{2} \operatorname{tr} \Big(D_z^2 \widehat{V}(t-, Z_{t-}, A_t)\, \sigma_t Q \sigma_t^* \Big)\, dt \\
&\quad + \frac{1}{t} \big(\langle w, f_t(Z_{t-}) \rangle_H - A_t \big)\, \partial_a \widehat{V}(t-, Z_{t-}, A_t)\, dt \\
&\quad + \int_E \Big[\widehat{V}\big(t, Z_{t-} + \eta_t(\zeta), A_t\big) - \widehat{V}(t-, Z_{t-}, A_t) - D_z \widehat{V}(t-, Z_{t-}, A_t)\, \eta_t(\zeta) \Big]\, \nu(d\zeta)\, dt \\
&\quad + D_z \widehat{V}(t-, Z_{t-}, A_t) \sigma_t\, dW_t \\
&\quad + \int_E \Big[\widehat{V}\big(t, Z_{t-} + \eta_t(\zeta), A_t\big) - \widehat{V}(t-, Z_{t-}, A_t) \Big]\, \widetilde{M}(d\zeta, dt).
\end{aligned}
$$

The very same martingale arguments as in the proof of Theorem 4.5 can now be used

to conclude the proof. □

6.2 Pricing Asian Options with POD

The payoff G of the Asian option depends on both the centered driving process Z and the average process A, which is a function of f. Corollary 3.3 shows that Z can be approximated with the POD method along its whole trajectory for $t \in [0,T]$. Thus, to approximate (6.6) with a low-dimensional PIDE, it remains to show that A and f can be accurately represented with the POD basis as well. To this end, recall that f is defined as a deterministic function of Z by (6.4). If we apply this function to the projected process $\mathcal{P}_d Z$ for arbitrary $t \in [0,T]$, we obtain

$$f_t(\mathcal{P}_d Z_t) = \sum_{k \in I} \langle f_0, e_k \rangle_H \; e^{\left\langle \int_0^t \gamma_s \, ds + \mathcal{P}_d Z_t, \, e_k \right\rangle_H} e_k \in H.$$

The following proposition is the foundation for generalizing the POD method to Asian options.

Proposition 6.4. *There is a constant $C > 0$ (depending on T) such that*

$$E\Big| \langle w, f_t(Z_t) \rangle_H - \langle w, f_t(\mathcal{P}_d Z_t) \rangle_H \Big| \leq C \, \|w\|_H \sqrt{\sum_{l=d+1}^{\dim(H)} \mu_l}$$

for every $t \in [0,T]$.

Proof. By definition of f_t, we get

$$\begin{aligned} & E\Big| \langle w, f_t(Z_t) \rangle_H - \langle w, f_t(\mathcal{P}_d Z_t) \rangle_H \Big| \\ & \quad = E\left| \sum_{k \in I} \langle w, e_k \rangle_H \; \langle f_0, e_k \rangle_H \left(e^{\left\langle \int_0^t \gamma_s \, ds + Z_t, \, e_k \right\rangle_H} - e^{\left\langle \int_0^t \gamma_s \, ds + \mathcal{P}_d Z_t, \, e_k \right\rangle_H} \right) \right| \\ & \quad \leq E\left[\sum_{k \in I} \left| \langle w, e_k \rangle_H \; \langle f_0, e_k \rangle_H \; e^{\int_0^t \langle \gamma_s, \, e_k \rangle_H \, ds} \left(e^{\langle Z_t, \, e_k \rangle_H} - e^{\langle \mathcal{P}_d Z_t, \, e_k \rangle_H} \right) \right| \right]. \end{aligned} \tag{6.8}$$

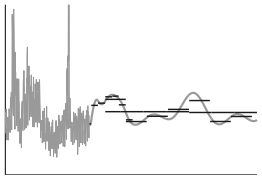
For the term depending on γ, we use Assumption 2.1 and obtain

$$\left| \int_0^t \langle \gamma_s, \, e_k \rangle_H \, ds \right| \leq \int_0^t \|\gamma_s\|_H \, ds \leq C_1 \left(\int_0^t \|\gamma_s\|_H^2 \, ds \right)^{\frac{1}{2}} \leq C_2,$$

with positive constants C_1, C_2 depending on T but not on t. Next, we apply the mean-value theorem to the exponential function and make use of the self-adjointness of the projection operator $\mathcal{P}_d$ for the estimate

$$\begin{aligned} \left| e^{\langle Z_t, \, e_k \rangle_H} - e^{\langle \mathcal{P}_d Z_t, \, e_k \rangle_H} \right| & \leq e^{\max\{\langle Z_t, \, e_k \rangle_H, \langle \mathcal{P}_d Z_t, \, e_k \rangle_H\}} \, |\langle Z_t - \mathcal{P}_d Z_t, \, e_k \rangle_H| \\ & \leq e^{\max\{\langle Z_t, \, e_k \rangle_H, \langle Z_t, \, \mathcal{P}_d e_k \rangle_H\}} \, \|Z_t - \mathcal{P}_d Z_t\|_H \end{aligned}$$

for every $k \in I$. Inserting these results into (6.8) and using the monotone convergence



theorem yields

$$E\Big|\langle w, f_t(Z_t)\rangle_H - \langle w, f_t(\mathcal{P}_d Z_t)\rangle_H\Big| \\ \leq C \sum_{k\in I} |\langle w, e_k\rangle_H \ \langle f_0, e_k\rangle_H| \ E\left[e^{\max\{\langle Z_t, e_k\rangle_H, \langle Z_t, \mathcal{P}_d e_k\rangle_H\}} \|Z_t - \mathcal{P}_d Z_t\|_H\right].$$

With the Cauchy–Schwarz inequality, we find

$$E\Big|\langle w, f_t(Z_t)\rangle_H - \langle w, f_t(\mathcal{P}_d Z_t)\rangle_H\Big| \\ \leq C \sum_{k\in\mathbb{N}} |\langle w, e_k\rangle_H \ \langle f_0, e_k\rangle_H| \left(E\left[e^{2\max\{\langle Z_t, e_k\rangle_H, \langle Z_t, \mathcal{P}_d e_k\rangle_H\}}\right]\right)^{\frac{1}{2}} \left(E\|Z_t - \mathcal{P}_d Z_t\|_H^2\right)^{\frac{1}{2}}.$$

For the first expectation, we use Proposition 2.4:

$$\begin{aligned} E\left[e^{2\max\{\langle Z_t, e_k\rangle_H, \langle Z_t, \mathcal{P}_d e_k\rangle_H\}}\right] &= E\left[\max\{e^{\langle Z_t, 2e_k\rangle_H}, e^{\langle Z_t, 2\mathcal{P}_d e_k\rangle_H}\}\right] \\ &\leq E\left[e^{\langle Z_t, 2e_k\rangle_H} + e^{\langle Z_t, 2\mathcal{P}_d e_k\rangle_H}\right] \\ &\leq C_3 e^{C_4 T} \end{aligned}$$

with positive constants C_3, C_4. The Cauchy–Schwarz inequality in $l^2(I)$ yields the following bound for the remaining sum in k:

$$\sum_{k\in I} |\langle w, e_k\rangle_H \ \langle f_0, e_k\rangle_H| \leq \|w\|_H \|f_0\|_H.$$

By Corollary 3.3, we thus have

$$\begin{aligned} E\Big|\langle w, f_t(Z_t)\rangle_H - \langle w, f_t(\mathcal{P}_d Z_t)\rangle_H\Big| &\leq C \|w\|_H \|f_0\|_H \left(E\|Z_t - \mathcal{P}_d Z_t\|_H^2\right)^{\frac{1}{2}} \\ &\leq C \|w\|_H \|f_0\|_H \sqrt{\sum_{l=d+1}^{\dim(H)} \mu_l}. \end{aligned} \tag{6.9}$$

□

Although $f_t(\mathcal{P}_d Z_t)$ is still an element of the possibly infinite-dimensional Hilbert space H, it can be computed from the d-dimensional object $\mathcal{P}_d Z_t$. This makes the approximation suitable for numerical computations. Similar to (6.1), we define the arithmetic average corresponding to $f_t(\mathcal{P}_d Z_t)$ by

$$A_t^d := \frac{1}{t}\int_0^t \langle w, f_s(\mathcal{P}_d Z_s)\rangle_H \, ds \in \mathbb{R}$$

for $t > 0$. Exactly like in (6.2), we set

$$A_0^d := \langle w, f_0(\mathcal{P}_d Z_0)\rangle_H = \langle w, f_0\rangle_H.$$

We find the following estimate for the approximation error.

Theorem 6.5. *There is a constant $C > 0$ (depending on T) such that*

$$E\left|A_t - A_t^d\right| \leq C \left\|w\right\|_H \sqrt{\sum_{l=d+1}^{\dim(H)} \mu_l}$$

for every $t \in [0,T]$.

Proof. By definition, $A_0^d = A_0$. For $t > 0$, we have

$$\begin{aligned} E\left|A_t - A_t^d\right| &= \frac{1}{t} E\left|\int_0^t \left\langle w, f_s(Z_s) - f_s(\mathcal{P}_d Z_s)\right\rangle_H ds\right| \\ &\leq \frac{1}{t} E \int_0^t \left|\left\langle w, f_s(Z_s) - f_s(\mathcal{P}_d Z_s)\right\rangle_H\right| ds. \end{aligned}$$

Using Fubini's theorem and applying Proposition 6.4 yields

$$\begin{aligned} E\left|A_t - A_t^d\right| &\leq \frac{1}{t} \int_0^t E\left|\left\langle w, f_s(Z_s) - f_s(\mathcal{P}_d Z_s)\right\rangle_H\right| ds \\ &\leq \frac{1}{t} \int_0^t C \left\|w\right\|_H \sqrt{\sum_{l=d+1}^{\dim(H)} \mu_l}\, ds. \end{aligned}$$

Since the integrand does no longer depend on the integration variable s, the proof is complete. □

In the time-homogeneous case, we obtain a t-dependent estimate for the approximation error.

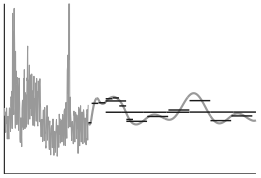
Corollary 6.6. *Suppose that Z is a time-homogeneous jump-diffusion process. Then there is a constant $C > 0$ (depending on T) such that*

$$E\left|A_t - A_t^d\right| \leq C \left\|w\right\|_H \sqrt{\frac{t}{T} \sum_{l=d+1}^{\dim(H)} \mu_l}$$

for every $t \in [0,T]$.

Proof. The proof is identical to that of Proposition 6.4 up to equation (6.9). Starting from there, we apply Corollary 3.4 and obtain

$$E\left|\left\langle w, S_t(Z_t)\right\rangle_H - \left\langle w, S_t(\mathcal{P}_d Z_t)\right\rangle_H\right| \leq C \left\|w\right\|_H \left\|f_0\right\|_H \sqrt{\frac{t}{T} \sum_{l=d+1}^{\dim(H)} \mu_l}.$$



We proceed as in the proof of Theorem 6.5 and find

$$E\left|A_t - A_t^d\right| \leq \frac{1}{t}\int_0^t C\,\|w\|_H \sqrt{\frac{s}{T}\sum_{l=d+1}^{\dim(H)} \mu_l}\,ds.$$

Since

$$\frac{1}{t}\int_0^t \sqrt{\frac{s}{T}}\,ds = \frac{2}{3}\sqrt{\frac{t}{T}},$$

the proof is complete. □

In the previous theorems, we have seen how to apply POD to the processes on which the payoff G of the Asian option depends: the centered process Z and the average A. Exactly as in the European case, we use these results to find a finite-dimensional approximation of the discounted option value $\widehat{V}$. For $t \in [0,T]$, we define

$$\widehat{V}_d(t,z,a) := e^{-rT} E\big[G(\mathcal{P}_d Z_T, A_T^d)\big|\mathcal{P}_d Z_t = z, A_t^d = a\big] \quad \text{for every } z \in U_d, a \in \mathbb{R}. \tag{6.10}$$

In contrast to the definition of $\widehat{V}$ in (6.5), the payoff is applied to the projected random variables $\mathcal{P}_d Z_T$ and A_T^d here instead of Z_T and A_T. Thus, $\widehat{V}_d$ is defined on the finite dimensional domain $[0,T] \times U_d \times \mathbb{R}$, which allows for numerical discretization. Similar to Theorem 6.6, we find that $\widehat{V}_d$ satisfies a PIDE. This PIDE is finite-dimensional.

Theorem 6.7. *Suppose that the approximated Asian option value $\widehat{V}_d$ defined in* (6.10) *is continuously differentiable with respect to t and twice continuously differentiable with respect to z and a. Then $\widehat{V}_d$ is a classical solution of the PIDE*

$$\begin{aligned} -\partial_t \widehat{V}_d(t,z,a) &= \frac{1}{2}\sum_{i,j=1}^{d} c_{ij}(t)\,\partial_i\partial_j \widehat{V}_d(t,z,a) + \frac{1}{t}\big(\langle w, f_t(z)\rangle_H - a\big)\,\partial_a \widehat{V}(t,z,a) \\ &\quad + \int_E \Big[\widehat{V}_d(t, z+\mathcal{P}_d\eta_t(y), a) - \widehat{V}_d(t,z,a) \\ &\qquad - \sum_{i=1}^{d} \langle \eta_t(y), p_i\rangle_H\,\partial_i \widehat{V}_d(t,z,a)\Big]\nu(dy), \end{aligned} \tag{6.11}$$

with time-dependent coefficients

$$c_{ij}(t) := \langle \sigma_t Q \sigma_t^* p_i, p_j\rangle_H, \quad i,j = 1,\dots,d,$$

and terminal condition

$$\widehat{V}_d(T,z,a) = e^{-rT} G(z,a)$$

for every $t \in (0,T)$, $z \in U_d$, and $a \in \mathbb{R}$.

Proof. This can be shown along the very same lines as in the proof of Theorems 5.8 and 6.3. The main difference is that we make use of a finite-dimensional version of Itô's

formula (see, e.g., [24, Prop. 8.19]). This yields finite sums of second derivatives instead of the trace operator. □

The value of the Asian option at time $t = 0$ is given by $\widehat{V}(0, 0, \langle w, f_0 \rangle_H)$, since $Z_0 = 0 \in H$ and $A_0 = \langle w, f_0 \rangle_H \in \mathbb{R}$ by definition. The solution of the finite-dimensional PIDE yields $\widehat{V}_d(0, 0, \langle w, f_0 \rangle_H)$. The following theorem states an upper bound of the approximation error for the option value. It is our main convergence result concerning Asian options.

Theorem 6.8. *There is a constant $C > 0$ (depending on T and w) such that*

$$\left| \widehat{V}(0, 0, \langle w, f_0 \rangle_H) - \widehat{V}_d(0, 0, \langle w, f_0 \rangle_H) \right| \leq C \sqrt{\sum_{l=d+1}^{\dim(H)} \mu_l}.$$

Proof. We start with the definition of $\widehat{V}$ and $\widehat{V}_d$ and make use of Assumption 4.1 to find

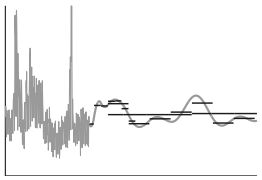
$$\begin{aligned} \left| \widehat{V}(0, 0, A_0) - \widehat{V}_d(0, 0, A_0^d) \right| &= e^{-rT} \left| E[G(Z_T, A_T) - E[G(\mathcal{P}_d Z_T, A_T^d)] \right| \\ &\leq e^{-rT} E \left[L_z^G \left\| Z_T - \mathcal{P}_d Z_T \right\|_H + L_a^G \left| A_T - A_T^d \right| \right] \\ &\leq e^{-rT} \max\{L_z^G, L_a^G\} \left(E \left\| Z_T - \mathcal{P}_d Z_T \right\|_H + E \left| A_T - A_T^d \right| \right). \end{aligned}$$

With the Cauchy–Schwarz inequality, we get

$$\left| \widehat{V}(0, 0, A_0) - \widehat{V}_d(0, 0, A_0^d) \right| \leq C \left[\left(E \left\| Z_T - \mathcal{P}_d Z_T \right\|_H^2 \right)^{\frac{1}{2}} + E \left| A_T - A_T^d \right| \right].$$

Applying Theorem 3.2 to $E \left\| Z_T - \mathcal{P}_d Z_T \right\|_H^2$ and Theorem 6.5 to $E \left| A_T - A_T^d \right|$ concludes the proof. □

There is an additional numerical difficulty when dealing with Asian options. The PIDE (6.11) for Asian options features no diffusion (second derivative) with respect to the arithmetic average a. This requires special attention. Equations of this kind are often termed "degenerate parabolic" PIDEs. A large number of authors has dealt with such problems, see, e.g., [5, 14, 19, 84] and the references therein. Since the dimension reduced equation is finite-dimensional, the numerical schemes and convergence results presented there can be applied directly. These include, among others, flux limiting methods, operator splitting, and difference-quadrature methods.



7 Bermudan Options

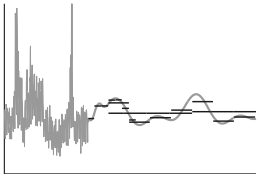
In this section, we extend the POD method to multidimensional Bermudan options with a finite number of exercise dates. While the dimension reduction for Bermudan options works just as well as in the case of European options, the option prices do not satisfy similar smoothness criteria. For European options, it is possible to solve the projected pricing problem (4.5) for relatively large dimensions, using Fourier or PIDE methods on sparse grids. Sparse grids, though, can be applied efficiently only for smooth functions. In the case of European options, the necessary differentiability properties of $\widehat{V}_d$ are ensured by the smoothing effect of the diffusive part of the process. For Bermudan options, however, a non-differentiable maximum function is involved at every exercise point. The resulting solutions are not sufficiently smooth for sparse grid convergence results, and the effect gets worse if the number of exercise points is increased. Moreover, the condition number of the linear equation systems of the corresponding time discretized PIDEs may increase substantially. Hence, the use of sparse grids in the context of Bermudan options poses considerable theoretical and practical challenges.

Using full grids instead, we are restricted to very low-dimensional POD approximations (dimension three or less). The results from such a coarse approximation are not always sufficiently accurate for the purpose of option pricing. Thus, we will not stop here, but rather use the low-dimensional result as a starting point for a fast-converging Monte Carlo (MC) method.

MC simulations for Bermudan options are often based on dynamic programming principles. The Snell envelope is obtained through backward recursion. The conditional expectation of future cashflows at each exercise point can be approximated using regression, an idea first introduced by Longstaff and Schwartz [60], see also [7, 53]. The option is exercised if and only if its current intrinsic value is larger than the expected future gain. This method yields an estimate of the optimal exercise strategy, and is thus an approximation from below for the fair price. A different approach uses a dual representation of the price, which allows for the computation of an upper bound [67]. For an overview of Bermudan MC pricing see [54] and the references therein.

The major drawback of all MC algorithms is their comparatively slow convergence rate. There are several techniques, subsumed under the term *variance reduction*, which can help obtaining more accurate results with fewer simulated paths. These include antithetic variables, importance sampling, and control variables [2, 36, 38]. We will focus on the latter and present an improved Longstaff-Schwartz algorithm using the POD approximation of the option price as the control variable.

The expectation of the low-dimensional POD approximation can be computed efficiently with any method suitable for low-dimensional Bermudan options. These include partial integro-differential equations (PIDEs) [29, 34], but also Fourier transform methods [22, 23, 58]. Once this is done, the solution can be used for two purposes: first, it can serve as a control variable. As the approximation is highly correlated with the full-dimensional price process, this results in a substantially decreased variance of the modified MC estimator. Second, the POD solution is a candidate for the minimizing martingale in Rogers' dual MC method [67]. We discuss both approaches. The dimension of the projection can be freely chosen. This allows for a trade-off between reduced variance and increased effort for the computation of the low-dimensional expectation.



By choosing the number of exercise dates sufficiently high, the methods presented here can, of course, be used to approximate American options with a continuum of exercise possibilities. Note that, in fact, whenever MC simulation is used for pricing American options, this actually amounts to an approximation with a Bermudan option on the MC time discretization grid.

7.1 Pricing Bermudan Options with POD

Since we will apply MC methods, which inherently require a discretization of the state space, we focus on the stock basket example throughout this section. We thus consider the asset price vector $S_t \in \mathbb{R}^n$, $t \in [0,T]$ for a set of n stocks, which we have introduced in Section 1.1. We write it as a function of the centered jump-diffusion Z:

$$S_t : \begin{cases} \mathbb{R}^n \to \mathbb{R}^n \\ Z_t \mapsto \left(S_0(1)\, e^{\int_0^t \gamma_s(1)\, ds + Z_t(1)}, \dots, S_0(n)\, e^{\int_0^t \gamma_s(n)\, ds + Z_t(n)} \right) \in \mathbb{R}^n, \end{cases}$$

where Z satisfies (2.10) as before.

A Bermudan option grants the holder the right to exercise at one of $N_{ex} \in \mathbb{N}$ admissible dates, which we denote by $0 \le t_1 < t_2 < \cdots < t_{N_{ex}} = T$. Let $\mathcal{T}(t,T)$ denote the set of all stopping times with values in $\{t_i | 1 \le i \le N_{ex} \text{ and } t_i \ge t\}$. For simplicity, we assume a constant interest rate $r > 0$. The discounted value $\widehat{V}$ of a Bermudan option which has not yet been exercised at time t is the solution of the optimal stopping problem

$$\widehat{V}(t,z) := \sup_{\tau \in \mathcal{T}(t,T)} E\big[e^{-r\tau} \widetilde{G}\big(S(Z_\tau)\big) \big| Z_t = z\big] \tag{7.1}$$

for $t \in [0,T]$, $z \in H$. The function $\widetilde{G} : \mathbb{R}^n \to \mathbb{R}$ is the payoff G written in terms of the asset price process S, i.e., $G(z) = \widetilde{G}(S(z))$. Throughout the section on Bermudan options, we will relax Assumption 4.1 and make the following, somewhat weaker assumption instead.

Assumption 7.1. *Suppose that the payoff function $\widetilde{G}$ is Lipschitz continuous in S on H with Lipschitz constant $L_{\widetilde{G}}$.*

The aim when pricing Bermudan options is to find the optimal exercise time for (7.1). It is well known that this can be done by backward dynamic programming: at time $t = T$ the value of the option is

$$\widehat{V}(T,z) = e^{-rT} \widetilde{G}\big(S(Z_T)\big). \tag{7.2}$$

For any previous exercise date t_i, $i = 0, \dots, N_{ex} - 1$, the value is

$$\widehat{V}(t_i,z) = \max\left\{ e^{-rt_i} \widetilde{G}\big(S(z)\big),\ E\left[\widehat{V}(t_{i+1}, Z_{t_{i+1}}) \big| Z_{t_i} = z \right] \right\}. \tag{7.3}$$

Hence, it is optimal to exercise at time t_i if and only if the intrinsic value $\widetilde{G}\big(S(z)\big)$ is larger than or equal to the expected discounted future cash flow (given the option is not

yet exercised). Computing the conditional expectations

$$E\left[\widehat{V}(t_{i+1}, Z_{t_{i+1}})\middle| Z_{t_i} = z\right]$$

for every exercise date is the basic challenge. The fair value of the option at time $t = 0$ is then given by $\widehat{V}(0,0)$.

Corollary 3.3 shows how to approximate every Z_t, $t \in [0,T]$, with POD. We use the POD projection $\mathcal{P}_d$ to define the function

$$\widehat{V}_d(t, z_d) := \sup_{\tau \in \mathcal{T}(t,T)} E\big[e^{-r\tau}\widetilde{G}\big(S(\mathcal{P}_d Z_\tau)\big)\big|\mathcal{P}_d Z_t = z_d\big] \tag{7.4}$$

for every $z_d \in \mathbb{R}^d$, $t \in [0,T]$. This is in fact the price process of a d-dimensional Bermudan option. If d is chosen sufficiently small, this price can be computed efficiently with Fourier or PIDE methods. In numerical experiments (compare Section 8), $d \leq 3$ gives good results. Since the payoff in (7.4) in general is different from the payoff in (7.1), the corresponding optimal stopping times need not be identical. Nevertheless, we can state an error estimate for the difference $\big|\widehat{V}_d(0,0) - \widehat{V}(0,0)\big|$. In order to show this convergence result, we will make use of the following lemma. It is concerned with the approximation error for a fixed stopping time.

Lemma 7.2. *Let $\tau \in \mathcal{T}(0,T)$ be a fixed stopping time and let $\mu_1 \geq \mu_2 \geq \ldots \geq 0$ be the eigenvalues of the covariance matrix $\mathcal{C}_{X_T}$. Then there exists a constant $C > 0$ (independent of d) such that*

$$\Big|E\big[e^{-r\tau}\widetilde{G}\big(S(\mathcal{P}_d Z_\tau)\big)\big] - E\big[e^{-r\tau}\widetilde{G}\big(S(Z_\tau)\big)\big]\Big| \leq C\sqrt{\sum_{l=d+1}^{n}\mu_l}$$

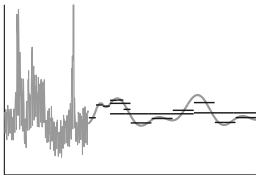
for every $d = 1, \ldots, n$.

Proof. Throughout the proof, we denote every constant factor depending on T but not on d by C, i.e., as before, the value of C is not fixed. Using Assumption 7.1 and the definition of S, we obtain

$$\begin{aligned}
&\Big|E\big[e^{-r\tau}\widetilde{G}\big(S(\mathcal{P}_d Z_\tau)\big)\big] - E\big[e^{-r\tau}\widetilde{G}\big(S(Z_\tau)\big)\big]\Big| \\
&\qquad \leq L_{\widetilde{G}}\, E\left\|S(\mathcal{P}_d Z_\tau) - S(Z_\tau)\right\|_{\mathbb{R}^n} \\
&\qquad = L_{\widetilde{G}}\, E\left\|\left(S_0(i)\, e^{\int_0^\tau \gamma_t(i)\,dt}\,\left|e^{(\mathcal{P}_d Z_\tau)(i)} - e^{Z_\tau(i)}\right|\right)_{i=1}^{n}\right\|_{\mathbb{R}^n}.
\end{aligned}$$

Since all norms on $\mathbb{R}^n$ are equivalent, we can use the 1-norm and obtain

$$\begin{aligned}
&\Big|E\big[e^{-r\tau}\widetilde{G}\big(S(\mathcal{P}_d Z_\tau)\big)\big] - E\big[e^{-r\tau}\widetilde{G}\big(S(Z_\tau)\big)\big]\Big| \\
&\qquad \leq C\, E\left[\sum_{i=1}^{n} S_0(i)\, e^{\left|\int_0^\tau \gamma_t(i)\,dt\right|}\,\left|e^{(\mathcal{P}_d Z_\tau)(i)} - e^{Z_\tau(i)}\right|\right].
\end{aligned} \tag{7.5}$$



For the term depending on γ, we use Assumption 2.1 and obtain

$$\left| \int_0^\tau \gamma_t(i)\, dt \right| \leq \int_0^T \|\gamma_t\|_{\mathbb{R}^n}\, dt \leq C \left(\int_0^T \|\gamma_t\|_{\mathbb{R}^n}^2\, dt \right)^{\frac{1}{2}} \leq C.$$

Next, we apply the mean-value theorem to the exponential function for the estimate

$$\left| e^{(\mathcal{P}_d Z_\tau)(i)} - e^{Z_\tau(i)} \right| \leq e^{\max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \left\| \mathcal{P}_d Z_\tau - Z_\tau \right\|_{\mathbb{R}^n}.$$

Inserting these estimates into (7.5) and applying the Cauchy–Schwarz inequality yields

(7.6)
$$\begin{aligned} &\left| E\big[e^{-r\tau} \widetilde{G}\big(S(\mathcal{P}_d Z_\tau)\big)\big] - E\big[e^{-r\tau} \widetilde{G}\big(S(Z_\tau)\big)\big] \right| \\ &\qquad \leq C \sum_{i=1}^n S_0(i) E\Big[e^{\max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \left\| \mathcal{P}_d Z_\tau - Z_\tau \right\|_{\mathbb{R}^n} \Big] \\ &\qquad \leq C \sum_{i=1}^n S_0(i) \Big(E\Big[e^{2\max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \Big] \Big)^{\frac{1}{2}} \Big(E \left\| \mathcal{P}_d Z_\tau - Z_\tau \right\|_{\mathbb{R}^n}^2 \Big)^{\frac{1}{2}}. \end{aligned}$$

In order to get rid of the stopping time τ, we will make use of Doob's inequality. The centered process Z is a martingale by construction, and so is $\mathcal{P}_d Z$, since $\mathcal{P}_d$ is a linear operator. The norm function is convex. Hence, $\left\| \mathcal{P}_d Z_\tau - Z_\tau \right\|_{\mathbb{R}^n}$ is a non-negative submartingale. Doob's inequality yields

$$E \left\| \mathcal{P}_d Z_\tau - Z_\tau \right\|_{\mathbb{R}^n}^2 \leq E\Big(\sup_{t \in [0,T]} \left\| \mathcal{P}_d Z_t - Z_t \right\|_{\mathbb{R}^n} \Big)^2 \leq 4E \left\| \mathcal{P}_d Z_T - Z_T \right\|_{\mathbb{R}^n}^2. \tag{7.7}$$

For the exponential term in (7.6), we find

$$e^{\max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \leq e^{(\mathcal{P}_d Z_\tau)(i)} + e^{Z_\tau(i)} = e^{\langle Z_\tau, \mathcal{P}_d u_i \rangle_{\mathbb{R}^n}} + e^{\langle Z_\tau, u_i \rangle_{\mathbb{R}^n}},$$

where u_i denotes the ith standard unit vector. Since the exponential function is convex, both $\big(e^{\langle Z_t, \mathcal{P}_d u_i \rangle_{\mathbb{R}^n}}\big)_{t \in [0,T]}$ and $\big(e^{\langle Z_t, u_i \rangle_{\mathbb{R}^n}}\big)_{t \in [0,T]}$ are submartingales. Using Young's and Doob's inequalities, we get

$$\begin{aligned} E\Big[e^{2\max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \Big] &\leq E\Big[\Big(e^{\langle Z_\tau, \mathcal{P}_d e_i \rangle_{\mathbb{R}^n}} + e^{\langle Z_\tau, e_i \rangle_{\mathbb{R}^n}} \Big)^2 \Big] \\ &\leq 2E\Big[\Big(e^{\langle Z_\tau, \mathcal{P}_d e_i \rangle_{\mathbb{R}^n}} \Big)^2 + \Big(e^{\langle Z_\tau, e_i \rangle_{\mathbb{R}^n}} \Big)^2 \Big] \\ &\leq 8E\Big[e^{2\langle Z_T, \mathcal{P}_d e_i \rangle_{\mathbb{R}^n}} + e^{2\langle Z_T, e_i \rangle_{\mathbb{R}^n}} \Big] \end{aligned}$$

With [45, Prop. 2.3], we obtain

$$E\Big[e^{2\max\{(\mathcal{P}_d Z_\tau)(i), Z_\tau(i)\}} \Big] \leq C. \tag{7.8}$$

Inserting (7.7) and (7.8) in (7.6) yields

$$\left|E\big[e^{-r\tau}\widetilde{G}\big(S(\mathcal{P}_d Z_\tau)\big)\big] - E\big[e^{-r\tau}\widetilde{G}\big(S(Z_\tau)\big)\big]\right| \leq C\left(E\left\|\mathcal{P}_d Z_T - Z_T\right\|_{\mathbb{R}^n}^2\right)^{\frac{1}{2}} \sum_{i=1}^{n} S_0(i).$$

Including the sum $\sum_{i=1}^{n} S_0(i)$ in the constant C and applying Theorem 3.2 to the remaining expectation concludes the proof. □

The following theorem states the main convergence result for Bermudan options.

Theorem 7.3. *Let $\mu_1 \geq \mu_2 \geq ... \geq 0$ be the eigenvalues of the covariance matrix $\mathcal{C}_{X_T}$. Then there exists a constant $C > 0$ such that*

$$\left|\widehat{V}_d(0,0) - \widehat{V}(0,0)\right| \leq C\sqrt{\sum_{l=d+1}^{n} \mu_l}$$

for every $d = 1, \ldots, n$.

Proof. Let

$$\tau_n := \operatorname{argsup}_{\tau\in\mathcal{T}(t,T)} E\big[e^{-r\tau}\widetilde{G}\big(S(Z_\tau)\big)\big]$$

and

$$\tau_d := \operatorname{argsup}_{\tau\in\mathcal{T}(t,T)} E\big[e^{-r\tau}\widetilde{G}\big(S(\mathcal{P}_d Z_\tau)\big)\big]$$

be optimal stopping times for $\widehat{V}(0,0)$ and $\widehat{V}_d(0,0)$, respectively. Then we have

$$E\big[e^{-r\tau_n}\widetilde{G}\big(S(Z_{\tau_n})\big)\big] \geq E\big[e^{-r\tau_d}\widetilde{G}\big(S(Z_{\tau_d})\big)\big] \tag{7.9}$$

and

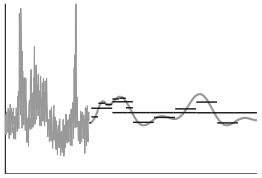
$$E\big[e^{-r\tau_d}\widetilde{G}\big(S(\mathcal{P}_d Z_{\tau_d})\big)\big] \geq E\big[e^{-r\tau_n}\widetilde{G}\big(S(\mathcal{P}_d Z_{\tau_n})\big)\big] \tag{7.10}$$

by construction. Moreover, we know from Lemma 7.2 that

$$\left|E\big[e^{-r\tau_x}\widetilde{G}\big(S(\mathcal{P}_d Z_{\tau_x})\big)\big] - E\big[e^{-r\tau_x}\widetilde{G}\big(S(Z_{\tau_x})\big)\big]\right| \leq C\sqrt{\sum_{l=d+1}^{n} \mu_l} \tag{7.11}$$

for $\tau_x \in \{\tau_n, \tau_d\}$. Combining (7.11) and (7.9), we find

$$\begin{aligned} E\big[e^{-r\tau_d}\widetilde{G}\big(S(\mathcal{P}_d Z_{\tau_d})\big)\big] &\leq E\big[e^{-r\tau_d}\widetilde{G}\big(S(Z_{\tau_d})\big)\big] + C\sqrt{\sum_{l=d+1}^{n} \mu_l} \\ &\leq E\big[e^{-r\tau_n}\widetilde{G}\big(S(Z_{\tau_n})\big)\big] + C\sqrt{\sum_{l=d+1}^{n} \mu_l}. \end{aligned}$$



On the other hand, using (7.10) and (7.11), we get

$$E\big[e^{-r\tau_d}\widetilde{G}\big(S(\mathcal{P}_d Z_{\tau_d})\big)\big] \geq E\big[e^{-r\tau_n}\widetilde{G}\big(S(\mathcal{P}_d Z_{\tau_n})\big)\big]$$
$$\geq E\big[e^{-r\tau_n}\widetilde{G}\big(S(Z_{\tau_n})\big)\big] - C\sqrt{\sum_{l=d+1}^{n}\mu_l}.$$

Together, these estimates yield

$$\Big|\widehat{V}_d(0,0) - \widehat{V}(0,0)\Big| = \Big|E\big[e^{-r\tau_d}\widetilde{G}\big(S(\mathcal{P}_d Z_{\tau_d})\big)\big] - E\big[e^{-r\tau_n}\widetilde{G}\big(S(Z_{\tau_n})\big)\big]\Big| \leq C\sqrt{\sum_{l=d+1}^{n}\mu_l}.$$

□

7.2 Variance-Reduced Least-Squares Monte Carlo

Pricing Bermudan options with MC simulations is more demanding than pricing European contracts. We will use the well established Longstaff–Schwartz least-squares MC method [60]. Starting with the terminal value (7.2) at time T, we iterate backwards ($i = N_{ex}-1, N_{ex}-2, \ldots, 1$) over all previous exercise dates using the recursion formula (7.3). We denote by $(\mathcal{F}_{t_i})_{i=1}^{N_{ex}}$ the natural filtration corresponding to the driving process Z. The conditional expectation

$$E\Big[\widehat{V}(t_{i+1}, Z_{t_{i+1}})\big|\mathcal{F}_{t_i}\Big] \tag{7.12}$$

is assumed to be a linear combination of a set of $\mathcal{F}_{t_i}$-adapted basis variables. This set of variables may, e.g., include the current state $Z_{t_i}(1), \ldots, Z_{t_i}(n)$ of the driving process, the values $S_{t_i}(1), \ldots, S_{t_i}(n)$ of the assets, and the value $P(Z_{t_i})$ of the function on which the option is written (the weighted average for index options, or the minimum or maximum for the corresponding put options). Polynomials in all of these quantities are also possible candidates for the regression. In practice, it turns out that the method is rather insensitive to the concrete choice of basis variables.

Once a set of variables is fixed, the basis variables are evaluated in each recursion step for each path of the MC simulation. The discounted future cashflow for each path when the option is not exercised is already known from the backward recursion. The conditional expectation (7.12) is then computed from simple linear regression *over all paths*. This is the major difference to European options: we make use of the information from all paths at the same time. Since the linear regression amounts to solving a symmetric linear equation system, the computational effort of the regression step is not linear in the number of paths. Moreover, it is harder to do these computations in parallel. The computational time of the whole least-squares algorithm, however, is often dominated by the simulation part (compare also the numerical experiments in Section 8).

As with all MC methods, a rather large number of paths is needed to obtain accurate approximations. For a discussion of convergence rates, see [6]. We denote the number of simulated paths by N. Let V_j^N, $j = 1, \ldots, N$, be the price computed for the jth

simulated path, using the exercise strategy given by least-squares MC. Once the number of paths and the exercise policy are fixed, these are iid copies of a random variable V^N. The MC estimator for the option price $\widehat{V}(0,0)$ is given by

$$\vartheta = \frac{1}{N}\sum_{j=1}^{N} V_j^N.$$

In order to improve the estimate, we will employ variance reduction with a control variable as described, e.g., in [36, chap. 1.3]. This amounts to finding a second random variable U^N, which is closely related to V^N, but whose expectation $E[U^N]$ can be computed much more efficiently. Then, we choose $\alpha \in \mathbb{R}$ and compute the new estimator

$$\vartheta_{vr}(\alpha) = \frac{1}{N}\sum_{j=1}^{N}\left[V_j^N + \alpha\left(U_j^N - E[U^N]\right)\right],$$

where U_j^N are iid copies of U^N. The expectation of $\vartheta_{vr}(\alpha)$ is obviously identical to that of ϑ, therefore no bias is introduced. The variance of the new estimator is given by

$$\mathrm{Var}(\vartheta_{vr}(\alpha)) = \frac{1}{N}\left(\mathrm{Var}(V^N) + 2\alpha\,\mathrm{Cov}(V^N,U^N) + \alpha^2\,\mathrm{Var}(U^N)\right).$$

The minimal possible variance

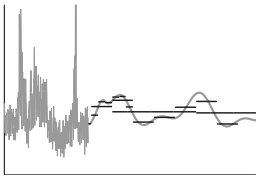
(7.13)
$$\vartheta_{vr}(\alpha^*) = \frac{1}{N}\left(\mathrm{Var}(V^N) - \frac{\mathrm{Cov}^2(V^N,U^N)}{\mathrm{Var}(U^N)}\right) = \frac{1}{N}\,\mathrm{Var}(V^N)\left(1-\mathrm{Corr}^2(V^N,U^N)\right)$$

is obtained for

$$\alpha^* = -\frac{\mathrm{Cov}(V^N,U^N)}{\mathrm{Var}(U^N)}.$$

This optimal value α^* cannot be calculated directly in practice, since at least the covariance $\mathrm{Cov}(V^N,U^N)$ is usually unknown. Thus, α^* has to be estimated. Since we simulate values of V^N and U^N anyway, this requires no additional effort. We can use the empirical estimates for $\mathrm{Cov}(V^N,U^N)$ and $\mathrm{Var}(U^N)$. Equation (7.13) shows that the variance of the improved estimator will be arbitrarily small, if the correlation $\mathrm{Corr}(V^N,U^N)$ of V^N and U^N is large.

We employ the dimension reduction presented in the previous section to obtain a suitable choice for U^N. The solution $\widehat{V}_d$ of the projected Bermudan pricing problem (7.4) converges to $\widehat{V}$ according to Theorem 7.3. Therefore, we set U^N equal to the approximated value of $\widehat{V}_d$ along each path. We can obtain this approximation with the very same regression method we use for $\widehat{V}$. Once we have simulated N paths for Z and approximated the pathwise value of $\widehat{V}$ with the least-squares method, computation of $\widehat{V}_d$ can be done with small additional effort for two reasons: first, we reuse the same paths and can thus skip the expensive simulation step. Second, since $\widehat{V}_d$ effectively depends only on the d-dimensional process $\mathcal{P}_d Z$, and $d \ll n$, we can reduce the number of basis variables for the regression significantly without loosing accuracy.



The main benefit of this choice of the control variable U^N is that any method suitable for low-dimensional Bermudan options can be used to compute the expectation $\widehat{V}_d(0,0)$. The steps required for this variance-reduced least-squares MC method are summarized in Algorithm 1.

Algorithm 1 Variance-reduced least-squares MC pricing

compute expectation $\widehat{V}_d(0,0) = \sup_{\tau\in\mathcal{T}(0,T)} E\big[e^{-r\tau}\widetilde{G}\big(S(\mathcal{P}_d Z_\tau)\big)\big]$ (PIDE or FFT)
simulate N paths $Z(1),\dots,Z(N)$
compute projected paths $\mathcal{P}_d Z(1),\dots,\mathcal{P}_d Z(N)$
for $j = 1,2,\dots,N$ **do** // terminal values
  $V_j^N = e^{-rT}\widetilde{G}\big(S(Z_T(j))\big)$
  $U_j^N = e^{-rT}\widetilde{G}\big(S(\mathcal{P}_d Z_T(j))\big)$
end for
for $i = N_{ex}-1, N_{ex}-2,\dots,1$ **do** // backward recursion
  regression with large basis set for $E\left[V_j^N \middle| Z_{t_i}\right]$
  regression with small basis set for $E\left[U_j^N \middle| \mathcal{P}_d Z_{t_i}\right]$
  for $j = 1,2,\dots,N$ **do** // pathwise exercise decision
    compute intrinsic values $v = e^{-rt_i}\widetilde{G}\big(S(Z_{t_i}(j))\big)$ and $u = e^{-rt_i}\widetilde{G}\big(S(\mathcal{P}_d Z_{t_i}(j))\big)$
    $V_j^N = \max\left\{v,\ E\left[V_j^N \middle| Z_{t_i}(j)\right]\right\}$
    $U_j^N = \max\left\{u,\ E\left[U_j^N \middle| \mathcal{P}_d Z_{t_i}(j)\right]\right\}$
  end for
end for
estimate mean μ_V and variance σ_V^2 of V^N
estimate mean μ_U and variance σ_U^2 of U^N
$\alpha = \frac{1}{N-1}\sum_{j=1}^N \frac{(V_j^N-\mu_V)(U_j^N-\mu_U)}{\sigma_V\sigma_U}$
$\theta_{vr}(\alpha) = \frac{1}{N}\sum_{j=1}^N \left[V_j^N + \alpha\left(U_j^N - \widehat{V}_d(0,0)\right)\right]$
return $\theta_{vr}(\alpha)$

There is a large variety of algorithms which can be used to compute the expectation $E[U^N] = \widehat{V}_d(0,0)$ numerically, most notably those based on PIDEs or on Fourier transforms. An overview of available methods can be found, e.g., in [24, 36]. The PIDE method is basically the same as discussed in Section 4, with the additional backward recursion (7.3) to obtain the Snell envelope. Alternatively, the function $\widehat{V}_d(t_i,z)$, $i = 1,\dots,N_{ex}$ can be represented in terms of Fourier transforms:

$$\widehat{V}_d(t_i,z) = \max\Bigg\{e^{-rt_i}\widetilde{G}\big(S(z)\big),\ \frac{e^{-rt_{i+1}}}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-iz^Tu}\,\varphi_{t_i,t_{i+1}}(-u)\int_{\mathbb{R}^d} e^{iu^Ty}\widetilde{G}\left(y+\int_{t_i}^{t_{i+1}}\gamma_s\,ds\right)dy\,du\Bigg\}$$

for $z\in\mathbb{R}^d$ and $i\in 1,\dots,N_{ex}-1$, where $\varphi_{t_i,t_{i+1}}$ denotes the characteristic function of the increment of Z over the interval $[t_i,t_{i+1}]$. In general, the payoff $\widetilde{G}$ is not integrable.

This can be remedied by either truncating the payoff or by subtracting a suitable smooth function. The Fourier transforms can then be approximated very efficiently with fast Fourier transforms (FFT). Since this is a well known method, we will not discuss it in detail here. Option pricing with Fourier methods is presented, e.g., in [22, 58].

7.3 Dual Method

Since the exercise policy found by linear regression is not necessarily optimal, the least-squares MC algorithm presented in the previous section actually computes an (arbitrarily precise) lower bound for the option price. The same value is of course also a lower bound for the value of the continuously exercisable American option with the same payoff function. The approach presented by Rogers [67] uses a duality argument to obtain an approximation from above. The Bermudan option price can be written as

$$\widehat{V}(0,0) = \inf_{M \in \mathcal{M}_0^1} E\left[\sup_{i=1,\dots,N_{ex}} \left(e^{-rt_i} \widetilde{G}\big(S(Z_{t_i})\big) - M_{t_i} \right) \right], \tag{7.14}$$

where $\mathcal{M}_0^1$ is the space of all martingales M satisfying $M_0 = 0$ and

$$E\left[\sup_{i=1,\dots,N_{ex}} |M_{t_i}| \right] < \infty.$$

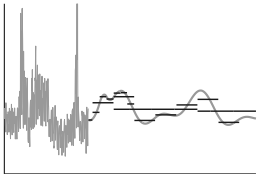
Rogers' idea is to pick a suitable martingale M and compute the expectation in (7.14) via MC simulation. Since the chosen martingale will in general not be optimal, the computed price is larger than the infimum and thus an upper bound. According to [40, Thm. 1], we could also insert a supermartingale into the dual formulation and still obtain an upper bound.

The choice of a "good" martingale $M \in \mathcal{M}_0^1$ is a delicate issue, since it is related to the increments of a hedging strategy for the option. In practice, however, any martingale related to the price of the option may yield remarkably accurate results. Since the projected price process is an approximation for the true price, we suggest setting

$$M_t := \begin{cases} \widehat{V}_d(t, \mathcal{P}_d Z_t) - \widehat{V}_d(0,0), & t \in [0,\tau], \\ e^{-r\tau} \widetilde{G}(S(\mathcal{P}_d Z_\tau)) - \widehat{V}_d(0,0), & t \in (\tau, T], \end{cases}$$

where τ denotes the optimal stopping time for the projected Bermudan option. Note that $\widehat{V}_d(\tau, \mathcal{P}_d Z_\tau) = e^{-r\tau} \widetilde{G}(S(\mathcal{P}_d Z_\tau))$ holds by construction. Moreover, $M_t + \widehat{V}_d(0,0)$ can be interpreted as the discounted value of a European option which grants the holder the same wealth as an optimally exercised Bermudan option at maturity. Thus, it is a martingale under the pricing measure. Algorithm 2 lists the computational steps of the POD-based dual method.

For practical applications, we are of course interested in bounds which are sufficiently sharp to serve as approximations of the true price. Therefore, we compare the dual MC method and the variance-reduced MC method with respect to computational speed and accuracy. In our numerical experiments (see Section 8), the dual method showed extremely fast convergence. A very small number of simulated paths for Z is sufficient



Algorithm 2 Duality based MC pricing

compute the solution $\widehat{V}_d$ on a space-time grid (PIDE or FFT)
simulate N paths paths $Z(1), \dots, Z(N)$
compute projected paths $\mathcal{P}_d Z(1), \dots, \mathcal{P}_d Z(N)$
for $j = 1, 2, \dots, N$ **do** // pathwise maximum
  initial value $v(j) = \widetilde{G}\big(S(0)\big)$
  $exercised\_projected = false$
  for $i = 1, 2, \dots, N_{ex}$ **do** // forward iteration
    intrinsic value $u = e^{-rt_i}\widetilde{G}\big(S(Z_{t_i}(j))\big)$
    if not *exercised_projected* **then**
      interpolate projected continuation value $M = \widehat{V}_d(t_i, \mathcal{P}_d Z_{t_i}(j)) - \widehat{V}_d(0,0)$
      intrinsic projected value $u_d = e^{-rt_i}\widetilde{G}\big(S(\mathcal{P}_d Z_{t_i}(j))\big)$
      if $u_d \geq M$ **then**
        $exercised\_projected = true$
      end if
    end if
    $v(j) = \max\{v(j), u - M\}$
  end for
end for
$\theta_{dual} = \frac{1}{N}\sum_{j=1}^{N} v(j)$
return θ_{dual}

to approximate the limit. In addition, there is no regression step. The individual paths can be processed completely in parallel. There are several caveats, though. In contrast to the variance reduction method, where we have used only the value $\widehat{V}_d(0,0)$, we have to compute the solution $\widehat{V}_d$ on a full space-time discretization grid in order to evaluate M. As before, this can be done with PIDE and FFT methods, but computing the full solution has several disadvantages. First, memory consumption is increased considerably. Second, since we have to truncate the computational domain to a compact subset of $\mathbb{R}^d$, the accuracy of the solution decreases when we approach the boundary of this subset. Moreover, the evaluation of M_t at arbitrary arguments requires interpolation and, thus, usually a finer discretization grid.

The most important drawback of the dual method is probably that the precision of the result depends on the chosen martingale M. If $\widehat{V}_d$ is too far from optimal, we cannot expect convergence to the true option value, no matter how many paths we simulate. In particular, we will see that for large numbers of assets with moderate correlation and many exercise possibilities, we would have to choose a large value for d in order to obtain a reasonably sharp bound. If, on the other hand, correlation is sufficiently high, the dual method is accurate and by far superior regarding computational speed.

8 Numerical Experiments

In this section, we will examine the performance of the presented pricing and hedging methods in simulation studies. It was a major part of this thesis to write efficient C++ code for this purpose. The resulting program contains methods for the simulation of time-inhomogeneous jump-diffusion processes, PIDE solving via finite elements and finite differences, POD-based dimension reduction, semi-definite quadratic optimization, least-squares Monte Carlo simulation, pricing with FFT, and many other mathematical utilities related to these topics. While large parts of the code have been written from scratch, the following open source C++ libraries have been employed for the implementation of specific tasks: the finite element framework *deal.II* [4], the general purpose linear algebra library *ALGLIB* [17], the geometric library *CGAL* [1] for interpolation tasks, and the fast Fourier transform implementation *FFTW* [35]. In order to make use of parallelization (for the MC simulation as well as the sparse grid combination technique), all experiments were run on a Linux workstation with eight Opteron CPU cores at 2.7 GHz.

8.1 Numerical PIDE Solution

This section contains a short overview of the numerical methods for solving finite-dimensional PIDEs of type (4.7). The same methods also apply to the hedging problem and, in combination with backward recursion, to Bermudan options. We will not go into details about convergence results for PIDE solvers but refer to recent literature instead. Finite difference methods for integro-differential equations are analyzed, e.g., in [25, 70], finite elements and wavelet compression techniques are described in [61, 81, 66].

Localization We consider PIDEs whose spatial domain is the whole of $\mathbb{R}^d$. The first step towards a numerical solution is therefore the localization to a bounded domain. To this end, we restrict the spatial part of the equation to a d-dimensional cuboid

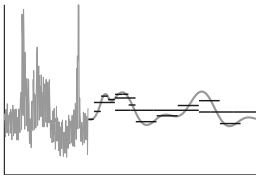
$$\Omega_R := [-R_1, R_1] \times [-R_2, R_2] \times \ldots \times [-R_d, R_d].$$

This simple domain can be described by a single vector $R = (R_1, \ldots, R_d) \in \mathbb{R}^d$. The probabilistic interpretation of the truncation is that we price a barrier option whose value is set to 0 if the process Z_L leaves Ω_R at any time before maturity. Under polynomial growth conditions for the payoff function G, one can show that the truncation error decreases exponentially with $\|R\|$ (cf., e.g., [81, Thm. 3.3.2]). The error is of course higher if the solution is evaluated closer to the boundary of Ω_R.

Analysis of the dependence of the localization error on the truncation parameter in one-dimensional problems shows that good results are achieved for R_1 greater than a certain multiple of the standard deviation of the jump-diffusion process at time T (cf., e.g., [24, Fig. 12.1]). By construction of the POD basis $(p_l)_{l=1}^d$, we have

$$\langle \mathcal{C}_{X_T} p_l, p_l \rangle_H = \mu_l \quad \text{for } l = 1, \ldots, d.$$

Hence, the eigenvalues μ_l describe the variance in direction of the POD vectors. This



suggests an adaptive truncation strategy, setting

$$R_l = C\sqrt{\mu_l}, \quad l = 1, \dots, d, \tag{8.1}$$

with a constant $C > 0$. This heuristic choice accounts for the decreasing variance in the coordinate directions (compare Proposition 3.5). It results in smaller domains Ω_R (compared to cubic domains) and allows for more accurate discretization results using the same number of grid points. We introduce artificial homogeneous Dirichlet boundary conditions and set

$$\widehat{V}_d^h(t,x) = 0 \quad \text{for every } t \in [0,T],\ x \in \partial\Omega_R,$$

where $\partial\Omega_R$ is the boundary of the domain and $\widehat{V}_d^h$ is the numerical approximation of $\widehat{V}_d$.

Discretization We solve the PIDE using a vertical method of lines, i.e., we first discretize the spatial operators and apply some time stepping for ordinary differential equations afterwards. Since we have already derived the variational formulation of the PIDE in Proposition 4.9, we can directly apply a finite element method to approximate the spatial derivatives in (4.9). Usually, finite elements have several advantages compared to finite differences. In particular, they allow for an easy discretization of geometrically complex domains, adaptive refinement, and higher-order approximations. Moreover, the theory of weak solutions allows for lower regularity assumptions than the classical differential operator.

However, in the specific setting of option pricing, these arguments are only partly valid. First, we may choose the shape of our computational domain arbitrarily due to localization. As described in the previous paragraph, we truncate the domain to a d-dimensional cuboid. Second, we have already shown that the European option price $\widehat{V}_d$ is a smooth classical solution to the PIDE (4.7). Third, despite the simple shape of the domain, a significant overhead is needed to compute and store the geometric information (neighbors, triangles, node ordering) for finite element discretizations.

Both finite elements and finite differences showed very similar accuracy and computational performance in all the numerical tests. Hence, we only report the results obtained with finite differences in the subsequent sections on numerical results. We define a regular but anisotropic grid g_α on Ω_R. The grid is described by a multiindex $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$, and the mesh size in each coordinate is given by $h_i := 2R_i/2^{\alpha_i}$, $i = 1, \dots, d$. The grid contains the points

$$x(\beta) := \big(-R_i + \beta_i h_i\big)_{i=1}^d \in \mathbb{R}^d, \quad \beta \in \{0, 1, \dots, 2^{\alpha_1}\} \times \cdots \times \{0, 1, \dots, 2^{\alpha_d}\}.$$

The corresponding discretized subspace is denoted $U_d^h \subset U_d \subset H$. The partial derivatives are approximated by central differences as follows:

$$\partial_{x_i} \widehat{V}_d\big(t, x(\beta)\big) \approx \frac{1}{2h_i}\big[\widehat{V}_d\big(t, x(\beta + u_i)\big) - \widehat{V}_d\big(t, x(\beta - u_i)\big)\big],$$

where u_i is the ith canonical unit vector. For the second derivatives, we have

$$\partial_{x_j}\partial_{x_i}\widehat{V}_d(t,x(\beta)) \approx \begin{cases} \frac{1}{4h_ih_j}\Big[\widehat{V}_d\big(t,x(\beta+u_i+u_j)\big) - \widehat{V}_d\big(t,x(\beta+u_i-u_j)\big) \\ \qquad -\widehat{V}_d\big(t,x(\beta-u_i+u_j)\big) + \widehat{V}_d\big(t,x(\beta-u_i-u_j)\big)\Big], & i \neq j, \\ \frac{1}{h_i^2}\Big[\widehat{V}_d\big(t,x(\beta+u_i)\big) - 2\widehat{V}_d\big(t,x(\beta)\big) + \widehat{V}_d\big(t,x(\beta-u_i)\big)\Big], & i = j. \end{cases}$$

The approximation of the nonlocal integral term will be addressed separately in the next paragraph.

After applying the finite difference scheme for the discretization in space, the resulting system of ordinary differential equations is solved with an appropriate time stepping scheme. Since the differential part of the PIDE is of parabolic type, we choose a discontinuous Galerkin method of order 1 for this purpose. Details on the topic (in particular error estimates) can be found in [77, Chap. 12]. Defining a partition $0 = t_0 < \cdots < t_{N_T} = T$ of $[0,T]$, we calculate a solution in the space

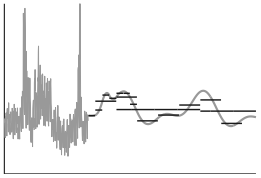
$$S_d^h := \big\{v \in \mathrm{L}^2(0,T;U_d^h);\ v\chi_{[t_{m-1},t_m)} \in \Pi_1(t_{m-1},t_m;U_d^h),\ m = 1,\ldots,N_T\big\},$$

where $\Pi_1(t_{m-1},t_m;U_d^h)$ is the space of polynomials of degree at most 1 having values in U_d^h. This method yields a stable algorithm, allowing for large time steps even in the presence of convection terms.

Nonlocal integral terms One of the main difficulties when solving a PIDE is the nonlocal nature of the integral term. In contrast to differential operators, this term involves the solution on the whole of $\mathbb{R}^d$. Since we have already introduced artificial zero boundary conditions, the solution can easily be continued with 0 outside the truncated domain Ω_R. However, numerical quadrature formulas will in general lead to full matrices. This is contradictory to the essential use of sparse matrices even for problems on relatively low-dimensional spaces. An efficient way to reduce the number of nonzero matrix entries is by using wavelet compression schemes. These make use of the fast decline of entries corresponding to wavelet basis functions with larger distance of their supports. Entries close to 0 are then discarded (cf. the references given at the beginning of this section).

Wavelets are the method of choice if the jump distribution in the additive model (2.1) covers large parts of the domain. The examples we are going to examine here, however, are rather factor models where for every $t \in [0,T]$, the jumps are restricted to one-dimensional subspaces spanned by $\eta_k(t) \in H$, $k = 1,\ldots,n_J \in \mathbb{N}$. If the number n_J of driving jump factors is low, a different approach is feasible as well. Instead of wavelet compression, direct numerical quadrature (Gauß–Kronrod or even Newton–Cotes methods) can be applied. We will now have a closer look at this specific case. We focus on European option pricing and use the equivalent time-homogeneous jump-diffusion process X_L defined in Section (2.3). The time-inhomogeneous case can be treated in the very same way.

Let $0 = t_0 < t_1 < \ldots < t_{N_T} = T$, $N_T \in \mathbb{N}$, be the time discretization grid. It is not necessary to compute the Lévy measure ν_L defined in (2.11) explicitly. Instead, we



define measures

$$\nu_L(t_k, t_l; B) := \frac{1}{t_l - t_k} \int_{t_k}^{t_l} \int_H \chi_B\big(\eta_s(y)\big)\, \nu(dy)\, ds,$$
$$\nu_L(t; B) := \int_H \chi_B\big(\eta_t(y)\big)\, \nu(dy)$$

for every $B \subset \mathcal{B}(H)$, $0 \le k < l \le N_T$, and $t \ge 0$. For each time step $[t_k, t_{k+1}]$ in the method of lines, we use integrals with respect to $\nu_L(t_k, t_{k+1}; \cdot)$, i.e., the equivalent Lévy measure for the current time interval. We are looking for an approximation of the integral

$$\begin{aligned}\int_H \Big[\widehat{V}_d(t, x + \mathcal{P}_d\zeta) - \widehat{V}_d(t, x) - \sum_{i=1}^{d} \langle \zeta, p_i\rangle_H\, \partial_{x_i}\widehat{V}_d(t, x)\Big]\, \nu_L(t_k, t_{k+1}; d\zeta)\\ = \int_H \widehat{V}_d(t, x + \mathcal{P}_d\zeta)\, \nu_L(t_k, t_{k+1}; d\zeta) - \lambda \widehat{V}_d(t, x)\\ - \sum_{i=1}^{d} \partial_{x_i}\widehat{V}_d(t, x) \int_H \langle \zeta, p_i\rangle_H\, \nu_L(t_k, t_{k+1}; d\zeta)\end{aligned}$$

in the PIDE (4.6). To reduce the computational effort for the method of lines, the time steps Δt are chosen rather large. In order to nevertheless achieve sufficient accuracy for the approximation of the above integrals, additional substeps of each interval $[t_k, t_{k+1}]$ are introduced. Let $n_s \in \mathbb{N}$ be the number of substeps for each major time step. Then we have

$$\begin{aligned}&\int_H \widehat{V}_d(t, x + \mathcal{P}_d\zeta)\, \nu_L(t_k, t_{k+1}; d\zeta)\\ &\qquad \approx \frac{1}{n_s} \sum_{j=1}^{n_s} \int_H \widehat{V}_d(t, x + \mathcal{P}_d\zeta)\, \nu_L\big(t_k + j\frac{t_{k+1} - t_k}{n_s}; d\zeta\big)\\ &\qquad = \frac{\lambda}{n_s} \sum_{j=1}^{n_s} \int_E \widehat{V}_d\Big(t, x + \mathcal{P}_d\eta_{t_k + j(t_{k+1} - t_k)/n_s}(y)\Big) P^Y(dy)\end{aligned}$$

for $0 \le k < n_T$. Similarly, we obtain

$$\int_H \langle \zeta, p_i\rangle_H\, \nu_L(t_k, t_{k+1}; d\zeta) \approx \frac{\lambda}{n_s} \sum_{j=1}^{n_s} \int_E \big\langle \eta_{t_k + j(t_{k+1} - t_k)/n_s}(y), p_i\big\rangle_H\, P^Y(dy).$$

The integrals with respect to P^Y can then be approximated with quadrature formulas. This requires interpolation of the integrands at quadrature points, which are not necessarily identical to grid points. Since we have assumed that jumps occur only along a relatively small subspace of Ω_R, only a small number of grid points is involved, and the corresponding matrices remain sparse. The idea is illustrated in Figure 4 for a single jump process, where P^Y is defined on $E = \mathbb{R}$ and the jumps are given by $\eta(t)y \in H$.

Note that the same formulas can be applied for the time-inhomogeneous PIDE. In

this case, however, we would choose smaller major time steps and $n_s = 1$ (no substeps).

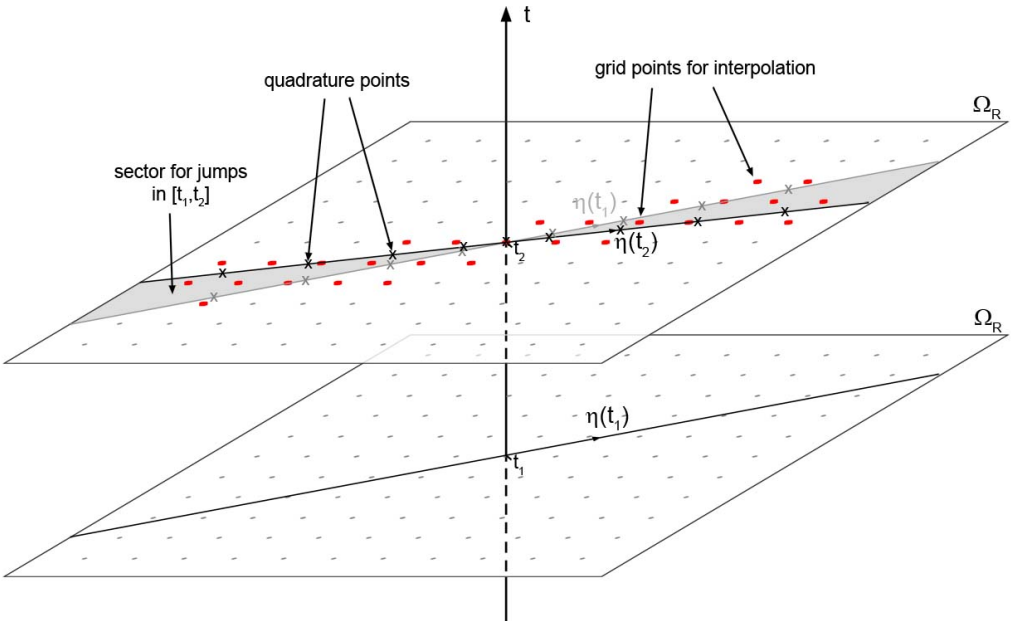


Figure 4: Grid points (on full grid) involved in the numerical quadrature of jump term integrals in the time interval $[t_1, t_2]$.

Sparse grids Besides the nonlocality of the integral term, the exponentially increasing computational complexity with increasing dimension d is the major numerical problem when solving the PIDE. This *curse of dimensionality* can be broken by using sparse grids. A comprehensive overview of this topic can be found in [20]. Figure 5 shows a sparse grid in two-dimensional space.

In particular, we make use of the combination technique [63]. Thus, we use a standard PIDE solver on a series of full, regular, but anisotropic grids. Instead of using all grids g_α with $\|\alpha\|_\infty \leq N_g \in \mathbb{N}$, only grids satisfying

$$N_g \leq \|\alpha\|_1 \leq N_g + d - 1 \tag{8.2}$$

for some $N_g \in \mathbb{N}$ are employed. We denote the approximation of $\widehat{V}_d$ on the grid g_α by $\widetilde{V}_d^\alpha$. An approximation $\widetilde{V}_d^{N_g}$, corresponding to a sparse grid solution, can then be obtained by linear combination as follows:

$$\widetilde{V}_d^{N_g}(x) := \sum_{k=1}^{d} (-1)^{k+1} \binom{d-1}{k-1} \sum_{\|\alpha\|_1 = N_g + d - k} \widetilde{V}_d^\alpha. \tag{8.3}$$

Since artificial zero boundary conditions are applied, grid points on the boundary $\partial\Omega_R$ can be ignored.

Due to the anisotropic truncation of the domain introduced in (8.1), an equal number of refinements in every coordinate would result in finer mesh widths for coordinates

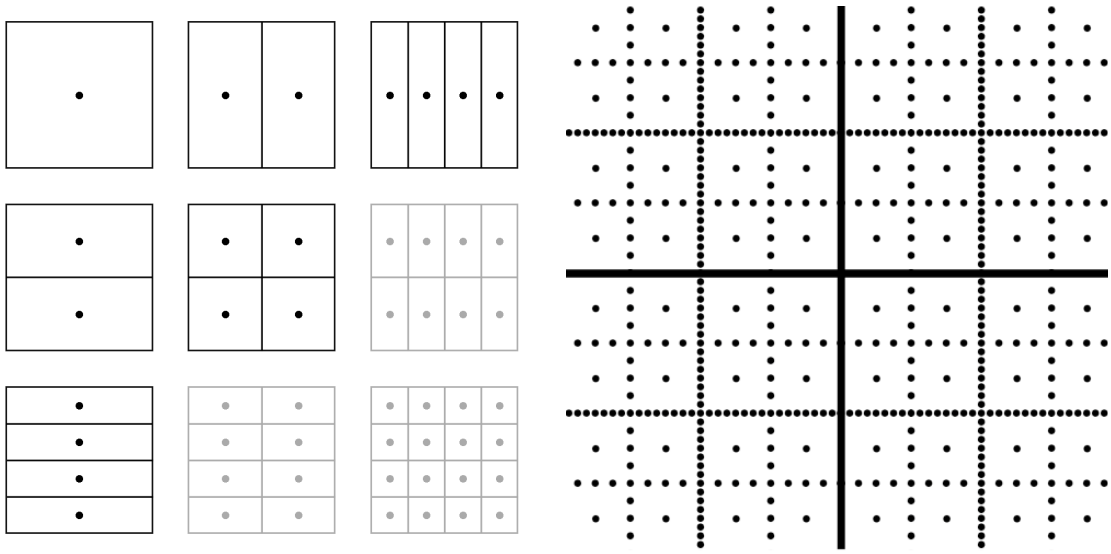


Figure 5: Construction of sparse tensor product (left) and sparse grid without boundary points (right) in $\mathbb{R}^2$.

which are in fact the least important ones. This mismatch can be remedied by introducing additional constraints on the multiindex of grids used. In order to achieve similar mesh widths, we demand

$$\alpha_i \leq N_g + \ln\left(\frac{R_i}{R_1}\right)/\ln(2), \quad i = 1, \ldots, d,$$

which is equivalent to $2^{\alpha_i} \leq 2^{N_g}\frac{R_i}{R_1}$. Since this yields a set of grids different from the one obtained by condition (8.2) alone, the corresponding coefficients in (8.3) have to be modified. For a detailed presentation of how to choose coefficients in anisotropic sparse grids, see [43].

8.2 Test Problems

The following two paragraphs give a precise description of the models used for the numerical experiments, one for stock basket options and one for electricity swaptions. They correspond to the motivating examples stated in Section 1. Note that both models can be interpreted as special cases of the general setting (2.1), (2.3). The effectiveness of the POD method depends on the correlation structure of the underlying assets. High correlation allows for a more efficient decrease of variance with fewer POD components. We thus investigate the performance of the dimension-reduction for different scenarios.

Stock basket options We examine baskets of n stocks. The initial value for the assets is $S_i(0) = 50$, $i = 1, \ldots, n$. We consider put options with strike $K = 50$. The driving stochastic process for our test problems is a jump-diffusion of the form (2.1) with time-constant volatility and jump distribution. Similar to [83], we include independent

jumps for each individual asset as well as common jumps for all assets. The common jumps are driven by a compound Poisson process with intensity λ_0 and are of fixed relative height η_0. The additional individual jumps of each asset price have intensity λ_i and relative height η_i, $i = 1, \ldots, n$. The price process of each asset $S_t(i)$ satisfies

$$\frac{dS_t(i)}{S_t(i)} = r dt + \sum_{j=1}^{n} \sigma_{ij} dW_j(t) + \eta_0 d[N_0(t) - \lambda_0 t] + \eta_i d[N_i(t) - \lambda_i t].$$

The risk-free interest rate is assumed to be constant at $r = 0.02$. The Brownian motions W_j, $j = 1, \ldots, n$, as well as the Poisson processes $N_0, N_1, \ldots, N_n$ are all independent. The entries $(\sigma_{ij})_{i,j=1}^n$ of the volatility matrix are chosen such that the covariance matrix $\mathcal{C}_D \in \mathbb{R}^{n \times n}$ of the diffusion part satisfies

$$\mathcal{C}_D(i,j) = \left(\sigma\sigma^T\right)_{ij} = 0.2\, e^{-\rho|i-j|}, \quad \text{for } i,j = 1, \ldots, n,$$

where $\rho \in \mathbb{R}$ is a parameter controlling the decay of correlation. The discounted value of every asset is a martingale under the pricing measure. The prices can be written as exponential jump-diffusion process as follows:

$$S_t(i) = S_0(i) \exp\left\{ \left(r - \frac{1}{2} \sum_{j=1}^{n} \sigma_{ij}^2 - \eta_0\lambda_0 - \eta_i\lambda_i \right) t + \sum_{j=1}^{n} \sigma_{ij} W_j(t) \right.$$
$$\left. + \ln\left(1+\eta_0\right) N_0(t) + \ln\left(1+\eta_i\right) N_i(t) \right\}, \quad i = 1, \ldots, n.$$

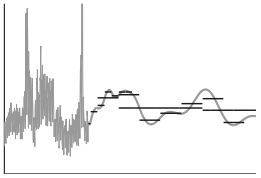
We set $\lambda_0 = \lambda_1 = \ldots = \lambda_n = 1$, $\eta_1 = \ldots = \eta_n = -0.05$. For the remaining correlation and jump parameters we use two different sets of values:

High correlation	$\rho = 0.1$	$\eta_0 = -0.15$,
Low correlation	$\rho = 0.4$	$\eta_0 = -0.10$.

Hence, we have a faster decaying correlation and less pronounced common jumps in the "low correlation" scenario.

Figure 6 shows the eigenvalues of the covariance matrix of Z_T in both scenarios. All values are divided by the largest eigenvalue μ_1 for normalization. The decay is exponential. A faster decay means higher correlation and, thus, usually better performance of the dimension reduction method. For comparison, the eigenvalues obtained from the empirical covariance matrix of the top 20 S&P 500 stocks are also plotted. The graph shows that these eigenvalues are between those obtained from the test problem with the two parameter sets described above. The explained variance in the low correlation scenario is mostly similar to that for the true stock data, but even smaller for the first few eigenvalues.

Electricity swaptions The second test problem is an an option on an electricity swap. We use the exponential additive model given in (2.3). The corresponding Hilbert space is $H = \mathrm{L}^2([T_1, T_2], \lambda_D)$, where $D = [T_1, T_2]$ is the delivery period of the swap and λ_D is the Lebesgue measure on D. For the diffusive part of the model, we use two



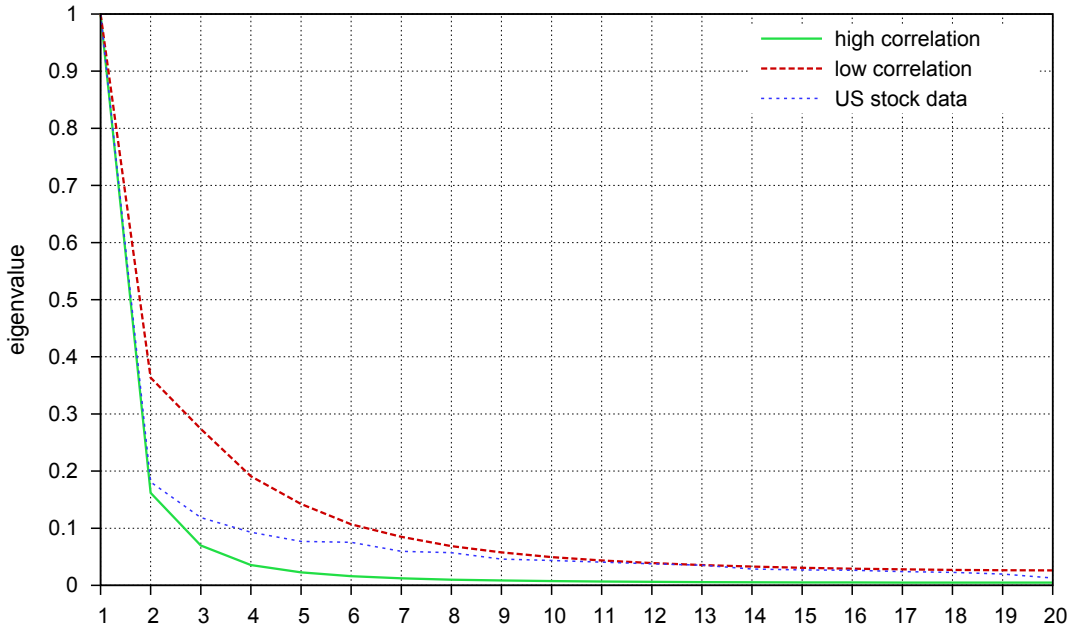


Figure 6: Eigenvalue decay for different correlation scenarios and S&P 500 stocks. The eigenvalues of the covariance matrices are ordered by size and normalized.

factors, similar to as in [51]. The driving jump-diffusion is

$$X_t(u) = \int_0^t \gamma_s(u)ds + \sum_{k=1}^{2} \int_0^t \sigma_k(s,u)\, dW_k(s) + \int_0^t \int_{\mathbb{R}} \eta_1(s,u) y\, \widetilde{M}_k(dy, ds).$$

The volatilities are given by

$$\sigma_1(t,u) \equiv 0.15, \quad \sigma_2(t,u) = 0.3\, e^{-1.4(u-t)}, \quad t \in [0,T], u \in D.$$

Moreover, we assume additional normally distributed jumps, which yield a *Merton model*. For the jumps, we use a compound Poisson process with intensity $\lambda_1 = 12$ and jump distribution $Y \sim \mathcal{N}(0.1, 0.1)$. The additional factor for dampening the jumps is

$$\eta_1(t,u) = 0.5 - 0.5\frac{u - T_1}{T_2 - T_1}, \quad t \in [0,T], u \in D.$$

Due to Proposition 2.6, the drift is given by

$$\gamma_t = -\frac{1}{2}\sum_{i=1}^{2} \sigma_i^2(t) - \lambda_1 \int_{\mathbb{R}} (e^{\eta_1(t)y} - 1 - \eta_1(t)y)\, P^Y(dy), \quad t \in [0,T].$$

We consider a swap with a delivery period of four weeks (28 days) maturing in one year, i.e., $T = T_1 = 1$, $T_2 = T_1 + 28/365$. The initial forward curve at time $t = 0$ is $S_0(u) \equiv 50$, $u \in D$, and the strike is $K = 50$. It remains to specify the discretization of the delivery

period. As we have a continuous forward curve model, we may use an arbitrary number of discretization points. On energy markets, monthly swaps on electricity are usually based on daily prices. Thus we will use exactly $n = 28$ components. We will further assume that there are 8 delivery hours per day. Setting $u_i = T_1 + (i-1)\frac{T_2-T_1}{28}$ for $i = 1, \ldots, 28$, we obtain

$$V(t) = e^{-rT} \cdot 8 \cdot 28 \cdot E\left[\left(\sum_{i=1}^{28} w(u_i; T_1, T_2) S_T(u_i) - K\right)^+ \middle| \mathcal{F}_t\right]$$

for the discounted price of a call swaption, without making any discretization error. The constant risk-free interest rate is $r = 0.02$.

8.3 Experiments for European Options

In this section, the results of the POD dimension reduction for European option pricing are presented. Since no analytical solutions are available, a large number of MC simulations were performed for each test problem to obtain a precise solution. All errors were computed using these MC reference values.

Results for basket option We consider a European put option on the average of a basket containing six stocks. For higher-dimensional baskets, see also the examples in Section 8.5 below. The discounted price of the basket option with maturity $T = 1.0$ at time $t \le T$ is

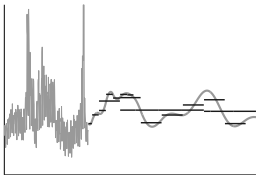
$$\widehat{V}(t) = e^{-rT} E\left[\left(\sum_{i=1}^{6} S_T(i) - K\right)^+ \middle| S_t\right].$$

We first examine the number of POD components needed to obtain a sufficiently good approximation. The eigenvalues μ_i of the covariance operator $\mathcal{C}_{X_T}$ defined in (2.9) are given in Table 1. The corresponding explained variability, defined by $\sum_{j=1}^{i} \mu_j \,/\, \sum_{j=1}^{6} \mu_j$, is also shown in the table. For the high correlation scenario, the eigenvalues decay faster and the explained variability is higher.

	low correlation		high correlation	
i	μ_i/μ_1	expl. var.	μ_i/μ_1	expl. var.
1	1.0000	0.6146	1.0000	0.8711
2	0.2805	0.7869	0.0734	0.9350
3	0.1403	0.8731	0.0286	0.9599
4	0.0859	0.9259	0.0182	0.9757
5	0.0649	0.9658	0.0146	0.9885
6	0.0557	1.0000	0.0132	1.0000

Table 1: Basket option – Eigenvalues and explained variability for both correlation scenarios.

The POD components are displayed in Figure 7. They resemble those known from fixed



income markets. In particular, the first three basis vectors represent the typical shift, tilt, and bend. Further components feature higher frequencies.

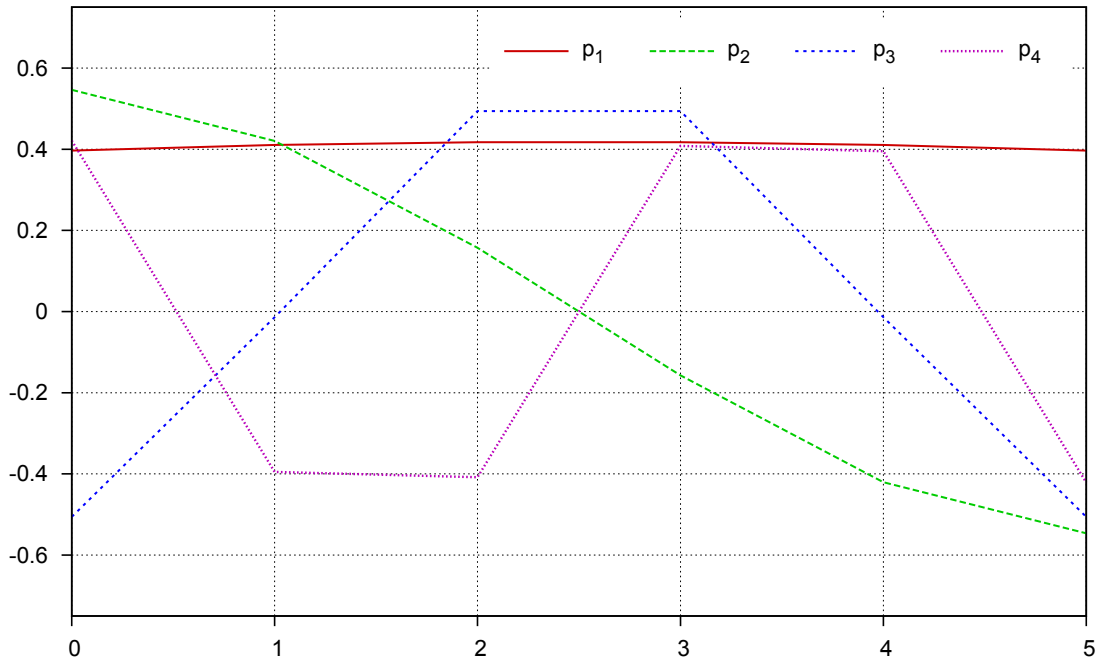


Figure 7: Basket option – The first four POD basis components.

We now examine the errors of the PIDE method. The "exact" reference solutions used here are the result from 10^7 MC paths with a standard deviation of 0.0018 and 0.0022 (estimated from 100 MC series) for low and high correlation, respectively. It turns out that, due to the stable discontinuous Galerkin method, the mesh size for the time variable has little influence on the PIDE results. Thus we do all computations with fixed, equidistant time steps of size $\Delta t = \frac{1}{10}$. The spatial grid, on the other hand, has an impact on the accuracy. Figure 8 displays the (signed) relative error for different dimensions d of the projected problem. The number N is the maximal number of discretization points in one coordinate and is always taken to be a power of 2. Two effects can be observed here. For each fixed value of N, the method converges to a certain limit when the dimension of the problem is increased. These limits, in turn, converge to the exact solution with increasingly fine meshes. Thus, we might accidentally get a very precise result for a low-dimensional computation on a coarse grid if the two errors (from dimension and mesh) happen to cancel each other. In practice, both the dimension and the number of grid points have to be chosen large enough in order to guarantee a precise result. With high correlation, $d = 3$ and $N = 2^8$ are sufficient to obtain relative errors below 1%. With low correlation, however, we have to use $d = 5$ POD components to achieve similar accuracy.

Finally, we have a look at the computational time needed for the PIDE method. To this end, we fix $N = 2^9$ and plot the error for various dimensions. Figure 9 displays the results; both y-axes (for time and error) have a logarithmic scale. The error decreases approximately exponentially with increasing dimension. The computational time, on

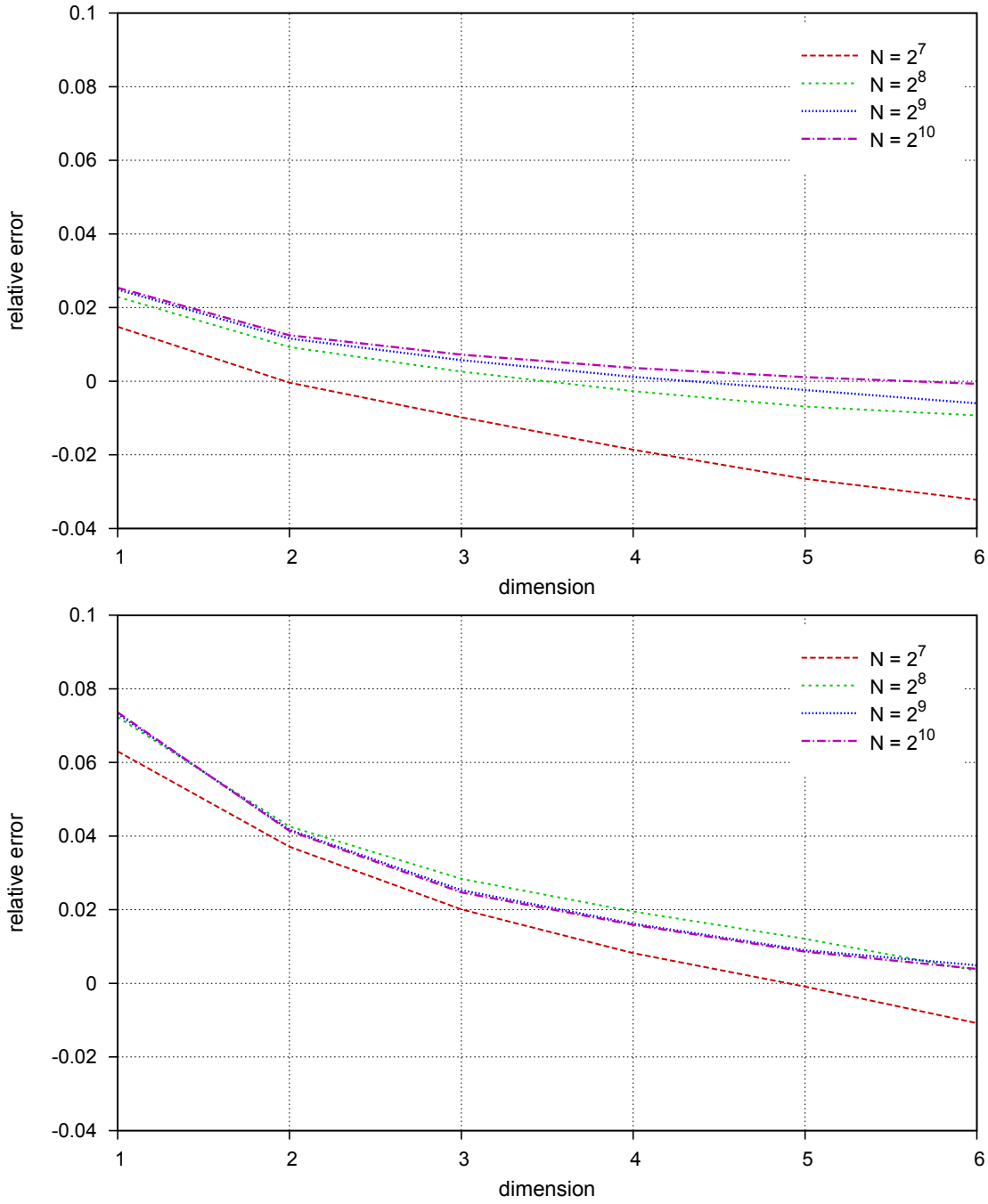


Figure 8: Basket option – Relative error of PIDE method with different meshes. Top: high correlation, bottom: low correlation.

the other hand, increases exponentially. Note that the solution of the problem without dimension reduction takes almost 3 minutes, even though we use the sparse grid combination technique. However, the increase of the projected problem dimension d by one increases the computational effort by only a factor of 4 to 5, despite the fact that we use up to $2^9 = 512$ grid points in each coordinate.

A reasonably precise solution (in practice within the bounds of the model error) takes a few seconds if the correlation is high and approximately 100 seconds if the correlation is low. While this is slightly faster than the MC method, it is not an extreme gain. The use of POD for variance reduction as described in Section 7 is therefore probably the more appealing way of applying the dimension reduction. The computational effort might be greatly reduced by using more sophisticated grids, featuring more grid points around the origin and fewer close to the boundary of the domain, which we will not consider here. Of course, the achieved accuracy depends on the correlation within the basket.

Results for electricity swaption In the test case for electricity swaptions, two POD basis components are already sufficient to explain almost 100% of the volatility. The sum of the remaining eigenvalues satisfies $\sum_{i\geq 3} \mu_i \approx 0$. This is of course due to the strong correlation between the price changes for different maturities $u \in [T_1, T_2]$, which makes the dimension reduction technique a particularly well suited method for this type of derivative. The forward curve defined on the delivery interval does not change its shape arbitrarily. The two POD components accurately describe the possible shape changes in our (rather simple) test setting. We are thus able to compute accurate prices by solving a two-dimensional PIDE.

Figure 10 displays the relative error and time of the PIDE method for different mesh widths. As was to be expected, the computational effort increases exponentially in the number of grid refinements (and thus linear in the total number of grid points). The relative error decreases exponentially. However, in contrast to the basket option, a very accurate solution can be computed for the electricity swaption within a fraction of a second. This is made possible by the very efficient dimension reduction from $n = 28$ to $d = 2$.

A comparison of the PIDE method and MC simulation is displayed in Figure 11. In addition, the result of a log-normal approximation is also included. This so called Lévy approximation yields a relative error of 4.0%, which is by far the largest of all the methods. MC simulation yields very good results for 10^6 paths and above. However, with $n = 28$ it takes considerably longer than in the case of the six-dimensional basket. The standard deviation after 10^6 simulations is 0.73 (estimated from 100 MC series). The PIDE solver is indeed the fastest and most accurate method for this second test problem.

8.4 Experiments for Hedging

In this section, we use the method developed in Section 5 to find hedging strategies for the monthly electricity swaption which we have just priced. The swaption can be hedged with any set of traded swaps. We start with a single swap with a delivery period covering the whole month (28 days). Then we add further swaps whose delivery periods

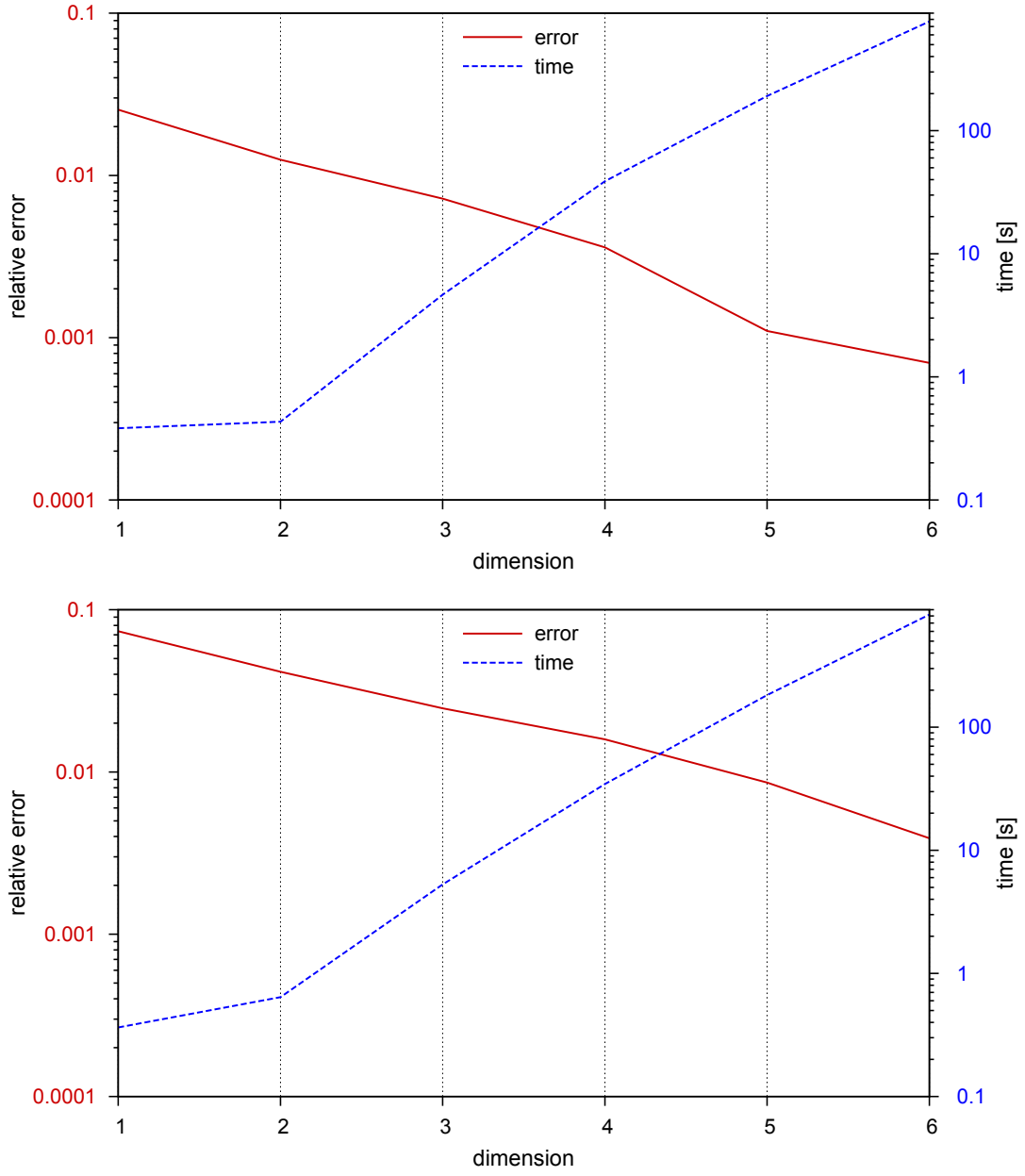


Figure 9: Basket option – Relative error and computational time of PIDE method with at most $N = 2^9$ grid points per coordinate. Top: high correlation, bottom: low correlation.

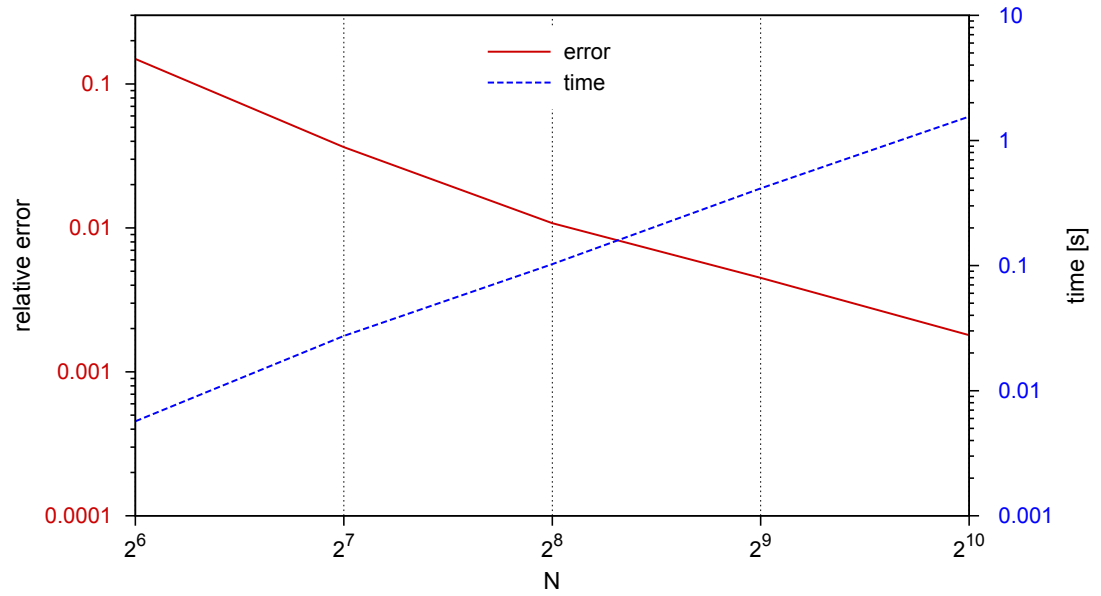


Figure 10: Swaption – Relative error and computational time of PIDE method for dimension $d = 2$.

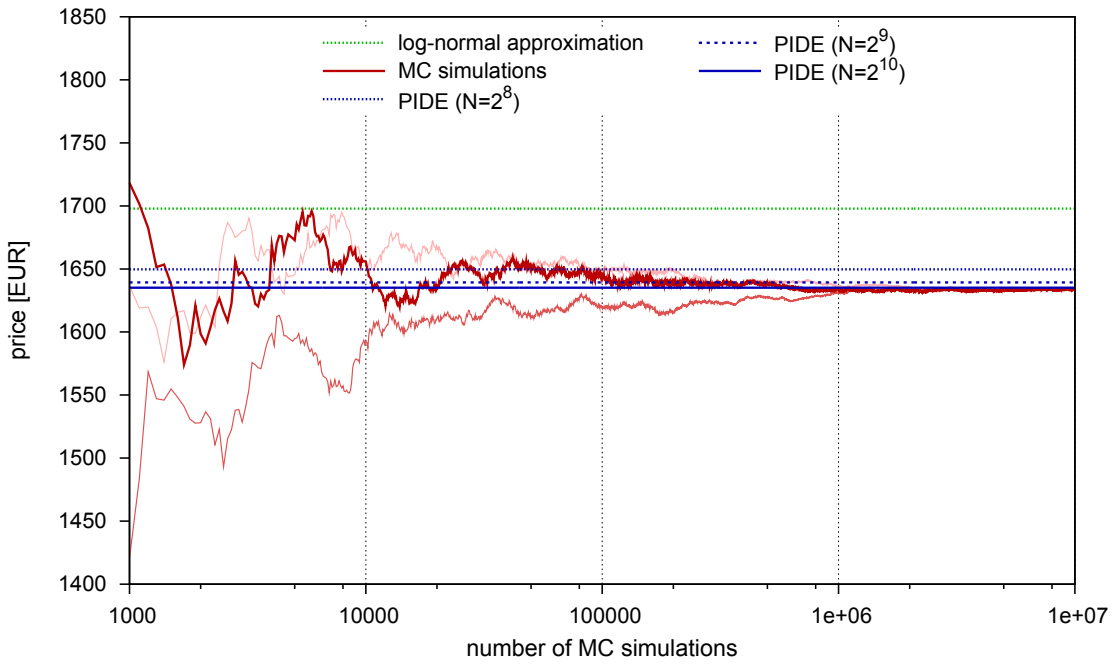


Figure 11: Swaption – PIDE solutions ($d = 2$), MC simulations, and log-normal approximation.

equal the weeks of the month. To simplify the notation, the corresponding strategies are denoted by capital letters as shown in Table 2. Moreover, we add the index "n-n" to the letter, if we consider only nonnegative strategies. Thus, e.g., $B_{\text{n-n}}$ denotes the nonnegative strategy using the monthly and the first weekly swap.

	delivery periods of included swaps				
	month	week 1	week 2	week 3	week 4
A	x				
B	x	x			
C	x	x	x		
D	x	x	x	x	
E		x	x	x	x
F	x	x	x	x	x

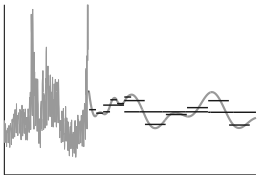
Table 2: Hedging portfolios used for numerical experiments.

For each set of swaps, the average hedging error is computed by MC simulation. To this end, 10000 stochastic paths for the forward curve are simulated. Along each path, the option prices and hedging strategies are calculated by solving the corresponding PIDE. The values of the portfolios (defined as the initial option value plus the increments given by the dynamics (5.8)) are then compared to the true terminal option values. The average of their differences is a good approximation for the unhedgeable component of the option price.

The Tikhonov regularization parameter described in Section 5.3 is set to $\delta = 10^{-10}$. The results are shown in Table 3. These errors equal the unhedgeable part of the risk, which is caused by two effects: the jumps in the model (which render even a one-dimensional market incomplete), and the fact that we use only a finite number of assets to hedge the infinite-dimensional forward curve. As was to be expected, the error decreases substantially if we add more swaps to the portfolio. A set of 4 swaps, however, is sufficient to obtain a very small error. This is of course due to the rather low number of driving factors in the test model. We also compare the terminal error with the initial option value for our test problem. (A relative error comparing with the option value at time T is ill-defined, since the option might be out of the money, yielding a division by zero.) Apparently, the cost for the additional nonnegativity constraint on the portfolio is high for portfolios which contain several swaps.

The next experiment is designed to study the influence of the regularization parameter. Table 4 shows the average hedging error with portfolio E for different values of δ. The error decreases monotonously with decreasing δ, until the regularizing effect is no longer sufficient and the error increases again at $\delta = 10^{-15}$ (which is close to machine precision). The error when using the projection method described in Section 5.3 is 18.25, which is about the same size as the best result achievable with Tikhonov regularization. We will continue to use regularization, however, since it is the more general concept, which also works for very ill-conditioned matrices M. We set $\delta = 10^{-10}$ for the remaining experiments.

Next, we examine the hedging strategies in more detail. In order to illustrate some of their features, we pick two sample paths from the MC simulation: one where the



portfolio	A	$A_{\text{n-n}}$	B	$B_{\text{n-n}}$	C	$C_{\text{n-n}}$
absolute error	215.81	217.58	173.01	178.75	30.90	175.74
error / initial value	0.132	0.133	0.106	0.109	0.019	0.108

portfolio	D	$D_{\text{n-n}}$	E	$E_{\text{n-n}}$	F	$F_{\text{n-n}}$
absolute error	18.86	176.70	18.77	160.47	18.87	161.88
error / initial value	0.012	0.108	0.011	0.098	0.012	0.099

Table 3: Average hedging errors for different portfolios. The bottom rows show the ratio of the terminal hedging error at time T and the option value at time 0 (the value of the latter is 1633.64).

δ	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}	10^{-11}	10^{-12}	10^{-13}	10^{-14}	10^{-15}
error	30.24	28.14	21.36	18.77	18.78	18.71	18.56	18.47	18.38	18.78

Table 4: Average hedging errors for different regularization parameters in portfolio E.

option is in the money (ITM) at maturity and one where it ends up out of the money (OTM). We first analyze the ITM case. The hedging strategy for portfolio A is shown in Figure 12, together with the swap rate of the underlying monthly contract. This case is similar to that of hedging a single stock in that the quantity of swaps held in the portfolio is a number in $[0, 1]$. The investment jumps simultaneously with the swap rate and in the same directions. There is, e.g., a jump of the swap rate at day 243. Since the option changes from OTM to ITM with this jump, this has a large impact on the optimal investment. The strategy approaches 1 when time goes to T, since the option ends in the money.

Figure 13 displays the strategy for portfolio E. The optimal investments in the weekly swaps are now real numbers, sometimes negative and of large absolute value. When approaching maturity, however, the strategy satisfies $|\theta_i| < 1$, $i = 1, \dots, 5$.

The chronological development of the hedging errors corresponding to different portfolios in the ITM case is shown in Figure 14. When adding more swaps to the portfolio, two effects can be seen. The error gets smaller, and it evolves more smoothly. Both the diffusion and the jump part of the error decrease for more diverse portfolios. While the value of the hedge portfolio is visibly different from the true option value for portfolio A, these two curves are almost indistinguishable for portfolio E. One can also observe the Samuelson effect of increasing volatility in the option price closer to maturity. The volatility of the hedging error, however, is very small during the last month before maturity. This is due to the fact that after a significant jump at day 318 the swap rate is far above the strike. The option will thus very likely end in the money. The hedging investment is then almost equal to 1 and no further hedging error is accumulated.

The strategies in the OTM case approach 0 close to maturity. This holds for the investment in the monthly swap as well as any weekly swaps. Once again, the correlation between the hedging strategy θ and swap rate is clearly visible, as shown in Figures 15 and 16. The terminal value of portfolio A is negative and thus slightly below the targeted value 0 (Figure 17). The behavior of the hedging errors is very similar to the ITM case. In particular, the error gets smaller and smoother when more swaps are included and is almost constant 0 for portfolio E.

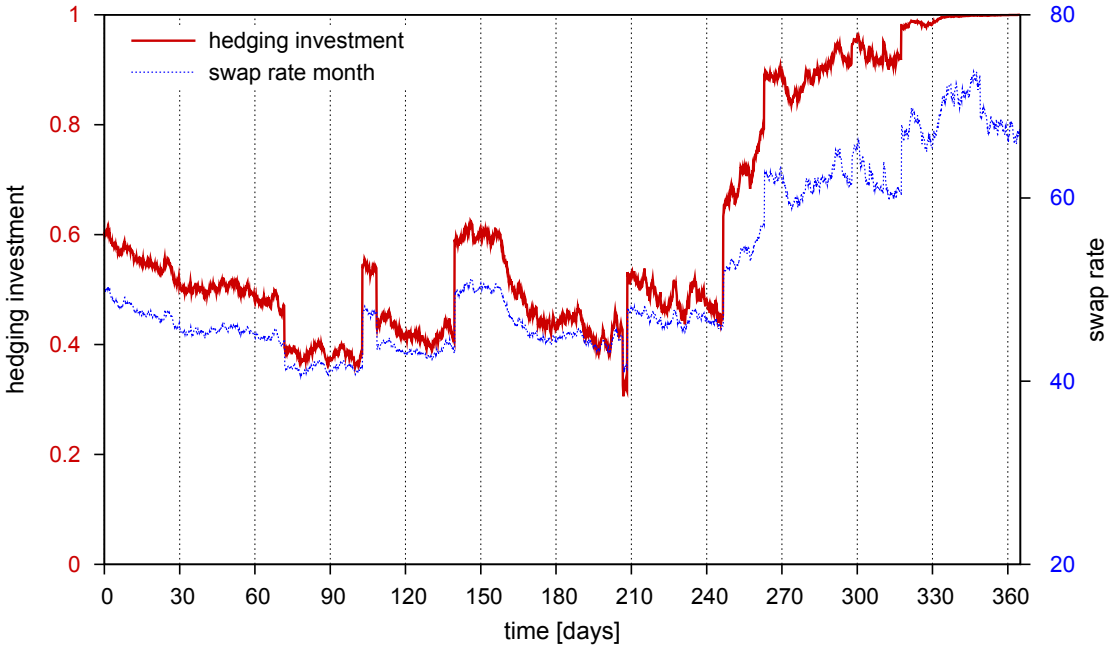


Figure 12: Swap rate and hedging strategy for portfolio A when the option ends ITM.

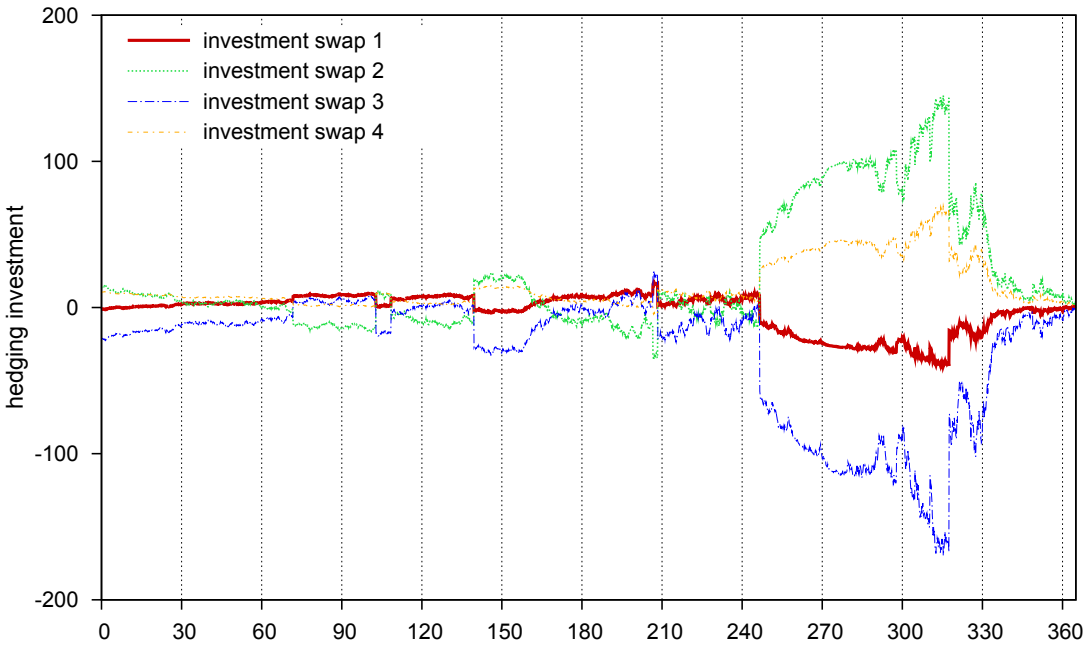
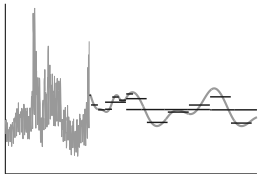


Figure 13: Hedging strategy for portfolio E when the option ends ITM.



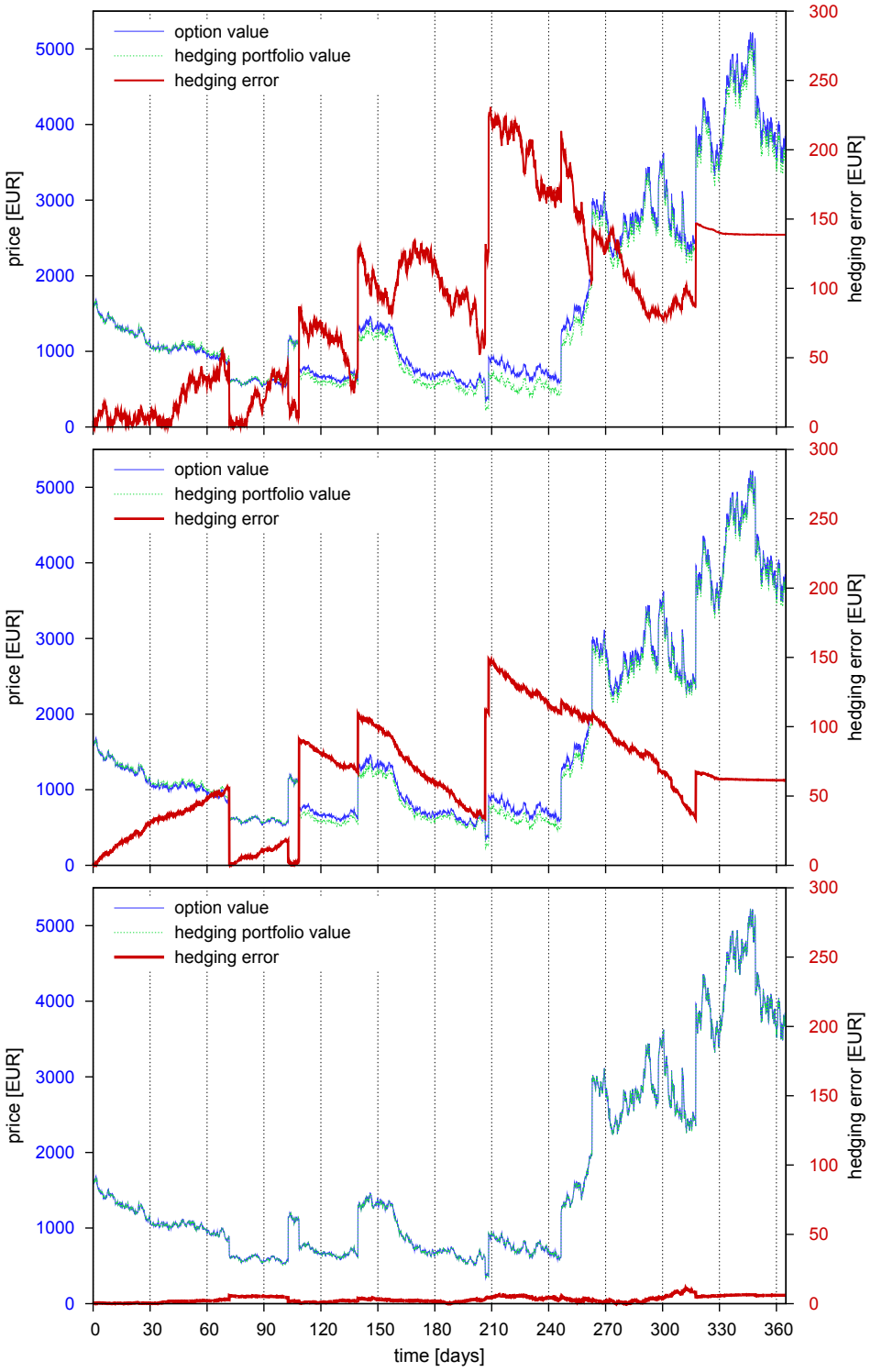


Figure 14: Option value, hedging portfolio value, and absolute hedging error for portfolios A, B, and E (top to bottom) when the option ends ITM.

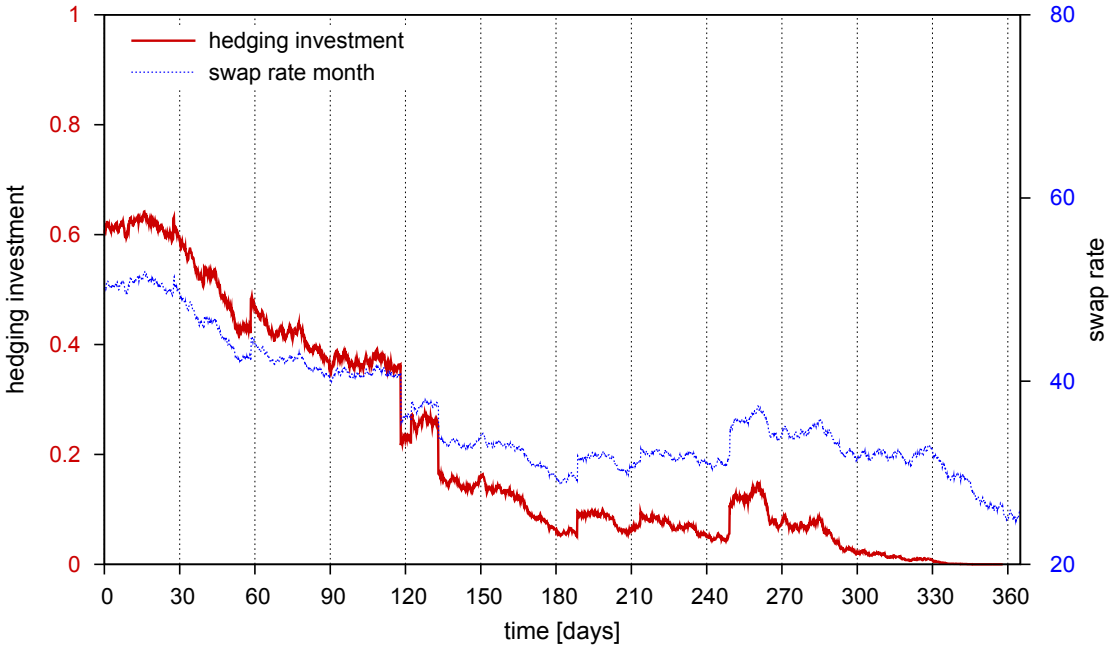


Figure 15: Swap rate and hedging strategy for portfolio A when the option ends OTM.

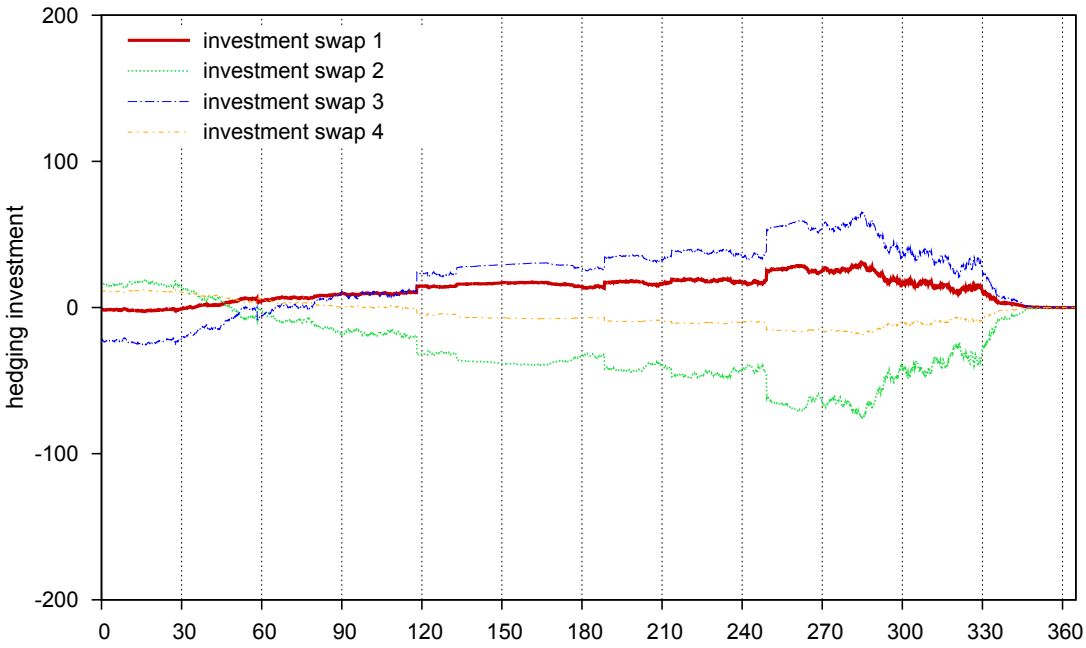
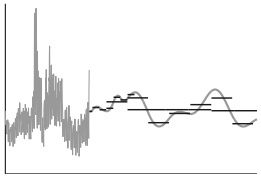


Figure 16: Hedging strategy for portfolio E when the option ends OTM.



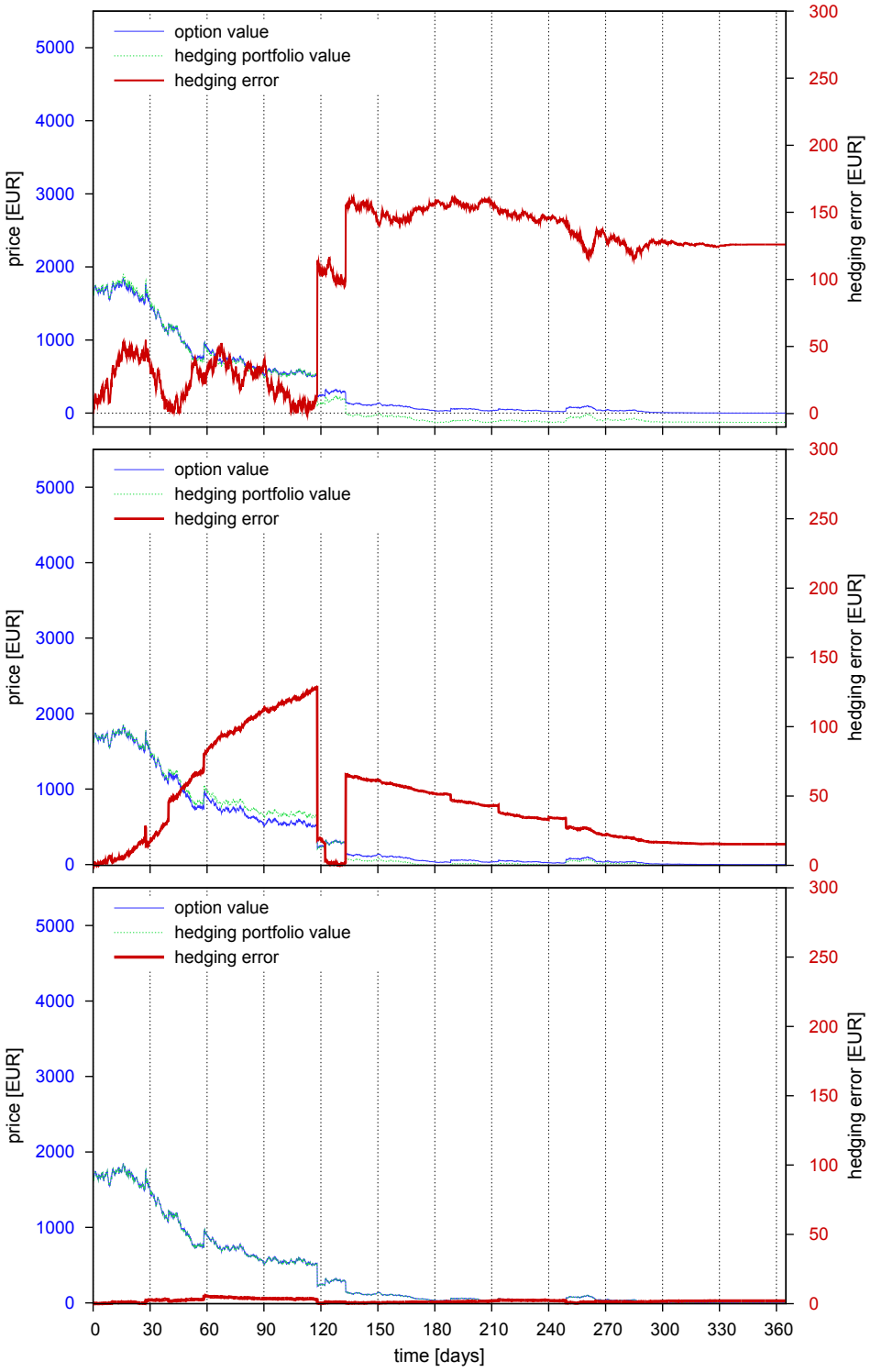


Figure 17: Option value, hedging portfolio value, and absolute hedging error for portfolio A, B, and E (top to bottom) when the option ends OTM.

8.5 Experiments for Bermudan Options

In this section, we apply the variance reduced and dual MC methods to stock basket options. Using different types of options, basket sizes, correlation parameters, and numbers of exercise dates, we show that the overall computational time can as a rule be reduced by at least 50%, often even by more than 80%. We consider baskets of $n = 10$, 20, and 30 assets. The admissible exercise dates are assumed to be equally spaced: $t_i = \frac{i}{N_{ex}}$, $i = 1, \ldots, N_{ex}$. Their total number is either $N_{ex} = 10$ or $N_{ex} = 100$. Of course, the variance reduced MC method can also be applied to high-dimensional European options, where the POD dimension which is necessary for sufficient accuracy is just too high. The variance reduced MC simulation can be interpreted as a means to further improve any low-dimensional PIDE result.

We compute the price of put options with strike $K = 50$ and value

$$\widehat{V}(0,0) = \sup_{\tau \in \mathcal{T}(0,T)} E\big[e^{-r\tau}\left(K - P\big(\tau, S(Z_\tau)\big)\right)^+\big],$$

where P is either the average

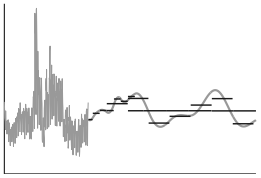
$$P_{avg}(t,S) = \frac{1}{n}\sum_{i=1}^{n} S(i)$$

or the minimum

$$P_{min}(t,S) = \min_{i=1,\ldots,n} S(i)$$

for $t \in [0,T]$. For the computation of $\widehat{V}_d$ both PIDE and FFT methods have been tested. They yield very similar results. Without using sparse grids, the PIDE solver does not scale any better than the Fourier method for higher dimensions. Since the FFT showed slightly superior accuracy on identical grids in our test cases, all of the results below refer to the FFT method. The grid refinement and domain truncation are chosen in such a way that the absolute error of $V_d(0,0)$ is well below 0.005 (half a cent). Usually, 2^6 grid points in each coordinate are sufficient to achieve this.

Computational Time Before we analyze the gain in precision obtained with variance reduction and dual MC, we examine the additional computational cost per path which is needed for these methods. If we fix the number N of MC paths, doing a plain least-squares MC is obviously less time consuming than computing additional control variables. This extra work is only worth the effort if we can reduce the number of paths significantly, so that the total computing time needed to achieve a given precision decreases. On the other hand, the dual method may take less time per path, since no regression is needed and the pathwise computation of maxima is inexpensive. Figure 18 gives an overview of the computational times needed for different parts of the algorithm. The simulation of paths is identical for all tested methods. The regression step for the control variables takes slightly less time than the regression for the original MC data, since it uses less basis variables. The set of basis variables for the full MC contains $n+3$ values $(Z(1), \ldots, Z(n), P, P^2, P^3)$, while the set for the control variable has only $d+3$ elements $((\mathcal{P}_d Z)(1), \ldots, (\mathcal{P}_d Z)(d), P, P^2, P^3)$. The cost of the FFT increases exponen-



tially in the dimension d. Using full grids with dimensions $d > 3$ implies computational times which are larger than the original MC method, even after taking the reduced number of paths into account.

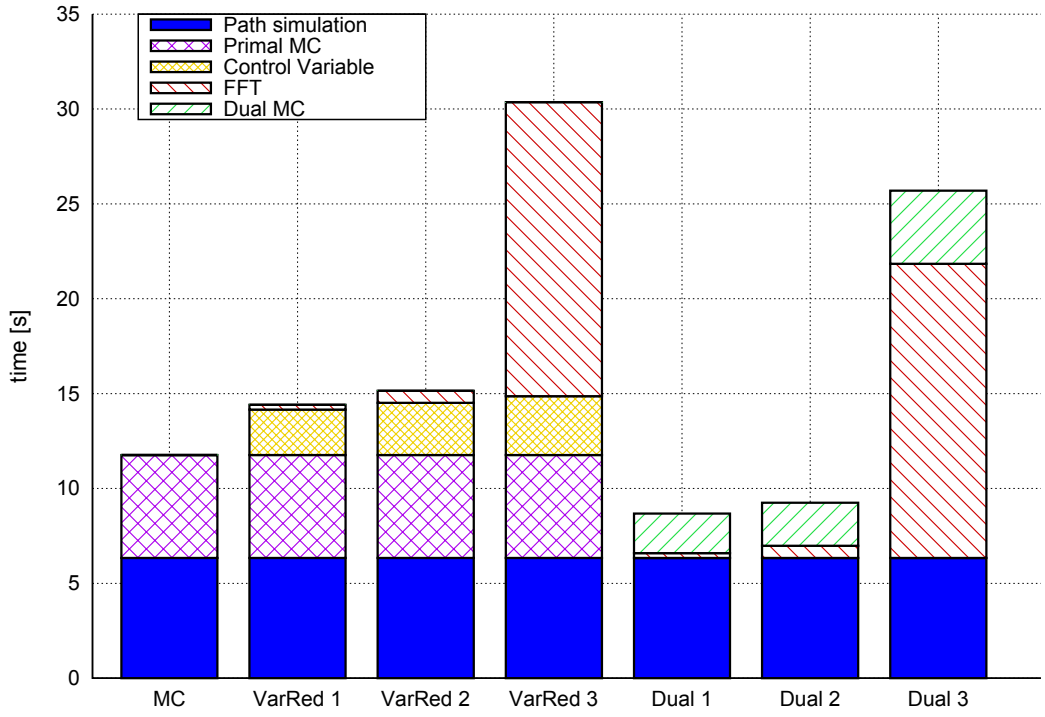


Figure 18: Computational times using variance reduction (VarRed) or dual MC (Dual) for $N = 50000$ paths. (Average put option, low correlation, $n = 30$, $N_{ex} = 100$.) The number behind the method indicates the projected dimension d.

Reduced Variance Since we have now seen how much additional time the variance reduction needs, the question is how much accuracy we gain from it. A measure for the accuracy is the variance of the MC prices. It can be used to obtain bounds on the precision, e.g., by Chebyshev's inequality. If the number of paths is sufficiently large and the computed exercise policies do not change substantially, when further paths are added, the variance is inversely proportional to the number of paths. Consequently, half the variance means that roughly half the number of simulated paths is sufficient for identical precision.

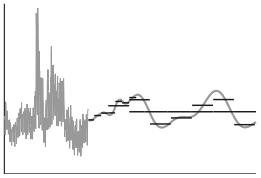
Figures 19 and 20 show results of the variance reduction and the dual MC for different settings. In general, the methods work better for the average option than for the minimum option. This is not surprising, because the average is captured better by the POD components than the minimum. The variance can of course be decreased further if the correlation is high. The number of exercise dates N_{ex} also has a small influence. More exercise points yield less effective variance reduction. The dual method converges extremely fast in every setting. Its accuracy, however, is only satisfying for highly correlated baskets and the average put option. In any other case, a higher value for d ($d > 3$)

and a finer grid for the FFT solution are needed to obtain a reasonable approximation with the dual method. This is usually not worth the computational effort. The variance reduction, on the other hand, always works, although its effect is hardly visible in the plot for the "worst case" of a minimum put option on lowly correlated assets.

Choice of Dimension The effectiveness of the variance reduction of course also depends on the dimension d of the projected problem. For the average option, it turns out that $d = 1$ already yields a substantial improvement over the plain least-squares MC method without variance reduction. Increasing the dimension gives only slightly smaller variances. Nevertheless, $d = 2$ is worth considering, since the additional computational effort is small. For the minimum put option, increasing d has a much larger impact. Adding further POD components decreases the variance significantly. Figure 21 illustrates this effect. In terms of overall computational effort, $d = 2$ turns out to be a good choice, although $d = 3$ is sometimes even better (depending on the number of assets and the efficiency of the method used to compute $V_d(0, 0)$).

In order to get a good estimate for the variance of the MC price, a large number of MC experiments (usually at least 1000), each with the given number of paths N, is necessary. A good approximation can be obtained by computing the variance *within* each set of N paths and dividing by N. In fact, this approximation is accurate if the exercise policy does not vary for different sets of simulated paths. In order to rule out effects due to a possible change of exercise policy between different sets of paths, we use the mean of this value over 100 MC experiments (with N paths each).

Variance and Time Ratios Tables 5 and 6 summarize a large number of computational results for average and minimum options, respectively. Each cell of these tables contains two numbers. The first one is the ratio of the variance σ^2_{vr} after variance reduction to the variance σ^2_{MC} of the plain least-squares MC method. The second one is the ratio of total computing time t_{vr} with variance reduction to computing time t_{MC} using plain MC for $N = 100000$ paths. While computational time increases with dimension d, the variance ratio decreases. The product of the two quantities gives a very rough estimate of the total computing time ratio, because the number of paths needed for a certain accuracy of the result decreases proportionally to the variance. Taking, e.g., the entry for $N_{ex} = 10$, $n = 10$, and $d = 2$ in Table 5, we obtain $0.04 \cdot 1.26 = 0.05$, which means that we can save about 95% of computing time with variance reduction of dimension 2. The entry for $N_{ex} = 100$, $n = 30$, and $d = 2$ in Table 6, on the other hand, yields $0.49 \cdot 1.19 = 0.58$, corresponding to a 42% lower computational cost due to variance reduction.



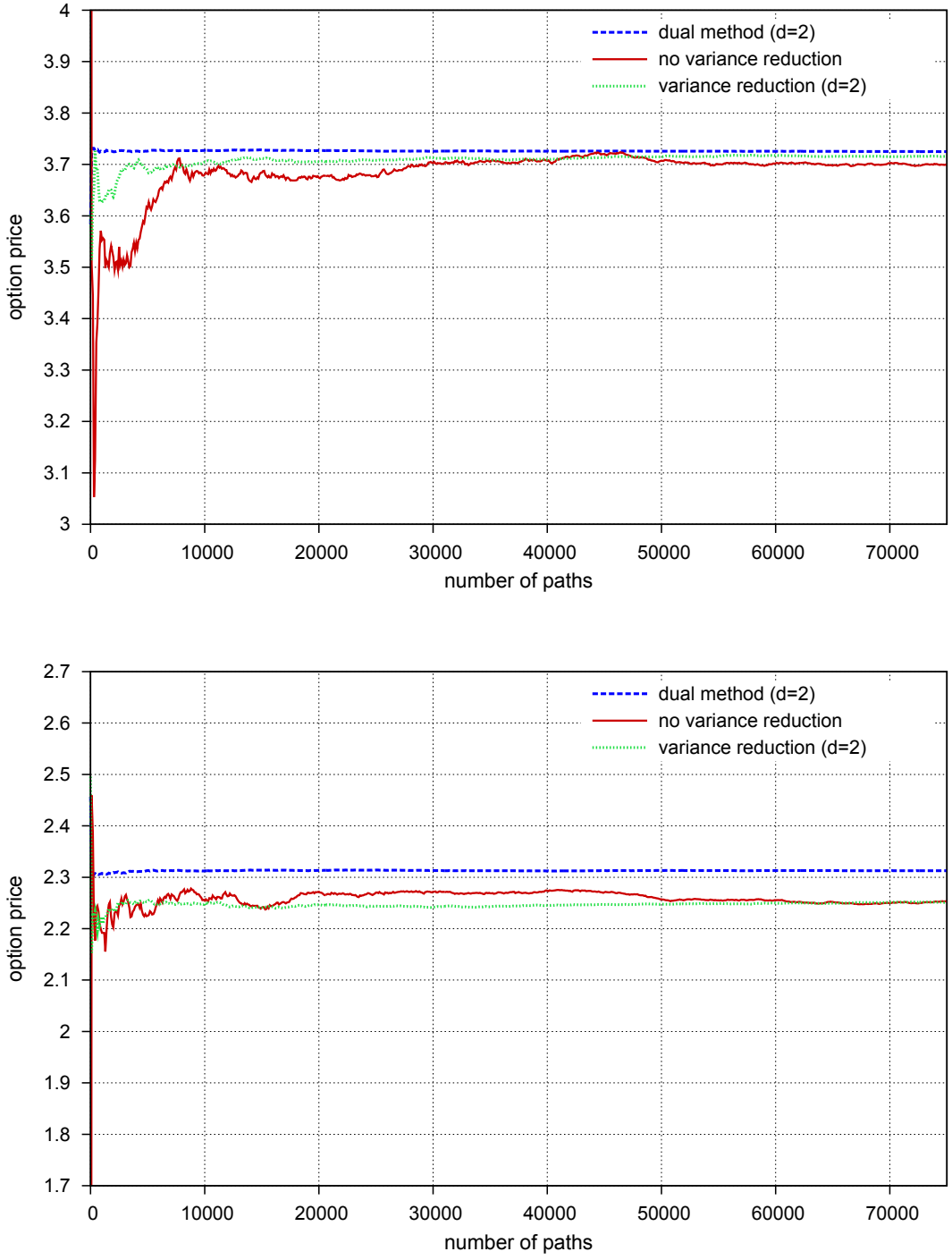


Figure 19: MC, variance reduced MC, and dual MC for *average put option* ($n = 30$, $d = 2$). Top: high correlation, $N_{ex} = 10$; bottom: low correlation, $N_{ex} = 100$.

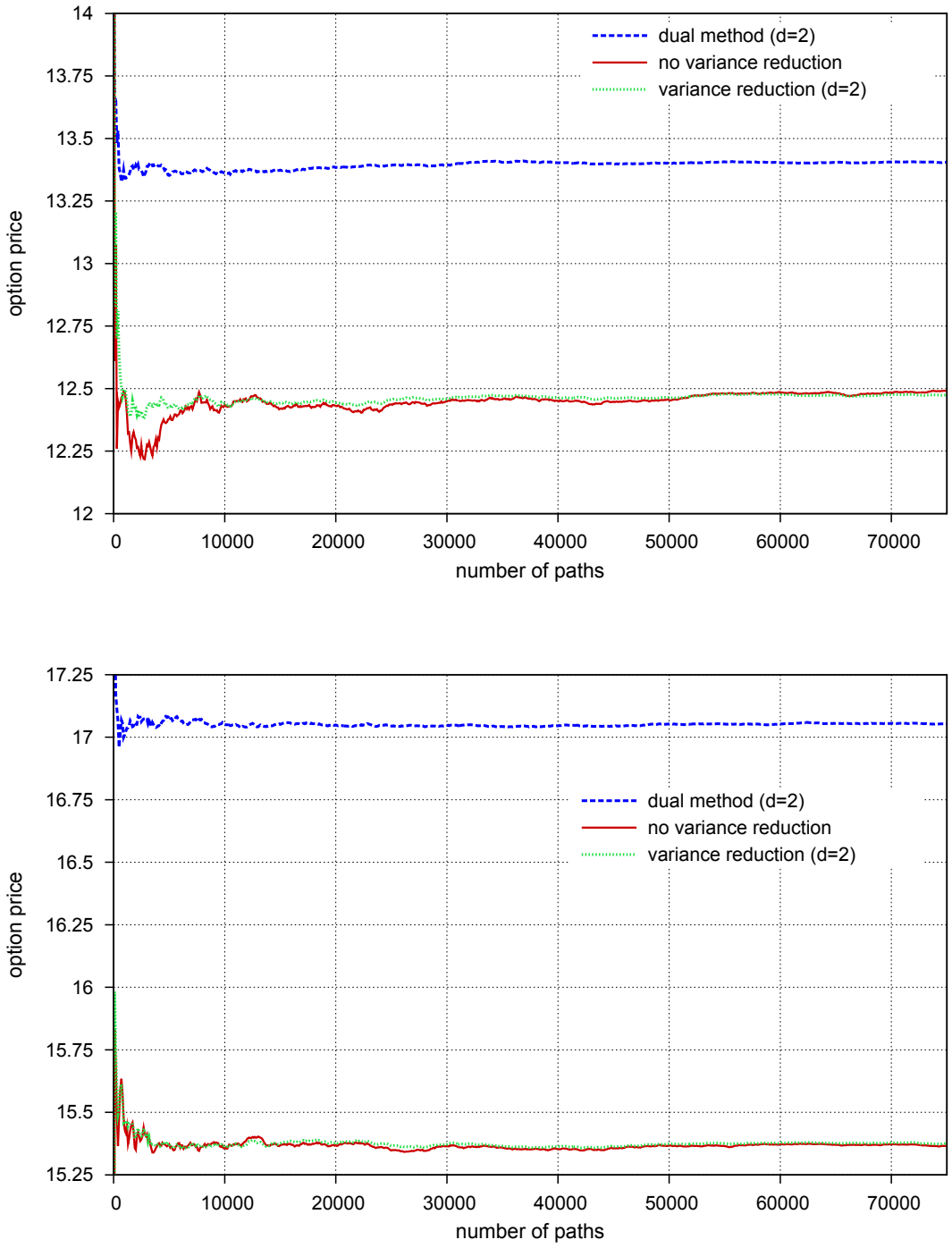


Figure 20: MC, variance reduced MC, and dual MC for *minimum put option* ($n = 30$, $d = 2$). Top: high correlation, $N_{ex} = 10$; bottom: low correlation, $N_{ex} = 100$.

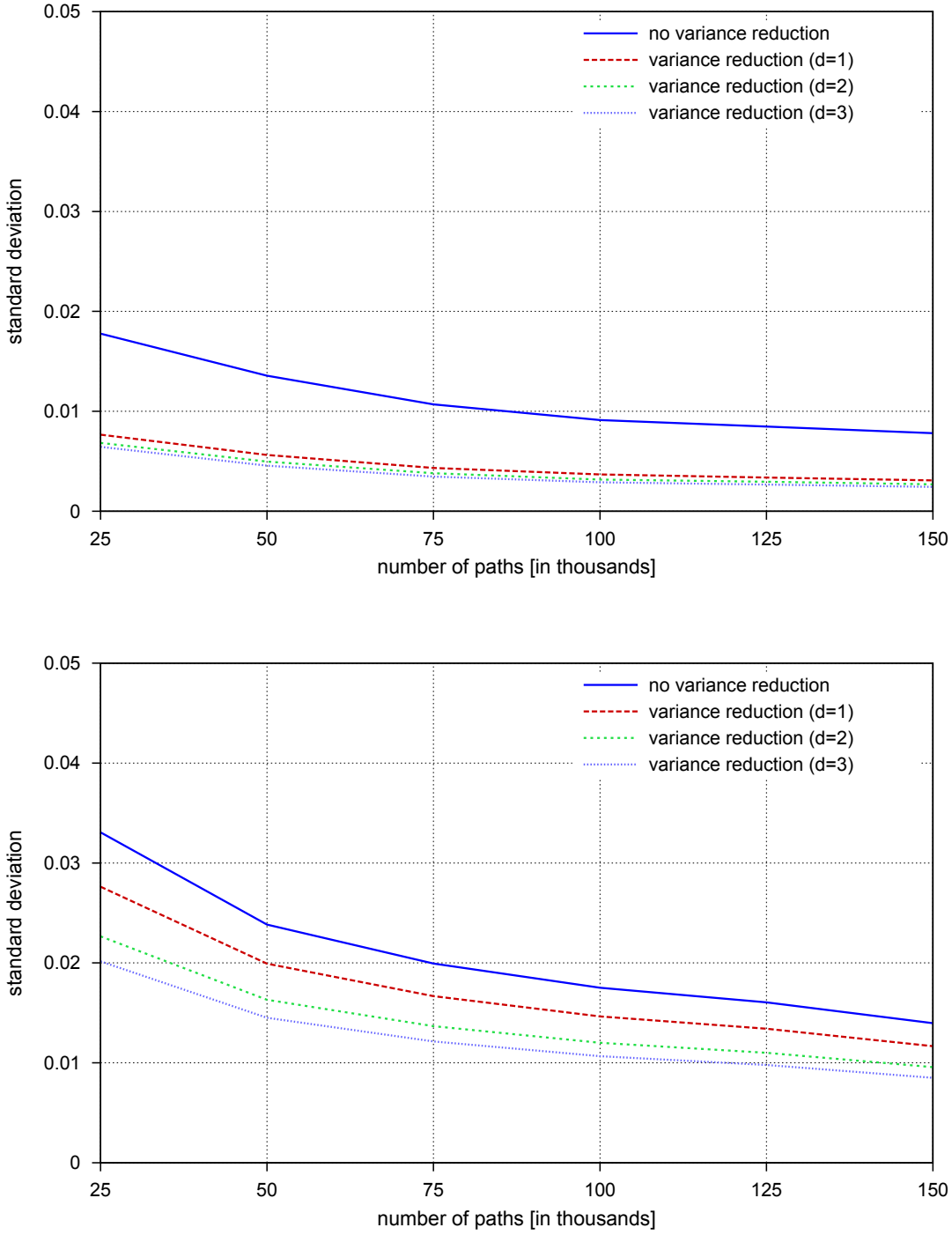


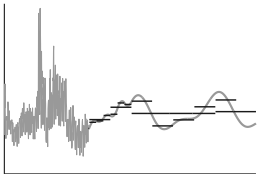
Figure 21: Standard deviation with and without variance reduction for $d = 1, 2, 3$ (low correlation, $n = 30$, $n_{ex} = 10$). Top: average put option; bottom: minimum put option.

		$N_{ex} = 10$			$N_{ex} = 100$		
	# assets	$d = 1$	$d = 2$	$d = 3$	$d = 1$	$d = 2$	$d = 3$
high corr	10	0.04|1.26	0.03|1.40	0.03|1.62	0.05|1.22	0.04|1.29	0.04|1.52
	20	0.06|1.24	0.05|1.31	0.03|1.63	0.08|1.21	0.06|1.30	0.05|1.65
	30	0.07|1.19	0.06|1.26	0.05|1.83	0.10|1.19	0.07|1.25	0.07|1.58
low corr	10	0.10|1.26	0.06|1.34	0.04|2.44	0.12|1.22	0.07|1.30	0.05|2.19
	20	0.14|1.22	0.10|1.29	0.08|2.40	0.15|1.21	0.10|1.28	0.08|2.03
	30	0.16|1.19	0.12|1.25	0.10|2.43	0.18|1.21	0.13|1.24	0.11|1.92

Table 5: Variance ratio $\frac{\sigma^2_{vr}}{\sigma^2_{MC}}$ and time ratio $\frac{t_{vr}}{t_{MC}}$ for *average put option* ($N = 100000$ paths).

		$N_{ex} = 10$			$N_{ex} = 100$		
	# assets	$d = 1$	$d = 2$	$d = 3$	$d = 1$	$d = 2$	$d = 3$
high corr	10	0.46|1.22	0.20|1.30	0.13|1.60	0.48|1.16	0.23|1.27	0.15|1.49
	20	0.52|1.19	0.26|1.26	0.14|1.57	0.54|1.17	0.29|1.24	0.20|1.55
	30	0.55|1.15	0.29|1.21	0.19|1.66	0.68|1.14	0.43|1.21	0.32|1.44
low corr	10	0.64|1.20	0.36|1.30	0.24|2.23	0.66|1.18	0.39|1.27	0.26|1.97
	20	0.69|1.17	0.44|1.25	0.32|2.09	0.70|1.16	0.47|1.24	0.35|1.81
	30	0.70|1.14	0.47|1.20	0.37|2.07	0.71|1.13	0.49|1.19	0.39|1.67

Table 6: Variance ratio $\frac{\sigma^2_{vr}}{\sigma^2_{MC}}$ and time ratio $\frac{t_{vr}}{t_{MC}}$ for *minimum put option* ($N = 100000$ paths).



References

[1] CGAL, Computational Geometry Algorithms Library. http://www.cgal.org.

[2] S. Asmussen and P. W. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer, New York, 2007.

[3] N. Audet, P. Heiskanen, J. Keppo, and I. Vehviläinen. Modeling electricity forward curve dynamics in the Nordic market. In D. W. Bunn, editor, *Modelling Prices in Competitive Electricity Markets*, chapter 12. Wiley Series in Financial Economics, Chichester, 2004.

[4] W. Bangerth, R. Hartmann, and G. Kanschat. deal.II – a general purpose object oriented finite element library. *ACM Trans. Math. Softw.*, 33(4):24/1–24/27, 2007.

[5] G. Barles. Convergence of numerical schemes for degenerate parabolic equations arising in finance theory. In *Numerical Methods in Finance*, Publ. Newton Inst., pages 1–21. Cambridge Univ. Press, Cambridge, 1997.

[6] D. Belomestny. Pricing Bermudan options using regression: optimal rates of convergence for lower estimates. *To appear in Finance and Stochastics*, 2010.

[7] D. Belomestny, A. Kolodko, and J. Schoenmakers. Regression methods for stochastic control problems and their convergence analysis. *SIAM J. Control Optim.*, 48(5):3562–3588, 2010.

[8] A. Bensoussan and J.-L. Lions. *Impulse Control and Quasi-Variational Inequalities*. Gauthier-Villars, Paris, 1984.

[9] F. E. Benth, Á. Cartea, and R. Kiesel. Pricing forward contracts in power markets by the certainty equivalence principle: explaining the sign of the market risk premium. *Journal of Banking and Finance*, 32(10):2006–2021, 2008.

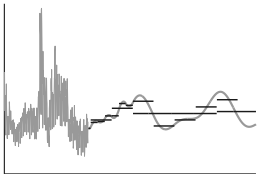
[10] F. E. Benth and S. Koekebakker. Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3):1116–1157, 2008.

[11] F. E. Benth, S. Koekebakker, and F. Ollmar. Extracting and applying smooth forward curves from average-based commodity contracts with seasonal variation. *Journal of Derivatives*, 52(15):62–66, 2007.

[12] F. E. Benth and T. Meyer-Brandis. The information premium for non-storable commodities. *Journal of Energy Markets*, 2(3):111–140, 2009.

[13] F. E. Benth, J. Šaltytė Benth, and S. Koekebakker. *Stochastic Modeling of Electricity and Related Markets*. World Scientific, Singapore, 2008.

[14] I. H. Biswas, E. R. Jakobsen, and K. H. Karlsen. Difference-quadrature schemes for nonlinear degenerate parabolic integro-PDE. *SIAM J. Numer. Anal.*, 48(3):1110–1135, 2010.



[15] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[16] O. Bobrovnytska and M. Schweizer. Mean-variance hedging and stochastic control: beyond the Brownian setting. *IEEE Transactions on Automatic Control*, 49(3), 2004.

[17] S. Bochkanov and V. Bystritsky. ALGLIB. http://www.alglib.net/.

[18] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer, Berlin, 3rd edition, 2008.

[19] M. Briani, C. L. Chioma, and R. Natalini. Convergence of numerical schemes for viscosity solutions to integro-differential degenerate parabolic problems arising in financial theory. *Numer. Math.*, 98:607–646, 2004.

[20] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.

[21] R. Carmona and M. Tehranchi. *Interest Rate Models: An Infinite-Dimensional Stochastic Analysis Perspective*. Springer, Berlin, 2006.

[22] P. Carr and D. B. Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2(4), 1999.

[23] C. Chiarella and A. Ziogas. American call options under jump-diffusion processes – A Fourier transform approach. *Applied Mathematical Finance*, 16(1):37–39, 2009.

[24] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall, Boca Raton, 2004.

[25] R. Cont and E. Voltchkova. A finite difference scheme for option pricing in jump diffusion and exponential Lévy models. *SIAM J. Numer. Anal.*, 43(4):1596–1626, 2005.

[26] R. Cont and E. Voltchkova. Integro-differential equations for option prices in exponential Lévy models. *Finance and Stochastics*, (9):299–325, 2005.

[27] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, Cambridge, 1992.

[28] G. Da Prato and J. Zabczyk. *Second Order Partial Differential Equations in Hilbert Spaces*. Cambridge University Press, Cambridge, 2004.

[29] Y. d'Halluin, P. Forsyth, and G. Labahn. A penalty method for American options with jump diffusion processes. *Numer. Math.*, 97(2):321–352, 2004.

[30] Y. d'Halluin, P. Forsyth, and G. Labahn. A semi-Lagrangian approach for American Asian options under jump diffusion. *SIAM Journal on Scientific Computing*, 27:315–345, 2005.

[31] J. Dieudonné. *Foundations of Modern Analysis*. Academic Press, New York, 1960.

[32] D. Duffie and H. R. Richardson. Mean-variance hedging in continuous time. *Ann. Appl. Probab.*, 1(1):1–15, 1991.

[33] L. C. Evans. *Partial Differential Equations.* American Mathematical Society, Graduate Studies in Mathematics, Volume 19, Providence, 1998.

[34] L. Feng and V. Linetsky. Pricing options in jump-diffusion models: An extrapolation approach. *Operations Research*, 56(2):304–325, 2008.

[35] M. Frigo and S. G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on "Program Generation, Optimization, and Platform Adaptation".

[36] G. Fusai and A. Roncoroni. *Implementing Models in Quantitative Finance: Methods and Cases.* Springer, Berlin, 2008.

[37] C. Geiger and C. Kanzow. *Theorie und Numerik Restringierter Optimierungsaufgaben.* Springer, Berlin, 2002.

[38] P. Glasserman. *Monte Carlo Methods in Financial Engineering.* Springer, New York, 2004.

[39] K. Glau. *Feynman-Kac-Darstellungen zur Optionspreisbewertung in Lévy-Modellen.* PhD thesis, Albert-Ludwigs-Universität Freiburg, 2010.

[40] M. B. Haugh and L. Kogan. Pricing American options: A duality approach. *Operations Research*, 52(2):258–270, 2004.

[41] E. Hausenblas. Existence, uniqueness and regularity of parabolic SPDEs driven by Poisson random measure. *Electron. J. Probab.*, 10:1496–1546, 2005.

[42] E. Hausenblas. A note on the Itô formula of stochastic integrals in Banach spaces. *Random Oper. Stochastic Equations*, 14(1):45–58, 2006.

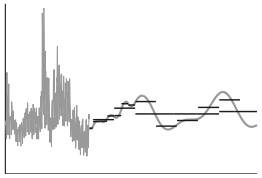
[43] M. Hegland, J. Garcke, and V. Challis. The combination technique and some generalisations. *Linear Algebra Appl.*, 420:249–275, 2007.

[44] P. Hepperger. Option pricing in Hilbert space valued jump-diffusion models using partial integro-differential equations. *SIAM Journal on Financial Mathematics*, 1:454–489, 2010.

[45] P. Hepperger. Hedging electricity swaptions using partial integro-differential equations. *Submitted for publication, preprint on http://www-m4.ma.tum.de*, 2011.

[46] P. Hepperger. Low-dimensional partial integro-differential equations for high-dimensional Asian options. *The Musiela Festschrift, Springer, accepted for publication, preprint on http://www-m4.ma.tum.de*, 2011.

[47] P. Hepperger. Numerical hedging of electricity swaptions using dimension reduction. *Submitted for publication, preprint on http://www-m4.ma.tum.de*, 2011.



[48] P. Hepperger. Pricing high-dimensional Bermudan options using variance-reduced Monte Carlo methods. *Journal of Computational Finance, accepted for publication, preprint on http://www-m4.ma.tum.de*, 2011.

[49] N. Hilber, N. Reich, C. Schwab, and C. Winter. Numerical methods for Lévy processes. *Finance and Stochastics*, 13:471–500, 2009.

[50] M. Kac. On distributions of certain Wiener functionals. *Transactions of the American Mathematical Society*, 65(1):1–13, 1949.

[51] R. Kiesel, G. Schindlmayr, and R. Börger. A two-factor model for the electricity forward market. *Quant. Finance*, 9(3):279–287, 2009.

[52] T. Kluge. *Pricing Swing Options and other Electricity Derivatives.* PhD thesis, University of Oxford, 2006.

[53] M. Kohler, A. Krzyżak, and N. Todorovic. Pricing of high-dimensional American options by neural networks. *Mathematical Finance*, 20:383–410, 2010.

[54] A. Kolodko and J. Schoenmakers. Iterative construction of the optimal Bermudan stopping time. *Finance and Stochastics*, 10:27–49, 2006.

[55] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numer. Math.*, 90(1):117–148, 2001.

[56] H. Kunita. Stochastic integrals based on martingales taking values in Hilbert space. *Nagoya Math. J.*, 38:41–52, 1970.

[57] J. P. Laurent and H. Pham. Dynamic programming and mean-variance hedging. *Finance Stoch.*, 3(1):83–110, 1999.

[58] A. L. Lewis. A simple option formula for general jump-diffusion and other exponential Lévy processes. In *Other Exponential Lévy Processes, Environ Financial Systems and OptionCity.net*, 2001.

[59] M. Loève. *Probability Theory II.* Springer, New York, 4th edition, 1978.

[60] F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: A simple least-squares approach. *The Review of Financial Studies*, 14:113–147, 2001.

[61] A.-M. Matache, T. von Petersdorff, and C. Schwab. Fast deterministic pricing of options on Lévy-driven assets. *ESAIM: Mathematical Modelling and Numerical Analysis*, 38(1):37–71, 2004.

[62] S. Peszat and J. Zabczyk. *Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach.* Cambridge University Press, Cambridge, 2007.

[63] C. Pflaum and A. Zhou. Error analysis of the combination technique. *Numer. Math.*, 84:327–350, 1999.

[64] H. Pham. On quadratic hedging in continuous time. *Math. Meth. Oper. Res.*, 51(2):315–339, 2000.

[65] P. E. Protter. *Stochastic Integration and Differential Equations.* Springer, Berlin, second edition, 2005.

[66] N. Reich. *Wavelet Compression of Anisotropic Integrodifferential Operators on Sparse Tensor Product Spaces.* PhD thesis, ETH Zürich, 2008.

[67] L. C. G. Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12(3):271–286, 2002.

[68] L. C. G. Rogers and Z. Shi. The value of an Asian option. *Journal of Applied Probability*, 32:1077–1088, 1995.

[69] B. Rüdiger and G. Ziglio. Itô formula for stochastic integrals with respect to compensated Poisson random measures on separable Banach spaces. *Stochastics*, (6):377–410, 2006.

[70] E. W. Sachs and A. K. Strauss. Efficient solution of a partial integro-differential equation in finance. *Appl. Numer. Math.*, 58(11):1687–1703, 2008.

[71] K. Sato. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, Cambridge, 1999.

[72] D. Schötzau and C. Schwab. Time discretization of parabolic problems by the hp-version of the discontinuous Galerkin finite element method. *SIAM J. Numer. Anal.*, 38(3):837–875, 2000.

[73] W. Schoutens. Exotic options under Lévy models: An overview. Technical report, UCS 2004-06, Leuven, Belgium.

[74] W. Schoutens. *Lévy Processes in Finance.* John Wiley & Sons, Chichester, 2003.

[75] C. Schwab and R. A. Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217:100–122, 2006.

[76] M. Schweizer. Mean-variance hedging for general claims. *Ann. Appl. Probab.*, 2(1):171–179, 1992.

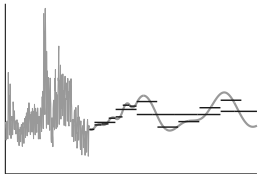
[77] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems.* Springer, Berlin, 2nd edition, 2006.

[78] R. A. Todor. Robust eigenvalue computations for smoothing operators. *SIAM J. Numer. Anal.*, 44(2):865–878, 2006.

[79] J. Večeř and M. Xu. Pricing Asian options in a semimartingale model. *Quantitative Finance*, 4:170–175, 2004.

[80] D. Werner. *Funktionalanalysis.* Springer, Berlin, 5th edition, 2005.

[81] C. Winter. *Wavelet Galerkin Schemes for Option Pricing in Multidimensional Lévy Models.* PhD thesis, ETH Zürich, 2009.



[82] J. Wloka. *Partial Differential Equations*. Cambridge University Press, Cambridge, 1987.

[83] G. Xu and H. Zheng. Approximate basket options valuation for a jump-diffusion model. *Insurance Math. Econom.*, 45(2):188–194, 2009.

[84] R. Zvan, P. Forsyth, and K. Vetzal. Robust numerical methods for PDE models of Asian options. *J. Computational Finance*, 1:39–78, 1998.