Universität Trier

Applications of the Adjoint Method in Stochastic Financial Modelling

Dissertation

zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften (Dr. rer. nat.)

Dem Fachbereich IV der Universität Trier
vorgelegt von

Bastian Peter Groß

Trier, Januar 2015

Gutachter: Prof. Dr. Ekkehard Sachs
Prof. Dr. Christian Bender

Tag der mündlichen Prüfung: 15. Mai 2015

Contents

German Summary (Zusammenfassung)

Die Bestimmung des fairen Preises eines Finanzkontraktes ist im Angesicht einerseits des wachsenden Marktes, andererseits der zunehmenden Komplexität von exotischen Optionen ein wesentlicher Bestandteil der Finanzmathematik. Der faire Preis ist maßgeblich von der Auszahlfunktion, die durch den Finanzkontrakt vorgegeben ist, und dem Finanzmarktmodell des Basiswertes, das von den Marktteilnehmern bestimmt wird, abhängig. Diese Arbeit beschäftigt sich damit, das Finanzmarktmodell, das die Bewegungen des Marktes widerspiegelt, so zu bestimmen, dass es zu beobachteten Marktpreisen passt. Diesen Vorgang, der mathematisch zu einem kleinsten Quadrate Problem führt, nennt man Kalibrierung.

In dieser Arbeit werden stochastische Finanzmarktmodelle, wie das einflussreiche Black-Scholes Modell oder das Heston Modell, betrachtet. Dabei spielt die Monte-Carlo Methode eine wichtige Rolle, bei der die Ausgänge einer Vielzahl von Simulationen, den sogenannten Monte-Carlo Simulationen, gemittelt werden. Diese Methode hat einen hohen Rechenaufwand und kann eine große Menge an Speicherplatz verbrauchen. Daher ist der erste Ansatz dieser Arbeit, genau diesen Rechenaufwand zunächst durch verbesserte Simulationsmethoden, den sogenannten stochastischen Prädiktor-Korrektor Verfahren, zu verringern. Die numerischen Ergebnisse dieser Arbeit bestätigen diese Verbesserung.

Neben dem Aufwand der Monte-Carlo Simulationen ist bei einer Gradienten-basierenden Kalibrierung vor allem die Berechnung des Gradienten bezüglich der zu bestimmenden Parameter bei der herkömmlichen Finiten-Differenzen Methode von hoher Komplexität, insbesondere, wenn die Anzahl der Parameter des Finanzmarktmodells hoch ist. Um diesen hohen Rechenaufwand zu umgehen, wird in dieser Arbeit die adjungierten Methode zur Berechnung des Gradienten verwendet. Zunächst wird diese adjungierte Methode im Allgemeinen vorgestellt und deren Anwendbarkeit auf das konkrete Kalibrierungsproblem des stochastischen Finanzmarktmodells gezeigt. Danach wird deren Implementierung ausgearbeitet und die numerischen Ergebnisse diskutiert. Dabei stellt sich heraus, dass die Kombination aus verbesserten Simulationsmethoden und der adjungierten Methode zu einer starken Verringerung des Rechenaufwands und des Speicherplatzverbrauchs führt. Damit hat sich die Rechenzeit der Kalibrierung deutlich reduziert.

Bei der adjungierten Methoden wird auch eine adjungierte stochastische Differentialgleichung berechnet. Das letzte Kapitel dieser Arbeit befasst sich mit dem Zusammenhang dieser adjungierten diskretisierten Gleichung zu der Lösung der rückwärtsstochastischen Differentialgleichungen, die aus der stochastischen Kontrolltheorie herrühren. Dabei stellt sich heraus, dass beide Ansätze im diskretisierten Fall zu demselben Ergebnis führen. Mit der Zusam-

menführung des adjungierten Ansatzes und der stochastischen Kontrolltheorie schließt diese Arbeit.

Bastian Groß
Trier, 2015

Acknowledgements (Danksagung)

Ich möchte mich an dieser Stelle bei all denen bedanken, die mich bei der Erstellung meiner Dissertation unterstützt haben.

Zuallererst bedanke ich mich an dieser Stelle herzlichst bei Herrn Prof. Dr. Ekkehard Sachs, der mich nicht nur während meiner ganzen Promotionszeit, sondern auch schon im Studium unterstützt und gefördert hat. Schon bei der Wahl meines Promotionsthemas sind Sie auf meine Stärken und Wünsche eingegangen und hatten bei Fragen und Diskussionsbedarf immer ein offenes Ohr. Das wissenschaftliche Forschen mit Ihnen hat mich akademisch weiterentwickeln lassen und mir Freude bereitet. Die Zeit als wissenschaftlicher Mitarbeiter bei Ihnen war eine sowohl aufbauende als auch verantwortungsvolle, in der ich viel Lernen aber auch Lehren durfte.

Außergewöhnlicher Dank gebührt auch meinem Zweitgutachter, Herrn Prof. Dr. Christian Bender, dessen wertvolle Expertise über rückwärtsstochastischer Differentialgleichungen und stochastischer Kontrolltheorie erst das letzte Kapitel dieser vorliegenden Arbeit ermöglicht haben. Ohne Ihren fruchtbaren akademischen Rat wäre dieser Teil meiner Dissertation nicht in der jetzigen Form entstanden. Die wissenschaftliche Zusammenarbeit mit Ihnen während der letzten Jahre war für mich Glück und Ehre zugleich.

Ebenso geht mein Dank an meine ehemaligen Kommilitonen und Kollegen die mich in den vergangenen Jahren an der Universität Trier mit bereichernden Diskussionen und einer belebenden Arbeitsatmosphäre in die richtigen Bahnen gelenkt haben: Bianca Schröder, Christina Jager, Christina Müller, Christina Schenk, Claudia Schillings, Julia Witzel, Kathrin Welker, Laura Somorowsky, Maike Thelen, Marina Schneider und ebenso Andre Lörx, Andreas Jung, Benedikt Wilbertz, Benjamin Rosenbaum, Bernd Perscheid, Christian Wagner, Christoph Käbe, Christoph Tasto, Dennis Sieg, Dirk Thomas, Duy Van Nguyen, Hans Franzen, Heinz Zorn, Jan Maruhn, Jona Schulz, Julian Wagner, Lukas Zimmer, Markus Born, Martin Rupp, Martin Siebenborn, Marvin Müller, Matthias Schu, Matthias Wagner, Nicolas Kenessy, Nils Langenberg, Roland Stoffel, Stefan Junglen, Stephan Schmidt, Thomas Kalmes, Thorben Schlierkamp, Timo Hylla, Ulf Friedrich, Yves Gessner, Xuancan Ye und auch die ganze Abteilung der Mathematik in Trier waren für mich immer ein Stütze in der Zeit an der Universität. Ich wünsche euch allen viel Glück und Erfolg bei euren weiteren beruflichen und privaten Wegen.

Nicht minder aufreibend waren die vergangenen Jahre für meine Familie und meine Freunde. Ihnen gilt mein besonderer Dank.

Chapter 1

Introduction

1.1 Motivation

The number of traded financial derivates increased rapidly over the last decades. Also the money value of those derivates leveled up as depicted in Figure 1.1. The number of traded financial contracts at the Chicago Board Options Exchange (CBOE)[1] increased from 284 million to 1.2 billion from 2003 to 2008 and stays nearly constant since that time. Almost the same progression is illustrated in the development at the German derivatives exchange named European Exchange (Eurex)[2] as shown in Figure 1.2.

Therefore, determining appropriate prices of financial derivates became an important element of financial engineering since Black, Scholes and Merton developed their pioneering formula. The value of a financial derivate depends on the value of some underlying. Since the movements of those underlying are unpredictable and carry certain risks, the dynamics of financial markets are often modeled with stochastic differential equations (SDEs). Modeling and calibrating such financial market models is a tough issue from a mathematical point of view because model calibration is often an ill-posed inverse problem in the sense of Hadamard (1902) as in Tikhonov and Arsenin (1977).

In this thesis, model calibration is about identifying the unknown market-implied parameters of a certain stochastic process that are governed by a stochastic differential equation which describes the dynamics of the underlying. This leads to a nonlinear least squares poblem. The calibration problem is based on the computation of a large number of sensitivities. Hence, every sophisticated calibration method relies on gradient-based optimization methods. The adjoint method, also known as the adjoint approach, is capable of accuratly and efficiently computing such sensitivities with respect to the model parameters.

This motivates the main topic of this thesis namely the application of the adjoint method in financial modeling. This work continues, extends and completes the research of Christoph Käbe (2010) concerning adjoint-based Monte-Carlo calibration.

Adjoint methods represent a wide range of applications of interests in different fields of computational engineering. For example, the adjoint method has been applied and implemented for optimal design of mechanical and aerodynamical systems (Giles (1997), Giles

[1] Data collected from https://www.cboe.com/data/AnnualMarketStatistics.aspx.
[2] Data collected from http://www.eurexchange.com/exchange-en/market-data/statistics/.

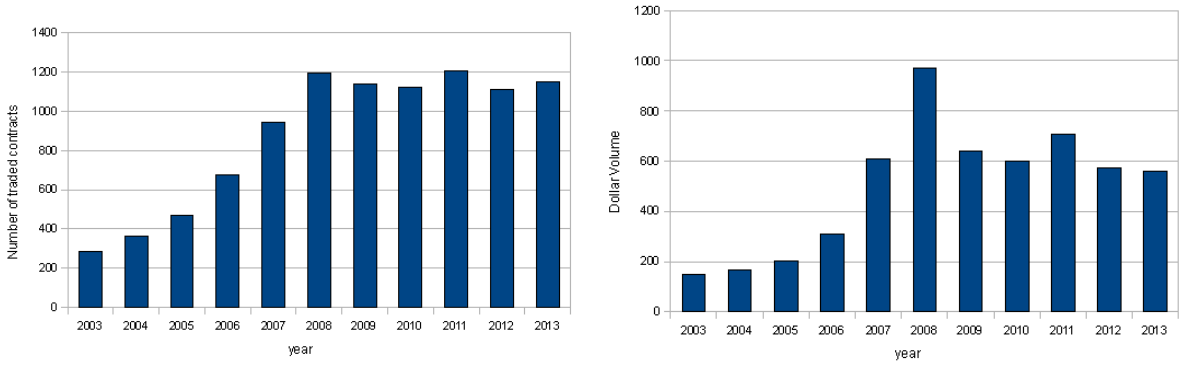

Figure 1.1: Total CBOE options trading summary: Number of traded contracts at the CBOE from 2003 to 2008 in million (left side) and total volume of this contracts in $ billion (right side).

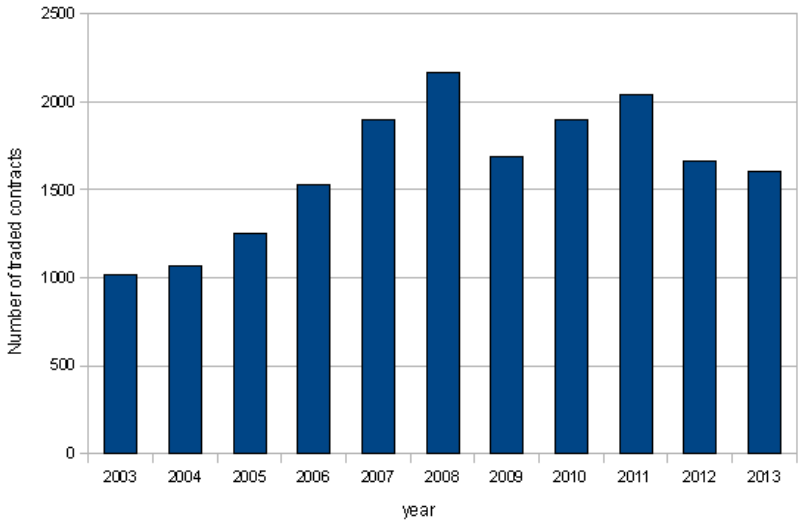

Figure 1.2: Number of traded contracts at the EUREX from 2003 to 2013 in million.

(2002), Giles and Pierce (2000), Giles and Süli (2002), Jameson (1988), Marchuk et al. (1996), Schillings et al. (2011), Schmidt et al. (2013) and Pironneau (1974)), optimal control of structural and mechanical systems (Carnarius et al. (2010), Giannakoglou and Papadimitriou (2008) and Haslinger and Mäkinen (2003)), air-conditioning, groundwater systems and weather forecasting (Courtier et al. (1993), Liu and Zhai (2007) and Michalak and Kitanidis (2004)), biomathematical modelling (Marchuk et al. (2005)) and computer graphics (Christensen (2003)). Altough Achdou and Pironneau (2005) and Giles and Glasserman (2006) explain how in practice the adjoint method can be used to compute option sensitivities, the use of adjoint methods in finance is still limited to only a few publications (Pironneau (2007), Capriotti and Giles (2012), Käbe et al. (2009), Lörx and Sachs (2012) and Lörx (2013)).

1.2 Outline

This thesis is divided into three main parts. The description of the calibration problem (Chapter 3), the numerical solution of this problem (4) and the connection to optimal stochastic control problems (Chapter 6). The introduction is finished by a short outline.

Chapter 2

Chapter 2 gives an introduction to the theoretical foundation of the adjoint method and provides tools to handle stochastic processes and calibrations problems. Principles of ordinary and stochastic calculus, option pricing and optimization are given.

Chapter 3

Chapter 3 introduces the calibration problem. Fitting model prices C to given market prices C_{obs} leads to the following abstract least squares formulation

$$\begin{aligned} &\min_{x \in X} f(x) := \tfrac{1}{2} \sum_{i=1}^{I} \left(C^i(x) - C^i_{\text{obs}} \right)^2 \\ &\text{where} \quad C^i(x) = \mathrm{e}^{-rT_i} \, \mathbb{E}\left[\max(S_{T_i}(x) - K_i, 0) \right]. \end{aligned} \tag{1.1}$$

The corresponding option price C can be computed by solving a stochastic differential equation via the Monte-Carlo method which seems to be preferred by most practitioners. Due to the fact that the Monte-Carlo method is expensive in terms of computational effort and required memory, more sophisticated stochastic predictor-corrector schemes are established in Section 3.3.2. The numerical advantages of these predictor-corrector schemes are presented and discussed in Section 5.1. The corresponding calibration problem is presented and further some mathematical challenges arising therein are briefly discussed.

Alternative statistical methods to estimate sensitivities of expectations are presented in Section 3.4. Their use for scientific applications is discussed, as well.

Chapter 4

The main part of this thesis is contained in Chapter 4. The adjoint method is applied to the calibration problem of Chapter 3 using tools from chapter 2. The corresponding discrete adjoint stochastic differential equation is figured out by using the implicit function theorem and the method of Lagrange multipliers. The discrete representation of the gradient is also proven.

The theoretical advantage of the adjoint method compared to the finite difference method is discussed in detail. In Section 4.3 it is shown that the computational effort of gradient calculation via the adjoint method is independent of the number of calibration parameters. Additionally, some techniques to reduce the calibration time are presented and briefly discussed.

Chapter 5

Numerical results are presented in Chapter 5 which is divided into two parts. First, different Monte-Carlo schemes namely the Euler-Maruyama scheme, the Milstein scheme and some stochastic predictor-corrector schemes are compared in terms of accuracy, speed of convergence and computational efficiency. The numerical results emphasize the advantage of the stochastic predictor-corrector schemes for multiple financial market models.

Secondly, the computation of the gradient via the adjoint method presented in Chapter 4 is implemented and its computer runtime is compared to the runtime of common methods. Applying the adjoint method to calibrate the model parameters produced the same accurate results compared with using the finite difference method. So the adjoint method is reliable. As testcase the Heston model

$$\begin{aligned} dS_t &= r(t)S_t dt + \sqrt{v_t}S_t \Delta W_t^1, \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}\Delta W_t^2, \\ \rho dt &= \mathrm{Cov}\left[dW_t^1, dW_t^2\right], \end{aligned}$$

is numerically solved using the proposed methods. These results numerically confirm the theoretical results of Section 4.3 and summarize the computational advantage of the adjoint method.

Chapter 6

Furthermore, Chapter 6 provides the connection to optimal stochastic control problems. The adjoint equation of Chapter 4 is the discretized solution Y of a stochastic differential equation with terminal condition as in the following abstract formulation

$$\begin{aligned} dY_t &= a(Y_t)dt + b(Y_t)dW_t, \quad t \in [0,T], \\ Y_T &= \xi. \end{aligned}$$

This solution of this problemis non-adapted. On the contrary, the solution of an optimal stochastic control problem leads to a pair of adapted processes (Y, Z) satisfying a backward stochastic differential equation

$$\begin{aligned} dY_t &= -f(Y_t, Z_t)dt + Z_t dW_t, \quad t \in [0,T], \\ Y_T &= \xi. \end{aligned}$$

Finally, this thesis presents the coincidence of this two approaches in the case of the discretized solutions of the adjoint equation and the backward stochastic differential equation. In particular, the discretized representations of the gradient match each other.

Chapter 2

Theoretical Background

In mathematics you don't understand things. You just get used to them.
— John von Neumann

This chapter introduces basic theory which will be frequently referred to throughout this thesis. Section 2.1 starts with some fundamentals of real and functional analysis which are often used. In addition, Section 2.2 contains an introduction to probability theory, stochastic processes, stochastic differential equations and gives a toolbox to handle such items. On the basis of this, Section 2.3 is intended to motivate the investigation of some important results of mathematical finance and option pricing. Finally, some fundamental concepts of numerical analysis, in particular optimization, will be explained in the last Section 2.4.

2.1 Elements of Real and Functional Analysis

In this first section some fundamentals of topological vector spaces, functional analysis and some concepts of differentiability are revised. The definitions are mostly taken from Bourbaki (1987), Debnath and Mikusinski (1998), Dieudonne (1969), Kantorovich and Akilov (1982) as well as Rudin (1976) and Rudin (1991). A detailed overview about functional analysis can be found in Meise and Vogt (1997). Additionally, fundamentals of linear algebra and related fields can be found in Golub and van Loan (1996).

Initially, assuming that topological spaces and metric spaces are known, definitions of some normed vector spaces that are important for applications in this thesis are given. Troughout this thesis, the term vector space will refer to a vector space over the real field $\mathbb{R}$.

Definition 2.1.1 (Banach space)**.** *A Banach space $(X, ||\cdot||_X)$ is a normed vector space which is complete with respect to its norm; this means that every Cauchy sequence is required to converge.*

Definition 2.1.2 (Hilbert space)**.** *A pair $(H, <\cdot,\cdot>_H)$ consisting of a vector space H and a bilinear map*

$$<\cdot,\cdot>_H \colon H \times H \to \mathbb{R}$$

named the inner product or scalar product on H is called inner product space. If the resulting normed space is complete, it is called a Hilbert space.

If the Hilbert space is clearly known, the subscript H will be omitted from the inner product and $<\cdot,\cdot>$ will be used for simplicitly. Obviously, Hilbert spaces are always Banach spaces.

Definition 2.1.3 (Separability). *A separable space X is a topological space which contains at least a countable dense subset in X.*

It is worth pointing out that separability is especially important in numerical analysis, since the convergence of many constructive and approximative algorithms can only be proved for separable spaces. Therefore, Banach spaces and Hilbert spaces are assumed to be separable spaces troughout this thesis.

Definition 2.1.4 (Dual Space). *The so-called dual space of a topological vector space X is the linear space X^* whose elements are the continuous linear functionals $l : X \to \mathbb{R}$.*

Note that since the scalar field $\mathbb{R}$ is always a complete space, the dual space of a normed space is always a Banach space (Debnath and Mikusinski, 1998, Theorem 1.6.5).

The following Theorem 2.1.5 establishes an important connection between a Hilbert space and its dual space.

Theorem 2.1.5 (Riesz Representation). *Let l be a bounded linear functional on an Hilbert space H. Then there exists a unique $x_0 \in H$ such that*

$$l(x) =< x, x_0 > \quad \forall\, x \in H.$$

Proof. (Debnath and Mikusinski, 1998, Theorem 3.10.1) □

Every Hilbert space H is reflexive, i.e. the natural map from H into its double dual space H^{**} is an isomorphism. Furthermore let T be a bounded operator on a Hilbert space H. Then for every fixed $x_0 \in H$ the functional defined on H by

$$x \mapsto < Tx, x_0 >$$

is a bounded linear functional on H. In consequence of the Riesz Representation Theorem 2.1.5 there exists a unique $y_0 \in H$ such that $< Tx, x_0 >=< x, y_0 >$ for all $x \in H$. The principal significance of this is that it allows to denote T^* as the operator which assigns every $x_0 \in H$ the according unique $y_0 \in H$, i.e.,

$$< Tx, y >=< x, T^*y > \quad \forall\, x, y \in H.$$

This leads to the following definition of the adjoint operator (Debnath and Mikusinski, 1998, Definition 4.4.1) which will play a crucial role in Chapters 4 and 6.

Definition 2.1.6 (Adjoint Operator). *Let T be a bounded operator on an Hilbert space H. The operator $T^* : H \to H$ defined by*

$$< Tx, y >=< x, T^*y > \quad \forall\, x, y \in H$$

is called the adjoint operator of T.

Up next some often used properties of functions on metric spaces are given.

Definition 2.1.7 (Lipschitz Continuity). *Let $(X, d_X), (Y, d_Y)$ be metric spaces. A map $f : X \to Y$ is called Lipschitz continuous if there exists a real constant $L \geq 0$ such that*

$$d_Y(f(x), f(y)) \leq L \; d_X(x, y) \quad \forall \; x, y \in X.$$

This L is called the Lipschitz constant for the function f. Furthermore, the function is called a contraction (contraction mapping) if its Lipschitz constant L is in the open intervall $(0, 1)$.

Theorem 2.1.8 (Banach Fixed-Point). *Let (X, d) be a non-empty complete metric space with a contraction mapping $T : X \to X$. Then T admits a unique fixed-point $x_* \in X$, this is $T(x_*) = x_*$. Furthermore, this fixed-point x_* can be constructed as follows: start with an arbitrary element $x_0 \in X$ and define a sequence $(x_n)_{n \in \mathbb{N}} \in X^{\mathbb{N}}$ by $x_n = T(x_{n-1})$, then this sequence converges with the fixed-point x_* as limit.*

Proof. (Debnath and Mikusinski, 1998, Theorem 1.8.1) □

Definition 2.1.9 (Fréchet Derivative). *Let $(X, ||\cdot||_X), (Y, ||\cdot||_Y)$ be Banach spaces, U some open subset of X and $x \in U$. A map $f : U \to Y$ is called Fréchet-differentiable at the point x, if there exists a bounded linear operator $f'_x : X \to Y$ such that*

$$\lim_{h \to 0+} \frac{\|f(x+h) - f(x) - f'_x(h)\|_Y}{\|h\|_X} = 0$$

converges uniformly with respect to an arbitrary $h \in X$ in bounded sets. Equivalently,

$$f(x+h) - f(x) = f'_x(h) + \epsilon_x(h)$$

$$\lim_{h \to 0+} \frac{\|\epsilon_x(h)\|_Y}{\|h\|_X} = 0.$$

f' is called Fréchet derivative or Fréchet differential of f. At last, $f : U \to Y$ is called Fréchet-differentiable on U, if f is Fréchet-differentiable at every point $x \in U$.

For a subset U of the Hilbert space $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$, the Fréchet derivative f' is the gradient of f denoted by ∇f. The gradient ∇f is a row or column vector depending on the context in which it is used

Corollary 2.1.10. *If the Fréchet derivative exists, it is unique.*

Proof. (Dieudonne, 1969, VIII 1. p. 149) □

Corollary 2.1.11. *If some map is Fréchet-differentiable on an open set U and its Fréchet derivative is bounded on U, it is Lipschitz continuous.*

Theorem 2.1.12 (Implicit Function Theorem). *Let X, Y, Z be Banach spaces, $G : Y \times X \to Z$ be a continuously differentiable map on the open neighbourhoods $\mathfrak{S}_{y^*} \times \mathfrak{S}_{x^*} \subset Y \times X$ of the point $(y^*, x^*) \in \mathfrak{S}_{y^*} \times \mathfrak{S}_{x^*}$ such that $G(y^*, x^*) = 0$. In addition, its partial derivative $G_y(y, x)$ is bijective in the open set containing (y^*, x^*).*

Then there exist a continuous map $y : \mathfrak{S}_{x^*} \to \mathfrak{S}_{y^*}$ *such that* $y^* = y(x^*)$ *and* $G(y(x), x) = 0$ *for all* $x \in \mathfrak{S}_{x^*}$. *The map* y *is unique, i.e., if* $y \in \mathfrak{S}_{y^*}$, $x \in \mathfrak{S}_{x^*}$ *and* $G(y, x) = 0$, *then* $y = y(x)$. *Furthermore,* y *is differentiable with*

$$y'(x) = -G_y(y(x), x)^{-1} G_x(y(x), x). \tag{2.1}$$

Proof. (Lusternik and Sobolev, 1965, Chapter VIII §8 Theorems 1-2) □

The function y is implicitly defined, hence the name of the Theorem.

2.2 Fundamentals of Stochastic Processes

Mathematical models that describe real world processes should have a probalistic component. Therefore, some principles of probability theory and the analysis of stochastic processes are introduced. Throughout this thesis, the existence of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ will be assumed, where Ω is the non-empty set where the random experiment takes place, $\mathcal{F}$ is a sigma algebra or σ-algebra of subsets of the set Ω and $\mathbb{P}$ an adequate probability measure over $\mathcal{F}$. $(E, \mathcal{E})$ is a measurable space, i.e., a set E equipped with a respective sigma algebra $\mathcal{E}$. If E is restricted to topological spaces, the sigma algebra $\mathcal{E}$ is assumed to be the Borel sigma algebra $\mathcal{B}(E)$, which is the sigma algebra generated by all open subsets of E. Usually $(E, \mathcal{E})$ is chosen as $(\mathbb{R}^d, \mathcal{B}^d)$, where $\mathcal{B}^d = \mathcal{B}(\mathbb{R}^d)$ is the Borel sigma algebra of $\mathbb{R}^d$.

Definition 2.2.1 (Measureability and Random Variable)**.** *Let* $(\Omega, \mathcal{F})$ *and* $(E, \mathcal{E})$ *be measurable spaces. Then a function* $g : \Omega \mapsto E$ *is measurable if for every set* $A \in \mathcal{E}$ *the preimage of* A *under* g *is in* $\mathcal{F}$, *i.e.*

$$g^{-1}(A) := \{\omega \in \Omega : g(\omega) \in A\} \in \mathcal{F}, \quad \forall\ A \in \mathcal{E}.$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* $(E, \mathcal{E})$ *a measurable space. A random variable is a measurable function* $X : \Omega \mapsto E$. *Then the expected value or the expectation is defined by*

$$\mathbb{E}[X] := \int_\Omega X d\mathbb{P} = \int_\Omega X(\omega)\mathbb{P}(d\omega). \tag{2.2}$$

Without exception probability spaces are assumed to be complete, this means $\mathcal{F}$ contains all $\mathbb{P}$-nullsets. In this manner, an event $A \in \mathcal{F}$ is said to happen almost surely (a.s.) if it happens with probability one, i.e., $\mathbb{P}(A) = 1$. All the equalities and inequalities involving random variables to appear in the sequel are understood to hold $\mathbb{P} - a.s.$.

Some arguments in this thesis involve infinite-dimensional vector spaces. To this end, L^p spaces are introduced. Let first $(\Omega, \mathcal{F}, \mu)$ be a measure space and $1 \leq p < \infty$. Then $L^p = L^p(\Omega) = L^p(\Omega, \mathcal{F}, \mu)$ consists of equivalence classes of measurable real functions f for which $|f|^p$ is integrable, i.e.

$$\int |f|^p d\mu < \infty,$$

where two measurable functions are equivalent if they are equal in the $\mu - a.s.$ sense. Then for $f \in L^p$ the following norm is used

$$||f||_p = \left(\int |f|^p d\mu \right)^{1/p} = \mathbb{E}[|f|^p]^{1/p}. \tag{2.3}$$

In the sequel, there will be no strong distinction between elements of L^p and representatives of their equivalence classes. In what follows, μ is restricted to be a probability measure. Hence $\mu = \mathbb{P}$ is σ-finite and according to the Riesz-Fischer Theorem (Meise and Vogt, 1997, Theorem 13.5) L^p is a separable Banach space for all $1 \leq p < \infty$.

Theorem 2.2.2. *The space L^p is complete and separable.*

Proof. (Billingsley, 1995, Theorem 19.1 and 19.2) □

The conditional expectation is defined as follow:

Definition 2.2.3 (Conditional Expectation)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X be an integrable random variable and $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra, then there exists an almost surely unique random variable denoted by $\mathbb{E}[X|\mathcal{G}]$, called the conditional expectation of X given $\mathcal{G}$, having the following two properties*

i) $\mathbb{E}[X|\mathcal{G}]$ *is $\mathcal{G}$-measurable and integrable.*

ii) $\mathbb{E}[X|\mathcal{G}]$ *satisfies the functional equation*

$$\int_G \mathbb{E}[X|\mathcal{G}] d\mathbb{P} = \int_G X d\mathbb{P}, \quad \forall\ G \in \mathcal{G}.$$

Let Y be another random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, then $\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)]$ is defined as the conditional expectation of X given Y.

The following lemma will list the basic properties of the conditional expectation.

Lemma 2.2.4 (Properties of Conditional Expectation)**.** *Let X be an integrable random variable and let $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ be sub-σ-algebras. $\mathbb{E}[\cdot|\mathcal{G}]$ may be seen as an operator on random variables that maps $\mathcal{F}$-measurable variables into $\mathcal{G}$-measurable ones.*

i) If $\mathcal{F}_0 = \{\emptyset, \Omega\}$ is the smallest possible σ-algebra, the following equation for the trivial conditional expectation holds

$$\mathbb{E}[X|\mathcal{F}_0] = \mathbb{E}[X], \tag{2.4}$$

and, therefore, the random variable denoted by the conditional expectation is constant.

ii) The tower property holds

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}]. \tag{2.5}$$

Since $\mathcal{F}_0 = \{\emptyset, \Omega\}$ is always a sub-σ-algebra of $\mathcal{G}$, also the law of total expectation is valid

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]. \tag{2.6}$$

iii) $\mathbb{E}[\cdot|\mathcal{G}]$ *is a projection by*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]. \tag{2.7}$$

iv) For $X \in L^p$, $\mathbb{E}[\cdot|\mathcal{G}]$ is a contraction by

$$||\mathbb{E}[X|\mathcal{G}]||_p \leq ||X||_p \quad 1 \leq p < \infty. \tag{2.8}$$

The norm $||\cdot||_p$ is defined by (2.3).

v) The conditional expectation has the property that it is the best approximation in the sense of later Theorem 2.4.12 insofar that $\mathbb{E}[X|\mathcal{G}]$ is the $\mathcal{G}$-measurable function that best approximates X as a variance-minimizing prediction. So, for any square integrable function h it holds

$$\mathbb{E}[(\mathbb{E}[X|Y] - X)^2] \leq \mathbb{E}[(h(Y) - X)^2] \tag{2.9}$$

with equality if and only if $h(Y) = \mathbb{E}[X|Y]$. This is in fact equivalent to the following orthogonality property

$$\mathbb{E}[X - \mathbb{E}[X|\mathcal{G}]|\mathcal{G}] = 0. \tag{2.10}$$

Proof. (Williams, 1991, Section 9.7 and 9.8) □

Definition 2.2.5 (Filtration)**.** *Let $I \subset \mathbb{R}$ be an index set. An increasing sequence of sigma algebras $\mathbb{F} = (\mathcal{F}_i)_{i \in I}$ is called filtration on a measurable space $(\Omega, \mathcal{F})$, i.e., for $i \leq j$ one has $\mathcal{F}_i \subseteq \mathcal{F}$ and $\mathcal{F}_i \subseteq \mathcal{F}_j$ for all $i, j \in I$.*

A probability space equipped with the filtration $\mathbb{F}$ is called a filtered probability space and denoted by $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. A filtered probability space is said to satisfy the usual conditions if $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{T}}$ is right-continuous and $\mathcal{F}_0$ contains all $\mathbb{P}$-nullsets. Hence, the filtration is a totally ordered set and assuming I represents a set of points in time, then heuristically one could interpret that the set $\mathcal{F}_t$ contains all information available up to time t. In many examples, the index set I is the natural numbers $\mathbb{N}$ or a time interval $\mathcal{T} := [0, T]$ or $[0, \infty)$.

In this thesis stochastic processes will be defined in the sense of (Øksendal, 2003, Definition 2.1.4).

Definition 2.2.6 (Stochastic Process)**.** *Consider measurable spaces $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ and a time interval $\mathcal{T} \subset \mathbb{R}^+$. A family $X = (X_t)_{t \in \mathcal{T}}$ of random variables*

$$X_t : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$$

is called a stochastic process.

In many cases stochastic processes exist trough Kolmogorov's Extension Theorem (Billingsley, 1995, Theorem 36.1). Furthermore, for a fixed event $\omega \in \Omega$, the map $X_t(\omega) : \mathcal{T} \to \mathbb{R}^d$ is called a path of the stochastic process. On the contrary, for a fixed time $t \in \mathcal{T}$ the random variable $X_t(\omega) : \Omega \to \mathbb{R}^d$ is called a realisation of the stochastic process.
The natural filtration in $\mathcal{F}$ with respect to a stochastic process $X = (X_i)_{i \in I}$ is defined as

$$\mathcal{F}^X = (\mathcal{F}_i^X)_{i \in I} = \left(\sigma \left\{ X_j^{-1}(A) \middle| j \in I, j \leq i, A \in \mathcal{E} \right\} \right)_{i \in I}, \tag{2.11}$$

i.e., the smallest sequence of sigma algebra on Ω that contains all preimages of X_j under $\mathcal{E}$-measurable subsets of $\tilde{\Omega}$ for times j up to i. All information concerning a stochastic process, and only that information, is available in its natural filtration.

Moreover, by extending Definition 2.2.3 having a stochastic process $(Y_t)_{t \in \mathcal{T}}$ and an integrable random variable X for abbreviation the expectation conditioned by the stochastic process $(Y_t)_{t \in \mathcal{T}}$ is given by $\mathbb{E}[X|\mathcal{F}_t^Y]$ for some $t \in \mathcal{T}$.

Definition 2.2.7 (Stopping Time). *For a given filtration $(\mathcal{F}_i)_{i \in I}$ a mapping*

$$\tau : \Omega \to I \cup \{+\infty\}$$

is called stopping time if $\{\tau \leq t\} := \{\omega \in \Omega : \tau(\omega) \leq t\} \subset \mathcal{F}_t$. Furthermore, for any stochastic process X and each stopping time τ the stochastic process X^τ denotes the process stopped in τ given by

$$X_t^\tau(\omega) := X_{\min(t, \tau(\omega))}(\omega) \quad \text{for } \omega \in \Omega.$$

Finally,

$$\mathfrak{T} = \{\tau | \tau \text{ is stopping time with } \tau < \infty\}$$

denotes the set of all stopping times which do not take the value $+\infty$.

Definition 2.2.8 (Adapted and Predictable Process). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{T}}$ be a filtration.*

i) A stochastic process $X = (X_t)_{t \in \mathcal{T}}$ is called adapted or non-anticipating with respect to the filtration $\mathbb{F}$ if each X_t is $\mathcal{F}_t$-measurable.

ii) A discrete-time stochastic process $X = (X_n)_{n \in \mathbb{N}}$ is called predictable with respect to the filtration $\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ if each X_n is $\mathcal{F}_{n-1}$-measurable.

Every stochastic process $X = (X_t)_{t \in \mathcal{T}}$ is adapted with respect to its own natural filtration $\mathcal{F}_t^X$ per definition. Furthermore, it follows from the definition of a stopping time τ that X^τ is adapted if X is adapted. An informal interpretation is that X is adapted if and only if, for every realisation and every time t, the random variable X_t is known only at time t.

Definition 2.2.9 (Martingale). *A stochastic process $M = (M_t)_{t \in \mathcal{T}}$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is called a martingale with respect to $\mathbb{F}$ if*

i) M is adapted,

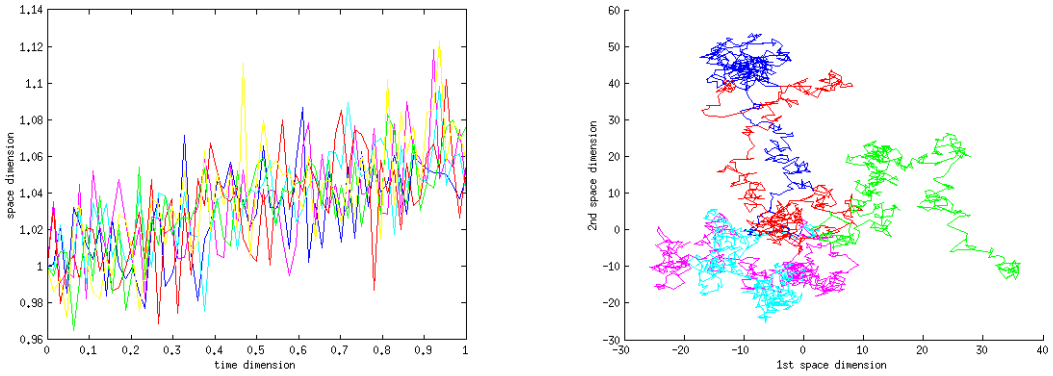

Figure 2.1: Some paths of the one-dimensional (left side) and the two-dimensional (right side) Brownian motion.

ii) $\mathbb{E}[|M_t|] < \infty$ *for all* $t \in \mathcal{T}$ *and*

iii) $M_s = \mathbb{E}[M_t|\mathcal{F}_s]$ *almost everywhere* $\forall\ 0 \leq s < t \leq T$.

Whether or not a given stochastic process $M = (M_t)_{t\in\mathcal{T}}$ is a martingale depends on the underlying probability measure $\mathbb{P}$. Inasmuch as the martingale property of M depends on a particular measure $\mathbb{P}$, one has to specify that M is a $\mathbb{P}$-martingale when confusion can arise. Following Föllmer and Schied (2004) a martingale M can be regarded as the mathematical formalization of a "fair game". This means that for every time s and for each time horizon $t > s$, the conditional expectation of the future gain $M_t - M_s$ is zero, given all the information available at s. By Doob's Stopping Theorem (see for instance Doob (1971)) the stopped process M^τ is a martingale for any stopping time τ if M is a martingale.

Some important examples for stochastic processes are the following:

Example 2.2.10 (Wiener Process)**.** *A Brownian motion or Wiener process is a stochastic process* $W = (W_t)_{t\in[0,T]}$ *with the properties*

i) The process starts at 0: $W_0 = 0\ \mathbb{P} - a.s.$.

ii) For any given times $0 \leq t_0 < t_1 < t_2 < \cdots < t_m \leq T$ *the increments* $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \ldots, W_{t_m} - W_{t_{m-1}}$ *are stochastically independent.*

iii) For all $0 \leq s < t$ *it is* $W_t - W_s \ \sim\ \mathcal{N}(0, t-s)$*, where* $\mathcal{N}(\mu, \sigma^2)$ *denotes the normal distribution with expected value* μ *and variance* σ^2.

iv) Paths are $\mathbb{P} - a.s.$ *continuous, this means the function* $t \to W_t$ *is almost surely everywhere continuous.*

Note that point *iv)* can be ommitted due to the Kolmogorov Continuity Theorem (Øksendal, 2003, Theorem 2.2.3). The Brownian motion achieved a high level of awareness by the paper of Einstein (1905).

Theorem 2.2.11 (Non-differentiability of the Wiener Process)**.** *Almost surely the Wiener process is nowhere differentiable.*

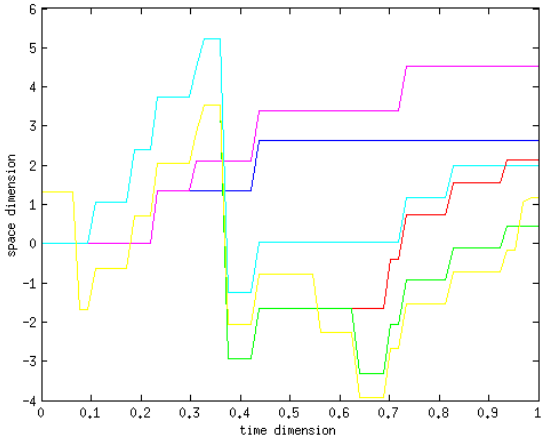

Figure 2.2: Some paths of the one-dimensional Poisson Process.

Proof. Karatzas and Shreve (2000) □

Example 2.2.12 (Poisson Process). *A Poisson process with intensity λ is a stochastic process $N = (N_t)_{t\in[0,T]}$ with the properties*

i) The process starts at 0: $N_0 = 0$ $\mathbb{P}-a.s.$.

ii) For any given times $0 \leq t_0 < t_1 < t_2 < \cdots < t_m \leq T$ the increments $N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, \ldots, N_{t_m} - N_{t_{m-1}}$ are stochastically independent.

iii) For all $0 \leq s < t$ it is $N_t - N_s \sim \mathcal{P}(\lambda(t-s))$, where $\mathcal{P}(\lambda)$ denotes the Poisson distribution with expected value λ and variance λ.

Example 2.2.13 (Lévy Process). *A Lévy process is a stochastic process $L = (L_t)_{t\in[0,T]}$ with the properties*

i) The process starts at 0: $L_0 = 0$ $\mathbb{P}-a.s.$.

ii) For any given times $0 \leq t_0 < t_1 < t_2 < \cdots < t_m \leq T$ the increments $L_{t_1} - L_{t_0}, L_{t_2} - L_{t_1}, \ldots, L_{t_m} - L_{t_{m-1}}$ are stochastically independent.

iii) For all $0 \leq s < t$ it is $L_t - L_s \sim L_{t-s}$, i.e. the increments are stationary.

iv) For any time $t \geq 0$ and an arbitray $\epsilon > 0$ it holds

$$\lim_{h\to 0+} \mathbb{P}(|L_{t+h} - L_t| > \epsilon) = 0.$$

Subsequently, stochastic (ordinary) differential equations are introduced. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space satisfying the usual conditions, on which a $\mathbb{R}^m$-valued standard Wiener process, $W = (W_t)_{0\leq t\leq T}$, is defined. Then the dynamic of a $\mathbb{R}^d$-valued stochastic process $X = (X_t)_{0\leq t\leq T}$ is the solution of a stochastic differential equation (SDE)

of the form established by Øksendal (2003):

$$\begin{aligned} dX_t &= a(t, X_t)dt + b(t, X_t)dW_t \quad t \in [0, T] \\ X_0 &= x_0. \end{aligned} \tag{2.12}$$

In this context a is called the deterministic continuous drift coefficient or the appreciation rate and b is called the continuous diffusion or volatility coefficient. These coefficient functions fulfill the usual properties of the existence and uniqueness of the solution of such an SDE pooled in the following theorem.

Theorem 2.2.14 (Existence and Uniqueness for Solutions of Stochastic Differential Equations). *Let the drift coefficient $a(\cdot, \cdot) : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and the diffusion coefficient $b(\cdot, \cdot) : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times m}$ be measurable functions satisfying the linear growth condition*

$$||a(t, x)|| + ||b(t, x)|| \leq C(1 + ||x||), \quad \forall\ x \in \mathbb{R}^d \text{ and } t \in [0, T] \tag{2.13}$$

for some constant $C > 0$ and further assume that the Lipschitz condition

$$||a(t, x) - a(t, y)|| + ||b(t, x) - b(t, y)|| \leq L||x - y||, \quad \forall\ x, y \in \mathbb{R}^d \text{ and } t \in [0, T] \tag{2.14}$$

for some Lipschitz constant $L > 0$ is satisfied. Then the stochastic differential equation with starting value Z

$$\begin{aligned} dX_t &= a(t, X_t)dt + b(t, X_t)dW_t \quad t \in [0, T] \\ X_0 &= Z \end{aligned} \tag{2.15}$$

has a pathwise unique t-continuous strong solution $X = (X_t)_{0 \leq t \leq T}$.

Proof. (Øksendal, 2003, Theorem 5.2.1 p.66ff) □

Note that this strong solution is an $\mathcal{F}_t^Z$-adapted stochastic process constructed from a certain version of W_t. Here uniqueness means that if X^1 and X^2 are two t-continuous solutions satisfying (2.15) then

$$X_t^1(\omega) = X_t^2(\omega) \quad \forall\ t \in [0, T], \quad a.s.. \tag{2.16}$$

Weak unique solutions are only identical in distribution. Similar conditions (2.13) and (2.14) could also be used for the proof of the existence and uniqueness of solutions of deterministic differential equations.

Due to the non-differentiability of the Wiener Process (see Theorem 2.2.11) there is no differentiation theory of stochastic processes. Therefore, one can rewrite the solution of above equation (2.12) to the integral formulation as X_t is a solution of the integral equation

$$X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s, \quad t \in [0, 1], \tag{2.17}$$

where $\int_0^t b(s, X_s)dW_s$ is a stochastic integral which leads either to an Itô integral, a (Fisk-)

Stratonovich integral or a forward integral in the sense of Russo and Vallois (1993). The suitable interpretation of the stochastic integral is crucial for its solution, but it will only be discussed briefly. Let $0 = t_0 \leq \ldots \leq t_N = T$, $N \in I\!N$ be a time discretization of the interval $[0, T] \subset \mathbb{R}$, then the stochastic integral is approximated by

$$\int_0^t f(s,\omega) dW_s(\omega) \approx \sum_{n=1}^{N} f(\tau_n, \omega) \left(W_{t_n} - W_{t_{n-1}} \right), \tag{2.18}$$

where $\tau_n \in [t_{n-1}, t_n]$. By choosing τ_n as t_n one obtains the Itô integral, on the other hand, by choosing τ_n as $\frac{t_n - t_{n-1}}{2}$ one gets the Stratonovich integral. The Russo-Vallois integral is briefly introduced in Russo and Vallois (1995). If not otherwise specified, stochastic integrals are meant to be Itô integrals, troughout this thesis.

Note that the symbolic notation of the SDE (2.12) and (2.17) includes the multi-dimensional case. Let $W_t = (W_t^1, \ldots, W_t^m)^\top$ be the m-dimensional vector of a Wiener process, $a : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ be a drift vector and $b : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^m$ be a volatility matrix, then the solution of (2.12) is denoted by $X_t = (X_t^1, \ldots, X_t^d)^\top$ and its componentwise integral representation is

$$X_t^i = X_0^i + \int_0^t a^i(s, X_s) ds + \sum_{\nu=1}^{m} \int_0^t b^{i,\nu}(s, X_s) dW_s^\nu, \quad t \in [0,1], \tag{2.19}$$

for a preassigned $\mathcal{F}_0$ measurable initial vector $X_0 \in \mathbb{R}^d$. Equipped with the integral formulation, an integral theory for SDE is established now. Assume $X = (X_t)_{0 \leq t \leq T}$ is an $\mathbb{R}$-valued stochastic process and $Y_t = g(t, X_t)$, where $g \in C^2([0,T] \times \mathbb{R}, \mathbb{R})$. In that case one has to adopt a stochastic chain rule to estimate $dY_t = d(g(t, X_t))$ via the well-known Itô's formula (Itô's rule, Itô's lemma) as in Øksendal (2003) or Protter (1991) or the following version of the Itô's formula in higher dimensions.

Theorem 2.2.15 (Itô's Lemma). *Let $X = (X_t)_{0 \leq t \leq T}$ be an $\mathbb{R}^d$-valued stochastic process, $F \in C^2(\mathbb{R}^d, \mathbb{R})$. Then Itô's formula or Itô's (chain) rule states*

$$F(X_t) - F(X_0) = \sum_{j=1}^{d} \int_0^t \frac{\partial F}{\partial x^j}(X_s) \mathrm{d}X_s^j + \frac{1}{2} \sum_{j,k=1}^{d} \int_0^t \frac{\partial^2 F}{\partial x^j \partial x^k}(X_s) \mathrm{d}[X^j, X^k]_s^c, \tag{2.20}$$

where

$$[X^j, X^k]_t^c = \lim_{\Delta t_n \to 0} \sum_{n=1}^{N} \left(X_{t_n}^j - X_{t_{n-1}}^j \right) \left(X_{t_n}^k - X_{t_{n-1}}^k \right)$$

is the quadratic variation (quadratic covariation) process, with $0 = t_1 < \ldots < t_N = t$ and $\Delta t_n = t_{n+1} - t_n$.

Proof. (Øksendal, 2003, Theorem 4.2.1) □

It is eminent to point out that for Stratonovich integral driven SDEs the stochastic chain rule is the same as the chain rule of ordinary calculus.

Lemma 2.2.16. *Let $G = (G_t)_{0\le t\le T}$ and $H = (H_t)_{0\le t\le T}$ be $\mathbb{R}^d$-valued stochastic processes with dynamics*

$$G_t = G_0 + \int_0^t g(s)ds + \int_0^t \gamma(s)dW_s,$$
$$H_t = H_0 + \int_0^t h(s)ds + \int_0^t \eta(s)dW_s.$$

Then,

$$\langle G_t, H_t\rangle = \langle G_0, H_0\rangle + \int_0^t \langle G_s, h(s)\rangle + \langle g(s), H_s\rangle + \langle \eta(s), \gamma(s)\rangle\, ds + \int_0^t \langle G_s, \eta(s)\rangle + \langle \gamma(s), H_s\rangle\, dW_s. \tag{2.21}$$

Lemma 2.2.16 is an extension of Corrollary 5.6 in Yong and Zhou (1999) and is the stochastic version of the integration-by-parts formula (cf. Elliott and Kohlmann (1989b)).

Proof. Applying the Itô Lemma (2.20) to $\langle G_t, H_t\rangle$ ($F : \mathbb{R}^{2d} \to \mathbb{R}$, $F((\cdot,\cdot)^\top) = \langle\cdot,\cdot\rangle$) one obtain above result. Let to this purpose $F : \mathbb{R}^{2d} \to \mathbb{R}$ be the Euclidian inner product for an extended vector $\widehat{x} := (x, y)^\top \in \mathbb{R}^{2d}$ with

$$F((x,y)^\top) = \langle x, y\rangle = \sum_{j=1}^{d} x_j y_j = \sum_{j=1}^{d} \widehat{x}_j \widehat{x}_{j+d} =: F(\widehat{x})$$

Then its derivates are

$$\frac{\partial F}{\partial \widehat{x}^j}(\widehat{x}) = \begin{cases} \widehat{x}_{j+d}, & 1 \le j \le d \\ \widehat{x}_{j-d}, & d+1 \le j \le 2d, \end{cases}$$

and

$$\frac{\partial^2 F}{\partial \widehat{x}^j \partial \widehat{x}^k}(\widehat{x}) = \begin{cases} \delta_{k,j+d}, & 1 \le j \le d \\ \delta_{k,j-d}, & d+1 \le j \le 2d \\ 0, & \text{else.} \end{cases}$$

By using the multi-dimensional Itô's Formula (2.20) one obtains via $(G_t, H_t)^\top = \widehat{X}_t$

$$\begin{aligned}\langle G_t, H_t\rangle &= F(\widehat{X}_t) \\ &= F(\widehat{X}_0) + \sum_{j=1}^{2d} \int_0^t \frac{\partial F}{\partial \widehat{X}^j}(\widehat{X}_s)\mathrm{d}\widehat{X}_s^j + \frac{1}{2}\sum_{j,k=1}^{2d} \int_0^t \frac{\partial^2 F}{\partial \widehat{X}^j \partial \widehat{X}^k}(\widehat{X}_s)\mathrm{d}[\widehat{X}^j, \widehat{X}^k]_s^c\end{aligned}$$

$$
\begin{aligned}
&= F(\widehat{X}_0) + \sum_{j=1}^{d} \int_0^t \widehat{X}_s^{j+d} \mathrm{d}\widehat{X}_s^j + \sum_{j=d+1}^{2d} \int_0^t \widehat{X}_s^{j-d} \mathrm{d}\widehat{X}_s^j \\
&+ \frac{1}{2} \left(\sum_{j=1}^{d} \int_0^t \mathrm{d}[\widehat{X}^j, \widehat{X}^{j+d}]_s^c + \sum_{j=d+1}^{2d} \int_0^t \mathrm{d}[\widehat{X}^j, \widehat{X}^{j-d}]_s^c \right) \\
&= F(\widehat{X}_0) + \sum_{j=1}^{d} \int_0^t \widehat{X}_s^{j+d} \mathrm{d}\widehat{X}_s^j + \sum_{j=d+1}^{2d} \int_0^t \widehat{X}_s^{j-d} \mathrm{d}\widehat{X}_s^j + \sum_{j=1}^{d} \int_0^t \mathrm{d}[\widehat{X}^j, \widehat{X}^{j+d}]_s^c \\
&= \langle G_0, H_0 \rangle + \sum_{j=1}^{d} \int_0^t H_s^j \mathrm{d}G_s^j + \sum_{j=1}^{d} \int_0^t G_s^j \mathrm{d}H_s^j + \sum_{j=1}^{d} \int_0^t \mathrm{d}[G^j, H^j]_s^c \\
&= \langle G_0, H_0 \rangle + \int_0^t \langle G_s, h(s) \rangle + \langle g(s), H_s \rangle + \langle \eta(s), \gamma(s) \rangle \, ds \\
&+ \int_0^t \langle G_s, \eta(s) \rangle + \langle \gamma(s), H_s \rangle \, dW_s,
\end{aligned}
$$

by using the multiplication table (Øksendal, 2003, Theorem 4.2.1)

$$dtdt = dW_t^j dt = dt dW_t^j = 0 \quad \text{and} \quad dW_t^j dW_t^k = dt\delta_{j,k} \quad \text{for all} \quad k, j.$$

This completes the proof and establishes the formula (2.21). □

Next, some spaces of stochastic processes (Yong and Zhou, 1999, page 32) are defined. Equipped with the definition of $L^p(\Omega)$ spaces given by equation (2.3) with $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}^d, \mathcal{B}^d, \lambda_{\mathcal{B}}^d)$, where $\lambda_{\mathcal{B}}$ denotes the Lebesgue measure the following spaces of stochastic processes are considered

$$
\begin{aligned}
L_{\mathcal{F}}^p(0, T; \mathbb{R}^d) := \Big\{ & f \in L^p(\mathbb{R}^d) : f \text{ is measurable and adapted to } \{\mathcal{F}_t\}_{t \in [0,T]} \\
& \text{and} \quad \mathbb{E}\left[\int_0^T f(t, \omega)^p dt \right] < \infty \Big\}
\end{aligned}
$$

$L_{\mathcal{F}}^2(0, T; \mathbb{R}^d)$ equipped with the inner product

$$< f, g > := \mathbb{E}[\int_0^T f(t, \omega)^\top g(t, \omega) dt]$$

is a Hilbert space (Yong and Zhou, 1999, page 32 (5.15)).

$$
\begin{aligned}
\mathcal{M}^2[0, T] := \big\{ & X \in L_{\mathcal{F}}^2(0, T; \mathbb{R}^n) : \text{X is a right-continuous } \{\mathcal{F}_t\}_{t \geq 0}\text{-martingale} \\
& \text{with } X(0) = 0, \mathbb{P} - a.s. \big\} \\
\mathcal{M}_c^2[0, T] := \big\{ & X \in \mathcal{M}^2[0, T] : t \mapsto X(t) \text{ is continuous}, \mathbb{P} - a.s. \big\}
\end{aligned}
$$

Equipped with the inner product

$$(X,Y)_{\mathcal{M}^2[0,T]} := \mathbb{E}\left[X_T^\top Y_T\right]$$

$(\mathcal{M}^2[0,T], (\cdot,\cdot)_{\mathcal{M}^2[0,T]})$ is a Hilbert space.
Next, let $\mathfrak{X}$ be the space of stochastic processes of the form

$$V_t = V_0 + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dW_s, \quad 0 \le t \le T,$$

with

$$\mathbb{E}\left[\int_0^T ||\mu(s)||^2 + ||\sigma(s)||^2 ds\right] < \infty.$$

Let $U \in \mathfrak{X}$ be another stochastic process of the form

$$U_t = U_0 + \int_0^t \nu(s)ds + \int_0^t \gamma(s)dW_s, \quad 0 \le t \le T,$$

then an inner product of $\mathfrak{X}$ is defined by

$$< V,U >_{\mathfrak{X}} = \mathbb{E}[V_0^\top U_0] + \mathbb{E}[\int_0^T \mu(s)^\top \nu(s) + \sigma(s)^\top \gamma(s) ds].$$

$(\mathfrak{X}, < \cdot,\cdot >_{\mathfrak{X}})$ is an Hilbert space. One notes that the properties of this inner product are deduced by the Euclidean scalar product. Moreover, the norm

$$|V|_{\mathfrak{X}} = \sqrt{< V,V >_{\mathfrak{X}}} = \left(||V_0||^2 + \mathbb{E}[\int_0^T ||\mu(s)||^2 + ||\sigma(s)||^2 ds]\right)^{1/2}$$

is equivalent to

$$||V||_{\mathfrak{X}} = \left(\mathbb{E}[||V_T||^2] + \mathbb{E}[\int_0^T ||\mu(s)||^2 ds]\right)^{1/2}.$$

$\mathfrak{X}_0$ is also introduced as the closed subspace of $\mathfrak{X}$ such that $V_0 = 0$.

The following Martingale Representation Theorem is concerned with how to represent a martingale $M_t \in \mathcal{M}^2[0,T]$ by a process $Z_t \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^n)$ with a fixed Wiener process.

Theorem 2.2.17 (Martingale Representation Theorem)**.** *Let $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ be a filtered probability space satisfying the usual conditions and $W_t = (W_t^1,\ldots,W_t^m)$ be a d-dimensional Wiener process. Suppose M_t is an $\mathbb{F}$-martingale with respect to $\mathbb{P}$ and $M_t \in \mathcal{M}^2[0,T]$ for all times $t \ge 0$. Then there exists a unique stochastic process $Z_s(\omega)$ such that $Z \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^n)$ for all $t \ge 0$ and*

$$M_t(\omega) = \mathbb{E}[M_0] + \int_0^t Z_s(\omega)dW_s \quad \mathbb{P}-a.s. \ \ \forall\ t \ge 0. \tag{2.22}$$

Proof. (Øksendal, 2003, Theorem 4.3.4) □

An interesting consequence of Theorem 2.2.17 follows by considering $\zeta \in L^2_{\mathcal{F}}(0,T;\mathbb{R})$ and the fact that $\mathbb{E}[\zeta|\mathcal{F}_t]$ is an $\mathbb{F}$-martingale. Thus, there exists a $Z \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^n)$ due to equation (2.22) such that

$$\mathbb{E}[\zeta|\mathcal{F}_t] = \mathbb{E}[\zeta] + \int_0^t Z_s dW_s \quad \forall\, t \in [0,T]. \tag{2.23}$$

In particular

$$\zeta = \mathbb{E}[\zeta] + \int_0^T Z_s dW_s. \tag{2.24}$$

This consequence will play a vital role in Chapter 6 concerning backward stochastic differential equations (BSDEs).

The following results are also needed. All homogeneous additive functionals of Brownian motions have zero (conditional) expectation.

Lemma 2.2.18. *For any $f \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^n)$ and stopping time σ and τ with $\sigma \leq \tau, \mathbb{P}-a.s.$,*

$$\mathbb{E}\left[\int_{s\wedge\sigma}^{t\wedge\tau} f(r)dW_r|\mathcal{F}_r\right] = 0, \quad \mathbb{P}-a.s., \tag{2.25}$$

in particular, for any $0 \leq s < t \leq T$,

$$\mathbb{E}\left[\int_s^t f(r)dW_r|\mathcal{F}_r\right] = 0, \quad \mathbb{P}-a.s., \tag{2.26}$$

Proof. (Yong and Zhou, 1999, Proposition 5.2 (5.24)) □

The following corollary enables the computation of second order terms of integrals of stochastic processes.

Corollary 2.2.19 (Itô Isometry)**.**

$$\mathbb{E}\left[\left(\int_0^T f(t)dW_t\right)^2\right] = \mathbb{E}\left[\int_0^T f^2(t)dt\right] \quad \forall\, f \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^n) \tag{2.27}$$

Proof. (Øksendal, 2003, Corollary 3.1.7 p. 29) □

In other words, the Itô integral is an isometry of the normed vector spaces with respect to the norms induced by the inner products.

Often, there is no analytical solution of an SDE known, so numerical approximation theory has to be initiated. To this end, let $X = (X_t)_{0\leq t\leq T}$ be an $\mathbb{R}^d$-valued stochastic process whose dynamic is described as stochastic differential equation of the form of equation (2.17). An approximation of such an $X = (X_t)_{0\leq t\leq T}$ on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ for some given time discretization

$0 = t_0 \leq \ldots \leq t_N = T$, $N \in \mathbb{N}$, is given by the piecewise constant stochastic process $Y^\Delta = (Y_t^\Delta)_{0\leq t\leq T}$

$$Y^\Delta = (Y_k \mathbb{1}_{(t_k, t_{k+1}]}(t))_{0\leq t\leq T}, \quad k = 0, \ldots, N-1, \tag{2.28}$$

where Δ denotes the maximal time step of the time discretization $\{t_0, t_1, \ldots, t_N\}$. Some examples of approximation schemes of this type will be introduced in Section 3.3. To measure the quality of such approximations, one needs the following two definitions.

Definition 2.2.20 (Definition of Strong Approximation). *Let $X = (X_t)_{0\leq t\leq T}$ be an $\mathbb{R}^d$-valued stochastic process. Then the time discretization Y^Δ is said to converge strongly to X at point T with order $\gamma > 0$ as $\Delta \downarrow 0$ if there exists a positive constant $C > 0$ such that*

$$\epsilon_S := \mathbb{E}\left[\left|X_T - Y_T^\Delta\right|\right] \leq C\Delta^\gamma \tag{2.29}$$

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

Note that the strong order of Y^Δ is developed as the error at a point $T > 0$, but for certain time step sizes Δ one observes that strong explicit schemes sometimes generate large errors on the interval $[0, T]$. This raises the following definition of weak approximation.

Definition 2.2.21 (Definition of Weak Approximation). *Let $X = (X_t)_{0\leq t\leq T}$ be an $\mathbb{R}^d$-valued stochastic process. Then Y^Δ is said to converge weakly to X with order $\beta > 0$ if there exists a positive constant $C > 0$ such that*

$$\epsilon_W := \left|\mathbb{E}\left[g(X_T)\right] - \mathbb{E}\left[g(Y_T^\Delta)\right]\right| \leq C\Delta^\beta \tag{2.30}$$

for any function $g \in C_p^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ and each $\Delta \in (0, \infty)$.

Remark 2.2.22. *$C_p^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ denotes the space of $2(\beta+1)$-continuously differentiable functions which, together with their partial derivatives of order up to $2(\beta+1)$, have polynomial growth. This means that for every $g \in C_p^{2(\beta+1)}(\mathbb{R}^d, \mathbb{R})$ there exist a constant $K > 0$ and some $r \in \mathbb{N}$, which depends on the function g, such that for all $x \in \mathbb{R}^d$ and any partial derivative $\frac{\partial^k}{\partial x^{[k]}} g(x, t)$ of order $k \leq 2(\beta+1)$ the following equation holds*

$$\left|\frac{\partial^k}{\partial x^{[k]}} g(x, t)\right| \leq K\left(1 + |x|^{2d}\right). \tag{2.31}$$

As one notices from the Definition 2.2.20, strong schemes provide approximations by single paths related to trajectories. Therefore, these schemes are used to solve problems such as hedge simulation and scenario analysis in finance. Otherwise one perceives from Definition 2.2.21 that weak schemes provide approximations of the probability measure related to the moments and in this way are appropriate for problems such as derivative pricing, expected utilities and risk measures, which will be the main applications in this thesis.

As a last point of this section, the following Law of Large Numbers is introduced as it is important in Monte-Carlo methods that are introduced and established by the seminal work of Metropolis and Ulam (1949).

Theorem 2.2.23 (Law of Large Numbers). *If $(X_n)_{n\in\mathbb{N}}$ is a sequence of stochastic independent and identically distributed random variables and $\mathbb{E}[X_1]$ exists, then*

$$\lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} X_m = \mathbb{E}[X_1] \quad \mathbb{P}-a.s.. \tag{2.32}$$

Proof. (Billingsley, 1995, Theorem 6.1 p. 85) □

Troughout this thesis $M \in \mathbb{N}$ denotes the number of Monte-Carlo simulations, realizations or trajectories. Note that for the number M of Monte-Carlo paths we get the following error estimation as in Glasserman (2004).

Corollary 2.2.24 (Monte-Carlo Error). *If $X = (X_n)_{n\in\mathbb{N}}$ is a sequence of stochastic independent and identically distributed random variables, then the Monte-Carlo error is estimated via*

$$e_M(X) \sim \frac{\hat{\sigma}_M(X)}{\sqrt{M}} \tag{2.33}$$

with variance

$$\hat{\sigma}_M(X) := \sqrt{\left(\frac{1}{M}\sum_{m=1}^{M} X_m^2\right) - \left(\frac{1}{M}\sum_{m=1}^{M} X_m\right)^2}. \tag{2.34}$$

Proof. (Jäckel, 2002, Section 2.7) □

A detailed overview on the mentioned topics can be found in Billingsley (1995), Bosq (2010), Jacod and Protter (2013), Karatzas and Shreve (2000), Karlin and Taylor (1975) and Shiryaev (1996) or Wengenroth (2008) for probability theory, stochastic calculus and Brownian motions, Cont and Tankow (2004), Karlin and Taylor (1981), Øksendal (2003) and Protter (1991) for stochastic differential equations, Chung and Williams (1990), Ikeda and Watanabe (1989) and the classical work of Kloeden and Platen (1999). For stochastic integration, approximation schemes and Monte-Carlo simulations the reader is refered to Duffy and Kienitz (2009), Glasserman (2004) or Jäckel (2002). Definitions and notations are mainly taken from Øksendal (2003) and Wengenroth (2008).

2.3 Financial Markets

On the basis of Section 2.2 some fundamental concepts of financial markets with volatil assets and their financial contracts will be explained in the following. For a more comprehensive treatment of the concepts, methods and results of mathematical modeling in finance see for instance Elliott and Kopp (2005), Föllmer and Schied (2004), Hakala and Wystup (2002), Hull (2008), Karatzas and Shreve (1998), Lamberton and Lapeyre (1996), Chapter 12 of Øksendal (2003), Shiryaev (1999), Shreve (2004) and Wilmott et al. (1993) and the references therein.

Definition 2.3.1 (Financial Market). *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space. A d-asset financial market model is an $\mathbb{F}$-adapted $d+1$-dimensional stochastic process $S_t = (S_t^0, S_t^1, \ldots, S_t^d)$ $t \in [0,T]$ of the form*

$$dS_t^0 = r(t,\omega)S_t^0 dt, \quad S_0^0 = 1 \quad \text{for } t \in [0,T] \tag{2.35}$$

called the numéraire and for $1 \leq i \leq d \in \mathbb{N}$

$$\begin{aligned} dS_t^i &= \mu_i(t,\omega)dt + \sum_{j=1}^{m} \sigma_{ij}(t,\omega)dW_t^j \\ &= \mu_i(t,\omega)dt + \sigma_i(t,\omega)dW_t, \quad S_i^0 = s_i \quad \text{for } t \in [0,T] \end{aligned} \tag{2.36}$$

called the risky assets or underlyings, where $W_t = (W_t^1, \ldots, W_t^m)$ is an m-dimensional Wiener process and σ_i is the i-th row of the volatility $d \times m$ matrix $[\sigma_{ij}]$.

To shorten notation, the initial values S_0^i, $i = 0, 1, \ldots, d$, will not be specified in detail.

A market is called to have no arbitrage if there is no possibility of gaining a positive amount of yield out of nothing almost surely. In this case, practitioners are used to say there is no "free lunch" (see Schachermayer (2008)). This is a main assumption of efficient markets.

Remark 2.3.2 (Efficient Market Hypothesis). *The following properties of financial markets are assumed to hold without exception in this thesis:*

i) The market is liquid, i.e., arbitrary amounts of assets are always available.

ii) Market participants can sell assets they do not hold. This is called short selling.

iii) It is possible to buy fractional quantities of assets.

iv) There are no transaction costs and no dividend yields

v) No arbitrage, i.e., no riskless returns.

Some assumptions may be contrary to intuition like the absence of transaction costs. Nevertheless, under these assumptions the considered financial market model can be introduced. The following Fundamental Theorem of Asset Pricing 2.3.9 translates these assumptions into mathematical language in the manner, that the absence of arbitrage opportunities is related to the existence of an equivalent martingale measure.

Definition 2.3.3 (Portfolio). *A portfolio is a measurable and $\mathbb{F}$-adapted stochastic vector process $\xi \in \mathbb{R}^{d+1}$ chosen by an investor in a d-asset financial model. Furthermore, a trading strategy is a predictable $\mathbb{R}^{d+1}$-valued process $\bar{\xi}$. A portfolio is self-financing if its changes in value are only due to changes of prices in the underlying financial model and do not require any additional capital.*

Definition 2.3.4 (Contigent Claim). *A contigent claim is a random variable C on the underlying probablity space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $0 \leq C < \infty \quad \mathbb{P} - a.s.$. A contigent claim C in a d-asset financial model is called financial derivative (financial contract or financial security) of the primary assets $S^0, S^1, \ldots, S^d$ if it is measurable with respect to the sigma-algebra $\sigma(S^0, \ldots, S^d)$.*

Such financial derivatives, like options, futures or swaps, have gained serious interest since the Chicago Board Options Exchange (CBOE) was founded in 1973, which is the first exchange to trade standardized stock options and is nowadays the world's largest options exchange. Subsequently, pioneering work of the mathematical basics of option pricing arises from the seventies, e.g. Black and Scholes (1973), Cox et al. (1979), Merton (1973) and Merton (1976). Many of the answers were obtained in a heuristic way by Bachelier (1900). Ensuing from these seminal academic introductory and much-cited publications, fundamental principles of derivative pricing are presented in the rest of this section. Aside from that, not only the volume of option markets, but also the sophistication of investments increased over the last decades as seen in the Introduction (Section 1.1). So more sophisticated and prudent models are conceived since then and presented below.

To price such options accurately, the following definitions and the Fundamental Theorem of Asset Pricing (Theorem 2.3.9) are needed.

Definition 2.3.5 (Replication Portfolio). *A contigent claim C is called replicable (attainable, hedgeable, redundant) in the financial market $(S_t)_{t\in[0,T]}$ if C can be replicated by some self-financing trading strategy $\bar{\xi} \in \mathbb{R}^{d+1}$. Such a portfolio strategy $\bar{\xi}$ is called a replicating or hedging portfolio for C.*

Definition 2.3.6 (Complete Market). *An arbitrage-free market model is called complete if every contigent claim is replicable.*

An equivalent formulation to the definition of complete markets is given in the following theorem.

Theorem 2.3.7. *An arbitrage-free market model $(S_t)_{t\in[0,T]}$ as in Definition 2.3.1 is complete if and only if the volatility matrix $\sigma(t,\omega)$ has a left inverse $\tilde{\sigma}(t,\omega)$ for almost all $(t,\omega) \in [0,T] \times \Omega$, i.e.*

$$\tilde{\sigma}(t,\omega)\sigma(t,\omega) = I_m \quad \text{for a. a.} \quad (t,\omega).$$

Proof. (Øksendal, 2003, Theorem 12.2.5 page 263f) □

Complete market models are precisely those models in which every contigent claim has a unique price.

Definition 2.3.8 (Equivalent Martingale Measure)**.** *A probability measure $\mathbb{Q}$ on some $(\Omega, \mathcal{F}_T)$ is called a martingale measure if the discounted price process X is a $\mathbb{Q}$-martingale. A martingale measure $\mathbb{P}^*$ is called an equivalent martingale measure or risk-neutral probability measure if it is equivalent to the original measure $\mathbb{P}$ on $\mathcal{F}_T$. The set of all equivalent martingale measures is denoted by $\mathcal{P}$.*

For simplicity the following theorem is presented in the case of finite discrete time as in Delbaen and Schachermayer (1994) or Schachermayer (1992).

Theorem 2.3.9 (Fundamental Theorem of Asset Pricing)**.**

i) The market model is arbitrage-free if and only if the set $\mathcal{P}$ of all equivalent martingale measures is non-empty.

ii) An arbitrage-free market model is complete if and only if there exists a unique equivalent martingale measure.

Proof.

i) (Föllmer and Schied, 2004, Theorem 5.17 page 232f)

ii) (Föllmer and Schied, 2004, Theorem 5.38 page 245f)

□

The Fundamental Theorem of Asset Pricing (Theorem 2.3.9) simply states that an arbitrage-free price of a European call option is given by the discounted expected future payoff under the equivalent martingale measure $\mathbb{Q}$. So let from now on $\mathbb{Q}$ be the equivalent martingale measure such that the discounted stock price $\mathrm{e}^{-rt} S_t$ is a martingale under $\mathbb{Q}$ and $\mathbb{E}$ be the corresponding expectation.

The choice of the financial market model which describes the dynamic of the underlying assets effects significantly the pricing of financial derivatives and option prices differs from that. So we introduce up next some stochastic asset dynamic models. For simplicity, one drops the numéraire as it is troughout this thesis given by the constant riskless rate of interest $r > 0$

$$dS_t^0 = rdt, \quad \text{for } t \in [0, T] \tag{2.37}$$

or may differ only in time

$$dS_t^0 = r(t)dt, \quad \text{for } t \in [0, T] \tag{2.38}$$

with deterministic $r(t) > 0$ for all $t \in [0, T]$.

Some discussion of the necessity of the riskless rate being positive occurs since negative interest rates are possible. Even though this is an interesting aspect of financial modeling, negative interest rates are not a part of this thesis.

Several dynamics of risky assets are described in the following financial market models:

- **Black-Scholes model.**

$$dS_t = r(t)S_t dt + \sigma(S,t)dW_t, \quad S_0, \tag{2.39}$$

where W_t is a Wiener process and $\sigma > 0$, hence, by Theorem 2.3.7 it is a complete market model.

However, it has been empirically proven, by among others, Dupire (1994) that due to its simplicity the classical Black-Scholes model (2.39) cannot properly capture the real market dynamics. The Black-Scholes model is, unfortunately, not suitable to adequately price and hedge exotic options (see Toivanen (2008)). That is the very reason why other sophisticated models are introduced and listed here.

- **Jump-diffusion model.**

$$dS_t = r(t)S_t dt + \sigma(S,t)dW_t + \gamma(S,t)JdN_t(\lambda), \quad S_0, \tag{2.40}$$

where $(N_t)_{0\leq t\leq T}$ is a Poisson process with intensity λ and random jump size J, see Cont and Tankow (2004). The processes $(W_t)_{0\leq t\leq T}$ and $(N_t)_{0\leq t\leq T}$ are stochasticly independent. The Random jump size J might have various types of probability distributions like the Meixner- or Gamma-distribution, see Schoutens (2003), but this thesis is restricted to the following two popular probability density functions by Merton (1976) and Kou (2002), respectively,

$$f_{\text{Merton}}(x) := \exp\left(\frac{1}{\sqrt{2\pi}\sigma_J}\,\mathrm{e}^{-\frac{(x-\mu_J)^2}{2\sigma_J^2}}\right) \tag{2.41}$$

$$f_{\text{Kou}}(x) := (1-p)\eta_1\,\mathrm{e}^{-\eta_1 x}\,\mathbb{1}_{\{x\geq 0\}} + p\eta_2 e^{-\eta_2 x}\mathbb{1}_{\{x<0\}} \tag{2.42}$$

Adding degrees of freedom to the Black-Scholes model (2.39) leads to the following financial market model developed by Heston (1993):

- **Stochastic volatility model.** The stochastic volatility model developed by Heston is a diffusion process with a stochastic process as diffusion coefficient.

$$\begin{aligned} dS_t &= r(t)S_t dt + \sqrt{v_t}S_t\Delta W_t^1, \quad S_0, \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}\Delta W_t^2, \quad v_0, \\ \rho dt &= \mathrm{Cov}\left[dW_t^1, dW_t^2\right], \end{aligned} \tag{2.43}$$

where $(v_t)_{0\leq t\leq\infty}$ is a mean-reverting process, the so-called Cox-Ingersoll-Ross (CIR) process in Cox et al. (1985). This is similar to the model of Bormetti et al. (2010). Several related stochastic volatility models of the term structure of interest rates are named after Black and Karasinski (1991), Hull and White (1987) or Stein and Stein (1991), respectively.

The empirical work of Bates (1996) indicates that extended financial market models are needed. Merging a stochastic volatility model with the jump aspect of (2.40) leads to the following financial market models:

- **Stochastic volatility jump-diffusion model.**

$$\begin{aligned} dS_t &= r(t)S_t dt + \sqrt{v_t}S_t \Delta W_t^1 + \gamma(S,t)JdN_t(\lambda), \quad S_0, \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}\Delta W_t^2, \quad v_0, \\ \rho dt &= \text{Cov}\left[dW_t^1, dW_t^2\right], \end{aligned} \tag{2.44}$$

- **Stochastic volatility jump-diffusion model with a stochastic jump intensity rate** The following model was developed by Fang (2000):

$$\begin{aligned} dS_t &= r(t)S_t dt + \sqrt{v_t}S_t \Delta W_t^1 + \gamma(S,t,J)dN_t(\lambda_t), \quad S_0, \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}\Delta W_t^2, \quad v_0, \\ d\lambda_t &= \kappa_\lambda(\theta_\lambda - \lambda_t)dt + \sigma_\lambda\sqrt{\lambda_t}\Delta W_t^3, \quad \lambda_0, \end{aligned} \tag{2.45}$$

- **Jump diffusion model with both price and volatility jumps** The following model was developed by Duffie et al. (2000):

$$\begin{aligned} dS_t &= r(t)S_t dt + \sqrt{v_t}S_t \Delta W_t^1 + \gamma(S,t,J)dN_t(\lambda), \quad S_0, \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}\Delta W_t^2 + \gamma(v,t)J_v dN_t(\lambda_v), \quad v_0, \\ \rho dt &= \text{Cov}\left[dW_t^1, dW_t^2\right], \end{aligned} \tag{2.46}$$

There is a broad field of different financial market models such as:

- **Model of Pricing of Catastrophe Insurance Options.** The stochastic process $(L_t)_{L\in[0,T]}$ representing the loss index of an insurance is modelled as in Biagini et al. (2008):

$$L_t = \sum_{j=1}^{N_t} Y_j A_{t-\tau_j}^j,$$

  where $(N_t)_{0\le t\le T}$ is a Poisson process with intensity $\lambda > 0$ and jump times τ_j, $j = 1, 2, \ldots$, that are stopping times. The positive independent and identically distributed random varibales Y_j are the initial loss estimates and $(A_t^j)_{0\le t\le T}$ are the reestimation martingales.

- **Model of Pricing of Electricity Options.** The non-storable nature of electricity, the infrequent consumption and the seasonal fluctuations cause pricing with a Markov regime-switching model (MRS) equipped with a deterministic seasonal component as in Janczura (2014)

$$P_t = g(t) + X_t,$$

  where g is the seasonal component and the dynamic of X_t is given by a 3-regime model and its base regime is given by the mean-reverting Vasicek model (Vasicek (1977))

$$X_{t,b} = (\alpha - \beta X_{t,b})dt + \sigma_b dW_t, \tag{2.47}$$

  as in Cartea et al. (2009).

The volume traded on the Power Derivatives Market of European Energy Exchange (EEX) amounted to 103.3 terawatt hours (TWh) in December 2013[1].

Payoff functions are the mathematical tool to handle financial contracts of options like "*a negotiable contract in which the bearer, for a certain sum of money (premium or option value), gives the buyer the right, not the liability, to demand within a specified time (maturity or expiration time) to purchase (call option) or sale (put option) by the bearer of a specified number of shares of a stock (underlying) at a fixed price (strike or exercise price)*" as in Stoll (1969). Therefore, often-used payoff functions are listed below.

This thesis will be restricted to the following financial contracts and their payoff functions as in Föllmer and Schied (2004). Here, K denotes the fixed strike price, T the fixed expiration date and the dynamics of the underlyings are driven by some market model $(S_t)_{t\in[0,T]}$. $\mathbb{E}[\cdot]$ is given by the risk-neutral equivalent martingale measure $\mathbb{Q}$, which represents the discount factor and may be evaluated by the famous Girsanov Theorem (Föllmer and Schied, 2004, Theorem 10.25).

- **European call and put options**

$$C(T, S_T) = \mathbb{E}\left[\max(S_T - K, 0)\right] = \mathbb{E}\left[\pi(S_T, K)\right], \tag{2.48}$$
$$P(T, S_T) = \mathbb{E}\left[\max(K - S_T, 0)\right] = \mathbb{E}\left[\pi(K, S_T)\right]. \tag{2.49}$$

- **European digital option**

$$C_{\text{digital}}(T, S_T) = P\mathbb{E}\left[\mathbb{1}_{\{S_T \geq K\}}\right]. \tag{2.50}$$

- **American Options**
  To price American Options, which may be exercised at any time up to expiration, one needs to solve an optimal stopping problem in the set of all stopping times in $[t, T]$, denoted by $\mathfrak{T}_t := \{\tau \in \mathfrak{T} | \tau \geq t\}$ as in Myneni (1992) or Christensen (2014), and to evaluate a conditional expectation as follow s

$$C_{\text{American}}(T, S) = \sup_{\tau\in\mathfrak{T}_t} \mathbb{E}\left[\max(S_\tau - K, 0)|\mathcal{F}_t\right], \tag{2.51}$$
$$P_{\text{American}}(T, S) = \sup_{\tau\in\mathfrak{T}_t} \mathbb{E}\left[\max(K - S_\tau, 0)|\mathcal{F}_t\right]. \tag{2.52}$$

This European and American option styles are called plain vanilla options. Now, some non-vanilla or exotic exercise rights are presented.

[1] Data collected from http://www.eurexchange.com/exchange-en/about-us/news/765426/.

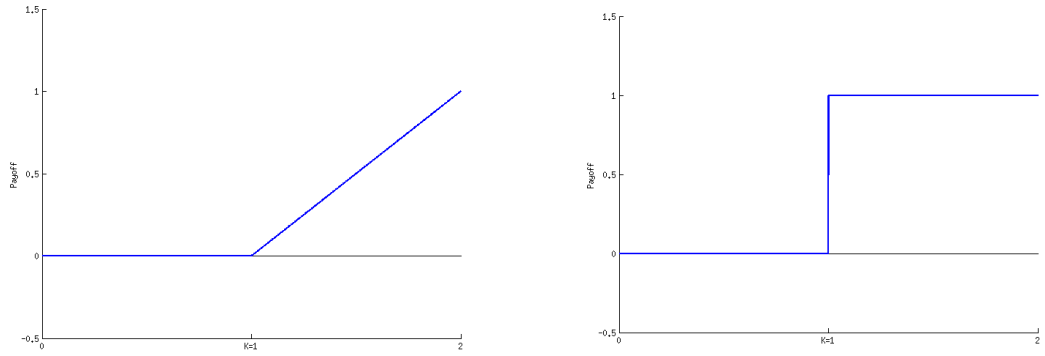

Figure 2.3: European call (left) and digital (right) payoff functions and gains.

- **European barrier (call) options**

$$C_{\text{up-and-out}}(T, S_T) = \mathbb{E}\left[\max(S_T - K, 0)\mathbb{1}_{\{S_t<B|0\leq t\leq T\}}\right] \tag{2.53}$$
$$= \mathbb{E}\left[\pi(S_T, K)\mathbb{1}_{\{S_t<B|0\leq t\leq T\}}\right], \tag{2.54}$$
$$C_{\text{up-and-in}}(T, S_T) = \mathbb{E}\left[\max(S_T - K, 0)\mathbb{1}_{\{S_t>B|\text{for not less than one } t\in[0,T]\}}\right] \tag{2.55}$$
$$= \mathbb{E}\left[\pi(S_T, K)\mathbb{1}_{\{S_t<B|\text{for not less than one } t\in[0,T]\}}\right], \tag{2.56}$$

  where $\mathbb{1}_A$ is the indicator function of a set A, being 1 when A occurs, and zero otherwise. The simplest way to price barrier options is to use a static replicating portfolio of vanilla options, see Carr et al. (1998) or Maruhn and Sachs (2006).

- **Lookback option (with floating strike)**

$$C_{\min}(T, S_T) = \mathbb{E}\left[\max(S_T - \min_{t\in[0,T]} S_t, 0)\right] = S_T - \min_{t\in[0,T]} S_t, \tag{2.57}$$
$$P_{\max}(T, S_T) = \mathbb{E}\left[\max(\max_{t\in[0,T]} S_t - S_T, 0)\right] = \max_{t\in[0,T]} S_t - S_T. \tag{2.58}$$

- **Asian or average options**

$$C_{\text{Asian}}(T, S) = \mathbb{E}\left[\max(S_T - A(T), 0)\right] = \mathbb{E}\left[\pi(A(T), K)\right], \tag{2.59}$$
$$P_{\text{Asian}}(T, S) = \mathbb{E}\left[\max(A(T) - S_T, 0)\right] = \mathbb{E}\left[\pi(K, A(T))\right]. \tag{2.60}$$

  There are numerous styles of Asian option depending on how to evaluate the average $A(T)$. The two most popular are the arithmetic average

$$\bar{A}(T) = \frac{1}{n}\sum_{i=1}^{n} S_{t_i} \tag{2.61}$$

and the geometric average

$$\hat{A}(T) = \left(\prod_{i=1}^{n} S_{t_i} \right)^{\frac{1}{n}}. \tag{2.62}$$

The following connection between these two averages exists

$$\hat{A}_T = \mathrm{e}^{\ln(\hat{A}_T)} = \exp\left(\frac{1}{n} \sum_{i=1}^{n} \ln\left(S_{t_i}\right) \right).$$

- **Basket option**

$$\hat{B}_t = \sum_{j=1}^{d} S_t^j. \tag{2.63}$$

The linear combination $\hat{B}$ of d-assets is called basket. A basket option is a call/put option or some path dependent option style, e.g. Asian, on such a basket.

This is by no means an exhaustive list of exotic options. The number of different styles of such options grows rapidly. As an example Bernhart et al. (2012), Bender, Schweizer and Zhuo (2013) and Bender and Dokuchaev (2014) introduce several methods for pricing swing options via backward stochastic differential equations, which are presented later in Section 6.2. Other methods to price such multiple exercise options, which is a hot topic, are given by Bender (2011) and Bender, Schoenmakers and Zhang (2013).

By the same token, options can be classified according to the underlying assets as well. Since the underlying could be all kinds of bonds, commodities, equities, futures, indices, private equities or real estates, for instance, Busch et al. (2013).

The simplest way to hedge plain vanilla options is given by the Put Call Parity, which plays an important role in the next Chapter 3.

Theorem 2.3.10 (Put Call Parity)**.**

$$C(t) - P(t) = S_t - \mathrm{e}^{-r(T-t)} K$$

Proof. (Stoll, 1969, (5) page 806) □

To summarize, the key requirements for utilizing options is calculating their fair value. Finding ways to efficiently solve this pricing problem or hedge such claims has been an active field of research in financial mathematics for more than fourty years and it continues to be a focus of modern financial engineering. This completes the introduction to option pricing and hedging.

2.4 Optimization in Function Spaces

Subsequently, important fundamentals of numerical optimization, optimization in an infinite-dimensional setting, inverse problems and necessary optimality conditions will be explained in this section. The notations and definitions are mostly taken from Bertsekas (1999), Bonnans and Shapiro (2000), Bonnans et al. (2006), Clarke (2013), Engl et al. (1996), Gill and Murray (1974), Hinze (2009), Kantorovich and Akilov (1982), Kelley (1995), Kosmol and Müller-Wichards (2014), Kushner (1972), Luenberger (1998), Nocedal and Wright (1999) and Zeidler (1995). The last books also provides a more detailed overview on different topics of optimization.

This section starts with some generalities on the unconstrained optimization problem

$$\min \phi(x), \tag{2.64}$$

where $\phi : X \to \mathbb{R}$ satisfies some smoothness conditions on the Banach space X.

Definition 2.4.1. *Let X be a Banach space, $U \subset X$ a convex subset and $\phi : U \to \mathbb{R}$ a functional. An element $x^* \in U$ is a minimal solution or local minimum of the functional ϕ, if and only if*

$$\phi(x^*) \leq \phi(x) \quad \forall\ x \in U. \tag{2.65}$$

The following theorem expresses the necessary optimality condition of the problem (2.64).

Theorem 2.4.2 (First Order Optimality Condition)**.** *If x^* is a minimal solution of ϕ and ϕ is continuously differentiable in an open neighborhood of U_{x^*}, x^* contained in U, then $\phi'(x^*) = 0$.*

Proof. (Kosmol and Müller-Wichards, 2014, Theorem 3.4.1) □

Consider the following abstract constrained optimization problem:

Let X, Y, U be Banach spaces, where X is the state space, Y the range space for the equality constraint and U is the space of parameters or controls.
In this setting let the objective function Φ and the constraint map g be given by

$$\Phi : X \times U \to \mathbb{R}, \qquad g : X \times U \to Y.$$

The following smoothness assumption on these maps is crucial:

Assumption 2.4.3. *Let the maps Φ and g be continuously Fréchet-differentiable on the space $X \times U$. Let g' denote the Fréchet-derivative with respect to both variables (x, u) whereas g_x and g_u are the partial derivatives with respect to states x and controls u, respectively. Let the partial derivatives g_x and g_u exist.*

The optimization problem under consideration is given by

$$\text{Minimize} \quad \Phi(x, u) \quad \text{over} \quad (x, u) \in X \times U \quad \text{such that} \quad g(x, u) = 0. \tag{2.66}$$

For applications where the constraint g is given by a differential equation for the variable x and the variable u is merely a control or design parameter, the following assumption is often satisfied. Note that g is mainly given by a stochastic differential equation and its approximation, respectively.

Assumption 2.4.4. *Assume that at $(x,u) \in X \times U$ the partial Fréchet-derivative*

$$g_x \text{ is surjective and an invertible map.}$$

Let Assumptions 2.4.3 and 2.4.4 be true at a point $(x_*, u_*) \in X \times U$. Then there exist - mainly due to the Implicit Function Theorem 2.1.12 - neighborhoods $B_X \subset X$ of x_* and $B_U \subset U$ of u_* and a Fréchet-differentiable function $s : B_U \to B_X$ such that

$$g(s(u), u) = 0 \quad \text{and} \quad g_x(s(u), u)s'(u) = -g_u(s(u), u). \tag{2.67}$$

Equation (2.67) is called the sensitivity equations and will play a crucial role in this thesis. Although the constraint qualification would hold if $g'(x,u)$ is only surjective, but the additional invertibility condition for $g_x(x,u)$ yields the applicability of the Implicit Function Theorem 2.1.12. Hence, Assumptions 2.4.3 and 2.4.4 have to be true at all points $(x,u) \in B_X \times B_U$. The map $s : B_U \to B_X$ is a solution map for the equality constraint. Under the assumptions of Theorem 2.1.12, the optimization problem (2.66) can be reformulated as a reduced unconstrained optimization problem.

Theorem 2.4.5. *Assume that $(x_*, u_*) \in X \times U$ is a local minimum of (2.66) and Assumptions 2.4.3 and 2.4.4 hold at (x_*, u_*). Then u_* is also a local minimum of the unconstrained optimization problem*

$$\textit{Minimize} \quad \phi(u) = \Phi(s(u), u) \quad \textit{over} \quad u \in \tilde{B}_U \tag{2.68}$$

for some neighborhood $\tilde{B}_U \in U$ and the map $\phi : \tilde{B}_U \to \mathbb{R}$.

Proof. (Nocedal and Wright, 1999, Chapter 12) □

In order to compute the Fréchet-derivative of ϕ, one can use the differentiability of s given by the Implicit Function Theorem 2.1.12 and obtain:

Lemma 2.4.6 (Gradient Representation)**.** *Let Assumptions 2.4.3 and 2.4.4 be true at a point $(x,u) \in X \times U$. Then*

$$\phi'(u)\Delta u = \Phi_x(s(u), u)\ s'(u)\Delta u + \Phi_u(s(u), u)\Delta u, \qquad \Delta u \in U, \tag{2.69}$$

where the Fréchet-derivative $s'(u)$ is defined through the sensitivity equations (2.67).

Proof. (Kantorovich and Akilov, 1982, Chapter VI) □

Note that (2.69) yields an explicit representation of the linear functional $c'(u) \in U^*$ due to the Riesz Representation Theorem 2.1.5. If U is a Hilbert space, a unique gradient

representation of the derivative $\phi'(u)$ exists due to Theorem 2.1.5. This changes if one rewrites the equation (2.69) in the following way. Note that the adjoint

$$s'(u)^* : X^* \to U^*$$

is defined on X^* and $\Phi_x \in X^*$, so that one obtains the following corollary which extends Lemma 2.4.6 to the uniqueness of the explicit gradient representation

Corollary 2.4.7. *Let Assumptions 2.4.3 and 2.4.4 be true at a point $(x,u) \in X \times U$. Then*

$$\phi'(u) = s'(u)^* \ \Phi_x(s(u),u) + \Phi_u(s(u),u) \in U^* \tag{2.70}$$

where the Fréchet-derivative $s'(u)$ is defined through (2.67).

The following theorem summarizes the fundamentals of the so-called adjoint approach in optimization problems. For further details, the reader is reffered to (Kantorovich and Akilov, 1982, Chapter XII of PART II: Functional Equations).

Theorem 2.4.8 (Derivatives with Adjoints)**.** *Let Assumptions 2.4.3 and 2.4.4 be true at a point $(x,u) \in X \times U$. Then*

$$\phi'(u) = g_u(s(u),u)^* l + \Phi_u(s(u),u) \in U^* \tag{2.71}$$

where $l \in Y^*$ is the unique solution of the adjoint equation

$$g_x(s(u),u)^* l = -\Phi_x(s(u),u), \tag{2.72}$$

equivalently, $l \in Y^*$ is uniquely defined through

$$l(y) = -\Phi_x(s(u),u) g_x(s(u),u)^{-1} y \quad \textit{for all } y \in Y. \tag{2.73}$$

Proof. The statement follows from rearranging the terms in equation (2.69)

$$\phi'(u)\Delta u = [s'(u)^* \Phi_x(s(u),u)]\Delta u + \Phi_u(s(u),u)\Delta u.$$

Note that the map $s'(u)$ in (2.69) or its dual $s'(u)^*$ in (2.70) is not needed in its full generality, but only $s'(u)^*\Phi_x(s(u),u)$.

From the Implicit Function Theorem 2.1.12 one knows that

$$s'(u) = -g_x(s(u),u)^{-1} g_u(s(u),u) \ \text{ and } \ s'(u)^* = -g_u(s(u),u)^* (g_x(s(u),u)^*)^{-1}.$$

Hence,

$$s'(u)^*\Phi_x(s(u),u) = -g_u(s(u),u)^* (g_x(s(u),u)^*)^{-1} \Phi_x(s(u),u) = g_u(s(u),u)^* l,$$

where $l \in Y^*$ solves the adjoint equation

$$g_x(s(u),u)^* l = -\Phi_x(s(u),u).$$

Similarly, one has

$$-\Phi_x(s(u),u)\Delta x = [g_x(s(u),u)^* l]\Delta x = l(g_x(s(u),u)\Delta x)$$

and with the surjectivity and invertibility of g_x

$$l(y) = -\Phi_x(s(u),u)g_x(s(u),u)^{-1}y.$$

This completes the proof and establishes the equations (2.71) and (2.72). □

The optimization problem, subject to a finite number of equality constraints is given by

$$\text{Minimize} \quad \Phi(x) \quad \text{over} \quad x \in X \quad \text{such that} \quad g_i(x) = 0,\ i = 1,...m. \tag{2.74}$$

To solve such problems one can uses Lagrange multipliers (cf. Bertsekas (1999)).

Definition 2.4.9 (Lagrange Function)**.** *Consider objective function Φ and constraints g from* (2.74)*. The Langrange function $\mathcal{L} : X \times Y \mapsto \mathbb{R}$ is given by*

$$\mathcal{L}(x,\lambda) := \Phi(x) + \sum_{i=1}^{m} \lambda_i g_i(x). \tag{2.75}$$

Theorem 2.4.10 (Karush-Kuhn-Tucker)**.** *Assume that Φ and g are continuously differentiable at some local minimum x^* of problem* (2.74)*. Additionally, let the linear independence constraint qualification (LICQ) hold at this x^*, e.g., the set $\{\nabla g_i(x^*)\}$ is linearly independent. Then there exists so-called Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ such that the following so called Karush-Kuhn-Tucker (KKT) conditions are satiesfied*

$$\begin{aligned} \nabla_x \mathcal{L}(x^*,\lambda^*) &= 0 \\ g_i(x^*) &= 0 \quad i = 1,...,m \\ \lambda_i^* &\geq 0 \quad i = 1,...,m. \end{aligned}$$

Proof. (Nocedal and Wright, 1999, Section 12.5) □

Theorem 2.4.11 (Lagrange Multiplier (Infinite-dimensional Case))**.** *If one forms the Lagrangian in the sense of* (2.66)

$$\mathcal{L}(x,u,l) = \Phi(x,u) + l(g(x,u)),$$

then the Lagrange multiplier l is defined by the equation $\mathcal{L}_x(s(u),u,l) = 0$ and the derivative $\phi'(u)$ is given by $\mathcal{L}_u(s(u),u,l)$.

Proof. (Kosmol and Müller-Wichards, 2014, Theorem 2.14.5) □

From the previous Theorem 2.4.8 it is obvious that the Lagrange multiplier l and the derivative $\phi'(u)$ are defined as in Theorem 2.4.11.

The resulting equations of the adjoint approach of Theorem 2.4.8 appear from the solution of the corresponding unrestricted optimization problem based on the Langrange function

$\mathcal{L} : X \times U \times Y \to \mathbb{R}$

$$\mathcal{L}(x, u, l) := \Phi(x, u) + l(g(x, u)) \tag{2.76}$$

which is given by

$$\text{Minimize} \quad \mathcal{L}(x, u, l) \quad \text{for} \quad (x, u, l) \in X \times U \times Y. \tag{2.77}$$

Adjoint equations could be interpreted as Lagrange multipliers. This issue is characterized in detail for instance by Alt (1990) or Noack and Walther (2007).

In order to evaluate conditional expectation such as in Property (2.9) one need following results on the uniqueness of best approximations.

Theorem 2.4.12 (Uniqueness of best approximation)**.** *Let S be a subspace of an Hilbert space H and $x \in H$. An element $y \in S$ is said to be the best approximation to x from S if*

$$||x - y|| \leq ||x - z|| \quad \forall\ z \in S.$$

If a best approximation of x from S exists, then it is unique and it is called the projection of x onto subspace S.

Proof. (Debnath and Mikusinski, 1998, Theorem 9.7.1) □

Chapter 3

Calibration Problem

Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind.

— George E. P. Box

Empirical Model-Building and Response Surfaces, Box and Draper (1987)

3.1 Motivation

The constant volatility σ in the Black-Scholes model (2.39) constitutes a one-to-one connection (Käbe, 2010, Lemma 2.23) to the call option price C given by the closed form solution developed by Black and Scholes (1973)

$$C(S,t) = S\,\mathcal{N}(d_1) - K\,\mathrm{e}^{-rt}\,\mathcal{N}(d_2), \quad S > 0, \quad t \in [0,T], \tag{3.1}$$

where $\mathcal{N}$ is the cumultative distribution function, abbreviated as cdf, of the standard normal distribution

$$\mathcal{N}(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d} \mathrm{e}^{-\frac{t^2}{2}}\,dt$$

and

$$d_1 = \frac{\ln(S/K) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}, \quad d_2 = d_1 - \sigma\sqrt{T-t}.$$

But real liquid market data, as shown in Figure 3.1, shows something different. Graphing implied volatilities against strike prices for a given maturity close to exercise results in a skewed curve instead of the expected flat curve. Therefore, one volatility corresponds to several call prices in real market data, which contradicts the bijectivity between option price C and volatility σ in the Black-Scholes model. This phenomenon is vividly called the volatility smile which was first explained by Dupire (1994).

Since Boyle (1977) proposed the use of Monte-Carlo methods for estimating option prices, there appear to be several methods which fit the volatility smile to market model prices. After suggesting local volatility by Dupire (1994) or local volatility with permitted jumps by Andersen and Andreasen (2000), local stochastic volatility models were developed by Heston (1993), Hull and White (1987), Stein and Stein (1991) or Bormetti et al. (2010) and have been extended to jump diffusion by Lipton (2002). According to Gatheral (2006), all these

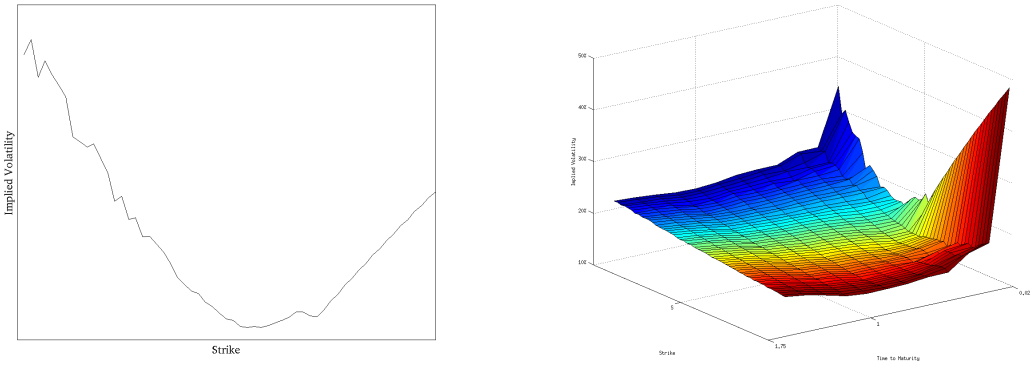

Figure 3.1: Implied volatility of the DAX in March 2014 as volatility smile (left) and volatility surface (right).

stochastic volatility models generate roughly the same shape of implied volatilities. Nevertheless, all these market models contain parameters which have to be adjusted in such a way that the model prices and their market dynamics coincide with the real market prices. Practitioners need to adapt their market models to the current market situation in a timely manner, i.e., the models have to be calibrated promptly to a set of liquidly traded standard instruments like plain vanilla options. This presents the main issue of the calibration problem in this thesis.

Modelling the volatility smile to achieve better pricing and hedging exotic options is an active area of research in quantitative finance. A detailed overview can be found in the seminal survey by Lörx and Sachs (2012) or the contemporary work by Sachs and Schneider (2014).

3.2 Calibration Problem of Financial Market Models

Problems of calibrating a financial market model to prices of liquidly traded instruments arise. SDEs play a vital role in these models. Hence, the main goal in this section is to gain an appropriate and inexpensive method of calibrating such models with side conditions to include the solution of the SDEs. Fitting a model to the implied risk-neutral distribution is the most efficient way of ensuring consistency with observed market prices. This consistency with the observed market dynamics implies stable model parameters and hedges (Käbe et al. (2009)). Without loss of generality, this thesis will focus on the calibration of an equity-type stock price model to European call options (2.48). The Notations and specifications are properly adopted from Käbe (2010). Nevertheless, the adjoint-based technique which will later be developed and presented in Chapter 4 is so common, that transferring this method onto other financial, biological or geophysical models will be facile.

For this purpose, let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and $(W^1, \ldots, W^m)$ be an m-dimensional Wiener process under an equivalent martingale measure $\mathbb{Q}$. One may be

interested in calibrating the parameters $x \in \mathbb{R}^P$ of a model, described by the d-dimensional system of stochastic differential equations in the sense of Käbe et al. (2009)

$$dX_t(x) = a(x, X_t(x))dt + b(x, X_t(x))dW_t \tag{3.2}$$

with initial value $X_0 \in (0, \infty)$ and time horizont $0 \leq t \leq T$. Here, $W_t = (W_t^1, \ldots, W_t^m)^\top$ is the vector of Wiener processes, drift $a : \mathbb{R}^P \times \mathbb{R}^d \to \mathbb{R}^d$ and volatility $b : \mathbb{R}^P \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ are functions satisfying the usual regularity conditions of Theorem 2.2.14 such that a solution of the stochastic differential equations (3.2) exists. The expression $a(x, X_t(x))dt$ denotes the componentwise integration $a^l(x, X_t(x))dt$ for $l = 1, \ldots, d$ and $b(x, X_t(x))dW_t$ is to be understood as

$$\sum_{\nu=1}^{m} b^{l,\nu}(x, X_t(x))dW_t^\nu, \quad l = 1, \ldots, d$$

as in Definition 2.3.1. Besides, the stock price S_t is identified in the following as the first component of the d-dimensional stochastic process $(Y_t)_{t\in[0,T]}$ as in Käbe et al. (2009), i.e.,

$$[X_t]_{l=1}^d = [S_t, X_t^2, ..., X_t^d]^\top, \quad t \in [0, T]. \tag{3.3}$$

In the settings (3.2), (3.3), the Fundamental Theorem of Asset Pricing 2.3.9 states that the arbitrage-free prices of the European call options are given by their discounted expected future payoffs under the equivalent martingale measure $\mathbb{Q}$

$$C^i(x) = \mathrm{e}^{-rT_i}\, \mathbb{E}_{\mathbb{Q}}\left[\max(S_{T_i}(x) - K_i, 0)\right], \tag{3.4}$$

where $\max(S_{T_i}(x) - K_i, 0)$ is assumed to be $\mathbb{Q}$-integrable $\forall\, i = 1, \ldots, I$ and $x \in X$. The properties of the model price function C^i will be decisive for the calibration procedure. In most applications the map $C^i : \mathbb{R}^P \to \mathbb{R}$ is continuously differentiable since the integral operator smooths out the non-differentiability of the maximum function, hence this thesis is focused on European call options.

The problem of calibrating the model prices (3.4) to the observed market prices C^i_{obs} is given by the following nonlinear least squares problem established by Käbe (2010)

$$\begin{aligned} &\min_{x\in X} f(x) := \tfrac{1}{2}\sum_{i=1}^{I}\left(C^i(x) - C^i_{\text{obs}}\right)^2 \\ \text{where}\quad &C^i(x) = \mathrm{e}^{-rT_i}\,\mathbb{E}_{\mathbb{Q}}\left[\pi(S_{T_i}(x) - K_i)\right] \\ \text{s.t.}\quad &dX_t(x) = a(x, X_t(x))dt + b(x, X_t(x))dW_t,\ X_0 > 0 \\ &0 \leq t \leq T, \quad T := \max_{i=1,\ldots,I} T_i, \end{aligned} \tag{P}$$

where $\pi(\xi) := \max(\xi, 0)$, and $X \subset \mathbb{R}^P$ is a suitable convex and compact set which for example may result from imposing box constraints $l_p \leq x_p \leq u_p$, $p = 1, ..., P$ on the model parameters x. Additional constraints may occur. For instance Feller (1951) established a method that is used to proof that the Cox-Ingersoll-Ross volatility process of the Heston model (2.43) stays

positive if the following condition (known as the Feller condition) is fulfilled:

$$2\kappa\theta \geq \sigma^2. \tag{3.5}$$

In this thesis an approximation of problem (P) is assumed via Monte Carlo in combination with an Euler-Maruyama scheme, Milstein-scheme or a predictor-corrector scheme of the SDEs (3.2) in time direction with step size Δt_n in the n-th step and with M as the amount of simulated trajectories. These stochastic schemes will be presented in the following Section 3.3 in detail.

3.3 Approximation schemes

Frequently stochastic differential equations (SDEs) describe the dynamics of financial and economic models. So the approximation of these dynamics is a vital issue in financial engineering. Additionally an appropriate approximation reduces expeditiously the approximation error whether strong or weak. Heretofore the numerical validations of the analytical strong convergence order of several approximation schemes are scarce in literature with the exception of the papers by Schmitz Abe and Shaw (2005), Abdulle et al. (2012) or Abdulle et al. (2014). So the results of this thesis and Groß (2009) show numerically that predictor-corrector schemes are adequate schemes.

Nevertheless, in the sequel an overview of the still rather limited literature on improved predictor-corrector schemes especially strong predictor-corrector methods is given. In Kloeden and Platen (1999) predictor-corrector schemes have been proposed as weak discrete time Monte-Carlo approximations. However, the results are limited to SDEs driven by Wiener processes. Bruti-Liberati and Platen (2005), Bruti-Liberati and Platen (2006a) first articulated predictor-corrector schemes for weak approximations in a jump driven framework. Weak convergence theorems are seldom, but some numerical results are presented. Furthermore for the strong discrete approximation of SDEs predictor-corrector schemes have been missing so far in the literature except for a paper by Bruti-Liberati and Platen (2008). A strong convergence theorem of the predictor-corrector Euler method in case of a Black-Scholes SDE is presented in this article. Also quite a few results of numerical stability (e.g. mean-square, state or asymptotic stability) using improved Monte-Carlo approximations in finance and other areas are presented by Bruti-Liberati and Platen (2008), Buckwar and Sickenberger (2011), Chalmers and Higham (2008), Li et al. (2013), Niu and Zhang (2012) and Platen and Shi (2008). Altough predictor-corrector schemes are applicable in financial engineering, their use for option pricing is still limited. In this chapter an adjoint-based technique to solve calibration problems of financial models similar to Käbe et al. (2009) is developed.

In conclusion some numerical results of the effort to implement a predictor-corrector scheme versus Euler-Maruyama are presented in the later Chapter 5.1. Numerical results about the benefit of the adjoint-technique are presented in Section 5.2.

3.3.1 Weak and Strong Approximations

In the following, different kinds of approximation schemes are considered in the one-dimensional case $d = 1$. Y_n denotes the approximation of X_{t_n} for some time discretization $0 = t_0 \leq \ldots \leq t_N = T$, $N \in I\!N$ in the sense of equation (2.28). The simplest stochastic scheme, for which $d = m = 1$ is fulfilled, is the Euler-Maruyama scheme which has the following form

$$Y_{n+1} = Y_n + a\Delta t_n + b\Delta W_n, \tag{3.6}$$

where $a = a(t_n, Y_n)$, $b = b(t_n, Y_n)$ and $\Delta t_n = t_{n+1} - t_n$ is the length of the time discretization interval. In conclusion

$$\Delta W_n = W_{t_{n+1}} - W_{t_n} \sim N(0, \Delta t_n)$$

is the nth-increment of the Wiener process on the interval $[t_n, t_{n+1}]$. This scheme has a well-known order of strong convergence $\gamma = 0.5$ and weak order $\beta = 1$, with reference to Kloeden and Platen (1999) (cf. Theorem 10.2.2 and Theorem 14.1.5).

For the more general case of $d, m > 1$ the ith component of the multi-dimensional Euler-Maruyama approximation of (2.19) is given by

$$Y^i_{n+1} = Y^i_n + a^i \Delta t_n + \sum_{\nu=1}^{m} b^{i,\nu} \Delta W^\nu_n, \quad i = 1, \ldots, d, \tag{3.7}$$

where $\Delta W^{\nu_1}_n$ and $\Delta W^{\nu_2}_n$ are independent for $\nu_1 \neq \nu_2$. In what follows, the simple formulation (3.6) will be written instead of the multi-dimensional formulation (3.7) when no confusion can arise.

Such a scheme together with the Monte Carlo method to evaluate the expectation leads to the following discretized optimization problem in the case of obtaining Euler-Maruyama scheme (3.6)

$$\begin{aligned} &\min_{x \in X} \hat{f}_{M,\Delta t}(x) := \tfrac{1}{2} \sum_{i=1}^{I} \left(\hat{C}^i_{M,\Delta t}(x) - C^i_{\text{obs}} \right)^2 \\ &\text{where} \quad \hat{C}^i_{M,\Delta t}(x) := \mathrm{e}^{-rT_i} \tfrac{1}{M} \textstyle\sum_{m=1}^{M} \left(\pi(s^m_{N_i} - K_i) \right) \\ &\text{s.t.} \qquad y^m_{n+1} = y^m_n + a(x, y^m_n)\Delta t_n + b(x, y^m_n)\Delta W^m_n, \\ &\qquad\qquad m = 1, \ldots, M, \quad n = 1, \ldots, N, \end{aligned} \tag{$\mathrm{P}^{EM}_{\mathrm{M},\Delta t}$}$$

with $s^m_{N_i}(x)$ defined as the first component of $y^m_{N_i}$, $i = 1, \ldots, I$, and $N = T/\Delta t$.

Next, an improved scheme proposed by Milstein (1975) is presented, which has an order of strong and weak convergence $\gamma = \beta = 1.0$ (Kloeden and Platen, 1999, Theorem 10.3.5 and Theorem 14.1.5). Its Itô formulation is given by

$$Y_{n+1} = Y_n + a\Delta t_n + b\Delta W_n + \frac{bb'}{2} \left((\Delta W_n)^2 - \Delta \right) \tag{3.8}$$

where b' describes the spatial derivatives $b'(t, Y) = \frac{\partial}{\partial Y} b(t, Y)$.

For the multi-dimensional Milstein scheme the reader is reffered to Section 10.3 of Kloeden and Platen (1999), for its implementation to Schmitz Abe (2010) and Schmitz Abe (2011).

3.3.2 Stochastic Predictor-Corrector Schemes

In a next step the formulation of a predictor-corrector scheme is given which is similar to Bruti-Liberati and Platen (2005) and Bruti-Liberati and Platen (2006b), whose order of strong or weak convergence depends highly on the chosen predictor and the chosen corrector. In order to show how a predictor-corrector scheme works, a so-called predictor is calculated at the nth time step by using an explicit approximation scheme. Then, this predicted value Y_{n+1}^- is J times improved by corrector steps, wich is an implicit scheme. Subsequently, using this predicted value and $(J-1)$th corrected value respectively instead of the unknown value to get the final computed solution Y_{n+1}. So the idea behind the predictor-corrector scheme is to use an appropriate combination of an explicit scheme and an implicit scheme to obtain a convergence characteristic.

Here and subsequently, the formulation of the Euler-Maruyama (3.6) or Milstein (3.8) scheme will be applied to calculate the predictor step

$$Y_{n+1}^- = Y_n + a\Delta t_n + b\Delta W_n + \alpha\frac{bb'}{2}\left((\Delta W_n)^2 - \Delta t_n\right).$$

The parameter $\alpha \in \{0,1\}$ indicates if the predictor is either Euler-Maruyama or Milstein scheme. Then the family of predictor-corrector schemes with strong order is given by the corrector

$$\begin{aligned} Y_{n+1} = Y_n &+ \left(\eta\underline{a}(t_{n+1}, Y_{n+1}^-) + (1-\eta)\underline{a}(t_n, Y_n)\right)\Delta t_n \\ &+ \left(\vartheta b(t_{n+1}, Y_{n+1}^-) + (1-\vartheta)b(t_n, Y_n)\right)\Delta W_n, \end{aligned} \tag{3.9}$$

where the corrected drift function is

$$\underline{a} = a - \vartheta bb'$$

and the parameters $\eta, \vartheta \in [0,1]$ are called the degree of implicitness in the drift and the diffusion coefficients, respectively. The implicit corrector can be seen as fixed-point iteration (see Theorem 2.1.8).

Remark 3.3.1. *1. For the case $\eta = \vartheta = 0$ and $\alpha = 0$ one obtains the explicit Euler-Maruyama scheme.*

2. For the case $\eta = 1$, $\vartheta = 0$ and $\alpha = 0$ one obtains the implicit Euler-Maruyama scheme.

3. For the case $\eta = \vartheta = 0$ and $\alpha = 1$ one obtains the Milstein scheme.

4. For the case $\eta = \vartheta = 1/2$ and $\alpha = 0$ one obtains the stochastic Heun scheme.

5. For the case $\eta = \vartheta = 1$ and $\alpha = 0$ one obtains the fully implicit PCEM-method.

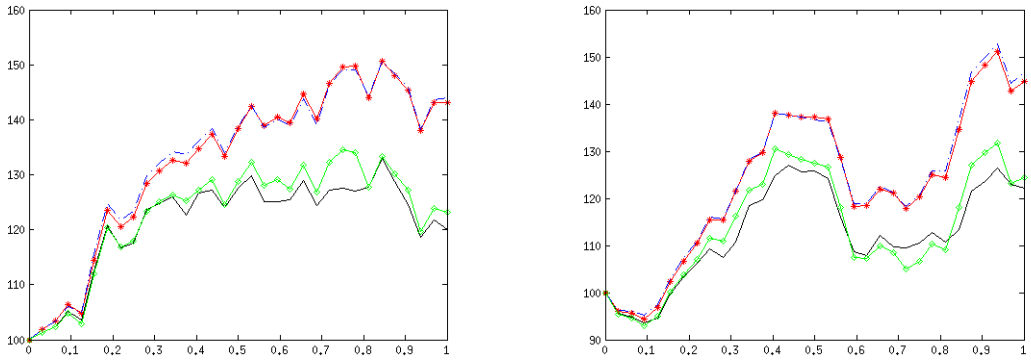

Figure 3.2: A path of a Wiener process (left side) and a Lévy process (right side) in black compared with their approximations (Euler-Maruyama in red, Milstein in blue and predictor corrector scheme in green).

6. *This approach includes the split-step backward Euler scheme proposed by Higham et al. (2002)*

In Section 5.1 it will be shown in what manner the strong order differs as related to the choice of number of corrector steps $k \in \mathbb{N}$, $\eta, \vartheta \in [0,1]$ and the chosen predictor scheme. Figure 3.2 is intended to motivate the investigations of stochastic predictor-corrector schemes. The Lévy process is a jump-diffussion process (2.40) with Merton jumps (2.41). Parameters that are used are from Chapter 5. A first review of this was made by Platen (1995).

Predictor-corrector schemes (3.9) lead analogously to $\mathrm{P}^{EM}_{\mathrm{M},\Delta t}$ which leads to the following discretized calibration problem

$$\begin{aligned} &\min_{x\in X} \hat{f}_{M,\Delta t}(x) := \tfrac{1}{2}\sum_{i=1}^{I}\left(\hat{C}^i_{M,\Delta t}(x) - C^i_{\mathrm{obs}}\right)^2, \\ &\text{where} \quad \hat{C}^i_{M,\Delta t}(x) := \mathrm{e}^{-rT_i}\,\tfrac{1}{M}\textstyle\sum_{m=1}^{M}\left(\pi(s^{m,J}_{N_i} - K_i)\right) \end{aligned} \tag{$\mathrm{P}^{PC}_{\mathrm{M},\Delta t}$}$$

$$\begin{aligned} \text{s.t. } y^{m,0}_{n+1} &= y^{m,J}_n + a(x, y^{m,J}_n)\Delta t_n + b(x, y^{m,J}_n)\Delta W^m_n \\ &\quad + \alpha\frac{b(x, y^{m,J}_n)b'(x, y^{m,J}_n)}{2}\left((\Delta W^m_n)^2 - \Delta t_n\right), \\ y^{m,j}_{n+1} &= y^{m,J}_n + \left(\eta\underline{a}(x, y^{m,j-1}_{n+1}) + (1-\eta)\underline{a}(x, y^{m,J}_n)\right)\Delta t_n \\ &\quad + \left(\vartheta b(x, y^{m,j-1}_{n+1}) + (1-\vartheta)b(x, y^{m,J}_n)\right)\Delta W^m_n \end{aligned}$$

$$m = 1, \ldots, M, \quad n = 1, \ldots, N,$$

for all $j = 1, \ldots, J$, where J denotes the number of corrector steps.

Other schemes like stochastic Runge-Kutta schemes can be found in Abdulle et al. (2014), Buckwar and Winkler (2007), Buckwar et al. (2010), Burrage and Tian (2002), Burrage et al.

(2004), Carletti et al. (2004), Debrabant and Rößler (2008), Komori (2007), Komori (2008), Rößler (2009) and Rößler (2010).

3.3.3 Smoothing Payoffs

Hereafter it might be tempting to immediately apply smooth optimization methods for the solution of the discretized problems $(\mathrm{P}^{EM}_{\mathrm{M},\Delta t})$ or $(\mathrm{P}^{PC}_{\mathrm{M},\Delta t})$. However, a closer look reveals that the payoff is not differentiable anymore since the integral operator previously smoothing the maximum function π was replaced by a discrete sum. In addition, the drift function a and the diffusion function b might not be continuously differentiable on all of $\mathbb{R}^P \times \mathbb{R}^d$. Käbe et al. (2009) observed in numerical experiments that optimization algorithms can get stuck in the resulting corners of the Monte Carlo estimator and consequently break down.

To prevent this smoothing out of the non-differentiability of the maximum function, the indicator function and the absolute value function with the help of a smoothing function π_ϵ, where $\epsilon > 0$ is an adequate smoothing parameter, is compulsory. Furthermore, one can smooth out potential non-differentiabilities of a, b to obtain appropriate modifications a_ϵ, b_ϵ in the manner of Käbe et al. (2009). Note that all payoff functions introduced in Section 2.3 $\pi \notin C^2$ and so $\pi \notin C_p^{2(\beta+1)}$ as well referring to Remark 2.2.22, so one obtains no weak approximation having such smooth functions. Nevertheless, weak solutions are mainly used in financial engineering. To solve this problem, smooth approximations of those payoff functions, in the manner of the following, could be used. Initially, let $\pi_\epsilon(x) \approx \max(x, 0)$ $\epsilon > 0$ be a first smoothing function for all $x \in \mathbb{R}$ in the following remark.

Remark 3.3.2.

$$\pi_\epsilon(x) = \frac{1}{\pi}\left(x \arctan\left(\frac{x}{\epsilon}\right) - \frac{\epsilon}{2}\log\left(\left(\frac{x}{\epsilon}\right)^2 + 1\right)\right) + \frac{1}{2}x \quad \forall x \in \mathbb{R} \tag{3.10}$$

$$\pi'_\epsilon(x) = \frac{1}{\pi}\left(\arctan(\frac{x}{\epsilon})\right) + \frac{1}{2} \quad \forall x \in \mathbb{R} \tag{3.11}$$

$$\pi''_\epsilon(x) = \frac{\epsilon}{\pi(\epsilon^2 + x^2)} \quad \forall x \in \mathbb{R} \tag{3.12}$$

1. *Function π_ϵ is a smoothed version of the ramp function* $\max(x, 0)$.

2. *Function π'_ϵ is a smoothed version of the Heaviside function*

$$H(x) = \begin{cases} 0 & , x \leq 0 \\ 1 & , x > 0. \end{cases} \tag{3.13}$$

3. *Function π''_ϵ is a smoothed version of the Dirac delta function*

$$\delta_0(x) = \begin{cases} 0 & , x \neq 0 \\ 1 & , x = 0. \end{cases} \tag{3.14}$$

One might also use a logistic sigmoid function such as $1/(1+\mathrm{e}^{-\frac{x}{\epsilon}})$ instead of π_ϵ to smooth

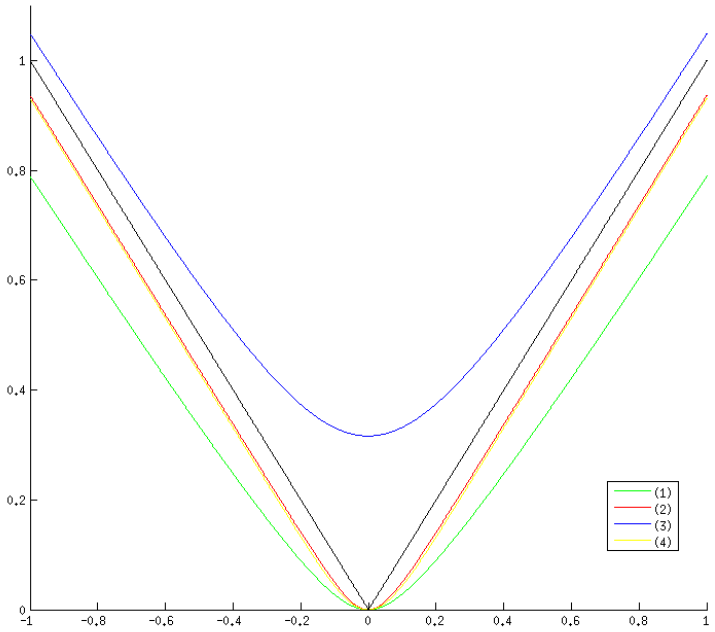

Figure 3.3: Smoothed absolute value functions with $\epsilon = 0.1$:
(1) $\frac{2}{\pi}\left(x \arctan\left(\frac{x}{\epsilon}\right) - \frac{\epsilon}{2}\log\left(\left(\frac{x}{\epsilon}\right)^2 + 1\right)\right)$ **(2)** $\frac{2}{\pi}x \arctan\left(\frac{x}{\epsilon}\right)$
(3) $\sqrt{x^2 + \epsilon}$ **(4)** $\epsilon \log\left(\mathrm{e}^{x/\epsilon} + \mathrm{e}^{-x/\epsilon}\right)$.

the function $\max(x, 0)$, that is known as ramp function. Polynomial approximations of the max function can be found in Käbe (2010).

Note that one is able to approximate an indicator function $\mathbb{1}_{[a,b]}$ of an interval $[a, b] \subset \mathbb{R}$ by using a smoothed version of the Heaviside function (3.13).

$$\mathbb{1}_{[a,b]} = H(x - a) - H(x - b) \approx \pi'_\epsilon(x - a) - \pi'_\epsilon(x - b)$$

Since the prices of financial derivatives are always nonnegative, a smoothed version of the absolute value function $|\cdot| : \mathbb{R} \to \mathbb{R}$ is needed, as well. The following smooth function provides an approximation of the absolute value function for all $x \in \mathbb{R}$.

$$\pi_\epsilon^{|\cdot|}(x) = \frac{2}{\pi}\left(x \arctan\left(\frac{x}{\epsilon}\right) - \frac{\epsilon}{2}\log\left(\left(\frac{x}{\epsilon}\right)^2 + 1\right)\right) \geq 0 \tag{3.15}$$

Another way to smooth the absolute value function is given by (Øksendal, 2003, Example 4.10) as follows

$$\pi_\epsilon^{|\cdot|}(x) = \begin{cases} |x| & , |x| \geq \epsilon \\ \frac{1}{2}(\epsilon + \frac{x^2}{\epsilon}) & , |x| < \epsilon. \end{cases}$$

In the setting of expected values it is important to keep in mind that non-differentiability on $\mathbb{P}$-nullsets is negligible. Nevertheless, the modifications above finally lead to the smoothed optimization problem in the case of the Euler-Maruyama scheme (3.6)

$$\begin{aligned} &\min_{x\in X} \hat{f}_{M,\Delta t,\epsilon}(x) := \tfrac{1}{2}\sum_{i=1}^{I}\left(\hat{C}^i_{M,\Delta t,\epsilon}(x) - C^i_{\text{obs}}\right)^2 \\ &\text{where}\quad \hat{C}^i_{M,\Delta t,\epsilon}(x) := \mathrm{e}^{-rT_i}\tfrac{1}{M}\textstyle\sum_{m=1}^{M}\left(\pi_\epsilon(s^m_{N_i,\epsilon} - K_i)\right) \\ &\text{s.t.}\quad\ y^m_{n+1,\epsilon} = y^m_{n,\epsilon} + a_\epsilon(x, y^m_{n,\epsilon})\Delta t_n + b_\epsilon(x, y^m_{n,\epsilon})\Delta W^m_n, \end{aligned} \tag{$\mathrm{P}^{EM}_{\mathrm{M},\Delta t,\epsilon}$}$$

$$m = 1,\dots,M, \quad n = 0,\dots,N-1,$$

with $s^m_{N_i,\epsilon}(x)$ defined as the first component of $y^m_{N_i,\epsilon}$. Otherwise the predictor-corrector scheme (3.9) leads to

$$\begin{aligned} &\min_{x\in X} \hat{f}_{M,\Delta t,\epsilon}(x) := \tfrac{1}{2}\sum_{i=1}^{I}\left(\hat{C}^i_{M,\Delta t,\epsilon}(x) - C^i_{\text{obs}}\right)^2, \\ &\text{where}\quad \hat{C}^i_{M,\Delta t,\epsilon}(x) := \mathrm{e}^{-rT_i}\tfrac{1}{M}\textstyle\sum_{m=1}^{M}\left(\pi_\epsilon(s^{m,J}_{N_i,\epsilon} - K_i)\right) \end{aligned} \tag{$\mathrm{P}^{PC}_{\mathrm{M},\Delta t,\epsilon}$}$$

$$\begin{aligned} \text{s.t.}\ y^{m,0}_{n+1,\epsilon} &= y^{m,J}_{n,\epsilon} + a_\epsilon(x, y^{m,J}_{n,\epsilon})\Delta t_n + b_\epsilon(x, y^{m,J}_{n,\epsilon})\Delta W^m_n \\ &\quad + \alpha\frac{b_\epsilon(x, y^{m,J}_{n,\epsilon})b'_\epsilon(x, y^{m,J}_{n,\epsilon})}{2}\left((\Delta W^m_n)^2 - \Delta t_n\right), \\ y^{m,j}_{n+1,\epsilon} &= y^{m,J}_{n,\epsilon} + \left(\eta\underline{a}_\epsilon(x, y^{m,j-1}_{n+1,\epsilon}) + (1-\eta)\underline{a}_\epsilon(x, y^{m,J}_{n,\epsilon})\right)\Delta t_n \\ &\quad + \left(\vartheta b_\epsilon(x, y^{m,j-1}_{n+1,\epsilon}) + (1-\vartheta)b_\epsilon(x, y^{m,J}_{n,\epsilon})\right)\Delta W^m_n \end{aligned}$$

$$m = 1,\dots,M, \quad n = 0,\dots,N-1,$$

for all $j = 1,\dots,J$, where J denotes the number of corrector steps. To facilitate notation from now on, $y^m_{n,\epsilon}$, $s^m_{n,\epsilon}$ is abbreviated by y_n, s_n and $y^{m,j}_{n,\epsilon}$, $s^{m,j}_{n,\epsilon}$ by $y^{m,j}_n$, $s^{m,j}_n$. Based on the assumed smoothness of the maps a_ϵ, b_ϵ and π_ϵ, one can deduce that the call price functions $\hat{C}^i_{M,\Delta t,\epsilon}$ as well as the objective $\hat{f}_{M,\Delta t,\epsilon}$ are continuously differentiable on X such that in particular their gradients exist. However, to keep notation simple the additional index ϵ is only used where necessary.

3.4 Alternative Approaches for Estimating Sensitivities

Calibration, gradient calculation and especially estimation of sensitivities of expectations commonly refered to as Greeks (i.e., the delta, the vega or the gamma) are certainly a broad and hot topic in financial engineering. In this context, statistical methods are often used to estimate Greeks. Besides statistical methods automatic differentiation is one way to estimate sensitivities. A detailed overview of this topic is given by Griewank and Walther (2008) or Griewank et al. (2012). Furthermore, all these methods are already used on the parameter identification of biological and geophysical models, as well. Allthough this is not the main

goal of this thesis, only a brief overview on other approaches is given in this section.

3.4.1 Likelihood Methods

The likelihoods methods are based on the relationship between the probability density function defined by the financial market model for the underlying price process and the calibration parameter. There exist various likelihood methods, such as the likelihood ratio method (LRM), maximum likelihood estimation (MLE), or singular value decomposition likelihood method (SVDL).

A detailed overview of the likelihood method can be found in the classical work by Broadie and Glasserman (1996) or the Section 7.3 of Glasserman (2004). A brief overview including a short discussion on its applicability is given in the following.

Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with probability density g_θ and θ being a parameter of this density. Let f be a real-valued function of X. Equipped with this notation, the expected function value is given by

$$\mathbb{E}_\theta[f(X)] = \int_{\mathbb{R}} f(x) g_\theta(x) dx. \tag{3.16}$$

By interchanging the order of differentiation and integration using Fubini's theorem (Billingsley, 1995, Theorem 18.3) one obtains

$$\begin{aligned} \frac{\partial}{\partial\theta}\mathbb{E}_\theta[f(X)] &= \int_{\mathbb{R}} f(x) \frac{\partial}{\partial\theta} g_\theta(x) dx \\ &= \int_{\mathbb{R}} f(x) \frac{\dot{g_\theta}(x)}{g_\theta(x)} g_\theta(x) dx \\ &= \mathbb{E}_\theta\left[f(X) \frac{\dot{g_\theta}(X)}{g_\theta(X)} \right], \end{aligned} \tag{3.17}$$

where $\dot{g_\theta}$ is written for $\frac{\partial}{\partial\theta} g_\theta$. It now follows from this equation (3.17) that the likelihood ratio or score function

$$f(X) \frac{\dot{g_\theta}(X)}{g_\theta(X)}$$

is an unbiased estimator of the derivative of $\mathbb{E}_\theta[f(x)]$ (Glasserman, 2004, Section 7.3.1).

The likelihood ratio method just like the adjoint technique leads to the exact derivative. However, the main limitation in the feasibility of likelihood ration method is that the probability density function g_θ has to be known for the financial market model. Even in the case that the probability function is known, the calibration parameter θ may not be a parameter of the density at all. For that reason, this method is only applicable for a few chosen financial market models as the Heston model in Aït-Sahalia and Kimmel (2007).

These likelihood method cover a wide range of applications:

- Apostolos and Skiadas (1995) use a maximum likelihood estimator (MLE) to estimate parameters of an SDE, by which the Greek electricity consumption is forecasted.
- In Janczura (2014) the expectation maximization (EM) algorithm of Dempster et al. (1977) searches for a local maximum of the likelihood function to fit given European energy exchange (EEX) data to the model prices of the Vasicek model 2.47.
- Parameter estimation by the EM method is also used in Horváth and Manini (2008). Gillespie's algorithm serves to simulate a network of stochastic reactions with given initial quantities and kinetic rate constants. In this paper the authors consider the estimation of the kinetic rate constants of the reactions based on a set of discrete observations generated by Gillespie's algorithm.
- LRM is used to estimate kinetic parameters of several biochemical reaction sytsems for modelling, i.e, the lambda phage switch Reinker et al. (2006).
- The work by Moummou et al. (2012) is concerned with the estimation of the parameters of interest in the drift coefficient of a stochastic Gompertz model (cf. Gutiérrez et al. (2005), Gutiérrez et al. (2006), Skiadas (2010)) with logarithmic therapeutic functions. They use a maximum likelihood method to estimate these parameters, specially of the tumor growth deceleration factor, achieved to date through numerically-based approaches. The proposed method in this paper provides explicit expressions for the statistical parameter estimators.

3.4.2 Direct Pathwise Derivatives

The pathwise methods are based on the relationship between the payoff π defined by the financial market model and its outcome for the underlying price process and the calibration parameter. A detailed overview of the pathwise method or pertubation analysis can be found in the Section 7.2 of Glasserman (2004).

For a short introduction of this pathwise derivative estimation, let $X^\theta = (X_t^\theta)_{t\in T}$ be a real-valued stochastic process indexed by the parameter $\theta \in \Theta \subset \mathbb{R}$ and $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For a fixed event $\omega \in \Omega$ the mapping $\theta \mapsto f(X^\theta)$ is a random function on Θ for some differentiable function f. Then,

$$f'(X^\theta) := \lim_{h\to 0} \frac{X^{\theta+h} - X^\theta}{h} \tag{3.18}$$

denotes the pathwise derivative of f at θ. Equation (3.18) states that the derivative exists with probability one. The expectation

$$\mathbb{E}\left[\frac{\partial}{\partial\theta} f(X^\theta)\right] = \frac{\partial}{\partial\theta}\mathbb{E}[f(X^\theta)] \tag{3.19}$$

is an unbiased estimator if the interchange of differentation and expectation is justified (Glasserman, 2004, Section 7.2.1).

For the estimation of the pathwise derivative one is referred to the Malliavin calculus, named after Paul Malliavin, which is also called the stochastic calculus of variations. A detailed overview on the Malliavin calculus is given by Nualart (2006) or Di Nunno et al. (2009). A similar approach is given by Cont and Fourniè (2013).

3.5 Biochemical and Geoscientific Applications

The problem of adjusting a biological or geoscientific model to physically observed measurements arises. Due to its variability the adjoint approach presented in this chapter is applicable for such parameter estimations.

- A stochastic model for replicators in catalyzed RNA-like polymers is presented and numerically solved in the paper by Rößler et al. (2009). The model consists of a system of reaction-diffusion equations describing the evolution of a population formed by RNA-like molecules with catalytic capabilities in a prebiotic process. A stochastic excitation with additive noise is introduced. To numerically solve the governing equations they apply the stochastic method of lines. A finite-difference reaction-diffusion system is constructed by discretizing the space and the associated stochastic differential system is numerically solved using a class of stochastic Runge-Kutta methods. Numerical experiments are carried out on a prototype of four catalyzed selfreplicator species along with an activated and an inactivated residue. Results are given only in two space dimensions.

- Applications to stochastic HIV-models like in Banks et al. (2008), where the extended Kalman filter is used to estimate the conditional expected value and the conditional covariance in order to identify unkown model parameters of their HIV-model, or Wai-Yuan (2002) or other phage-bacteria interaction models like Carletti (2002) or Carletti (2006) are possible.

- Stochastic optimal control is about the optimizing drug-control history, maximizing efficacy of the drug while minimizing its side effects and cost as established by Stengel and Ghigliazza (2004). Furthermore, Chen et al. (2005) use a stochastic differential equation model for quantifying transcriptional regulatory network in Saccharomyces cerevisiae.

- Dowd (2011) proclaimed that parameter estimation for stochastic dynamic oceanic systems is a core problem for the environmental and ecological sciences. This ecological study considers parameter estimation for a simple nonlinear numerical model of marine biogeochemistry. They present a nonlinear stochastic differential equation based model for estimating parameters from ocean measurements collected at a coastal ocean observatory. The ecosystem components considered are: phytoplankton (P) and inorganic

nutrients (N). Its dynamics are described by

$$dP_t = \frac{N}{k_N}(\hat{\gamma}(t) - \Delta\gamma)P_t - \lambda P_t^2 + \sigma dW_t^P$$
$$dN_t = \frac{N}{k_N}(\hat{\gamma}(t) - \Delta\gamma)P_t + \lambda P_t^2 + \sigma dW_t^N.$$

The parameters of interest are the ecological parameters k_N, λ and the statistical parameter σ. $\hat{\gamma}(t)$ denotes the seasonal photosynthetic rate and $\Delta\gamma$ its mean.

- Due to Alkhatib and King (2014) uncertainty in surfactant-polymer flooding is an important challenge to the wide-scale implementation of enhanced oil recovery process. Thus, it is essential to have the ability to quantify this uncertainty in an efficient manner. Monte Carlo simulation is the traditional uncertainty quantification approach that is used for quantifying parametric uncertainty. However, the convergence of Monte Carlo simulation is relatively low, requiring a large number of realizations to converge. This study proposes the use of the probabilistic collocation method in parametric uncertainty quantification for surfactant-polymer flooding using four synthetic reservoir models. Four sources of uncertainty were considered: the chemical flood residual oil saturation, surfactant and polymer adsorption, and the polymer viscosity multiplier. The output parameter approximated is the recovery factor. The output metrics were the input-output model response relationship, the probability density function, and the first two moments. These were compared with the results obtained from Monte Carlo simulation over a large number of realizations. Two methods for solving for the coefficients of the output parameter polynomial chaos expansion are compared: Gaussian quadrature and linear regression.

Chapter 4

Adjoint Equations of Improved Monte-Carlo Schemes

Civilization is the process of reducing the infinite to the finite.
— Oliver Wendell Holmes

4.1 Stochastic Dynamics in Finance

The main area for the use of stochastic differential equations and their appropriate approximations in financial engineering is the pricing of financial derivatives and the estimation of sensitivities with respect to designated parameters, which are widely used to measure risk or to calibrate financial market models. Those derivatives are financial contracts and instruments whose values are derived from the value of an underlying. To this purpose let $S = (S_t)_{0 \leq t \leq T}$ be an $\mathbb{R}^d$-valued price process of some underlying and π a payoff function, which is defined by the type of the financial contract and a strike price K as introduced in Section 2.3. The prospective price of a financial derivative is then determined by the expected value of the payoff function π applied to the underlying process S.

4.2 Calibration via Predictor-Corrector Monte-Carlo Method

4.2.1 Calculation of the Gradient

To solve the nonlinear optimization problem with equality constraints ($\mathrm{P}^{PC}_{\mathrm{M},\Delta t,\epsilon}$) of the previous chapter, one needs to compute either the forward difference approximation

$$\nabla f(x) \approx \left[\frac{f(x + h\mathbf{e}_p) - f(x)}{h} \right]_{p=1}^{P}, \tag{4.1}$$

the backward difference approximation

$$\nabla f(x) \approx \left[\frac{f(x) - f(x - h\mathbf{e}_p)}{h} \right]_{p=1}^{P} \tag{4.2}$$

or the central difference approximation

$$\nabla f(x) \approx \left[\frac{f(x+h\mathbf{e}_p) - f(x-h\mathbf{e}_p)}{2h}\right]_{p=1}^{P}, \tag{4.3}$$

where $\mathbf{e}_p$ denotes the p-th unit vector $(0,...,0,1,0,...,0)^\top \in \mathbb{R}^P$ - a simple, but yet expensive way (cf. Section 7.1 of Glasserman (2004)). Especially in cases when the calculation of $f(x)$ is expensive, as it is in a Monte-Carlo framework, this method results in a high computational effort because P additional evaluations of f are necessary. Even though the structure of $f(x)$ allows to compute the derivatives for all $i = 1,\dots,I$ in one sweep by differentiating $s^m_{N_i}$ with respect to x, the total effort of a finite difference approximation is substantial. Furthermore, an inadequate choice of h can lead to some other problems, while in applications of Monte-Carlo methods the variability in estimates of function values ranges.

Lemma 4.2.1 (Finite Difference Approximation). *Let $\hat{f}_{M,\Delta t,\epsilon} : \mathbb{R}^P \to \mathbb{R}$ be the objective functional of $(\mathrm{P}^{EM}_{\mathrm{M},\Delta t,\epsilon})$ or $(\mathrm{P}^{PC}_{\mathrm{M},\Delta t,\epsilon})$,*

$$\hat{f}_{M,\Delta t,\epsilon}(x) = \frac{1}{2}\sum_{i=1}^{I}\left(\mathrm{e}^{-rT_i}\,\frac{1}{M}\sum_{m=1}^{M}\Big(\pi_\epsilon(s^{m,J}_{N_i,\epsilon}(x) - K_i)\Big) - C^i_{\mathrm{obs}}\right)^2.$$

For an arbitrary $h > 0$ the central difference estimator

$$\left[\frac{f(x+h\mathbf{e}_p) - f(x-h\mathbf{e}_p)}{2h}\right]_{p=1}^{P} \tag{4.4}$$

is an approximation of the gradient $\nabla\hat{f}_{M,\Delta t,\epsilon}$ with order $\mathcal{O}(h^2)$.

Proof. (Glasserman, 2004, Subsection 7.1.1) □

One should recall that problems $(\mathrm{P}^{EM}_{\mathrm{M},\Delta t,\epsilon})$ and $(\mathrm{P}^{PC}_{\mathrm{M},\Delta t,\epsilon})$ are nonlinear least squares problems with a special structure. In the end, a nonlinear optimization problem with equality constraints is obtained

$$\begin{aligned} &\text{Minimize } \hat{f}_{M,\Delta t,\epsilon}(y,x) \\ &\text{s.t. } g_n(y,x) = 0, \quad n = 1,\dots,N. \end{aligned} \tag{ECP}$$

To apply Lagrange multipliers one rewrites the side conditions to equality constraints in the following notations.

- $\mathrm{P}^{EM}_{\mathrm{M},\Delta t,\epsilon}$:

$$g^m_n(y,x) = -y^m_{n,\epsilon} + y^m_{n-1,\epsilon} + a_\epsilon(x, y^m_{n-1,\epsilon})\Delta t_{n-1} + b_\epsilon(x, y^m_{n-1,\epsilon})\Delta W^m_{n-1}$$

$$m = 1,\dots,M, \quad n = 1,\dots,N.$$

- $\mathrm{P}^{PC}_{\mathrm{M},\Delta t,\epsilon}$:

$$\begin{aligned}
g_n^{m,0}(y,x) &= -y_{n,\epsilon}^{m,0} + y_{n-1,\epsilon}^{m,J} + a_\epsilon(x, y_{n-1,\epsilon}^{m,J})\Delta t_{n-1} + b_\epsilon(x, y_{n-1,\epsilon}^{m,J})\Delta W_{n-1}^m \\
&\quad + \alpha \frac{b_\epsilon(x, y_{n-1,\epsilon}^{m,J}) b_\epsilon'(x, y_{n-1,\epsilon}^{m,J})}{2} \left((\Delta W_{n-1}^m)^2 - \Delta t_{n-1} \right), \\
g_n^{m,j}(y,x) &= -y_{n,\epsilon}^{m,j} + y_{n-1,\epsilon}^{m,J} + \left(\eta \underline{a}_\epsilon(x, y_{n,\epsilon}^{m,j-1}) + (1-\eta)\underline{a}_\epsilon(x, y_{n-1,\epsilon}^{m,J}) \right) \Delta t_{n-1} \\
&\quad + \left(\vartheta b_\epsilon(x, y_{n,\epsilon}^{m,j-1}) + (1-\vartheta) b_\epsilon(x, y_{n-1,\epsilon}^{m,J}) \right) \Delta W_{n-1}^m
\end{aligned}$$

$$m = 1, \ldots, M, \quad n = 1, \ldots, N, \quad j = 1, \ldots, J, \quad \alpha, \eta, \vartheta \in [0,1].$$

Note that y_n^m, g_n^m and $g_n^{m,j}$ are d-dimensional vectors for all $m = 1, \ldots, M, n = 1, \ldots, N, j = 0, \ldots, J$. To shorten notation the dimension is choosen as $d = 1$ when no confusion can arise. In what follows, it will be seen how the calculation can be sped up using an adjoint-based approach. To this purpose the side conditions are written in terms of a Lagrange functional $\mathcal{L} : \mathbb{R}^{N+P+N} \to \mathbb{R}$ defined by

$$\mathcal{L}(y, x, \lambda) = \hat{f}(y, x) + \sum_{m=1}^{M} \sum_{n=1}^{N} \lambda_n^{m\top} g_n^m(y, x) \tag{4.5}$$

and $\mathcal{L} : \mathbb{R}^{JN+P+JN} \to \mathbb{R}$ defined by

$$\mathcal{L}(y, x, \lambda) = \hat{f}(y, x) + \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{j=0}^{J} \lambda_n^{m,j\top} g_n^{m,j}(y, x) \ , \tag{4.6}$$

respectively. Here $\lambda_1^m, \ldots, \lambda_N^m$ for all $1 \le m \le M$ and $\lambda_1^{m,0}, \lambda_1^{m,1}, \ldots, \lambda_N^{m,J}$ for all $1 \le m \le M,\ 0 \le j \le J$ are the Lagrange multipliers.

For notional simplification, the constraint function $G : \mathbb{R}^{N+P} \to \mathbb{R}^N$ is defined by

$$G(y, x) := (g_1(y, x), \ldots, g_N(y, x)) \tag{4.7}$$

and $G : \mathbb{R}^{JN+P} \to \mathbb{R}^{JN}$ is defined by

$$G(y, x) := (g_1^0(y, x), g_1^1(y, x), g_1^2(y, x), \ldots, g_1^J(y, x), g_2^0(y, x), g_2^1(y, x), \ldots, g_N^J(y, x)), \tag{4.8}$$

respectively. So one can write the constraints of the Problem (ECP) in the more compact form

$$G(y, x) = 0, \tag{4.9}$$

which leads to a main condition of the following well-known linear version of the Implicit Function Theorem 2.1.12, which is repeated here for the convenience of the reader.

Proposition 4.2.2 (Implicit Function Theorem (Linear Version)). *Let $G : \mathbb{R}^{N+P} \to \mathbb{R}^N$ be a continuously differentiable linear function and the Jacobian matrix $\nabla_y G \in \mathbb{R}^{N\times N}$ be non singular, then there corresponds to every vector $x^* \in \mathfrak{S}_{x^*} \subset \mathbb{R}^P$ a unique vector $y^* \in \mathfrak{S}_{y^*} \subset \mathbb{R}^N$ such that $G(y^*, x^*) = 0$. This y^* can be computed from x^* via*

$$y^* = -\left(\nabla_y G\right)^{-1} \nabla_x G x^*.$$

Furthermore, one obtains implicitly the gradient of the unique function $y : \mathfrak{S}_{x^*} \to \mathfrak{S}_{y^*}$ for all $x \in \mathfrak{S}_{x^*}$ by

$$\nabla y(x) = -\left(\nabla_y G(y(x), x)\right)^{-1} \nabla_x G(y(x), x). \tag{4.10}$$

Lemma 4.2.3 (Surjectivity). *The constraint gradients of $\mathrm{P}^{EM}_{\mathrm{M},\Delta t,\epsilon}$ and $\mathrm{P}^{PC}_{\mathrm{M},\Delta t,\epsilon}$ in direction of y are linearly independent, thereby $\nabla_y G$ is non singular for all $y \in \mathbb{R}^N$ and $x \in \mathbb{R}^P$.*

Proof. To show that $\nabla_y g_1(y, x), \ldots, \nabla_y g_N(y, x)$ are linearly independent one builds the $N\times N$ matrix

$$\begin{aligned}\nabla_y G(y, x) &= \begin{pmatrix} \vdots & \vdots & & \vdots \\ \nabla_y g_1(y, x) & \nabla_y g_2(y, x) & \cdots & \nabla_y g_N(y, x) \\ \vdots & \vdots & & \vdots \end{pmatrix} \\ &= \begin{pmatrix} -1 & \widetilde{g}_1 & 0 & \ldots & \ldots & 0 \\ 0 & -1 & \widetilde{g}_2 & 0 & \ldots & 0 \\ \vdots & 0 & -1 & \widetilde{g}_3 & 0\ldots & 0 \\ \vdots & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \widetilde{g}_{N-1} \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix}\end{aligned}$$

for all $(y, x) \in \mathbb{R}^N \times \mathbb{R}^P$. Since $\nabla_y G$ is an upper triangular matrix and hence $\det(\nabla_y G) = \prod_{i=1}^N \nabla_y G_{i,i} = (-1)^N \neq 0$, one gets that $\nabla_y G$ is non singular (Golub and van Loan, 1996, §2.1), thus the linearly independence of the columns of G which are $\nabla_y g_1(y, x), \ldots, \nabla_y g_N(y, x)$ follows. The proof of the linear independence of $\nabla_y g_1^0(y, x), \ldots, \nabla_y g_N^J(y, x)$ and that the $JN \times JN$ matrix $\nabla_y G$ is invertible is similar. □

Lemma 4.2.3 leads to a proposition of the Implicit Function Theorem 4.2.2 for all $(y, x) \in \mathbb{R}^N \times \mathbb{R}^P$. Henceforth, the following Lemma 4.2.4 shows coherence between the Lagrange multipliers and the adjoint equation. The introductive paper by Giles and Pierce (2000) established this approach in a deterministic airfoil design optimization.

Lemma 4.2.4. *Let $G : \mathbb{R}^{N+P} \to \mathbb{R}^N$ and $f : \mathbb{R}^{N+P} \to \mathbb{R}$ be continuously differentiable functions such that $G(y(x), x) = 0$ and $\nabla_y G(y(x), x)$ is non singular. Furthermore, let $\lambda \in \mathbb{R}^N$ be a vector corresponding to $x \in \mathbb{R}^P$ defined by*

$$\begin{aligned} \lambda &:= -\nabla_y f(y(x), x)\left(\nabla_y G(y(x), x)\right)^{-1} \\ \Leftrightarrow \quad -\nabla_y G(y(x), x)^\top \lambda^\top &= \nabla_y f(y(x), x)^\top, \end{aligned} \tag{4.11}$$

then λ is uniquely defined. In addition, let $\phi : \mathbb{R}^P \to \mathbb{R}$ be a function defined by

$$\phi(x) := f(y(x), x), \tag{4.12}$$

then it holds

$$\nabla\phi(x) = \lambda\nabla_x G(y(x), x) + \nabla_x f(y(x), x). \tag{4.13}$$

Proof. The uniqueness of λ follows directly from the fact that $\nabla_y G(y(x), x)$ is invertible and its definition by equation (4.11). The gradient of ϕ is obtained by using the chain rule and the Implicit Function Theorem 4.2.2 in combination with the definition of λ in equation (4.11)

$$\begin{aligned}\nabla\phi(x) &= \nabla_y f(y(x), x)\nabla y(x) + \nabla_x f(y(x), x)\\ &= -\nabla_y f(y(x), x)\left(\nabla_y G(y(x), x)\right)^{-1}\nabla_x G(y(x), x) + \nabla_x f(y(x), x)\\ &= \lambda\nabla_x G(y(x), x) + \nabla_x f(y(x), x).\end{aligned}$$

This shows the proposition and establishes the formula (4.13). □

In what follows, the dimension of the stochastic process d is free to choose. The vectors $\lambda, y \in \mathbb{R}^{JdN}$ are understood as

$$\lambda = ((\lambda_1^0)^\top, (\lambda_1^1)^\top, (\lambda_1^2)^\top, \ldots, (\lambda_1^J)^\top, (\lambda_2^0)^\top, (\lambda_2^1)^\top, \ldots, (\lambda_N^J)^\top)^\top$$

or

$$y = ((y_1^0)^\top, (y_1^1)^\top, (y_1^2)^\top, \ldots, (y_1^J)^\top, (y_2^0)^\top, (y_2^1)^\top, \ldots, (y_N^J)^\top)^\top,$$

respectively, where $\lambda_n^j \in \mathbb{R}^d$ and $y_n^j \in \mathbb{R}^d$ for all $n = 1, \ldots, N, j = 0, \ldots, J$. The following smoothness assumption on the deterministic continuous drift coefficient a_ϵ, the volatility coefficient b_ϵ and the payoff function π_ϵ is crucial:

Assumption 4.2.5. *Let the deterministic maps a_ϵ, b_ϵ and the payoff π_ϵ be continuously Fréchet-differentiable on the space $\mathbb{R}^{JdN+P}$ or $\mathbb{R}^{JdN}$, respectively. Let b'_ϵ and π'_ϵ denote the Fréchet-derivative with respect to the variable x. Furthermore, $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ are the partial derivatives with respect to parameters x and discretization scheme y, respectively.*

Equipped with this notations and framework, the adjoint approach or the adjoint technique via Monte-Carlo methods is established by the following main theorem. This theorem provides an efficient algorithm to calculate the gradient of the objective function.

Theorem 4.2.6 (The Case of Predictor Corrector Discretization)**.** *Let Assumption 4.2.5 be true and $\hat{f}_{M,\Delta t,\epsilon} : \mathbb{R}^{JdN+P} \to \mathbb{R}$ be the real valued objective function for maturity T_i and strike price K_i for all $i = 1, \ldots, I$*

$$\begin{aligned}\hat{f}_{M,\Delta t,\epsilon}(y, x) &:= \frac{1}{2}\sum_{i=1}^{I}\left(\hat{C}^i_{M,\Delta t,\epsilon}(y, x) - C^i_{\text{obs}}\right)^2 \\ &= \frac{1}{2}\sum_{i=1}^{I}\left(\mathrm{e}^{-rT_i}\,\frac{1}{M}\sum_{m=1}^{M}\left(\pi_\epsilon(s^m_{N_i} - K_i)\right) - C^i_{\text{obs}}\right)^2\end{aligned} \tag{4.14}$$

with predictor-corrector discretization

$$\begin{aligned} y_{n+1,\epsilon}^{m,0} &= y_{n,\epsilon}^{m,J} + a_\epsilon(x, y_{n,\epsilon}^{m,J})\Delta t_n + b_\epsilon(x, y_{n,\epsilon}^{m,J})\Delta W_n^m \\ &\quad + \alpha \frac{b_\epsilon(x, y_{n,\epsilon}^{m,J}) b_\epsilon'(x, y_{n,\epsilon}^{m,J})}{2} \left((\Delta W_n^m)^2 - \Delta t_n\right) \qquad (4.15) \\ y_{n+1,\epsilon}^{m,j} &= y_{n,\epsilon}^{m,J} + \left(\eta \underline{a}_\epsilon(x, y_{n+1,\epsilon}^{m,j-1}) + (1-\eta)\underline{a}_\epsilon(x, y_{n,\epsilon}^{m,J})\right)\Delta t_n \\ &\quad + \left(\vartheta b_\epsilon(x, y_{n+1,\epsilon}^{m,j-1}) + (1-\vartheta) b_\epsilon(x, y_{n,\epsilon}^{m,J})\right)\Delta W_n^m, \qquad (4.16) \end{aligned}$$

$$\begin{aligned} &y_0^{m,J} = Y_0, \quad n = 0, \dots, N-1, \quad m = 1, \dots, M, \\ &j = 1, \dots, J, \quad N := \max_{i=1,\dots,I} N_i, \quad \alpha, \eta, \vartheta \in [0,1]. \end{aligned}$$

Then, y is given by $y(x)$ and the gradient of $f(x) := \hat{f}_{M,\Delta t,\epsilon}(y(x), x)$ for all $x \in \mathbb{R}^P$ is computed regardless of $y \in \mathbb{R}^{JdN}$, maturity T_i and strike price K_i $i = 1, \dots, I$ via

$$\begin{aligned} \nabla f(x) = \frac{1}{M} \sum_{m=1}^{M} \sum_{n=1}^{N} \Bigg(& (\lambda_n^{m,0})^\top \Bigg[\frac{\partial}{\partial x} a_\epsilon(x, y_{n-1}^{m,0})\Delta t_{n-1} + \frac{\partial}{\partial x}\left(b_\epsilon(x, y_{n-1}^{m,0})\Delta W_{n-1}^m\right) \\ & + \alpha \frac{\partial}{\partial x} \frac{b_\epsilon(x, y_{n,\epsilon}^{m,J}) b_\epsilon'(x, y_{n,\epsilon}^{m,J})}{2} \left((\Delta W_n^m)^2 - \Delta t_n\right) \Bigg] \qquad (4.17) \\ & + \sum_{j=1}^{J} (\lambda_n^{m,j})^\top \Bigg[\left(\eta \frac{\partial}{\partial x} \underline{a}_\epsilon(x, y_n^{m,j-1}) + (1-\eta)\frac{\partial}{\partial x}\underline{a}_\epsilon(x, y_{n-1}^{m,J}) \right)\Delta t_n \\ & + \left(\vartheta \frac{\partial}{\partial x}\left(b_\epsilon(x, y_n^{m,j-1})\Delta W_n^m\right) + (1-\vartheta)\frac{\partial}{\partial x}\left(b_\epsilon(x, y_{n-1}^{m,J})\Delta W_n^m\right)\right)\Bigg]\Bigg), \end{aligned}$$

where $\lambda_n^{m,j} \in \mathbb{R}^d$ results from the backward adjoint equations

$$\lambda_n^{m,j} = \left[\eta \frac{\partial}{\partial y}\underline{a}_\epsilon(x, y_n^{m,j})\Delta t_{n-1} + \vartheta \frac{\partial}{\partial y}\left(b_\epsilon(x, y_n^{m,j})\Delta W_{n-1}^m\right)\right]^\top \lambda_n^{m,j+1} \qquad (4.18)$$

$$j = J-1, J-2, \dots, 1, 0, \quad m = 1, \dots, M, \quad n = N, N-1, \dots, 1$$

and

$$\begin{aligned} \lambda_n^{m,J} = & \Bigg[I + \frac{\partial}{\partial y} a_\epsilon(x, y_n^{m,J})\Delta t_n + \frac{\partial}{\partial y}\left(b_\epsilon(x, y_n^{m,J})\Delta W_n^m\right) \\ & + \alpha \frac{\partial}{\partial y} \frac{b_\epsilon(x, y_{n,\epsilon}^{m,J}) b_\epsilon'(x, y_{n,\epsilon}^{m,J})}{2} \left((\Delta W_n^m)^2 - \Delta t_n\right) \Bigg]^\top \lambda_{n+1}^{m,0} \qquad (4.19) \\ & + \sum_{j=1}^{J} \left[I + (1-\eta)\frac{\partial}{\partial y}\underline{a}_\epsilon(x, y_n^{m,j})\Delta t_n + (1-\vartheta)\frac{\partial}{\partial y}\left(b_\epsilon(x, y_n^{m,j})\Delta W_n^m\right)\right]^\top \lambda_{n+1}^{m,j} \end{aligned}$$

$$+ \mathbb{1}_{\{n=N_i\}} \mathrm{e}^{-2rT_i} \left[\pi'_\epsilon(s_n^{m,J}(x) - K_i), 0, \ldots, 0\right] \left(\frac{1}{M} \sum_{m=1}^{M} \pi_\epsilon(s_n^{m,J} - K_i) - \mathrm{e}^{rT_i} C^i_{\mathrm{obs}} \right)$$

$$m = 1, \ldots, M, \quad n = N, N-1, \ldots, 1$$

with

$$\lambda_N^{m,J} = \mathrm{e}^{-2rT_I} \left[\left(\pi'_\epsilon(s_N^{m,J} - K_I) \right), 0, \ldots, 0 \right] \left(\tfrac{1}{M} \textstyle\sum_{m=1}^{M} \pi_\epsilon(s_N^{m,J} - K_I) - \mathrm{e}^{rT_I} C^I_{\mathrm{obs}} \right)$$

$$m = 1, \ldots, M \tag{4.20}$$

as the initial vector of the backward iteration.

Figure 4.1 impressively demonstrates that jumps of the adoint equation (4.19) occur at the maturities T_i.

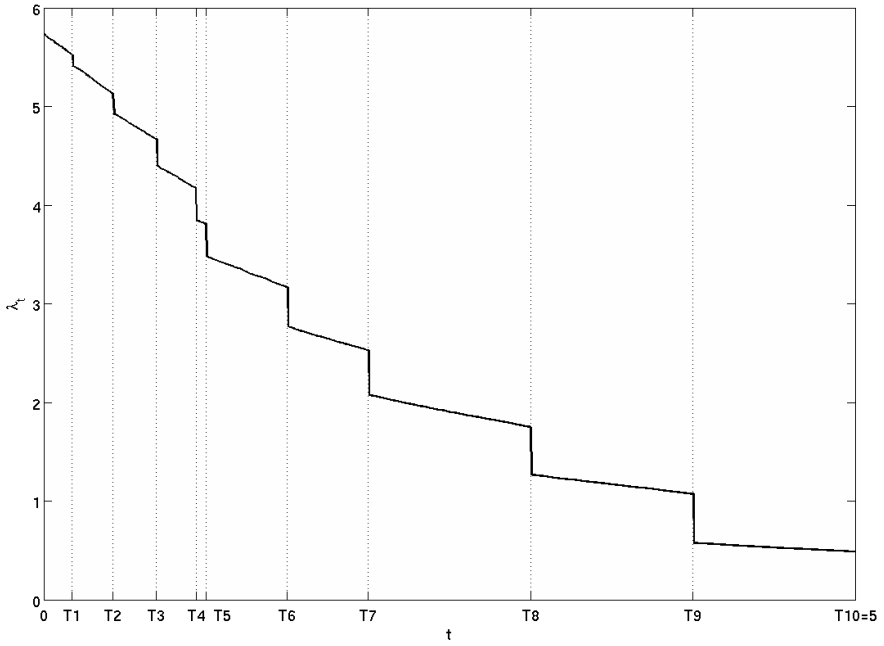

Figure 4.1: An example solution $(\lambda_t)_{0 \le t \le 5}$ of an adjoint equation.

Proof. First, the predictor-corrector discretization of (4.16) is rewritten to equality constraints $g_n^j(y, x) = 0$ for all $n = 1, \ldots, N$ and $j = 0, \ldots, J$ to fit the assumptions of Lemma 4.2.4. The Lagrange function is given by equation (4.6) as

$$\mathcal{L}(y, x, \lambda) = \hat{f}_{M,\Delta t,\epsilon}(y, x) + \sum_{n=1}^{N} \sum_{j=0}^{J} {\lambda_n^j}^\top g_n^j(y, x), \tag{4.21}$$

and the adjoint equation is obtained by solving the following root equation

$$\nabla_y \mathcal{L}(y,x,\lambda) = \nabla_y \hat{f}_{M,\Delta t,\epsilon}(y,x) + \sum_{n=1}^{N}\sum_{j=0}^{J} {\lambda_n^j}^\top \nabla_y g_n^j(y,x) \stackrel{!}{=} 0. \tag{4.22}$$

To this purpose, the gradient $\nabla_y \hat{f}_{M,\Delta t,\epsilon}(y,x)$ of (4.14) is computed by using the chain rule

$$\begin{aligned}
\nabla_y \hat{f}_{M,\Delta t,\epsilon}(y,x) &= \sum_{i=1}^{I} \nabla_y \hat{C}^i_{M,\Delta t,\epsilon}(y,x) \left(\hat{C}^i_{M,\Delta t,\epsilon}(y,x) - C^i_{\text{obs}}\right) \\
&= \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{I} \mathrm{e}^{-2rT_i} \frac{\partial}{\partial y}\pi_\epsilon(s_{N_i}^{m,J} - K_i) \left(\frac{1}{M}\sum_{m=1}^{M} \pi_\epsilon(s_{N_i}^{m,J} - K_i) - \mathrm{e}^{rT_i} C^i_{\text{obs}}\right) \\
&= \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{I} \mathbf{e}_{N_i} \mathrm{e}^{-2rT_i} \pi'_\epsilon(s_{N_i}^{m,J} - K_i) \left(\frac{1}{M}\sum_{m=1}^{M} \pi_\epsilon(s_{N_i}^{m,J} - K_i) - \mathrm{e}^{rT_i} C^i_{\text{obs}}\right) \\
&= \frac{1}{M}\sum_{m=1}^{M} \begin{pmatrix}
0 \\ \vdots \\ 0 \\
\mathrm{e}^{-2rT_1} \pi'_\epsilon(s_{N_1}^{m,J} - K_1) \left(\frac{1}{M}\sum_{m=1}^{M} \pi_\epsilon(s_{N_1}^{m,J} - K_1) - \mathrm{e}^{rT_1} C^1_{\text{obs}}\right) \\
0 \\ \vdots \\ 0 \\
\mathrm{e}^{-2rT_2} \pi'_\epsilon(s_{N_2}^{m,J} - K_2) \left(\frac{1}{M}\sum_{m=1}^{M} \pi_\epsilon(s_{N_2}^{m,J} - K_2) - \mathrm{e}^{rT_2} C^2_{\text{obs}}\right) \\
0 \\ \vdots \\ \vdots \\ 0 \\
\mathrm{e}^{-2rT_I} \pi'_\epsilon(s_{N}^{m,J} - K_I) \left(\frac{1}{M}\sum_{m=1}^{M} \pi_\epsilon(s_{N}^{m,J} - K_I) - \mathrm{e}^{rT_I} C^I_{\text{obs}}\right)
\end{pmatrix}.
\end{aligned}$$

As of now, the proof is considered by a manner of trajectories. So one gets with the notation in equation (4.8) from the root equation (4.22) and by using Lemma 4.2.4 the following system of linear equation with an upper triangular Jacobian matrix

$$\begin{aligned}
& -\nabla_y \hat{f}_{M,\Delta t,\epsilon}(y,x) = \sum_{n=1}^{N}\sum_{j=0}^{J} {\lambda_n^j}^\top \nabla_y g_n^j(y,x) \\
\Leftrightarrow \quad & -\nabla_y \hat{f}_{M,\Delta t,\epsilon}(y,x) = \nabla_y G(y,x)\lambda
\end{aligned}$$

$$\Leftrightarrow \quad \begin{pmatrix} \widetilde{f}_1 \\ \widetilde{f}_2 \\ \vdots \\ \vdots \\ \widetilde{f}_I \end{pmatrix} = \begin{pmatrix} \widetilde{G}_1 & \widetilde{G}_1^J & 0 & \cdots & 0 \\ 0 & \widetilde{G}_2 & \widetilde{G}_2^J & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \widetilde{G}_{N-1}^J \\ 0 & 0 & \cdots & 0 & \widetilde{G}_N \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \vdots \\ \lambda_N \end{pmatrix}, \tag{4.23}$$

where $\widetilde{f}_i = \left(0, \ldots, 0, -\mathrm{e}^{-2rT_i}\, \pi'_\epsilon(s_{N_i}^{m,J} - K_i) \left(\frac{1}{M}\sum_{m=1}^M \pi_\epsilon(s_{N_i}^{m,J} - K_i) - \mathrm{e}^{rT_i}\, C_{\text{obs}}^i\right)\right)^\top \in \mathbb{R}^{N_i}$ for all $i = 1, \ldots, I$, $n = 1, \ldots, N-1$ and the block submatrices on the main diagonal of the Jacobian matrix are defined for all $n = 1, \ldots, N$ as

$$\widetilde{G}_n = \begin{pmatrix} -I & \widetilde{g}_0^n & 0 & \cdots & 0 \\ 0 & -I & \widetilde{g}_1^n & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \widetilde{g}_{J-1}^n \\ 0 & 0 & \cdots & 0 & -I \end{pmatrix}, \tag{4.24}$$

with $\widetilde{g}_j^n = \left[\eta \frac{\partial}{\partial y} \underline{a}_\epsilon(x, y_n^j)\Delta t_n + \vartheta \frac{\partial}{\partial y}\left(b_\epsilon(x, y_n^j)\Delta W_n\right)\right]$ for all $j = 0, \ldots, J-1$. The submatrices on the first diagonal above characterize the fixed part of the fixed-point iteration following the corrector steps and are defined for all time steps $n = 1, \ldots, N-1$ as

$$\widetilde{G}_n^J = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ddots & 0 \\ \widetilde{g}_{J,0}^n & \widetilde{g}_{J,1}^n & \cdots & \widetilde{g}_{J,J}^n \end{pmatrix}, \tag{4.25}$$

with

$$\begin{aligned} \widetilde{g}_{J,0}^n = \Bigg[& I + \frac{\partial}{\partial y} a_\epsilon(x, y_n^J)\Delta t_n + \frac{\partial}{\partial y}\left(b_\epsilon(x, y_n^J)\Delta W_n\right) \\ & + \alpha \frac{\partial}{\partial y} \frac{b_\epsilon(x, y_{n,\epsilon}^J) b'_\epsilon(x, y_{n,\epsilon}^J)}{2} \left((\Delta W_n)^2 - \Delta t_n\right) \Bigg] \\ \text{and } \widetilde{g}_j^n = & \left[I + (1-\eta)\frac{\partial}{\partial y} \underline{a}_\epsilon(x, y_n^j)\Delta t_n + (1-\vartheta)\frac{\partial}{\partial y}\left(b_\epsilon(x, y_n^j)\Delta W_n\right)\right], \end{aligned}$$

for all corrector steps $j = 1, \ldots, J$. This establishes the representation of the adjoint equations in (4.18) and (4.19).

From the last line in the matrix of equation (4.23) one obtains for every path $m \in \{1, \ldots, M\}$

$$\lambda_N^{m,J} = \mathrm{e}^{-2rT_I} \left[\left(\pi'_\epsilon(s_N^{m,J}(x) - K_I)\right), 0, ..., 0\right] \left(\frac{1}{M}\sum_{m=1}^M \pi_\epsilon(s_N^{m,J} - K_I) - \mathrm{e}^{rT_I}\, C_{\text{obs}}^I\right).$$

Therefore, one finally gets the following adjoint backward equations for all Lagrange multipliers $\lambda_n^j \in \mathbb{R}^d$ for all $n = N, N-1, \dots, 1$ and $j = J, J-1, \dots, 0$,

$$\lambda_n^{m,j} = \left[\eta \frac{\partial}{\partial y} \underline{a}_\epsilon(x, y_n^{m,j})\Delta t_n + \vartheta \frac{\partial}{\partial y}\left(b_\epsilon(x, y_n^{m,j})\Delta W_n^m\right)\right]^\top \lambda_n^{m,j+1}$$

$$j = J-1, J-2, \dots, 1, 0, \quad m = 1, \dots, M, \quad n = N, N-1, \dots, 1 \tag{4.26}$$

$$\begin{aligned}
\lambda_n^{m,J} = \Bigg[& I + \frac{\partial}{\partial y} a_\epsilon(x, y_n^{m,J})\Delta t_n + \frac{\partial}{\partial y}\left(b_\epsilon(x, y_n^{m,J})\Delta W_n^m\right) \\
& + \alpha \frac{\partial}{\partial y} \frac{b_\epsilon(x, y_{n,\epsilon}^{m,J}) b_\epsilon'(x, y_{n,\epsilon}^{m,J})}{2}\left((\Delta W_n^m)^2 - \Delta t_n\right)\Bigg]^\top \lambda_{n+1}^{m,0} \\
& + \sum_{j=1}^{J}\left[I + (1-\eta)\frac{\partial}{\partial y}\underline{a}_\epsilon(x, y_n^{m,J})\Delta t_n + (1-\vartheta)\frac{\partial}{\partial y}\left(b_\epsilon(x, y_n^{m,J})\Delta W_n^m\right)\right]^\top \lambda_{n+1}^{m,j} \\
& + \mathbb{1}_{\{n=N_i\}} \mathrm{e}^{-2rT_i}\left[\pi_\epsilon'(s_n^{m,J}(x) - K_i), 0, \dots, 0\right]\left(\frac{1}{M}\sum_{m=1}^{M}\pi_\epsilon(s_n^{m,J} - K_i) - \mathrm{e}^{rT_i} C_{\text{obs}}^i\right)
\end{aligned}$$

$$n = N-1, N-2, \dots, 1, \quad i \in \{1, \dots, I\}, \quad m = 1, \dots, M.$$

Applying the previous equations, which correspond to (4.19), and in combination with the existence of the inverse of $\nabla_y C(y, x)$ by Lemma 4.2.3, one uses Lemma 4.2.4 to obtain the gradient of $f(x) := \hat{f}_{M,\Delta t,\epsilon}(y(x), x)$ by

$$\begin{aligned}
\nabla f(x) &= \nabla_x L(y, x, \lambda) \\
&= \nabla_x \hat{f}_{M,\Delta t,\epsilon}(y, x) + \sum_{n=1}^{N} \lambda_n \nabla_x c_n(y, x) \\
&= \sum_{n=1}^{N}\Bigg((\lambda_n^0)^\top\left[\frac{\partial}{\partial x} a_\epsilon(x, y_{n-1}^0)\Delta t_{n-1} + \frac{\partial}{\partial x}\left(b_\epsilon(x, y_{n-1}^0)\Delta W_{n-1}\right)\right] \\
&\quad + \sum_{j=1}^{J}(\lambda_n^j)^\top\Bigg[\left(\eta\frac{\partial}{\partial x}\underline{a}_\epsilon(x, y_n^{j-1}) + (1-\eta)\frac{\partial}{\partial x}\underline{a}_\epsilon(x, y_{n-1}^J)\right)\Delta t_n \\
&\quad + \left(\vartheta\frac{\partial}{\partial x}\left(b_\epsilon(x, y_n^{j-1})\Delta W_n\right) + (1-\vartheta)\frac{\partial}{\partial x}\left(b_\epsilon(x, y_{n-1}^J)\Delta W_n\right)\right)\Bigg]\Bigg),
\end{aligned}$$

since $\nabla_x \hat{f}_{M,\Delta t,\epsilon}(y, x)$ vanishes. This completes the proof and establishes formula (4.17) for the gradient. □

An economic interpretation of the adjoint equation can be made in the sense of Duffie et al. (1996), El Karoui et al. (1997) or Bender and Kohlmann (2008). Thus, the dynamics of the value of the replicating portfolio $\bar{\xi}$ as in Definition 2.3.6 are given by a backward stochastic dynamics such as the adjoint equation. For the case of robust replication see Bender et al.

(2008) or Tikanmäki (2013).

The numerical effort of this method consists of d backward solves compared to P forward solves for the finite difference method of Lemma 4.2.1 in either instance for N time steps and M trajectories. In particular, when P dominates the dimension d, for instance when the underlying model parameters are chosen time dependent (see later Section 5.2), the adjoint technique reduces the numerical effort substantially. It is all about computational efficiency. This will be explained in Section 4.3 in detail.

Detailed overviews on the adjoint technique in several fields of research can be found in Achdou and Pironneau (2005), Arridge and Schweiger (1998), Bosse et al. (2014), Capriotti and Giles (2012), Giles and Pierce (2000), Giles and Süli (2002), Jameson (1988), Käbe et al. (2009), Lörx (2013), Marchuk et al. (1996) and Pironneau (2007). Parameter identification methods arising in option pricing can be found in Schulze (2002). Moreover, some issues and results on calibration of financial market models including jumps using model reduction methods can be found in Sachs and Schu (2008), Sachs and Schu (2010), Sachs and Schu (2013) and Sachs et al. (2014).

4.2.2 Special cases of Theorem 4.2.6

Theorem 4.2.7 (The Case of Euler-Maruyama Discretization)**.** *Let Assumption 4.2.5 be true and $\hat{f}_{M,\Delta t,\epsilon} : \mathbb{R}^{dN+P} \to \mathbb{R}$ be the real valued objective function for maturity T_i and strike price K_i for all $i = 1, \dots, I$*

$$\begin{aligned} \hat{f}_{M,\Delta t,\epsilon}(y,x) &:= \frac{1}{2}\sum_{i=1}^{I}\left(\hat{C}^i_{M,\Delta t,\epsilon}(y,x) - C^i_{\text{obs}}\right)^2 \\ &= \frac{1}{2}\sum_{i=1}^{I}\left(\mathrm{e}^{-rT_i}\frac{1}{M}\sum_{m=1}^{M}\left(\pi_\epsilon(s^m_{N_i} - K_i)\right) - C^i_{\text{obs}}\right)^2 \end{aligned} \tag{4.27}$$

with Euler-Maruyama discretization

$$\begin{aligned} &y^m_{n+1} = y^m_n + a_\epsilon(x, y^m_n)\Delta t + b_\epsilon(x, y^m_n)\Delta W^m_n, \\ &y^m_0 = Y_0, \quad n = 0, \dots, N-1, \quad m = 1, \dots, M, \quad N := \max_{i=1,\dots,I} N_i. \end{aligned} \tag{4.28}$$

Then the gradient of $f(x) := \hat{f}_{M,\Delta t,\epsilon}(y(x), x)$ for all $x \in X$ can be computed regardless of $y \in \mathbb{R}^{dN}$, maturity T_i and strike price K_i $i = 1, \dots, I$ via

$$\nabla f(x) = \frac{1}{M}\sum_{m=1}^{M}\sum_{n=1}^{N}(\lambda^m_n)^\top\left[\frac{\partial}{\partial x}a_\epsilon(x, y^m_{n-1})\Delta t_{n-1} + \frac{\partial}{\partial x}\left(b_\epsilon(x, y^m_{n-1})\Delta W^m_{n-1}\right)\right], \tag{4.29}$$

where $\lambda_n^m \in \mathbb{R}^d$ *results from the backward adjoint equation*

$$\lambda_n^m = \left[I + \frac{\partial}{\partial y} a_\epsilon(x, y_n^m) \Delta t_n + \frac{\partial}{\partial y} \left(b_\epsilon(x, y_n^m) \Delta W_n^m \right) \right]^\top \lambda_{n+1}^m \tag{4.30}$$
$$+ \mathbb{1}_{\{n=N_i\}} \, \mathrm{e}^{-2rT_i} \left[\pi_\epsilon'(s_n^m(x) - K_i), 0, \ldots, 0 \right] \left(\frac{1}{M} \sum_{m=1}^{M} \pi_\epsilon(s_n^m - K_i) - \mathrm{e}^{rT_i} \, C_{\text{obs}}^i \right)$$

$$n = N-1, N-2, \ldots, 1, \quad i \in \{1, 2, \ldots, I\}, \quad m = 1, \ldots, M,$$
$$\lambda_N^m = \mathrm{e}^{-2rT_I} \left[\pi_\epsilon'(s_N^m - K_I), 0, \ldots, 0 \right] \left(\tfrac{1}{M} \textstyle\sum_{m=1}^{M} \pi_\epsilon(s_N^m - K_I) - \mathrm{e}^{rT_I} \, C_{\text{obs}}^I \right) \in \mathbb{R}^d.$$

This adjoint-based Monte-Carlo method is described in the simpler case for the calibration of implied volatilities in a similar way by Cangiani (2000).

Example 4.2.8 (Black-Scholes Model)**.** *The dimensions of the calibration problem in the case of the Black-Scholes model* (2.39) *are given as* $d = P = m = 1$*, the parameter space as* $X = (0, \infty) \subset \mathbb{R}$*, the calibration parameter is denoted as* $x = \sigma$ *and* $y = y_1$*. The smoothed drift function is given by* $a_\epsilon : X \times \mathbb{R} \to \mathbb{R}$ *and the smoothed volatility function by* $b_\epsilon : X \times \mathbb{R} \to \mathbb{R}$

$$a_\epsilon(x, y) := r y_1$$
$$b_\epsilon(x, y) := \sigma y_1.$$

In order to implement the gradient (4.17) *and the adjoint equations* (4.18) *and* (4.19)*, the following partial derivatives are necessary:*

$$\frac{\partial}{\partial x} a_\epsilon(x, y) = 0$$
$$\frac{\partial}{\partial x} b_\epsilon(x, y) = y_1$$
$$\frac{\partial}{\partial y} a_\epsilon(x, y) = r$$
$$\frac{\partial}{\partial y} b_\epsilon(x, y) = \sigma.$$

Example 4.2.9 (Heston Model)**.** *The dimensions of the calibration problem in the case of the Heston stochastic volatility model* (2.43) *are given as* $d = m = 2$ *and* $P = 5$*, the parameter space as* $X = (0, \infty) \times (0, \infty) \times \mathbb{R} \times (0, \infty) \times [-1, 1] \subset \mathbb{R}^5$*, the calibration parameters are denoted as* $x = (v_0, \kappa, \theta, \sigma, \rho)^\top \in X$*. By applying the simple transformation* $\tilde{v}_t := v_t / v_0$ *and a Cholesky decomposition (Golub and van Loan, 1996, Theorem 4.2.5) of the correlation matrix one obtains the equivalent model dynamics*

$$\begin{aligned} dS_t &= r(t) S_t dt + \sqrt{v_0 \tilde{v}_t^+} S_t d\tilde{W}_t^1, \; S_0 \\ d\tilde{v}_t &= \kappa \left(\frac{\theta}{v_0} - \tilde{v}_t^+ \right) dt + \frac{\sigma}{\sqrt{v_0}} \sqrt{\tilde{v}_t^+} \left(\rho d\tilde{W}_t^1 + \sqrt{1 - \rho^2} d\tilde{W}_t^2 \right), \; \tilde{v}_0 = 1, \end{aligned}$$

where $(\tilde{W}_t^1)_{t\in[0,T]}$ *and* $(\tilde{W}_t^2)_{t\in[0,T]}$ *are uncorrelated Wiener processes. With* $y = (y_1, y_2)^\top = (S, \tilde{v})^\top$ *the drift function* $a : X \times \mathbb{R}^2 \to \mathbb{R}^2$ *and the volatility function* $b : X \times \mathbb{R}^2 \to \mathbb{R}^2 \times \mathbb{R}^2$ *are defined as*

$$a(x,y) := \begin{pmatrix} r(t)y_1 \\ \kappa\left(\frac{\theta}{v_0} - y_2^+\right) \end{pmatrix}$$
$$b(x,y) := \begin{pmatrix} v_0 y_2^+ y_1 & 0 \\ \sigma\rho & \sigma\sqrt{1-\rho^2} \end{pmatrix}.$$

Obvioulsy, the maps a *and* b *are not continuously differentiable on* $X \times \mathbb{R}^2$*. Therefore, the smoothing function defined in Remark 3.3.2 is used to eliminate the non-differentiability and to fulfill the smoothness Assumption 4.2.5*

$$a_\epsilon(x,y) := \begin{pmatrix} r(t)y_1 \\ \kappa\left(\frac{\theta}{v_0} - \pi_\epsilon(y_2)\right) \end{pmatrix}$$
$$b_\epsilon(x,y) := \begin{pmatrix} v_0 \pi_\epsilon(y_2) y_1 & 0 \\ \sigma\rho & \sigma\sqrt{1-\rho^2} \end{pmatrix}.$$

For the implementation of the gradient (4.17) *and the adjoint equations* (4.18) *and* (4.19)*, the following partial derivatives are necessary:*

$$\frac{\partial}{\partial x} a_\epsilon(x,y) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -\kappa\frac{\theta}{v_0^2} & \frac{\theta}{v_0} - \pi_\epsilon(y_2) & \frac{\kappa}{v_0} & 0 & 0 \end{pmatrix}$$
$$\frac{\partial}{\partial x} b_\epsilon(x,y)\Delta W = \begin{pmatrix} \pi_\epsilon(y_2) y_1 \Delta W^1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho\Delta W^1 + \sqrt{1-\rho^2}\Delta W^2 & \sigma\left(\Delta W^1 - \left(\frac{\rho}{\sqrt{1-\rho^2}}\right)\Delta W^2\right) \end{pmatrix}$$
$$\frac{\partial}{\partial y} a_\epsilon(x,y) = \begin{pmatrix} r(t) & 0 \\ 0 & -\kappa\pi_\epsilon'(y_2) \end{pmatrix}$$
$$\frac{\partial}{\partial y} b_\epsilon(x,y)\Delta W = \begin{pmatrix} v_0\pi_\epsilon(y_2)\Delta W^1 & v_0\pi_\epsilon'(y_2) y_1 \Delta W^1 \\ 0 & 0 \end{pmatrix}.$$

For some further information about the calibration of the Heston stochastic volatility model one is emphasized to read Käbe et al. (2009), Kahl and Jäckel (2006) or Mikhailov and Nögel (2003).

4.3 Computational Complexitiy and Efficiency

Any sophisticated optimization algorithm needs at least gradient information. Thereby, such methods require an efficient computation of the first derivative. Unless analytically evalutation of the gradient is possible, the simplest numerical way is the use of the finite difference method (FDM) (4.4). However, since the Monte-Carlo method to calculate the objective function is very expensive on matters such as computational time and occupancy of working memory, FDM is not recommended.

The analysis of the computational effort of the adjoint approach developed in Theorem 4.2.6 and its comparison to the derivative evalutation via finite differences is addressed in this section. Therefore, it is crucial to remark the computational complexity of the computation of the objective gradient via the finite difference approach.

Remark 4.3.1 (Finite Differences Method). *The calculation of $\hat{f}_{M,\Delta t,\epsilon}$ due to Lemma 4.2.1 requires for every approximated path $s^{m,J}_{N_i,\epsilon}(x)$ of the underlying d-dimensional SDE $N = \max_{i\in I} N_i$ calculations of the stochastic scheme, that needs Q flops for every timestep. To evaluate the expected value via Monte-Carlo method, M of such paths have to be simulated. Moreover, the gradient calculation according to formula* (4.4) *needs the estimation of $f(x + h\mathbf{e}_p)$ and in addition $f(x - h\mathbf{e}_p)$ for $p = 1, \dots, P$. In the aggregate this all is pooled together to a total flop count of order*

$$\mathcal{O}((P+1)dQMN). \tag{4.31}$$

Thus, the computational effort of the FDM scales linearly in the number of calibration parameters P. Hence, it would be worthwhile to have a method which is less expensive. This will be provided by the use of more sophisticated schemes combined with the adjoint approach introduced with Theorem 4.2.6 as it is pointed out in the following remark.

Remark 4.3.2 (Adjoint Technique). *Let p_j, $j = 1, \dots, P$, be the calibration parameters and $u_j := |\mathrm{supp}(p_j)|$, $j = 1, \dots, P$. Then $U = \sum_{j=1}^{P} u_j$ is the total number of timesteps affected by the calibration routine. Equipped with this, the calculation of the forward mode as in Theorem 4.2.6 requires for every approximated path $s^{m,J}_{N_i,\epsilon}(x)$ as in above Remark 4.3.1 $\mathcal{O}(dQMN)$ flops.*
The backward mode 4.19 needs the same amount of flops as the forward mode plus some additions for the maturities. This results in $\mathcal{O}(dQMN) + I$ flops. Furthermore, the gradient calculation due to equation (4.17) *needs $\mathcal{O}(dQMU)$ flops.*
In the aggregate this all is pooled together to a total flop count of order

$$\mathcal{O}(dQM(2N+U)) + I. \tag{4.32}$$

Corollary 4.3.3 (Computational Reduction of the Adjoint Technique). *If on uses "local support" of the calibration parameters ($U = pN$), then according to Remark 4.3.2 the gradient computation via adjoint technique is of order*

$$\mathcal{O}(dQMN) \tag{4.33}$$

total flops.

So the key point of this chapter is that the computational effort of gradient calculation via the adjoint technique is independent of the number of calibration parameters P. Subsequent numerical results confirm this advantage of the Monte-Carlo based adjoint technique with regard to the computation time. For this see later Section 5.2.2.

4.4 Alternative Approaches to Reduce Computational Time

Generating and Storing Random Numbers, Variance Reduction and Parallelization

Generating and storing random numbers is a crucial issue in term of computational efficiency (cf. Gentle (2004)). The simulation and storing of Wiener increments requires a lot of system memory, i.e. random access memory (RAM), as one has to store ΔW_n^m for every simulation $m = 1, \dots, M$ and time step $n = 1, \dots, N$. Therefore, MN independently and normally distributed random numbers have to be stored. Considering for instance the pricing of an Asian call option (2.59) with a two-dimensional stochastic volatility model 2.43, maturity of 2 years, a time step every bank day and $1,000,000$ simulations requires the simulation of one billion normally distributed random numbers. This is approximately 8 GB stored in double precision and will exhaust the random access memory of a desktop PC. However, since the transfer rate of reading on a hard drive is limited, this is no alternative. This is precisely why regenerating random numbers outperformes generating and storing as shown in Section 6.3 of Käbe (2010). To this end, one needs random number generators that exactly replicate sequences of normally distributed samples.

For the further acceleration of computation time of Monte-Carlo methods via parallelization on graphics processing units (GPUs) the reader will be referred to following articles Bradley et al. (2011), Bradley et al. (2012), Dai et al. (2010), Fatone et al. (2012), Joshi (2010), Lee et al. (2010), Peng et al. (2011) and Kolb and Pharr (2005). The parallel computing platform and application programming interface (API) developed by NVIDIA© named Compute Unified Device Architecture (CUDA©)[1] is used in most of the articles. GPUs are also usable for quantization methods, see Pagès and Wilbertz (2010), Wilbertz (2005) or more detailed in Pagès and Wilbertz (2012).

Hence the variance $\hat{\sigma}_M(X)$ dominates the Monte-Carlo error (see Definition 2.2.24), variance reduction methods, such as antithetic variates, control variates, importance sampling or Latin hypercube sampling, arises. An introduction to variance reduction methods and a survey of the computational improvement of them is for instance given in Boyle et al. (1997) or Chapter 4 of Glasserman (2004).

Furthermore, multilevel Monte Carlo methods (abbreviated "MLMC") were developed by Giles and Szpruch (2013), Giles et al. (2009), Giles (2015) and modified by Debrabant and Rößler (2015) to handle multi-dimensional SDEs and reduce the computational complexity.

[1] NVIDIA Developer https://developer.nvidia.com/cuda-zone.

Chapter 5

Numerical Results

Nature is wont to hide herself.
— Heraclitus

In this chapter numerical results are presented and illustrated. It is split into two parts. First, Section 5.1 illustrates the improvement of the predictor-corrector scheme concerning the computational performance and the rate of convergence. Section 5.2 is then devoted to a detailed analysis of the speed up obtained for the calibration of the Heston model by applying the adjoint technique introduced in the main Theorem 4.2.6.

5.1 Monte-Carlo Schemes: Rate of Convergence

Some results concerning the accuracy of an approximate solution are given, which are calculated via the predictor-corrector schemes and compared with the Euler-Maruyama scheme and the Milstein scheme as in Higham (2001). Henceforth some numerical tests emphasize that $\eta = \vartheta = \frac{1}{2}$ is an appropriate choice, this leads to a stochastic Heun's scheme which is a predictor-corrector scheme with the explicit Euler-Maruyama scheme as predictor and the trapezoidal method of an implicit Euler-Maruyama scheme as corrector.

To compute the results of this subsection the following parameters are used, if not otherwise specified, $M = 100,000$, maturity $T = 1$, $\mu = 0.08$, volatility $\sigma = 0.16$, $\lambda = 0.5$, $p = 0.55$, $\eta_1 = 10$, $\eta_2 = 15$, $\mu_J = -0.05$, $\sigma_J = 0.09$, $\nu_0 = 0.4$, $\theta = 0.25$, $\kappa = 0.5$, $\sigma_\nu = 1.5$ and correlation coefficient $\rho = -0.75$.

All codes are implemented in MATLAB® and all numerical test runs are realized on a desktop PC with an Intel® Xeon® CPU E5620 with 2.4GHz and 2GB random access memory (RAM).

Previous research by Giles et al. (2013) and Schmitz Abe and Shaw (2005) has analysed the computational efficiency when using the Euler-Maruyama discretization and also demonstrates an improved efficiency using the Milstein discretization with the improved strong convergence. An extensive comparison of these schemes with different predictor-corrector schemes is presented here. In particular, Table 5.6 clearly illustrates the improved strong convergence of the predictor corrector schemes for the case of the Black-Scholes model (2.39).

Black-Scholes Model					
Δt	Euler-Maruyama scheme	Milstein scheme	predictor-corrector schemes		
			Preco1E1	Preco2E1	Preco1M1
1	1.0205	1.0203	0.0585	0.02918	0.0585
$\frac{1}{2}$	0.5355	0.5354	0.0275	0.00933	0.0275
$\frac{1}{4}$	0.2746	0.2745	0.0132	0.00329	0.0132
$\frac{1}{8}$	0.1391	0.1390	0.0065	0.00132	0.0065
$\frac{1}{16}$	0.0701	0.0700	0.0032	0.00051	0.0032
$\frac{1}{32}$	0.0352	0.0351	0.0016	0.00025	0.0016
$\frac{1}{64}$	0.0176	0.0176	0.0008	0.00012	0.0008
$\frac{1}{128}$	0.0088	0.0088	0.0004	0.00006	0.0004

Table 5.1: Black-Scholes: weak approximation error $\left|\mathbb{E}\left[\max(X_T - K, 0)\right] - \mathbb{E}\left[\max(Y_T^{\Delta t} - K, 0)\right]\right|$ with 1,000,000 trajectories and volatility $\sigma = 0.05$.

Furthermore, Tables 5.1, 5.2 and 5.3 show the comparsion of the weak error ϵ_W for different discretization schemes regarding different volatilities $\sigma = 0.05$, $\sigma = 0.5$ and $\sigma = 0.95$. Additionally, in Figures 5.1, 5.5 and 5.3 log-log plots of the weak convergence error ϵ_W as defined in Definition 2.2.21 versus different time steps $\Delta t = \left\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512}, \frac{1}{1024}\right\}$ are graphed for Euler-Maruyama, Milstein and the following predictor-corrector schemes: predictor (Euler-Maruyama) and one-time and two-times corrector step, predictor (Milstein) and one-time and two-times corrector step and finally predictor (Euler-Maruyama and Milstein) and one-time corrector step of order 1.5 or the θ-scheme proposed by Schmitz Abe (2010). These are labeled "Euler-M", "Milstein", "Preco1E1", "Preco2E1", "Preco1M1", "Preco2M1", "Preco1E1.5" and "Preco2M1.5", "Preco1Mθ", respectively. Note that in Figures 5.1, 5.5 and 5.3 the achieved numerical weak order β describes the slope of the lines and C describes the point where the slope and the y-axis intersect. Reference slopes for $\beta = 1/2$ and $\beta = 1$ are plotted in each figure.

However, a closer look on Tables 5.1, 5.2 and 5.3 reveals that the weak error ϵ_W becomes slightly worse when increasing the volatility. Nevertheless, the predictor-corrector schemes seem to be the best approximation in terms of smaller errors and computational efficiency, especially in the case of a Milstein scheme as predictor and a single corrector step.

In real market data the volatility σ dominates the deterministic drift term represented by the risk-free interest rate r. In Table 5.4 the weak error ϵ_W of a European call option on the Dax index from March 2014 are presented. The riskfree interest rate is chosen as $r = 0.01$, the implied volatility as $\sigma = 0.16$, the initial stock price as $S_0 = 9,500$ and the strike price as $K = 10,000$ with maturity $T = 1$. Figure 5.4 graphically illustrates these results.

To further reinforce the above results Table 5.5 and the corresponding Figure 5.5 show the

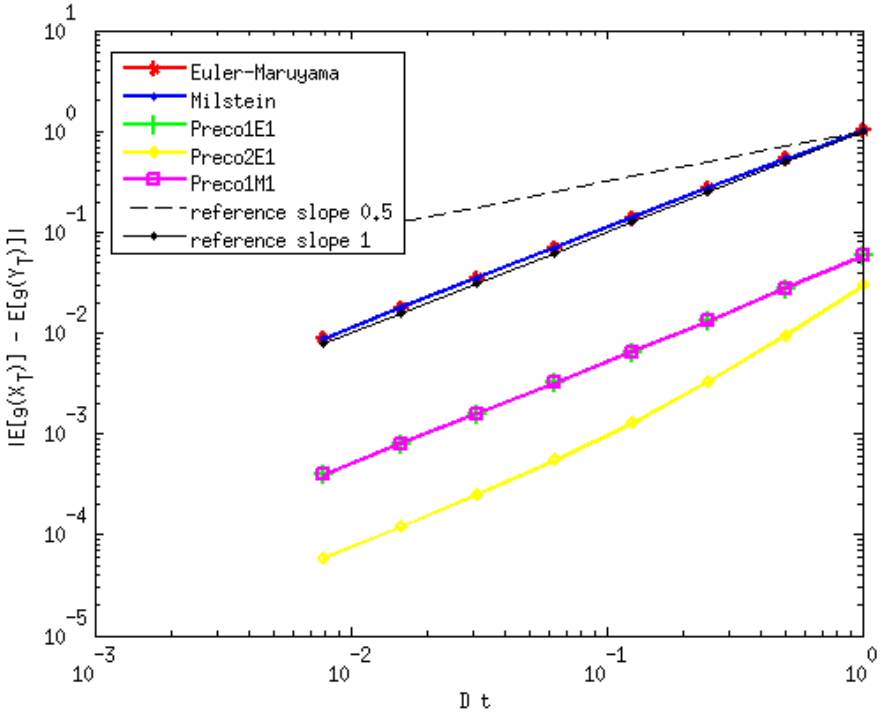

Figure 5.1: Plotted log **weak approximation error against** log **step size** Δt **of Table 5.1.**

competitive advantage of the predictor-corrector schemes over the Euler-Maruyama and the Milstein scheme in the case of the jump-diffusion model (2.40) with Merton jump size (2.41).

	Black-Scholes Model				
Δt	Euler-Maruyama scheme	Milstein scheme	predictor-corrector schemes		
			Preco1E1	Preco2E1	Preco1M1
1	3.7423	2.4948	0.9131	0.6796	0.1703
$\frac{1}{2}$	2.1940	1.4627	0.5362	0.4508	0.1470
$\frac{1}{4}$	1.0671	0.7681	0.2874	0.2481	0.0804
$\frac{1}{8}$	0.5993	0.3995	0.1489	0.1277	0.0399
$\frac{1}{16}$	0.3040	0.2027	0.0754	0.0650	0.0197
$\frac{1}{32}$	0.1525	0.1017	0.0379	0.0325	0.0097
$\frac{1}{64}$	0.0764	0.0509	0.0187	0.0165	0.0048
$\frac{1}{128}$	0.0383	0.0255	0.0094	0.0084	0.0024

Table 5.2: Black-Scholes: weak approximation error $\left|\mathbb{E}\left[\max(X_T - K, 0)\right] - \mathbb{E}\left[\max(Y_T^{\Delta t} - K, 0)\right]\right|$ with 1,000,000 trajectories and volatility $\sigma = 0.5$.

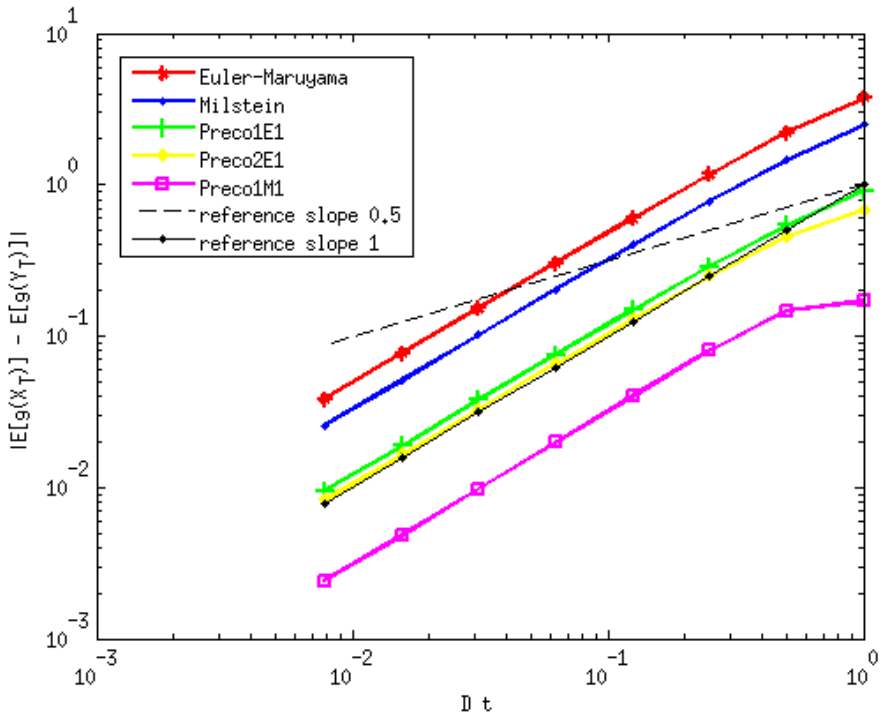

Figure 5.2: Plotted log weak approximation error against log step size Δt of Table 5.2.

	Black-Scholes Model					
Δt	Euler-Maruyama scheme	Milstein scheme	predictor-corrector schemes			
			Preco1E1	Preco2E1	Preco1M1	Preco1Mθ
$\frac{1}{4}$	2.5479	1.0068	1.2424	0.9240	1.1994	0.2200
$\frac{1}{8}$	1.5155	0.5992	0.6511	0.4188	0.5813	0.1928
$\frac{1}{16}$	0.7175	0.3221	0.3383	0.1891	0.2833	0.1162
$\frac{1}{32}$	0.3309	0.1662	0.1729	0.0874	0.1383	0.0632
$\frac{1}{64}$	0.1608	0.0816	0.0828	0.0426	0.0683	0.0300
$\frac{1}{128}$	0.0797	0.0414	0.0415	0.0210	0.0340	0.0156
$\frac{1}{256}$	0.0390	0.0205	0.0217	0.0101	0.0170	0.0076

Table 5.3: Black-Scholes: weak approximation error $\left|\mathbb{E}\left[\max(X_T - K, 0)\right] - \mathbb{E}\left[\max(Y_T^{\Delta t} - K, 0)\right]\right|$ with 1,000,000 trajectories and volatility $\sigma = 0.95$.

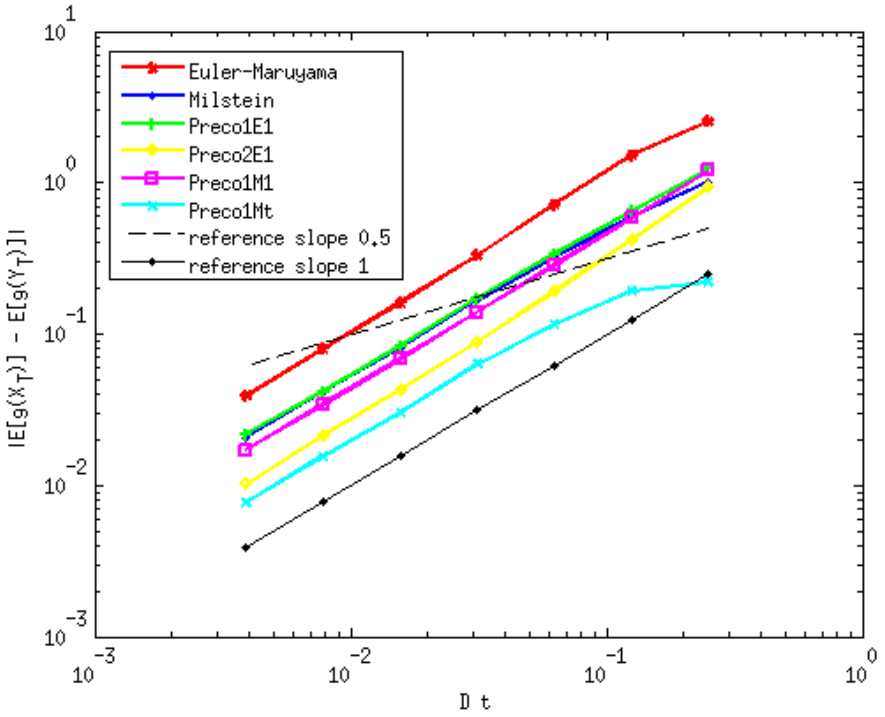

Figure 5.3: Plotted log weak approximation error against log step size Δt of Table 5.3.

	Black-Scholes Model of an index option (DAX)				
Δt	Euler-Maruyama scheme	Milstein scheme	predictor-corrector schemes		
			Preco1E1	Preco2E1	Preco1M1
1	12.0723	3.9976	1.9039	0.6497	0.7918
$\frac{1}{2}$	7.7697	2.5574	1.1093	0.3943	0.4854
$\frac{1}{4}$	4.1655	1.4071	0.5963	0.2132	0.2628
$\frac{1}{8}$	2.1749	0.7346	0.3060	0.1104	0.5813
$\frac{1}{16}$	1.0857	0.3747	0.1569	0.0561	0.0691
$\frac{1}{32}$	0.5336	0.1895	0.0802	0.0283	0.0349
$\frac{1}{64}$	0.2617	0.0945	0.0396	0.0141	0.0174
$\frac{1}{128}$	0.1288	0.0473	0.0197	0.0071	0.0087
$\frac{1}{256}$	0.0600	0.0237	0.0100	0.0035	0.0044

Table 5.4: Black-Scholes: weak approximation error $\left|\mathbb{E}\left[\max(X_T - K, 0)\right] - \mathbb{E}\left[\max(Y_T^{\Delta t} - K, 0)\right]\right|$ of an index option (DAX) with 1.000.000 trajectories.

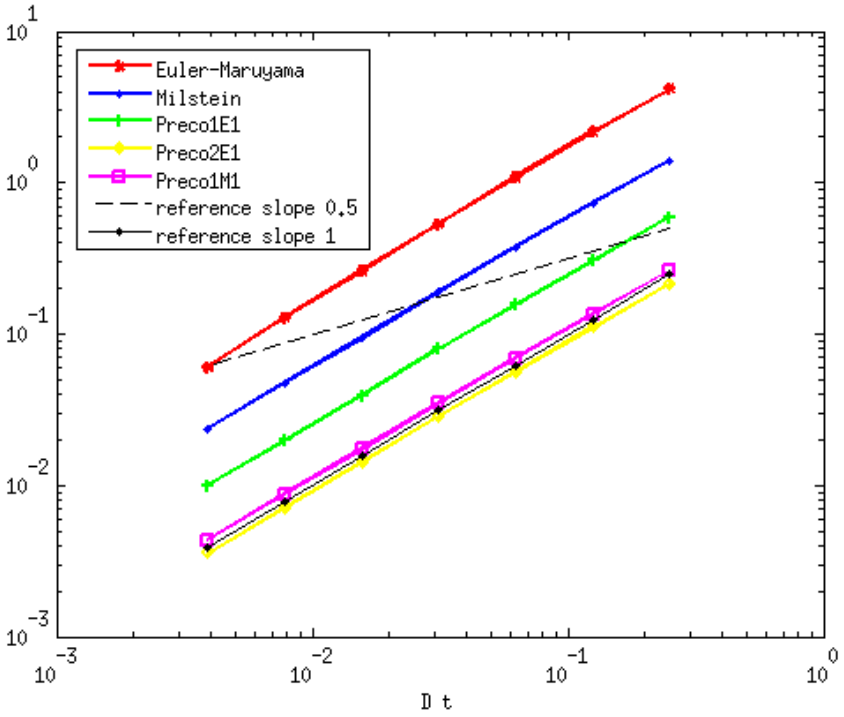

Figure 5.4: Plotted log weak approximation error against log step size Δt of Table 5.4.

	JUMP DIFFUSION (MERTON)				
Δt	Euler-Maruyama scheme	Milstein scheme	predictor-corrector schemes		
			Preco1E1	Preco2E1	Preco1M1
$\frac{1}{2}$	1.0974	0.7399	0.1824	0.1294	0.1542
$\frac{1}{4}$	0.7677	0.3954	0.1119	0.0844	0.0950
$\frac{1}{8}$	0.5411	0.2147	0.0720	0.0565	0.0610
$\frac{1}{16}$	0.3825	0.1208	0.0482	0.0385	0.0404
$\frac{1}{32}$	0.2699	0.0713	0.0330	0.0265	0.0274
$\frac{1}{64}$	0.1910	0.0444	0.0229	0.0185	0.0189
$\frac{1}{128}$	0.1353	0.0288	0.0161	0.0129	0.0132
$\frac{1}{256}$	0.0956	0.0193	0.0113	0.0091	0.0092
$\frac{1}{512}$	0.0674	0.0132	0.0080	0.0064	0.0065
$\frac{1}{1024}$	0.0477	0.0092	0.0056	0.0045	0.0045

Table 5.5: Jump diffusion (Merton): weak approximation error $\left|\mathbb{E}\left[\max(X_T - K, 0)\right] - \mathbb{E}\left[\max(Y_T^{\Delta t} - K, 0)\right]\right|$ with 1,000,000 trajectories.

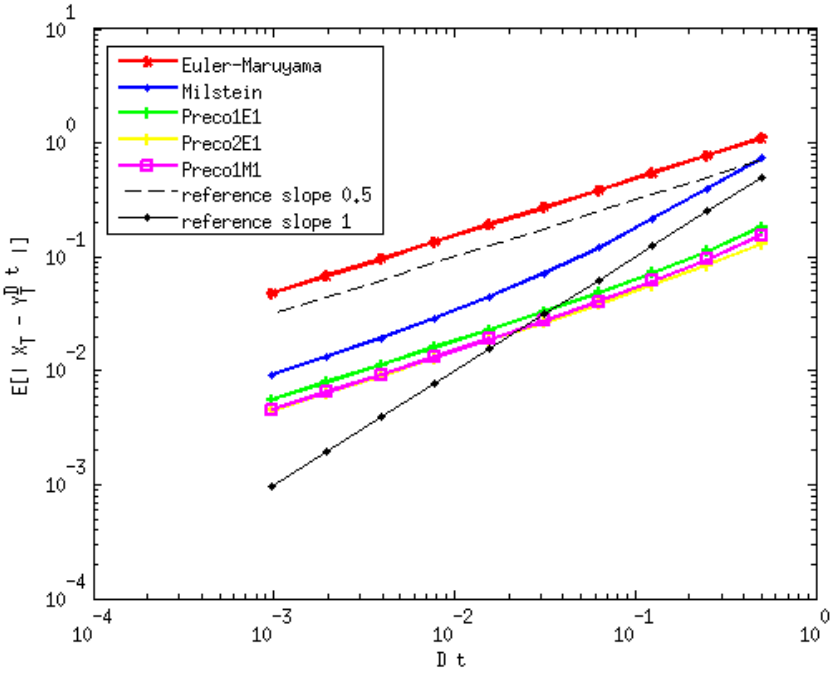

Figure 5.5: Plotted log weak approximation error against log step size Δt of Table 5.5.

Black-Scholes Model					
Δt	Euler-Maruyama scheme	Milstein scheme	predictor-corrector schemes		
			Preco1E1	Preco2E1	Preco1M1
$\frac{1}{2}$	2.4771	0.8009	0.2211	0.1340	0.1133
$\frac{1}{4}$	1.8196	0.4255	0.1523	0.0805	0.0687
$\frac{1}{8}$	1.3142	0.2207	0.1070	0.0460	0.0393
$\frac{1}{16}$	0.9408	0.1125	0.0761	0.0253	0.0213
$\frac{1}{32}$	0.6728	0.0568	0.0543	0.0139	0.0112
$\frac{1}{64}$	0.4763	0.0285	0.0384	0.0078	0.0058
$\frac{1}{128}$	0.3378	0.0142	0.0272	0.0046	0.0029
$\frac{1}{256}$	0.2397	0.0072	0.0193	0.0028	0.0015

Table 5.6: Black-Scholes: strong approximation error $\mathbb{E}\left[\left|X_T - Y_T^{\Delta t}\right|\right]$ with 100,000 trajectories.

A tool how to estimate the rate of strong convergence γ or weak convergence β numerically by using the logarithm of the equation (2.29) in Definition 2.2.20 or the equation (2.30) in Definition 2.2.21 is given for small step sizes Δ_i in what follows

$$\begin{aligned} \epsilon_S(\Delta_i) &= \mathbb{E}\left[\left|X_T - Y_T^{\Delta_i}\right|\right] = C\Delta_i^{\gamma} \\ \Rightarrow \quad \log(\epsilon_S(\Delta_i)) &= \log\left(\mathbb{E}\left[\left|X_T - Y_T^{\Delta_i}\right|\right]\right) = \log(C) + \gamma\log(\Delta_i) \end{aligned}$$

$i = 1, \ldots, R$. In the case of the results in Table 5.6 one has $r = 8 > 2 = |\{\gamma, \log(C)\}|$. So one obtains an overdetermined system of linear equations wherefore one has to solve a linear least squares regression by choosing the following model function to get the positive constant C and the strong convergence order γ or weak convergence order β, respectively,

$$m(t_i) = x_0 + x_1 t_i = \log(C) + \gamma\log(\Delta_i) = \log(\epsilon_S(\Delta_i)), \quad i = 1, \ldots, R.$$

This is equivalent to the following linear least squares problem

$$\min_{x_0, x_1} \left\| \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_R \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} - \begin{pmatrix} b_1 \\ \vdots \\ b_R \end{pmatrix} \right\|_2 = \min_{x_0, x_1} \|Ax - b\|_2 , \tag{5.1}$$

where $t_i := \log(\Delta_i)$ and $b_i := \log(\epsilon_S(\Delta_i))$ for all $i = 1, \ldots, R$. In this case the solutions $x_0 = \log(C)$ and $x_1 = \gamma$ of (5.1) are estimated by

$$\begin{aligned} x_0 &= \bar{b} - x_1\bar{t} \\ x_1 &= \frac{\sum_{i=1}^n (b_i - \bar{b})(t_i - \bar{t})}{\sum_{i=1}^n (t_i - \bar{t})^2}, \end{aligned}$$

where $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$ and $\bar{b} = \frac{1}{n}\sum_{i=1}^{n} b_i$ are arithmetic mean values (see also Nocedal and Wright (1999)).

Numerical results of estimating the rates of convergence, i.e, γ and β and the positive constant C for several financial market models (e.g. the Black-Scholes model (2.39), the Heston model (2.43) and some jump-diffusion models (2.40)) are listed in Table 5.7. This results also numerically confirm the theoretical rates of convergence presented in Section 3.3.2 and impressively demonstrates the advantage of the stochastic predictor-corrector scheme.

Figure 5.6 corresponds to Table 5.6 and shows a log-log plot of the strong convergence error ϵ_S as defined in Definition 2.2.20 versus different time steps $\Delta t = \left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}\right\}$ for Euler-Maruyama, Milstein and the following predictor-corrector schemes: predictor (Euler-Maruyama) and one-time and two-times corrector step, predictor (Milstein) and one-time and two-times corrector step and finally predictor (Euler-Maruyama and Milstein) and one-time corrector step of order 1.5. These are labeled as above.

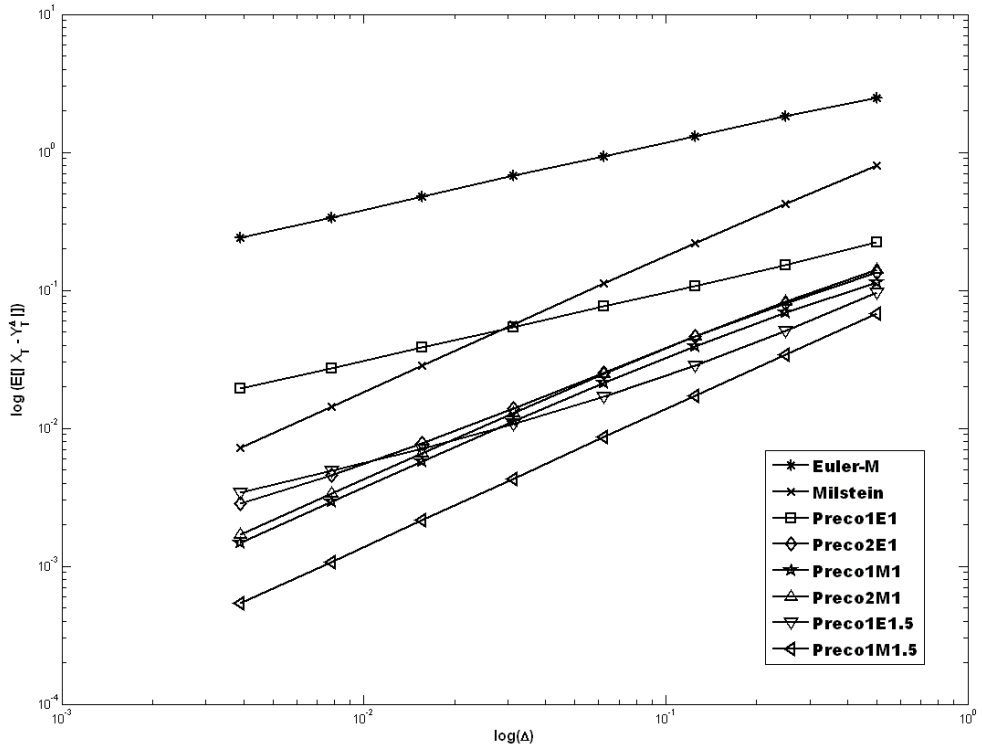

Figure 5.6: Plotted log **strong approximation error against** log **step size** Δt**.**

Note that in Figure 5.6 the achieved numerical strong order γ is described by the slope of the lines and the constant C describes the point where the slope and the y-axis intersect.

Model			strong			weak			
Model	**Scheme**		γ	C	$\epsilon_S(1/2^{10})$	β	C	$\epsilon_W(1/2^{10})$	time/quotient
Black-Scholes	Euler-M	$\alpha = 0,\ J = 0$	0.4835	3.5474	$1.1941e^{-1}$	0.9547	0.7936	$9.488e^{-4}$	100%
	Milstein	$\alpha = 0.7,\ J = 0$	0.5263	1.3249	$3.5839e^{-2}$	0.9679	0.5759	$7.0832e^{-4}$	113%
	Milstein	$\alpha = 1,\ J = 0$	0.9754	1.6404	$1.8274e^{-3}$	0.9973	0.8336	$8.2324e^{-4}$	112%
	Preco1E1	$\alpha = 0,\ J = 1$	0.4995	0.3065	$9.6040e^{-3}$	0.9802	0.1644	$1.7474e^{-4}$	120%
	Preco2E1	$\alpha = 0,\ J = 2$	0.8117	0.2401	$1.2467e^{-3}$	0.9925	0.1628	$1.6432e^{-4}$	140%
	Preco1M1	$\alpha = 1,\ J = 1$	0.9022	0.2397	$3.7246e^{-4}$	1.0178	0.0164	$1.3373e^{-5}$	131%
	Preco2M1	$\alpha = 1,\ J = 2$	0.9179	0.2946	$4.2407e^{-4}$	0.9853	0.0176	$1.9407e^{-5}$	152%
Heston	Euler-M	$\alpha = 0,\ J = 0$	0.4752	4.2741		0.9762	1.104		
	Milstein	$\alpha = 1,\ J = 0$	0.9654	1.8434		0.9894	0.8624		
	Preco1E1	$\alpha = 0,\ J = 1$	0.4958	0.4503		0.9842	0.2266		
	Preco2E1	$\alpha = 0,\ J = 2$	0.8238	0.4158		0.9905	0.2317		
Jump-diffusions (Kou)	Euler-M	$\alpha = 0,\ J = 0$	0.4956	4.3093	$1.3774e^{-1}$	0.9200	0.8850	$1.4672e^{-3}$	
	Milstein	$\alpha = 1,\ J = 0$	0.5878	2.2454	$3.8402e^{-2}$	0.9885	1.3702	$1.8005e^{-3}$	
	Preco1E1	$\alpha = 0,\ J = 1$	0.5754	0.8014	$1.9685e^{-3}$	0.9993	0.7527	$6.1671e^{-4}$	
	Preco2E1	$\alpha = 0,\ J = 2$	0.5660	0.7487	$1.9207e^{-3}$	1.0172	0.4340	$3.7323e^{-4}$	
Jump-diffusions (Merton)	Euler-M	$\alpha = 0,\ J = 0$	0.5019	1.5476	$4.7809e^{-2}$	0.9135	0.5001	$9.7563e^{-4}$	
	Milstein	$\alpha = 1,\ J = 0$	0.7012	0.9601	$9.0261e^{-3}$	0.9992	0.8623	$8.0292e^{-4}$	
	Preco1E1	$\alpha = 0,\ J = 1$	0.5489	0.2364	$5.6456e^{-3}$	1.0015	0.2883	$3.2650e^{-4}$	
	Preco2E1	$\alpha = 0,\ J = 2$	0.5347	0.1756	$4.6732e^{-3}$	1.0285	0.1510	$9.8894e^{-5}$	
	Preco1M1	$\alpha = 1,\ J = 1$	0.5573	0.2020	$4.5035e^{-3}$	1.0178	0.2573	$2.3101e^{-4}$	
	Preco2M1	$\alpha = 1,\ J = 2$	0.5358	0.1758	$4.2770e^{-3}$	1.0361	0.1390	$8.6103e^{-5}$	

Table 5.7: Estimation of numerical order of convergence (strong and weak).

In order to estimate the strong approximation error $\mathbb{E}\left[\left|X_T - Y_T^{\Delta t}\right|\right]$ one needs an exact evaluation of X_T. In the case of the well-known Heston model there is no exact solution of X_T known, so one applies the following Theorem 5.1.1 to determinate the order of convergence γ as in Table 5.7.

Theorem 5.1.1 (Rate of Convergence without Knowing of the Exact Solution)**.** *If there exists a stochastic process $(X_t)_{0\leq t\leq T}$ and a corresponding time discretization scheme $Y^{\Delta t}$ that has the following strong convergence property*

$$\mathbb{E}\left[\left|X_T - Y_T^{\Delta t}\right|\right] \leq C_1 \Delta t^{\gamma} \tag{5.2}$$

for all time steps $\Delta t \in (0, \infty)$ with $C_1 > 0$. Then, there exists a positive constant C_2 with

$$\mathbb{E}\left[\left|Y_T^{\Delta t} - Y_T^{2\Delta t}\right|\right] \leq C_2 \Delta t^{\gamma}, \tag{5.3}$$

where $C_2 = C_1\left(1 + 2^{\gamma}\right)$ and this constant does not depend on timestep $\Delta t \in (0, \infty)$. Nevertheless, C_2 depends on the strong order of convergence γ and the positive constant C_1.

Proof. If equation (5.2) holds for all $\Delta t \in (0, \infty)$, then it must be true for $2\Delta t \in (0, \infty)$ as well. So one obtains

$$\mathbb{E}\left[\left|X_T - Y_T^{2\Delta t}\right|\right] \leq C_1 (2\Delta t)^{\gamma}. \tag{5.4}$$

By using the triangle inequality and combining equations (5.2) and (5.4) one gets

$$\begin{aligned}
\mathbb{E}\left[\left|Y_T^{\Delta t} - Y_T^{2\Delta t}\right|\right] &= \mathbb{E}\left[\left|X_T - Y_T^{\Delta t} - X_T + Y_T^{2\Delta t}\right|\right] \\
&\leq \mathbb{E}\left[\left|X_T - Y_T^{\Delta t}\right|\right] + \mathbb{E}\left[\left|X_T - Y_T^{2\Delta t}\right|\right] \\
&\leq C_1 \Delta t^{\gamma} + C_1 (2\Delta t)^{\gamma} \\
&= C_1\left(1 + 2^{\gamma}\right) \Delta t^{\gamma} \\
&=: C_2 \Delta t^{\gamma}.
\end{aligned}$$

This completes the proof and establishes the formula (5.3). □

A very similar result for the weak convergence property

$$\left|\mathbb{E}\left[g(X_T)\right] - \mathbb{E}\left[g(Y_T^{\Delta})\right]\right| \leq C \Delta^{\beta}$$

may be proved in the same way as Theorem 5.1.1. Theorem 5.1.1 is very similar to Theorem 1 by Schmitz Abe and Shaw (2005) or Theorem 1 by Schmitz Abe and Giles (2008). Some results of Theorem 5.1.1 for the Heston model are shown in Table 5.7.

The following example shows how the predictor-corrector scheme performs compared to the Euler-Maruyama scheme. To illustrate this some prices of plain vanilla and exotic options (e.g. path-dependent Asian option as in (2.59) and (2.60) or barrier options as in (2.54) and (2.56)) are listed in Tables 5.8 and 5.9. The reader is also reffered to Groß (2009).

Example 5.1.2. *The following results concern the pricing of different options via Monte-Carlo methods and will also confirm the theoretical results of Corollary 2.2.24. The initial stock price is chosen as* $S_0 = 95$ €*, the strike price as* $K = 100$ €*, the time to maturity as* $T = 1$*, the risk-free interest rate as* $r = 0.04$ *and the volatility of the Black-Scholes model* (2.39) *as constant* $\sigma^2 = 0.16$*. The closed form solution given by equation* (3.1) *yields* $C(0, S_0) = 5.56753$ €*. Using the call put parity of Theorem 2.3.10 one gets* $P(0, S_0) = 6.64647$ €*. The barriers are chosen as* $B_{up} = 150$ € *and* $B_{down} = 70$ € *in this example.*

option prices (€)	**100**	**1,000**	**10,000**	**100,000**	**1,000,000**
$C(0, S_0)$	6.0240	5.7106	5.5251	5.5482	5.5579
deviation from 5.56753 €	8.20%	2.57%	0.76%	0.35%	0.17%
$P(0, S_0)$	6.2431	6.7576	6.6039	6.6717	6.6559
deviation from 6.64647 €	4.56%	1.67%	0.64%	0.38%	0.14%
$C_{\text{digital}}(0, S_0)$	0.4900	0.4506	0.4191	0.4211	0.4222
$C_{\text{up-and-out}}(0, S_0)$	5.8106	5.8294	5.2517	5.2128	5.2266
$C_{\text{down-and-out}}(0, S_0)$	6.0240	5.8106	5.5151	5.5382	5.5579
$C_{\text{up-and-in}}(0, S_0)$	0.0038	0.1946	0.2634	0.3254	0.3373
$C_{\text{Asian}}(0, S_0)$	2.0827	2.1574	2.1433	2.1608	2.2106
$P_{\text{Asian}}(0, S_0)$	4.6283	4.9555	4.9616	4.9972	5.0355
$C_{\min}(0, S_0)$	9.0124	10.2082	9.5874	9.6776	9.6921
$P_{\max}(0, S_0)$	14.5507	13.4616	13.9920	13.9133	13.9048
cputime	0.00025	0.00257	0.02267	0.22407	2.23274

Table 5.8: Option prices computed using Euler-Maruyama scheme.

option prices (€)	**100**	**1,000**	**10,000**	**100,000**	**1,000,000**
$C(0, S_0)$	5.8162	5.5171	5.5396	5.5632	5.5676
deviation from 5.56753 €	4.47%	0.91%	0.50%	0.08%	0.01%
$P(0, S_0)$	6.4428	6.7548	6.6462	6.6433	6.6469
deviation from 6.64647 €	3.06%	1.63%	0.15%	0.05%	0.01%
$C_{\text{digital}}(0, S_0)$	0.4394	0.4328	0.4198	0.4227	0.4229
$C_{\text{up-and-out}}(0, S_0)$	5.6567	5.4702	5.2357	5.2218	5.2203
$C_{\text{down-and-out}}(0, S_0)$	5.9451	5.6258	5.5757	5.5615	5.5609
$C_{\text{up-and-in}}(0, S_0)$	0.1483	0.2504	0.2972	0.3105	0.3471
$C_{\text{Asian}}(0, S_0)$	2.0962	2.1831	2.2220	2.2505	2.2548
$P_{\text{Asian}}(0, S_0)$	4.9830	5.0450	5.1028	5.2107	5.2143
cputime	0.00037	0.00372	0.03584	0.35739	3.57822

Table 5.9: Option prices computed using predictor-corrector scheme.

The deviation from option prices is given by

$$\frac{\left|C(0, S_0) - \mathbb{E}\left[\max(Y_T^{\Delta t} - K, 0)\right]\right|}{C(0, S_0)}. \tag{5.5}$$

So far, predictor-corrector schemes are useful approximations in terms of accuracy and computational speed. Moreover, the computational efficiency of the adjoint approach presented in Theorem 4.2.6 to solve the calibration problem ($\mathrm{P}^{PC}_{\mathrm{M},\Delta t,\epsilon}$) will be analyzed in the next section.

5.2 Efficiency of the Adjoint-based Monte-Carlo Calibration

5.2.1 Numerical Calibration Results

First the chosen market data and additional settings for this section are introduced. For all test cases the routines calibrate the financial market model to a set of 100 European call options on the S&P 500 index taken from Andersen and Brotherton-Ratcliffe (1997). The volatilty surface is illustrated in Table 5.10 and Figure 5.7 in the form of implied volatilities. As in Andersen and Brotherton-Ratcliffe (1997) one chooses the risk-free interest rate as $r = 0.06$, the dividend yield $\delta = 0.0262$ and assumes that the initial stock price is normalized to $S_0 = 1$.

K \ T	0.175	0.425	0.695	0.94	1	1.5	2	3	4	5
0.85	0.190	0.177	0.172	0.171	0.171	0.169	0.169	0.168	0.168	0.168
0.90	0.168	0.155	0.157	0.159	0.159	0.160	0.161	0.161	0.162	0.164
0.95	0.133	0.138	0.144	0.149	0.150	0.151	0.153	0.155	0.157	0.159
1.00	0.113	0.125	0.133	0.137	0.138	0.142	0.145	0.149	0.152	0.154
1.05	0.102	0.109	0.118	0.127	0.128	0.133	0.137	0.143	0.148	0.151
1.10	0.097	0.103	0.104	0.113	0.115	0.124	0.130	0.137	0.143	0.148
1.15	0.120	0.100	0.100	0.106	0.107	0.119	0.126	0.133	0.139	0.144
1.20	0.142	0.114	0.101	0.103	0.103	0.113	0.119	0.128	0.135	0.140
1.30	0.169	0.130	0.108	0.100	0.099	0.107	0.115	0.124	0.130	0.136
1.40	0.200	0.150	0.124	0.110	0.108	0.102	0.111	0.123	0.128	0.132

Table 5.10: Market data: Implied volatilities for S&P 500 index options.

As in Figure 3.1 the market data for the S&P 500 index options and the graphical illustration of Figure 5.7 show the typical volatility smile known from many empirical studies especially the options that are close to exercise maturity.

The Heston model (2.43) is chosen as the test case of calibrating via Theorem 4.2.6

$$\begin{aligned} dS_t &= r(t)S_t dt + \sqrt{v_t} S_t dW_t^1, \quad S_0, \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t} dW_t^2, \quad v_0, \\ \rho dt &= \mathrm{Cov}\left[dW_t^1, dW_t^2\right]. \end{aligned}$$

The call option prices corresponding to this model are calibrated to a set of call options that are computed via a common Monte-Carlo approach considering the volatility surface listed in Table 5.10. This is a distinction as to the work of Käbe et al. (2009). Since the parameter $x^1 = v_0$ is the start value of the stochastic variance differential equation, the Heston dynamics at first sight do not seem to fit into the general model framework of equation (3.2). However, the simple transformation $\tilde{v}_t := v_t/v_0$, and a Cholesky decomposition yields the equivalent model dynamics as in Example 4.2.9

$$\begin{aligned} dS_t &= (r-\delta)S_t dt + \sqrt{v_0 \tilde{v}_t^+} S_t d\tilde{W}_t^1, \quad S_0 = 1, \\ d\tilde{v}_t &= \kappa\left(\frac{\theta}{v_0} - \tilde{v}_t^+\right)dt + \frac{\sigma}{\sqrt{v_0}}\sqrt{\tilde{v}_t^+}\left(\rho d\tilde{W}_t^1 + \sqrt{1-\rho^2} d\tilde{W}_t^2\right), \ \tilde{v}_0 = 1, \end{aligned} \tag{5.6}$$

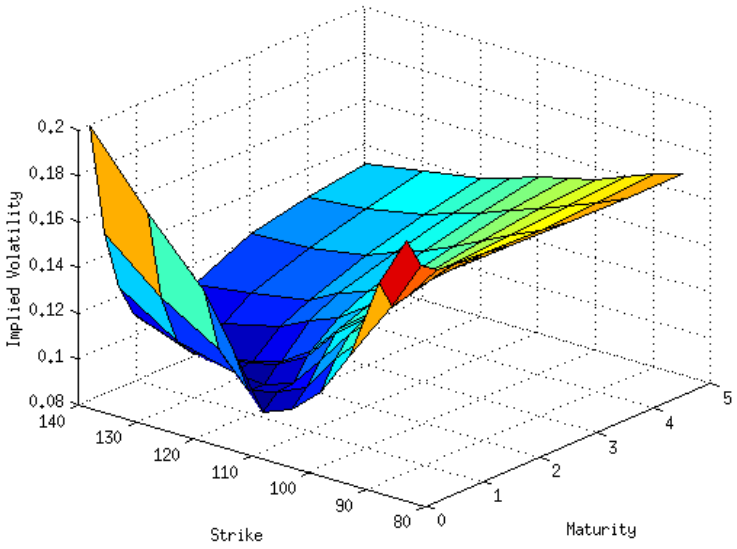

Figure 5.7: Volatility Smile: Graphical illustration of the implied volatilities for S&P 500 index options from Table 5.10.

with start values $(S_0, \tilde{v}_0)^\top$ that are independent of the model parameters $x = (v_0, \kappa, \theta, \sigma, \rho)^\top \in U \subset \mathbb{R}_+{}^4 \times [-1, 1]$. The set U of feasible parameters is described by the following box constraints and the Feller condition (3.5):

$$\begin{aligned} 0.0001 \leq v_0 &\leq 2 \\ 0.05 \leq \kappa &\leq 2 \\ 0.0001 \leq \theta &\leq 2 \\ 0.0001 \leq \sigma &\leq 4 \\ -0.995 \leq \rho &\leq 0.995 \\ \sigma^2 &\leq 2\kappa\theta, \end{aligned} \tag{5.7}$$

hence, by Theorem 2.3.7, it is a complete financial market model. The selected lower and upper bounds assure the non-emptiness, compactness and convexity of U. Equipped with this, the adjoint equation and the gradient is given in Example 4.2.9. Table 5.11 contains the calibration results for four different collocations of Monte-Carlo paths M and discretization time steps Δt and compares the resulting least square errors

$$\hat{f}_{M,\Delta t,\epsilon} = \frac{1}{2} \sum_{i=1}^{I} \left(\hat{C}^i_{M,\Delta t,\epsilon}(y, x) - C^i_{\text{obs}} \right)^2$$

based on the Monte-Carlo function evaluations with smoothing parameter $\epsilon = 0.01$ (see Remark 3.3.2). The parameters of calibration problem $\mathrm{P}^{PC}_{\mathrm{M},\Delta t,\epsilon}$ are chosen as $\eta = \vartheta = \alpha = 0$ in this first example.

The calibration routine is implemented with the MATLAB® function `fmincon` from the Optimization Toolbox[1] and is started with initial values $v_0 = 0.16$, mean reversion speed of volatility $\kappa = 0.8$, long run mean volatility $\theta = 0.16$, volatility of volatility $\sigma = 0.4$ and correlation $\rho = -0.75$. Since one calibrates five parameters to fit 100 market prices, the problem is clearly underdetermined and ill-posed in the sense of Hadamard (1902).

To ensure that the Wiener processes $(\tilde{W}_t^1)_{t\in[0,T]}$ and $(\tilde{W}_t^2)_{t\in[0,T]}$ are uncorrelated the MATLAB© random number stream function `RandStream.create`[2] is used. As pseudo random number generator (PRNG) is the combined multiple recursive generator (`mrg32k3a`) chosen. A detailed overview on this random number generator can be found in L'Ecuyer (1996). It is also very easy to implement on parallel GPUs as established by Giles and Bradley et al. (2012) because of its inherent structure of long streams and substreams, which can produce a deterministic sequence of random numbers. In particular, this is an important feature allowing each Monte-Carlo simulation to be exactly replicated. The idea of regenerating the Wiener increments every time they are needed is also discussed in Section 4.4. An alternative implementation is proposed by Saito and Matsumoto (2008).

x	$M = 10,000$ $\Delta t = 0.01$	$M = 100,000$ $\Delta t = 0.01$	$M = 100,000$ $\Delta t = 0.005$	$M = 500,000$ $\Delta t = 0.005$
v_0	0.1654	0.1617	0.1577	0.1553
κ	1.9674	1.8181	1.8224	1.8390
θ	0.2684	0.2593	0.2510	0.2427
σ	0.2936	0.3612	0.3816	0.3893
ρ	-0.6159	-0.6147	-0.6252	-0.6312
$\hat{f}_{M,\Delta t,0.01}$	3.5777e-05	2.0919e-05	2.0906e-05	2.0783e-05

Table 5.11: Calibration results for the case of the Heston model with several discretization time steps and numbers of Monte-Carlo simulations corresponding to the implied volatility surface in 5.10.

Table 5.11 impressively demonstrates the convergence of the solutions of the calibration problem $\mathrm{P}^{PC}_{\mathrm{M},\Delta t,\epsilon}$ as one increases the number of Monte-Carlo paths M and reduces the discretization step size Δt. More particularly, it appears that the sample with $M = 100,000$ and $\Delta t = 0.01$ leads already to a small least square error. Therefore, for practical applications, the accuracy in this case is certainly sufficient. Hence, in what follows, the number of Monte-Carlo paths is chosen as $M = 100,000$ and the discretization step size as $\Delta t = 0.01$.

Due to different surfaces of option prices resulting from miscellaneous approaches of computing them, to be more precise closed-form versus Monte-Carlo solution, the results of Table 5.11 and Table 5.12 differs slightly from the calibration results of Table 2 in Käbe et al. (2009). Also the choice of the smoothed functions results in different solutions of the calibration problem. Käbe et al. (2009) uses polynomial approximations of the max func-

[1] https://de.mathworks.com/help/optim/ug/fmincon.html.
[2] https://de.mathworks.com/help/matlab/ref/randstream.html.

tion. The results in this chapter are computed via the smoothed version of the ramp function $\max(x, 0)$ of Remark 3.3.2, i.e., function (3.10). However, Table 5.12 compares the solutions of $\mathrm{P}^{PC}_{\mathrm{M},\Delta t,\epsilon}$ with different initial values. These results reveal that the stationary point of Table 5.11 is attractive for almost all initial vectors.

x	initial	end	initial	end	initial	end	initial	end
v_0	0.16	0.1613	0.25	0.1556	0.5	0.1549	1	0.1554
κ	0.8	1.8181	0.5	1.8426	1.5	1.7716	1.5	1.7293
θ	0.16	$\rightarrow$ 0.2592	0.25	$\rightarrow$ 0.2424	0.5	$\rightarrow$ 0.2437	1	$\rightarrow$ 0.2436
σ	0.25	0.3612	0.1	0.4101	0.09	0.3773	0.5	0.3664
ρ	-0.75	-0.6147	-0.5	-0.6280	-0.25	-0.6329	0.25	-0.6376
$\hat{f}$	2.092e-05		2.077e-05		2.086e-05		2.104e-05	
i	24		43		65		86	

Table 5.12: Calibration results for the case of the Heston model with several initial values.

Additionally, the calibration of the Heston model to a set of 510 European call options on the DAX index corresponding to the Figure 3.1 is presented here. The calibration parameters as well as the initial values are choosen as above. Table 5.13 compares also the solutions of the calibration routine for the case of the Euler-Maruyama scheme with the predictor-corrector scheme in both, the finite difference approach and the adjoint approach. An interesting observation in both cases of real market data is the negativity of the correlation ρ between returns of the underlying S and volatility $\tilde{v}$. Nonetheless, it is often observed that if the price of the underlying drops down, the volatility will increase.

scheme approach x	Euler-Maruyama FDM $\Delta t = 0.01$	Euler-Maruyama Adjoint $\Delta t = 0.01$	predictor-corrector FDM $\Delta t = 0.02$	predictor-corrector Adjoint $\Delta t = 0.02$
v_0	0.0266	0.0266	0.0254	0.0254
κ	1.9973	1.9971	1.9798	1.9797
θ	0.0490	0.0491	0.0490	0.0490
σ	0.1527	0.1527	0.1508	0.1509
ρ	-0.8713	-0.8714	-0.8886	-0.8886
$\hat{f}_{M,\Delta t,0.01}$	2.4825e-03	2.4837e-03	2.2852e-03	2.2877e-05
cputime	427.513	419.953	276.313	256.363

Table 5.13: Calibration results for the case of the Heston model in the case of the DAX index options.

.

The resulting solutions, displayed in Table 5.13, are almost identical and only differ slightly with regard to the approach, i.e., FDM or adjoint approach. The speedup resulting from the

predictor-corrector scheme in comparison with the Euler-Maruyama scheme is based on the smaller time step. Table 5.7 emphasizes the selection of a smaller time step size Δt in the case of the predictor-corrector scheme since it produces smaller weak errors ϵ_W.

5.2.2 Computational Effort of the Adjoint Technique

To fully assess the potential speedup of the adjoint-based Monte-Carlo calibration in comparison to alternative computation methods for the gradient, the modified Heston model (5.6) is extended by introducing time-dependent parameters $\kappa(t), \theta(t), \sigma(t)$ and $\rho(t)$ for $0 \leq t \leq T$ as established by Benhamou et al. (2010) and Käbe et al. (2009). Piecewise constant parameters on $[t_{b-1}, t_b)$, $b = 1, ..., B$ are chosen for this calculation, i.e.,

$$\begin{aligned} \kappa(t) &= \kappa_b, \quad t \in [t_{b-1}, t_b), \quad b = 1, ..., B \quad \text{and} \quad \kappa(T) = \kappa_B, \\ \theta(t) &= \theta_b, \quad t \in [t_{b-1}, t_b), \quad b = 1, ..., B \quad \text{and} \quad \theta(T) = \theta_B, \\ \sigma(t) &= \sigma_b, \quad t \in [t_{b-1}, t_b), \quad b = 1, ..., B \quad \text{and} \quad \sigma(T) = \sigma_B, \\ \rho(t) &= \rho_b, \quad t \in [t_{b-1}, t_b), \quad b = 1, ..., B \quad \text{and} \quad \rho(T) = \rho_B, \end{aligned} \tag{5.8}$$

where $0 = t_0 < t_1 < ... < t_B = T$ is a suitable discretization of the time interval $[0, T]$ into $B \in I\!N$ subintervals (Hull (2008)). In the examples below the equidistant points $t_0, t_1, ..., t_B$ with $t_b = b\frac{T}{B}$ are chosen. For the time-dependent parameters the notation of a vector $x \in \mathbb{R}^P$ can be retained by arranging the elements of x in the following way

$$x = (v_0, \kappa_1, ..., \kappa_B, \theta_1, ..., \theta_B, \sigma_1, ..., \sigma_B, \rho_1, ..., \rho_B)^\top \in \mathbb{R}^P = \mathbb{R}^{4B+1}. \tag{5.9}$$

This only changes the calculation of the adjoint equation slightly insofar that one has to replace the previously constant x_i by its corresponding value on the subinterval.

scheme		Euler-Maruyama $\alpha = 0, J = 0$ $\Delta t = 0.01$		Milstein $\alpha = 1, J = 0$ $\Delta t = 0.01$		predictor-corrector $\alpha = 0, J = 1$ $\Delta t = 0.02$		
B	P	FDM	Adjoint	FDM	Adjoint	FDM	Adjoint	Speedup ratio
1	5	18.57	18.25	21.02	20.62	11.37	10.42	1.0175
2	9	32.78	18.37	37.17	20.78	19.99	10.47	1.7844
3	13	47.26	18.35	53.40	20.75	28.82	10.45	2.5755
4	17	61.96	18.44	70.23	20.72	37.79	10.51	3.3601
5	21	78.61	18.33	88.84	20.79	47.95	10.44	4.2886
10	41	152.24	18.24	172.04	20.71	92.86	10.49	8.3465
20	81	302.35	18.59	341.79	20.92	184.43	10.46	16.2641
50	201	824.27	18.51	931.44	21.00	502.80	10.45	44.5311
100	401	1636.66	18.32	1838.26	20.76	992.26	10.44	89.3373

Table 5.14: Calculation time (in seconds) for the gradient of the objective function $\hat{f}_{M,\Delta t,\epsilon}$ for a changing number B of time intervals $[t_{b-1}, t_b)$ and resulting $P = 4B + 1$ number of parameters by using different schemes.

Table 5.14 summarizes the computational effort for alternative computation methods for the gradient. It compares the calculation time for the gradient of the objective function $\hat{f}_{M,\Delta t,\epsilon}$ computed via the Euler-Maruyama scheme, the Milstein scheme and the predictor-corrector scheme (stochastic Heun's scheme). The finite difference method (FDM) as well as the adjoint approach are presented. These results are also illustrated in Figure 5.8 and also numerically confirm the theoretical results of Section 4.3.

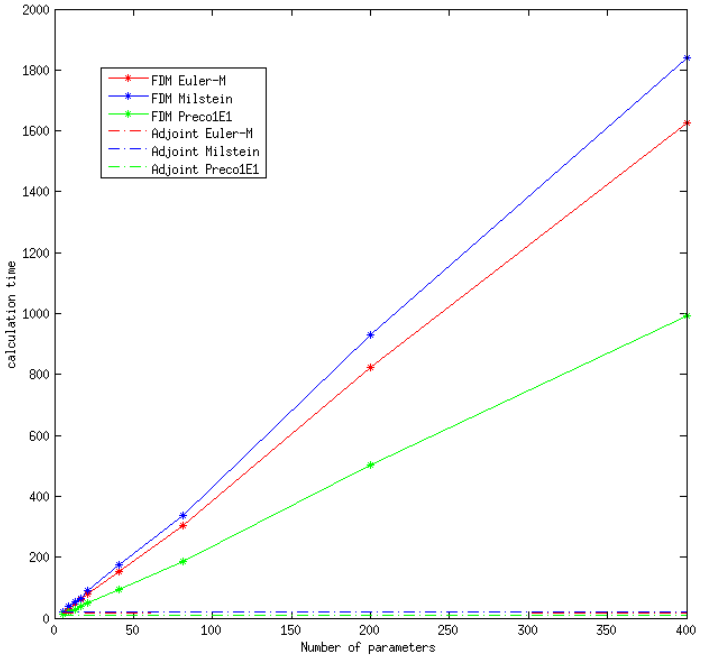

Figure 5.8: Calculation time (in seconds) for the gradient plotted against the number of parameters.

The presented adjoint-based Monte-Carlo calibration take advantage of two facts. On the one hand due to Benhamou et al. (2010) for time-dependent parameters there is no analytical formula and one usually has to perform Monte-Carlo simulations to price options. On the other hand it is worthwhile noting that the speedup resulting from the adjoint technique strongly depends on the number of parameters of the financial market model. In this setting the number of time intervals B allows one to easily analyze this dependence in more detail. Since one does not want to distort the analysis with varying numbers of iterations for changing B, one measures the required time for the pure gradient calculation based on finite differences and the adjoint equation in Table 5.14. As expected, the calculation time for the finite difference method increases with an almost linear rate in the number of subintervals B, whereas the time required by the adjoint method nearly stays constant in all cases, i.e. the Euler-Maruyama scheme, the Milstein scheme and the predictor-corrector scheme. Thus the speedup increases with an increasing number of subintervals. Käbe (2010) has already shown

this facts. Moreover, the predictor-corrector scheme combined with the adjoint approch is the fastest one of all considered schemes. For this very reason it is not required to obtain a reasonably good approximation in time, because the predictor-corrector scheme does not necessarily require small time steps to perform excellent calibration results.

In summary, the results of the predictor-corrector scheme are promising in the sense that they provide a significant acceleration of the calibration of financial market models, along with the adjoint method. The use of stochastic predictor-corrector schemes in the Monte-Carlo framework of pricing options and calibrating financial market models should attract attention in the financial research community.

Chapter 6

Coincidence to Optimal Control Theory

Life can only be understood backwards; but it must be lived forwards.
— Søren Kierkegaard

This chapter deals with the question of how the the numerical solution of SDE constrained optimal control problems of the so called *discretize-then-optimize* approach of Chapter 4 coincide with solution of the opposed *optimize-then-discretize* approach. Both approaches are applied to calculate adjoint information which generates the required derivatives of the calibration functional.

6.1 First-Optimize-then-Discretize

In the research community of optimization problems governed by partial differantial equations (PDEs) a thought-provoking question arises from linking the *discretize-then-optimize* approach with the *optimize-then-discretize* approach as in Collis and Heinkenschloss (2002), Hinze (2009), Hinze and Tröltzsch (2010) and Hinze and Rösch (2012). Solutions to optimization problems with PDE constraints inherit special properties. The associated state solves the PDE which in the optimization problem takes the role of a equality constraint, and this state together with the associated control solves an optimization problem, i.e., together with multipliers satisfies first-order necessary optimality conditions (2.4.2). This approach will be transfered below to stochastic differential equations.

This chapter will bridge the class of problems dealt with here is a static parameter identification calibration problem with SDE constraints to backward stochastic differential equations (BSDEs) arising from optimal control problems. To solve this static optimal control problem an adjoint technique is presented for an efficient evaluations of the gradient of the calibration function in gradient-based optimization algorithms. Adjoint equation may considered as Lagrange multiplier in the infinite-dimensional case.

In the *discretize-then-optimize* approach of Theorem 4.2.6, such as in Käbe et al. (2009) or Groß and Sachs (2012), the calibration problem is first discretized, using for instance the Euler-Maruyama discretization (e.g. Theorem 4.2.7) of a stochastic differential equation (SDE). Afterwards the resulting finite dimensional optimization problem is solved. On the contrary, in the *optimize-then-discretize* approach one first solves the infinite dimensional

optimality system, involving the (forward) stochastic differential equation as well as the (adjoint) backward stochastic differential equation (BSDE). One finally discretizes this optimality system using standard schemes for both the forward and the adjoint backward equations. Betimes these approaches lead to different results. Following Collis and Heinkenschloss (2002) one will use discrete adjoint equations and discrete gradient equations in notation to mean that these are the adjoint and gradient equations for the discretized problem. One will also use the phrases discretized adjoint equations and discretized gradient equations to refer to discretizations of the adjoint equation and gradient equation, respectively.

The main result of this section is to show that there is no gap between the discrete gradient equations of the *discretize-then-optimize* approach and the discretized gradient equations of the *optimize-then-discretize* approach for some standard schemes. Nevertheless, there is a difference between the discretized adjoint equations and its counterpart, namely, the discretizations of the adjoint backward equation as in Kohlmann and Zhou (2000). Insofar as the discrete adjoint equations of Section 4 are (non-adapted) solutions of linear stochastic differential equations with some terminal condition. However by the *optimize-then-discretize* approach one gets a discretizations of an adjoint linear backward stochastic differential equation as we see in section 6.3 and what is also a result of the similar approach of stochastic control problems (e.g. Bahlali et al. (2007), Baras et al. (1988), Duffie and Skiadas (1994), Elliott and Kohlmann (1989d), El Karoui et al. (1997), Kunita (1982), Mou and Yong (2007), Yong (2010) or Yong and Zhou (1999)). This section shows how this leads in an elegant manner to the same gradient representation of both approaches. To this end, one has to use the same method of Section 2.4 to determine the discretized gradient representation and the discrete gradient representation.

In Section 6.4 results of the previous sections are used to calibrate a financial market model to observed market prices. Finally, this chapter concludes this thesis in the manner of Bender et al. (2014).

6.2 Introductory Notes on BSDEs

First some important properties of backward stochastic differential equation (BSDE) are established followed by some general results. The standard work on BSDEs is Yong and Zhou (1999). Their is a strong connection between BSDEs and the Martingale Representation Theorem 2.2.17 (cf. Elliott and Kohlmann (1989c)). A dual representation in terms of martingales was first suggested by Rogers (2002) and then established by Haugh and Kogan (2004), Belomestny et al. (2009), Belomestny et al. (2013), Bender, Schoenmakers and Zhang (2013) and Rogers (2007).

To discuss the connection between the *optimize-then-discretize* approach and non-anticipating or adapted BSDEs, which were introduced as adjoint equations in stochastic control by Bismut in the 1970's, i.e., Bismut (1973), Bismut (1978) and later by Haussmann (1981), Elliott (1977) and Kushner (1972) and generalized by Pardoux and Peng (1990) the following framework is given.

6.3 Stochastic Adjoint Equation

6.3.1 Framework of Calibration with SDE

Let the following calibration problem be considered for given market data $x_{obs} \in \mathbb{R}^n$

$$\min \frac{1}{2}\|\mathbb{E}[\pi(X_T(u))] - x_{obs}\|^2$$

where $X_\cdot$ is a solution of the stochastic integral equation

$$X_t(u) = X_0 + \int_0^t a(s, X_s(u), u)ds + \int_0^t b(s, X_s(u), u)dW_s, \tag{6.1}$$

$u \in \mathbb{R}^p$ is the calibration parameter and $\pi : \mathbb{R}^d \to \mathbb{R}^n$ is a payoff-function.

Assumption 6.3.1. *Let deterministic functions*

$$a : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}^d, \quad b : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$$

be measurable in $t \in [0,\infty)$ and $a(\cdot,0), b(\cdot,0) \in L^2(0,T;\mathbb{R}^d)$ for any $T > 0$. Furthermore, there exists a Lipschitz constant $L > 0$ such that

$$\|a(t,x) - a(t,y)\| + \|b(t,x) - b(t,y)\| \le L\|x - y\|$$

for all $t \in [0,\infty)$, $x, y \in \mathbb{R}^d$.

Since for the computation of adjoints, integration by parts is an essential step in the derivation, Lemma 2.2.16 will play a crucial role in what follows.

Assumption 6.3.2. *For each parameter vector $u \in \mathbb{R}^p$ let $a(\cdot,\cdot,u)$ and $b(\cdot,\cdot,u)$ satisfy Assumption 6.3.1. Moreover, let $a(t,X,\cdot)$ and $b(t,X,\cdot)$ be continuously differentiable in the variable u for each $t \in [0,T]$ and $X \in \mathbb{R}^d$. Furthermore, let $a(t,\cdot,u)$ and $b(t,\cdot,u)$ be continuously differentiable in the variable X for each $t \in [0,T]$ and $u \in \mathbb{R}^p$. The partial Fréchet-derivatives $a_X(t,\cdot,u)$ and $b_X(t,\cdot,u)$ are Lipschitz. Additionally, let $\pi : \mathbb{R}^d \to \mathbb{R}^n$ be continuously differentiable.*

To cast this problem into the framework of Section 2.4, one has to set

$$X = \mathfrak{X}, \quad Y = \mathfrak{X}_0, \quad U = \mathbb{R}^p,$$

and consider u and X as independent variables

$$\Phi(X,u) = \frac{1}{2}\|\mathbb{E}[\pi(X_T)] - x_{obs}\|^2 \tag{6.2}$$

coupled through the equality constraint $g : \mathfrak{X} \times U \mapsto \mathfrak{X}_0$

$$g(X,u) = X_\cdot - X_0 - \int_0^\cdot a(s, X_s, u)ds - \int_0^\cdot b(s, X_s, u)dW_s = 0. \tag{6.3}$$

The Fréchet-derivative $g_x(\cdot) : \mathfrak{X} \times U \mapsto \mathcal{L}(\mathfrak{X} \times U, \mathfrak{X}_0)$ is given by the partial Fréchet-derivatives in the next lemma.

Lemma 6.3.3. *Let Assumption 6.3.2 hold. Then the map $g : \mathfrak{X} \times U \mapsto \mathfrak{X}_0$ as defined in (6.3) is continuously Fréchet-differentiable and its partial Fréchet-derivatives are given by*

$$g_X(X,u)\Delta X = \Delta X_\cdot - \int_0^\cdot a_X(s,X_s)\Delta X_s ds - \int_0^\cdot b_X(s,X_s)\Delta X_s dW_s, \Delta X \in \mathfrak{X}, \tag{6.4}$$

$$g_u(X,u)\Delta u = -\int_0^\cdot a_u(s,X_s)\Delta u ds - \int_0^\cdot b_u(s,X_s)\Delta u dW_s, \quad \Delta u \in U. \tag{6.5}$$

Proof. First one shows, that g is continuously Fréchet-differentiable with respect to a $X \in \mathfrak{X}$. To this let be $\Delta X \in \mathfrak{X}$ with $|\Delta X|_{\mathfrak{X}} \to 0$ and note that a, b are already Fréchet-differentiable with using the Taylor formula

$$a(s, X_s - \Delta X_s) = a(s,X_s) - a_x(s,X_s)\Delta X_s + o(\Delta X_s)$$

or

$$b(s, X_s - \Delta X_s) = b(s,X_s) - b_x(s,X_s)\Delta X_s + o(\Delta X_s),$$

respectively. So one obtains using Definition 2.1.9

$$\begin{aligned}
&|g(X - \Delta X, u) - g(X,u) - g_X(X,u)\Delta X|_{\mathfrak{X}} \\
&= \Big(||X - X_0 - X_0 + \Delta X - X + X_0 - \Delta X||^2 \\
&+ \mathbb{E}\left[\int_0^T ||a(s, X_s - \Delta X_s) - a(s,X_s) + a_x(s,X_s)\Delta X_s||^2 ds \right] \\
&+ \mathbb{E}\left[\int_0^T ||b(s, X_s - \Delta X_s) - b(s,X_s) + b_x(s,X_s)\Delta X_s||^2 ds \right] \Big)^{1/2} \\
&= |\epsilon_X(\Delta X)|_{\mathfrak{X}}
\end{aligned}$$

as long as

$$g_X(x,u)\Delta X = \Delta X_\cdot - \int_0^\cdot a_x(s,X_s)\Delta X_s ds - \int_0^\cdot b_x(s,X_s)\Delta X_s dW_s.$$

Note that the Fréchet-differentiable with respect to $u \in U$ of the constraint $g : \mathfrak{X} \times U \mapsto \mathfrak{X}_0$ is very similar.

$$|g(X, u - \Delta u) - g(X,u) - g_u(X,u)\Delta u|_{\mathfrak{X}} = |\epsilon_X(\Delta u)|_{\mathfrak{X}}$$

□

Next one has to check the Assumption 2.4.4 concerning the surjectivity of the constraint.

Lemma 6.3.4. *Let Assumption 6.3.2 hold. Then for each $u \in U$ the map $g_X : \mathfrak{X} \mapsto \mathfrak{X}_0$ as defined in (6.4) is surjective and invertible.*

Proof. Due to the fact that the existence and uniqueness holds for the linear SDEs (Theorem 2.2.14), there exists a unique $\Delta X \in \mathfrak{X}$ for any given stochastic process $V \in Y = \mathfrak{X}_0$ with

$$g_X(X,u)\Delta X = \Delta X_\cdot - \int_0^\cdot a_x(s,X_s,u)\Delta X_s ds - \int_0^\cdot b_x(s,X_s,u)\Delta X_s dW_s = V. \tag{6.6}$$

□

Finally one checks the Fréchet-derivative for the objective function Φ.

Lemma 6.3.5. *The map $\Phi : \mathfrak{X}\times U \mapsto \mathbb{R}$ defined in (6.2) is continuously Fréchet-differentiable and its partial Fréchet-derivatives are given by*

$$\Phi_X(X,u)\Delta X = (\mathbb{E}[\pi(X_T)] - x_{obs})^\top \mathbb{E}[\pi'(X_T)\Delta X_T] \quad \textit{for all} \quad \Delta X \in \mathfrak{X}, \tag{6.7}$$

$$\Phi_u(X,u)\Delta u = 0 \quad \textit{for all} \quad \Delta u \in U. \tag{6.8}$$

Proof. Since Φ is independent of parameters u, one has $\Phi_u(X,u)\Delta u = 0$. Further it can be shown, that Φ is continuously Fréchet-differentiable with respect to an $X \in \mathfrak{X}$. To this let be $\Delta X \in \mathfrak{X}$ and ϵ_X be a map with $\mathbb{E}[\epsilon_X(\Delta X)] \to 0$

$$\begin{aligned}
&|\Phi(X-\Delta X,u) - \Phi(X,u) - \Phi_x(X,u)\Delta X| \\
&= \left|\frac{1}{2}\|\mathbb{E}[\pi(X_T-\Delta X_T)] - x_{obs}\|^2 - \frac{1}{2}\|\mathbb{E}[\pi(X_T)] - x_{obs}\|^2 - \Phi_x(X,u)\Delta X\right| \\
&= \left|\frac{1}{2} < \mathbb{E}[\pi(X_T-\Delta X_T)], \mathbb{E}[\pi(X_T-\Delta X_T)] > - < \mathbb{E}[\pi(X_T-\Delta X_T)], x_{obs} > \right. \\
&\left. - \frac{1}{2} < \mathbb{E}[\pi(X_T)], \mathbb{E}[\pi(X_T)] > + < \mathbb{E}[\pi(X_T)], x_{obs} > - \Phi_x(X,u)\Delta X\right| \\
&= \left| - < \mathbb{E}[\pi(X_T)] - x_{obs}, \mathbb{E}[\pi'(X_T)\Delta X(T)] > - \Phi_x(X,u)\Delta X + \mathbb{E}[o(\Delta X_T)]\right| \\
&= |\epsilon_X(\Delta X)|,
\end{aligned}$$

where

$$\pi(X_T - \Delta X_T) = \pi(X_T) - \pi'(X_T)\Delta X_T + o(\Delta X_T)$$

by using the Taylor formula. □

This establishes the framework of the following section.

6.3.2 Adjoint Equation defined with an Adapted BSDE

In this section the infinite-dimensional theory of Section 2.4 is used to derive the adjoint equation as a solution of a backward stochastic differential equation. The proof of the following theorem follows the same strategy as for optimal control problems with ordinary or partial differential equations. Although the linear functional representing the Lagrange multiplier lies in a dual space, one proposes a particular representation or "ansatz" for this functional and then prove that it this particular functional satisfies all the requirements in equation (2.73) and due to the uniqueness it is the desired linear functional.

Theorem 6.3.6. *Let Assumption 6.3.2 hold. The function* $\phi : \mathbb{R}^p \to \mathbb{R}$ *defined as*

$$\phi(u) := \Phi(X(u), u) = \frac{1}{2}\|\mathbb{E}[\pi(X_T(u))] - x_{obs}\|^2,$$

where $X(u)$ *solves a stochastic differential equation*

$$X_t(u) = X_0 + \int_0^t a(s, X_s(u), u)ds + \int_0^t b(s, X_s(u), u)dW_s, \tag{6.9}$$

has a gradient representation

$$\nabla\phi(u) = \mathbb{E}\left[\int_0^T P_s^\top a_u(s, X_s(u), u) + Z_s^\top b_u(s, X_s(u), u)ds\right] \tag{6.10}$$

where the pair $(P(\cdot), Z(\cdot))$ *is the solution of on the linear adjoint BSDE*

$$\begin{aligned} dP_t &= -(a_x(t, X_t(u))^\top P_t + b_x(t, X_t(u))^\top Z_t)dt + Z_t dW_t \quad t \in [0, T] \\ P_T &= \pi'(X_T)^\top(\mathbb{E}[\pi(X_T(u))] - x_{obs}). \end{aligned} \tag{6.11}$$

Note that the dimension of the Wiener process $m = 1$ is chosen here for simplicity and comparability. Otherwise, equation (6.11) turns into

$$dP_t = -\left(a_x(t, X_t(u))^\top P_t + \sum_{\nu=1}^m b_x^\nu(t, X_t(u))^\top Z_t^\nu\right)dt + Z_t dW_t \quad t \in [0, T] \tag{6.12}$$

corresponding to (2.19). Similarly, the gradient representation (6.10) turns into

$$\nabla\phi(u) = \mathbb{E}\left[\int_0^T P_s^\top a_u(s, X_s(u), u) + \sum_{\nu=1}^m Z_s^{\nu\top} b_u^\nu(s, X_s(u), u)ds\right]. \tag{6.13}$$

Assumption 6.3.2 ensures the set of assumptions in Bender and Steiner (2013), and thus the existence and uniqueness of an adapted solution (6.11) is assured as in Ma and Yong (1999) and Bouchard and Touzi (2004). Notation is based on Zhang (2001).

Proof. The Lagrange multiplier $l \in Y^\star$ is uniquely defined in equation (2.73). This means that for all $V \in Y$

$$l(V) = -C_X(X(u), u)\Delta X$$

where ΔX is defined as the solution of

$$g_X(X(u), u)\Delta X = V.$$

For this purpose, with Lemma 6.3.5 one has for all $V \in \mathfrak{X}_\mathfrak{o}$ using equation (6.7)

$$l(V) = (E[\pi(X_T(u))] - x_{obs})^\top \mathbb{E}[\pi'(X_T(u))\Delta X_T], \tag{6.14}$$

where $\Delta X \in \mathfrak{X}$ using (6.6) is the solution of

$$\Delta X_t - \int_0^t a_x(s, X_s)\Delta X_s ds - \int_0^t b_x(s, X_s)\Delta X_s dW_s = V_t, \quad t \in [0, T]. \tag{6.15}$$

At this point, one has to assume that $l \in (\mathfrak{X}_o)^*$ is represented by some generator $\lambda \in L^2_{\mathcal{F}}(0, T; \mathbb{R}^d)$, that is a drift of an stochastic process, which is represented via following backward stochastic differential equation (BSDE) with a pair $(P(\cdot), Z(\cdot)) \in \mathcal{M}^2[0, T] \times L^2_{\mathcal{F}}(0, T; \mathbb{R}^{d\times m})$ (see also (Yong and Zhou, 1999, Definition 2.1 page 349), (Mou and Yong, 2007, (3.2) on page 545) or (Yong, 2010, (3.11) on page 4132))

$$P_t = P_T + \int_t^T \lambda(s)ds - \int_t^T Z_s dW_s, \tag{6.16}$$

such that for any stochastic process $V \in \mathfrak{X}_0$

$$l(V) := \mathbb{E}[\int_0^T V_t^\top \lambda(t)dt] + \mathbb{E}[V_T^\top P_T]. \tag{6.17}$$

The proof is in a sense constructive, since it will be shown in the sequel, how the unknown generator $\lambda(\cdot)$ and the unknown terminal condition P_T have to be chosen such that equation (6.14) holds.
To this end, one omits for the rest of the proof the dependence of X on u and set for brevity

$$Y_\cdot := \int_0^\cdot a_x(s, X_s)\Delta X_s ds + \int_0^\cdot b_x(s, X_s)\Delta X_s dW_s = \Delta X_\cdot - V_\cdot \in \mathfrak{X}_o. \tag{6.18}$$

By using the sensitivity equation (6.15) one obtains

$$\begin{aligned} l(V) &= \mathbb{E}[\int_0^T V_t^\top \lambda(t)dt] + \mathbb{E}\left[V_T^\top P_T\right] \\ &= \mathbb{E}\left[\int_0^T \left(\Delta X_t - \int_0^t a_x(s, X_s)\Delta X_s ds - \int_0^t b_x(s, X_s)\Delta X_s dW_s\right)^\top \lambda(t)dt\right] + \mathbb{E}\left[V_T^\top P_T\right] \\ &= \mathbb{E}[\int_0^T (\Delta X_t - Y_t)^\top \lambda(t)dt] + \mathbb{E}\left[V_T^\top P_T\right] \end{aligned}$$

In order to rewrite the term $\int_0^T Y_t^\top \lambda(t)dt$ Itô's formula (2.21) is used by setting

$$\begin{array}{cccccccccccc} G_t & = & Y_t, & G_0 & = & 0, & g(s) & = & a_x(s, X_s)\Delta X_s, & \gamma(s) & = & b_x(s, X_s)\Delta X_s, \\ H_t & = & P_t, & H_0 & = & P_0, & h(s) & = & \lambda(s), & \eta & = & Z_s. \end{array} \tag{6.19}$$

Using $Y \in \mathfrak{X}_o$, Lemma 2.2.18 and equation (6.18) one finally obtains

$$l(V) = \mathbb{E}[\int_0^T \Delta X_t^\top \lambda(t)dt + Y_T^\top P_T - Y_0^\top P_0 - \int_0^T (a_x(t, X_t)\Delta X_t)^\top P_t dt$$

$$
\begin{aligned}
&- \int_0^T (b_x(t, X_t)\Delta X_t)^\top Z_t dt - \int_0^T ((b_x(s, X_s)\Delta X_s)^\top P_t + Y_t^\top Z_t) dW_t] + \mathbb{E}[V_T^\top P_T] \\
&= \mathbb{E}[\int_0^T \Delta X_t^\top \left(\lambda(t) - a_x(t, X_t)^\top P_t - b_x(t, X_t)^\top Z_t)\right) dt] + \mathbb{E}[\Delta X_T^\top P_T]
\end{aligned}
$$

Recall that on the other hand the linear functional l by equation (6.14) should satisfy

$$
l(V) = (\mathbb{E}[\pi(X_T)] - x_{obs})^\top \mathbb{E}[\pi'(X_T)\Delta X_T] = \mathbb{E}[\Delta X_T^\top \pi'(X_T)^\top (\mathbb{E}[\pi(X_T)] - x_{obs})] \tag{6.20}
$$

To satisfy equation (6.14) the following equations have to hold

$$
0 = \mathbb{E}\left[\int_0^T \Delta X_t^\top \left(\lambda(t) - (a_x(t, X_t)^\top P_t) - (b_x(t, X_t)^\top Z_t)\right) dt\right] \text{ and} \tag{6.21}
$$

$$
P_T = \pi'(X_T)^\top (\mathbb{E}[\pi(X_T)] - x_{obs}). \tag{6.22}
$$

Since the generator λ and the terminal condition P_T in in the definition (6.17) of l are still free to choose, the previous equation (6.20) is true, if

$$
\begin{aligned}
\lambda(t) &= a_x(t, X_t)^\top P(t) + b_x(t, X_t)^\top Z(t), \mathbb{P} - a.s. \quad t \in [0, T] \\
P_T &= \pi'(X_T)^\top (\mathbb{E}[\pi(X_T)] - x_{obs})
\end{aligned}
$$

holds. And so the following linear backward stochastic differential equation with a pair $(P(\cdot), Z(\cdot)) \in \mathcal{M}^2[0, T] \times L^2_{\mathcal{F}}(0, T; \mathbb{R}^{d\times m})$ with terminal condition is defined

$$
\begin{aligned}
dP_t &= -(a_x(t, X_t)^\top P_t + b_x(t, X_t)^\top Z_t)dt + Z_t dW_t \quad t \in [0, T] \\
P_T &= \pi'(X_T)^\top (\mathbb{E}[\pi(X_T)] - x_{obs}),
\end{aligned} \tag{6.23}
$$

or, equivalently, corresponding to (6.16)

$$
P_t = \pi'(X_T)^\top (\mathbb{E}[\pi(X_T)] - x_{obs}) + \int_t^T a_x(s, X_s)^\top P_s + b_x(s, X_s)^\top Z_s ds - \int_t^T Z_s dW_s.
$$

Then (6.14) holds and since the linear functional l from Theorem 2.4.8 is unique, this is the representation of it.

Next, the representation for the gradient of the objective function is considered. One obtains from Theorem 2.4.8 and equation (6.5) that

$$
\begin{aligned}
\nabla\phi(u)^\top \Delta u &= l(g_u(X(u), u)\Delta u) + \Phi_u(X(u), u)^\top \Delta u \\
&= l\left(\int_0 a_u(s, X_s, u)\Delta u ds + \int_0 b_u(s, X_s, u)\Delta u dW_s\right) + 0 \\
&= \mathbb{E}[\int_0^T \left(\int_0^t a_u(s, X_s, u)\Delta u ds + \int_0^t b_u(s, X_s, u)\Delta u dW_s\right)^\top \lambda(t) dt] \\
&+ \mathbb{E}[\left(\int_0^T a_u(s, X_s, u)\Delta u ds + \int_0^T b_u(s, X_s, u)\Delta u dW_s\right)^\top P_T] \\
&= \mathbb{E}[\int_0^T \widetilde{Y}_t^\top \lambda(t) dt + \widetilde{Y}_T^\top P_T]^\top \Delta u
\end{aligned}
$$

by setting $\widetilde{Y}_t := \int_0^t a_u(s, X_s, u)ds + \int_0^t b_u(s, X_s, u)dW_s$. If one uses Itô's formula (2.21) again, Lemma 2.2.18 and $Y \in \mathfrak{X}_\mathfrak{o}$, one finally obtains

$$\begin{aligned}
\nabla\phi(u)^\top \Delta u &= \mathbb{E}\left[\int_0^T \widetilde{Y}_t^\top \lambda(t)dt + \widetilde{Y}_T^\top P_T\right]^\top \Delta u \\
&= \mathbb{E}\left[\int_0^T a_u(s, X_s, u)^\top P_s ds + \int_0^T b_u(s, X_s, u)^\top P_s dW_s \right. \\
&+ \left. \int_0^T b_u(s, X_s, u)^\top Z_s ds + \int_0^T \widetilde{Y}_t^\top Z_s dW_s - \widetilde{Y}_0^\top P_0\right]^\top \Delta u \\
&= \mathbb{E}\left[\int_0^T a_u(s, X_s, u)^\top P_s + b_u(s, X_s, u)^\top Z_s ds\right]^\top \Delta u
\end{aligned}$$

This completes the proof and establishes the formula (6.10). □

An alternative approach to derive such a BSDE is via Pontryagin neccesary conditions for optimality or the stochastic maximum principle and its Hamiltonian solution as overviewed by Bahlali et al. (2007) and Peng (1992). Roughly speaking this is done as follows by defining the Hamiltonian as in Bahlali et al. (2007)

$$\mathcal{H}(t, X, u, P) = P_t b(t, X, u) - h(t, X, u) \tag{6.24}$$

and assuming its solution given by $(\hat{X}, \hat{u})$. Consider the adapted pair of solution (P_t, Z_t) of the BSDE

$$\begin{aligned}
dP_t &= -\mathcal{H}_x(t, \hat{X}_t, \hat{u}_t, P_t)dt + Z_t dW_t \quad t \in [0, T] \\
P_T &= g_x(\hat{X}_T).
\end{aligned} \tag{6.25}$$

Under some smoothness assumptions the stochastic maximum principles states that

$$\max_{u \in U} \mathcal{H}(t, \hat{X}_t, u, P_t) = \mathcal{H}_u(t, \hat{X}_t, \hat{u}_t, P_t) \quad a.e.\ t \in [0, T], \ \ \mathbb{P} - a.s. \tag{6.26}$$

Among others Elliott and Kohlmann (1989a) used only this approach to get an adjoint stochastic equation. By contrast, Theorem 6.3.6 presents another way in the sense of Section 2.4 by using the adjoint approach.

6.4 Discretization: Adjoint Stochastic Differential Equation

Back to the specific calibration problem. In order to calibrate financial market model to the market prices C^i_{obs} with strikes K_i, maturities T_i and european payoff $\max(S_{T_i}(x) - K_i, 0)$ at time T_i, $i = 1, \ldots, I$ one obtains the following nonlinear least squares problem as P of Chapter 3

$$\begin{aligned} &\min_{x\in X} \phi(u) := \min_{u\in U} \tfrac{1}{2} \sum_{i=1}^{I} \left(C^i(u) - C^i_{\text{obs}}\right)^2 \\ \text{where} \quad &C^i(u) = e^{-rT_i} \mathbb{E}_Q \left[\pi(S_{T_i}(u) - K_i)\right] \\ \text{s.t.} \quad &dY_t(u) = a(x, Y_t(u))dt + b(x, Y_t(u))dW_t, \ Y_0 > 0 \\ &0 \le t \le T, \quad T := \max_{i=1,\ldots,I} T_i, \end{aligned} \tag{P}$$

where $\pi(\xi) := \max(\xi, 0)$, and $X \subset \mathbb{R}^P$ is a suitable convex and compact set which for example may result from imposing box constraints $l_p \le x_p \le u_p$, $p = 1, ..., P$ on the model parameters u as in Chapter 3. For further information about this issue the reader is reffered to Käbe et al. (2009) and Lörx (2013).

So a calibration requires at least gradient information for the objective function ϕ with respect to parameters $u \in U = \mathbb{R}^P$. To get this gradient information one needs to discretize above SDE and BSDE (6.11) to end up estimating gradient (6.10). For discretization of such BSDEs the reader is referred to Bender and Denk (2007), Bender and Zhang (2008), Bouchard and Touzi (2004) or Ma et al. (2002). Given a partition $0 = t_0 < ... < t_N = T$ this suggests naturally for the BSDE (6.11)

$$\begin{aligned} dP_t &= -(a_x(t, X_t)^\top P_t + b_x(t, X_t)^\top Z_t)dt + Z_t dW_t \quad t \in [0, T] \\ P_T &= \pi'(X_T)^\top (\mathbb{E}[\pi(X_T(u))] - x_{obs}) \end{aligned}$$

the following discretized approximation assuming that conditional expectations can be estimated:

$$P_{t_N} = \pi'(Y_N)^\top (\mathbb{E}[\pi(Y_N)] - x_{obs}) \tag{6.27}$$

$$Z_{t_n} = \mathbb{E}^n \left[P_{t_{n+1}} \frac{\Delta W_n}{\Delta t_n} \right] \tag{6.28}$$

$$P_{t_n} = \mathbb{E}^n \left[P_{t_{n+1}} + (a_x(t_n, Y_n)^\top P_{t_{n+1}} + b_x(t_n, Y_n)^\top Z_{t_n}) \Delta t_n \right], \tag{6.29}$$

where $\mathbb{E}^n[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{t_n}]$ is the conditional expectation, $m = 1$ the dimension of the Wiener process and $(Y_n)_{n\in\{0,\ldots,N\}}$ the corresponding approximation of the forward process $(X_t)_{0\le t\le T}$ via a standard discretization scheme of Sections 3.3.1 and 3.3.2.

The seminal Longstaff-Schwartz algorithm presented in the pioneering work Longstaff and Schwartz (2001) gives an advice how to estimate conditional expectation in order to approximate the solution of an BSDE. Second-order approximation schemes can be found in Crisan and Manolarakis (2014). The above scheme is not directly implementable. Special care needs to be taken for the computation of the involved conditional expectations as described

in Crisan et al. (2010). Schemes with high order of convergence can be found in Gobet and Labart (2007). An survey of the accuracy and complexity of such algorithms can be found in Lemor et al. (2006). The rate of convergence is of order $\mathcal{O}(\Delta t)$.

By combining equations (6.29) with (6.28) and using the tower property and taking out what is known (Lemma 2.2.4) one obtains

$$P_{t_n} = \mathbb{E}^n \left[P_{t_{n+1}} + a_x(t_n, Y_n)^\top P_{t_{n+1}} \Delta t_n + b_x(t_n, Y_n)^\top P_{t_{n+1}} \Delta W_n \right]. \tag{6.30}$$

This discretized approximation corresponds to equation (4.30). Afterwards the usual discretization of the gradient representation (6.10)

$$\nabla \phi(u) = \mathbb{E}\left[\int_0^T P_s^\top a_u(s, X_s, u) + Z_s^\top b_u(s, X_s, u) ds \right]$$

looks like

$$\mathbb{E}\left[\sum_{n=0}^{N-1} P_{t_{n+1}}^\top a_u(t_n, Y_n, u) \Delta t_n + Z_{t_n}^\top b_u(t_n, Y_n, u) \Delta t_n \right]. \tag{6.31}$$

Subsequently, to compare this *optimize-then-discretize* gradient (6.31) with the corresponding *discretize-then-optimize* gradient (4.29) of Theorem 4.2.7 the following corollary is presented.

Corollary 6.4.1. *The discrete gradient* (4.29) *does not differ from the discretized gradient* (6.31).

Proof. By using equations (6.28) and (6.29) and, in particular, the tower property (2.5) to eliminate the conditional expectation one obtains

$$\begin{aligned}
\nabla \phi(u) &= \mathbb{E}\left[\sum_{n=0}^{N-1} P_{t_{n+1}}^\top a_u(t_n, Y_n, u) \Delta t_n + Z_{t_n}^\top b_u(t_n, Y_n, u) \Delta t_n \right] \\
&= \mathbb{E}\left[\sum_{n=0}^{N-1} P_{t_{n+1}}^\top a_u(t_n, Y_n, u) \Delta t_n + \mathbb{E}^n \left[P_{t_{n+1}}^\top \frac{\Delta W_n}{\Delta t_n} \right] b_u(t_n, Y_n, u) \Delta t_n \right] \\
&= \mathbb{E}\left[\sum_{n=0}^{N-1} P_{t_n+1}^\top a_u(t_n, Y_n, u) \Delta t_n + P_{t_n+1}^\top b_u(t_n, Y_n, u) \Delta W_n \right]
\end{aligned}$$

Furthermore, by approximating the expectation via Monte-Carlo method and making use of the fact that $P_{t_n} = \mathbb{E}^n[\lambda_n]$, $Z_{t_n} = \mathbb{E}^n\left[\lambda_{n+1} \frac{\Delta W_n}{\Delta t_n}\right]$ and $y_n = Y_n$ pathwise for all $n = 1, \ldots, N$ one finally obtains the corresponding discrete gradient (4.29) as

$$\frac{1}{M} \sum_{m=1}^{M} \sum_{n=1}^{N} (\lambda_n^m)^\top \left[\frac{\partial}{\partial u} a(t_{n-1}, y_{n-1}^m, u) \Delta t_{n-1} + \frac{\partial}{\partial u} b(t_{n-1}, y_{n-1}^m, u) \Delta W_{n-1}^m \right]$$

□

This result concludes this thesis in an elegant manner. Although, the discretization of the solution of an BSDEs need conditional expectations, a closer look reveals, that the expensive computations of conditional expectations are not necessary in the case of gradient-based calibration routines. In conclusion, Theorem 4.2.6 cleverly achieves a gradient-based calibration method only with the most necessary computations. This fortifies the adjoint method in calibration of stochastic financial models.

List of Tables

List of Figures

Bibliography

Abdulle, A., Cohen, D., Vilmart, G. and Zygalakis, K. (2012). High weak order methods for stochastic differential equations based on modified equations, *SIAM Journal on Scientific Computing* **34**(3): A1800–A1823.

Abdulle, A., Vilmart, G. and Zygalakis, K. (2014). High order numerical approximation of the invariant measure of ergodic SDEs, *SIAM Journal on Numerical Analysis* **52**(4): 1600–1622.

Achdou, Y. and Pironneau, O. (2005). *Computational Methods for Option Pricing*, Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM).

Aït-Sahalia, Y. and Kimmel, R. (2007). Maximum likelihood estimation of stochastic volatility models, *Journal of Financial Economics* **83**(2): 413–452.

Alkhatib, A. and King, P. (2014). Robust quantification of parametric uncertainty for surfactant-polymer flooding, *Computational Geosciences* **18**(1): 77–101.

Alt, W. (1990). The Lagrange-Newton method for infinite-dimensional optimization problems, *Numerical Functional Analysis and Optimization* **11**(3-4): 201–224.

Andersen, L. and Andreasen, J. (2000). Jump-diffusion processes: Volatility smile fitting and numerical methods for option pricing, *Review of Derivatives Research* **4**(3): 231–262.

Andersen, L. and Brotherton-Ratcliffe, R. (1997). The equity option volatility smile: an implicit finite-difference approach, *Journal of Computional Finance* **1**: 5–38.

Apostolos, G. N. and Skiadas, C. H. (1995). Forecasting the electricity consumption by applying stochastic modelling techniques: The case of Greece, *in* J. Janssen, C. H. Skiadas and C. Zopounidis (eds), *Advances in Stochastic Modelling and Data Analysis*, Springer Netherlands, pp. 85–100.

Arridge, S. and Schweiger, M. (1998). A gradient-based optimisation scheme for optical tomography, *Optics Express* **2**(6): 213–226.

Bachelier, L. (1900). Théorie de la spéculation, *Annales scientifiques de l'École Normale Supérieure* **3**(17): 21–86.

Bahlali, K., Djehiche, B. and Mezerdi, B. (2007). On the stochastic maximum principle in optimal control of degenerate diffusions with Lipschitz coefficients, *Applied Mathematics and Optimization* **56**(3): 364–378.

Banks, H. T., David, J. and Tran, H. (2008). HIV model analysis under optimal control based treatment strategies, *Technical Report CRSC-TR08-07*, Center for Research in Scientific Computation, Raleigh, US-NC.

Baras, J. S., Elliott, R. J. and Kohlmann, M. (1988). The conditional adjoint process, *in* A. Bensoussan and J. Lions (eds), *Analysis and Optimization of Systems*, Vol. 111 of *Lecture Notes in Control and Information Sciences*, Springer Berlin Heidelberg, pp. 654–662.

Bates, D. (1996). Jump and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options, *The Review of Financial Studies* **9**: 69–107.

Belomestny, D., Bender, C. and Schoenmakers, J. (2009). True upper bounds for Bermudan products via non-nested Monte Carlo, *Mathematical Finance* **19**(1): 53–71.

Belomestny, D., Schoenmakers, J. and Dickmann, F. (2013). Multilevel dual approach for pricing American style derivatives, *Finance and Stochastics* **17**(4): 717–742.

Bender, C. (2011). Primal and dual pricing of multiple exercise options in continuous time, *SIAM Journal of Financial Mathematics* **2**(1): 562–586.

Bender, C. and Denk, R. (2007). A forward scheme for backward SDEs, *Stochastic Processes and their Applications* **117**(12): 1793–1812.

Bender, C. and Dokuchaev, N. (2014). A first-order BSPDE for swing option pricing, *Mathematical Finance* .

Bender, C., Groß, B. P. and Sachs, E. W. (2014). Adjoint SDE for calibration, *Preprint* .

Bender, C. and Kohlmann, M. (2008). Optimal superhedging under non-convex constraints - a BSDE approach, *International Journal of Theoretical and Applied Finance* **11**(04): 363–380.

Bender, C., Schoenmakers, J. and Zhang, J. (2013). Dual representations for general multiple stopping problems, *Mathematical Finance* .

Bender, C., Schweizer, N. and Zhuo, J. (2013). A primal-dual algorithm for BSDEs, *arXiv preprint arXiv:1310.3694* .

Bender, C., Sottinen, T. and Valkeila, E. (2008). Pricing by hedging and no-arbitrage beyond semimartingales, *Finance and Stochastics* **12**(4): 441–468.

Bender, C. and Steiner, J. (2013). A posteriori estimates for backward SDEs, *SIAM ASA Journal on Uncertainty Quantification* **1**(1): 139–163.

Bender, C. and Zhang, J. (2008). Time discretization and markovian iteration for coupled FBSDEs, *The Annals of Applied Probability* **18**(1): 143–177.

Benhamou, E., Gobet, E. and Miri, M. (2010). Time dependent Heston model, *SIAM Journal on Financial Mathematics* **1**(1): 289–325.

Bernhart, M., Pham, H., Tankov, P. and Warin, X. (2012). Swing options valuation: A bsde with constrained jumps approach, *in* R. A. Carmona, P. Del Moral, P. Hu and N. Oudjane (eds), *Numerical Methods in Finance*, Vol. 12 of *Springer Proceedings in Mathematics*, Springer Berlin Heidelberg, pp. 379–400.

Bertsekas, D. P. (1999). *Nonlinear Programming*, number 2nd in *Optimization and Computation Series*, 2 edn, Athena Scientific, Belmont, Massachusetts, USA.

Biagini, F., Bregman, Y. and Meyer-Brandis, T. (2008). Pricing of catastrophe insurance options written on a loss index with reestimation, *Insurance: Mathematics and Economics* **43**(2): 214–222.

Billingsley, P. (1995). *Probability and Measure*, Wiley Series in probability and mathematical statistics, 3 edn, John Wiley & Sons.

Bismut, J.-M. (1973). Conjugate convex functions in optimal stochastic control, *Journal of Mathematical Analysis and Applications* **44**(2): 384–404.

Bismut, J.-M. (1978). An introductory approach to duality in optimal stochastic control, *SIAM Review* **20**(1): 62–78.

Black, F. and Karasinski, P. (1991). Bond and option pricing when short rates are lognormal, *Financial Analysts Journal* **47**(4): 52–59.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* **81**(3): 637–654.

Bonnans, J. F., Gilbert, J. C., Lemarechal, C. and Sagastizabal, C. A. (2006). *Numerical Optimization*, 2nd edn, Springer-Verlag.

Bonnans, J. F. and Shapiro, A. (2000). *Pertubation Analysis of Optimization Problems*, Springer Series in Operation Research, Springer-Verlag.

Bormetti, G., Cazzola, V. and Delpini, D. (2010). Option pricing under Ornstein-Uhlenbeck stochastic volatility: a linear model, *International Journal of Theoretical and Applied Finance* **13**: 1047–1063.

Bosq, D. (2010). *Linear Processes in Function Spaces: Theory and Applications*, Vol. 149 of *Lecture Notes in Statistics*, Springer-Verlag.

Bosse, T., Gauger, N., Griewank, A., Günther, S., Kaland, L. and et al. (2014). Optimal design with bounded retardation for problems with non-separable adjoints, *International Series of Numerical Mathematics* **165**: 67–84.

Bouchard, B. and Touzi, N. (2004). Discrete time approximation and Monte-Carlo simulation of backward stochastic differential equations, *Stochastic Processes and their Applications* **111**(2): 175–206.

Bourbaki, N. (1987). *Topological vector spaces - Elements of mathematics*, Springer-Verlag, Berlin.

Box, G. E. P. and Draper, N. R. (1987). *Empirical Model-building and Response Surfaces*, Wiley.

Boyle, P. (1977). Options: A Monte Carlo approach, *Journal of Financial Economics* **4**(3): 323–338.

Boyle, P., Broadie, M. and Glasserman, P. (1997). Monte Carlo methods for security pricing, *Journal of Economic Dynamics and Control* **21**(8-9): 1267–1321.

Bradley, T., Chong, J., Dixon, M. F. and Keutzer, K. (2012). Monte Carlo based financial market Value-at-Risk estimation on GPUs, *GPU Computing Gems; Edition: Jade* pp. 337–358.

Bradley, T., du Toit, J., Giles, M., Tong, R. and Woodhams, P. (2011). Parallelisation techniques for random number generators, *GPU Computing Gems, Emerald Edition*, Morgan Kaufmann, pp. 231–246.

Broadie, M. and Glasserman, P. (1996). Estimating security price derivatives using simulation, *Management Science* pp. 269–285.

Bruti-Liberati, N. and Platen, E. (2005). On the strong approximation of jump-diffusion processes, *Technical Report 157*, Quantitative Finance Research Centre: University of Technology, Sidney.

Bruti-Liberati, N. and Platen, E. (2006a). Approximation of jump diffusions in finance and economics, *Technical Report 176*, Quantitative Finance Research Centre: University of Technology, Sidney.

Bruti-Liberati, N. and Platen, E. (2006b). On weak predictor-corrector schemes for jump-diffusion processes in finance, *Technical Report 179*, Quantitative Finance Research Centre: University of Technology, Sidney.

Bruti-Liberati, N. and Platen, E. (2008). Strong predictor-corrector Euler methods for stochastic differential equations, *Technical Report 222*, Quantitative Finance Research Centre: University of Technology, Sidney.

Buckwar, E., Rößler, A. and Winkler, R. (2010). Stochastic Runge–Kutta methods for ito SODEs with small noise, *SIAM Journal of Scientific Computing* **32**(4): 1789–1808.

Buckwar, E. and Sickenberger, T. (2011). A comparative linear mean-square stability analysis of Maruyama- and Milstein-type methods, *Mathematics and Computers in Simulation* **81**(6): 1110–1127.

Buckwar, E. and Winkler, R. (2007). Improved linear multi-step methods for stochastic ordinary differential equations, *Journal Computational and Applied Mathematics* **205**(2): 912–922.

Burrage, K., Burrage, P. M. and Tian, T. (2004). Numerical methods for strong solutions of stochastic differential equations: An Overview, *Proceedings: Mathematical, Physical and Engineering Sciences*, Vol. 460 - 2041, The Royal Society, pp. 373–402.

Burrage, K. and Tian, T. (2002). Predictor-corrector methods of Runge-Kutta type for stochastic differential equations, *SIAM Journal on Numerical Analysis* **40**(4): 1516–1537.

Busch, M., Korn, R. and Seifried, F. T. (2013). Optimal consumption and investment for a large investor: An intensity-based control framework, *Mathematical Finance* **23**(4): 687–717.

Cangiani, A. (2000). *Implied volatility estimation using adjoint monte carlo methods*, Master's thesis, St Hugh's College - University Oxford.

Capriotti, L. and Giles, M. (2012). Adjoint Greeks made easy, *Risk-London* **25**(9): 92–102.

Carletti, M. (2002). On the stability properties of a stochastic model for phage-bacteria interaction in open marine enviroment, *Mathematical Bioscience* **175**: 117–131.

Carletti, M. (2006). Numerical simulation of stochastic differential problems in the bioscience, *Journal of Computional and Applied Mathematics* **186**: 422–440.

Carletti, M., Burrage, K. and Burrage, P. M. (2004). Numerical simulation of stochastic ordinary differential equations in biomathematical modeling, *Mathematics and Computers in Simulation* **64**: 271–277.

Carnarius, A., Thiele, F., Özkaya, E. and Gauger, N. R. (2010). Adjoint approaches for optimal flow control, *American Institute of Aeronautics and Astronautics Paper* **5088**.

Carr, P., Ellis, K. and Gupta, V. (1998). Static hedging of exotic options, *The Journal of Finance* **53**(3): 1165–1190.

Cartea, A., Figueroa, M. G. and Geman, H. (2009). Pricing of catastrophe insurance options written on a loss index with reestimation, *Applied Mathematical Finance* **16**(2): 103–122.

Chalmers, G. D. and Higham, D. (2008). Asymptotic stability of a jump-diffusion equation and its numerical approximation, *SIAM Journal of Scientific Computing* **21**(2): 1141–1155.

Chen, K.-C., Wang, T.-Y., Tseng, H.-H., Huang, C.-Y. and Kao, C.-Y. (2005). A stochastic differential equation model for quantifying transcriptional regulatory network in Saccharomyces cerevisiae, *Life Science: Bioinformatics* **21**(12): 2883–2890.

Christensen, P. H. (2003). Adjoints and importance in rendering: An overview, *IEEE Transactions on Visualization and Computer Graphics: Pixar Animation Studios, Seattle, WA, USA* **9**(3): 329–340.

Christensen, S. (2014). A method for pricing American options using semi-infinite linear programming, *Mathematical Finance* **24**(1): 156–172.

Chung, K. L. and Williams, R. J. (1990). *Introduction to Stochastic Integration*, Birkhäuser.

Clarke, F. (2013). *Functional Analysis, Calculus of Variations and Optimal Control*, Vol. 264 of *Graduate Texts in Mathematics*, Springer.

Collis, S. S. and Heinkenschloss, M. (2002). Analysis of the streamline upwind/Petrov Galerkin method applied to the solution of optimal control problems, *Technical Report TR02-01*, Computational and Applied Mathematics CAAM Technical Report of the Rice University, Houston, TX, USA.

Cont, R. and Fourniè, D.-A. (2013). Functional Ito calculus and stochastic integral representation of martingales, *Annals of Probability* **41**(1): 109–133.

Cont, R. and Tankow, P. (2004). *Financial Modelling with Jump Process*, Financial Mathematics Series, 1st edn, Chapman/Hall.

Courtier, P., Derber, J., Errico, R., Louis, J.-F. and Vukicevic, T. (1993). Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology, *Tellus A* **45**(5): 342–357.

Cox, J. C., Ingersoll, Jonathan E, J. and Ross, S. A. (1985). A theory of the term structure of interest rates, *Econometrica* **53**(2): 385–407.

Cox, J. C., Ross, S. A. and Rubinstein, M. (1979). Option pricing: A simplified approach, *Journal of Financial Economics* **7**(3): 229–263.

Crisan, D. and Manolarakis, K. (2014). Second order discretization of backward SDEs and simulation with the cubature method, *The Annals of Applied Probability* **24**(2): 652–678.

Crisan, D., Manolarakis, K. and Touzi, N. (2010). On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights, *Stochastic Processes and their Applications* **120**(7): 1133 – 1158.

Dai, B., Peng, Y. and Gong, B. (2010). Parallel option pricing with BSDE method on GPU, *Grid and Cooperative Computing (GCC), 2010 9th International Conference on*, pp. 191–195.

Debnath, L. and Mikusinski, P. (1998). *Introduction to Hilbert Spaces with Applications*, Academic Press.

Debrabant, K. and Rößler, A. (2008). Classification of stochastic Runge-Kutta methods for the weak approximation of stochastic differential equations, *Mathematics and Computers in Simulation* **77**(4): 408–420.

Debrabant, K. and Rößler, A. (2015). On the acceleration of the multi-level Monte Carlo method, *Journal of Applied Probability* **52**(2): 307–322.

Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing, *Mathematische Annalen* **300**(1): 463–520.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B. Methodological* **39**(1): 1–38.

Di Nunno, G., Øksendal, B. and Proske, F. (2009). *Malliavin Calculus for Lèvy Processes with Applications to Finance*, Universitext, Springer Berlin Heidelberg.

Dieudonne, J. (1969). *Foundations of Modern Analysis*, Academic Press.

Doob, J. L. (1971). What is a martingale?, *The American Mathematical Monthly* **78**(5): 451–463.

Dowd, M. (2011). Estimating parameters for a stochastic dynamic marine ecological system, *Environmetrics* **22**(4): 501–515.

Duffie, D., Pan, J. and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* **68**(6): 1343–1376.

Duffie, D., Schroder, M. and Skiadas, C. (1996). Recursive valuation of defaultable securities and the timing of resolution of uncertainty, *The Annals of Applied Probability* **6**(4): 1075–1090.

Duffie, D. and Skiadas, C. (1994). Continuous-time security pricing: A utility gradient approach, *Journal of Mathematical Economics* **23**(2): 107–131.

Duffy, D. J. and Kienitz, J. (2009). *Efficient and Robust Monte Carlo Methods in Financial Engineering - Design and Implementation in C++*, Wiley Finance.

Dupire, B. (1994). Pricing with a smile, *Risk* (1): 18–20.

Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, *Annalen der Physik* **322**(8): 549–560.

El Karoui, N., Peng, S. and Quenez, M. C. (1997). Backward stochastic differential equations in finance, *Mathematical Finance* **7**(1): 1–71.

Elliott, R. J. (1977). The optimal control of a stochastic system, *SIAM Journal on Control and Optimization* **15**(5): 756–778.

Elliott, R. J. and Kohlmann, M. (1989a). The adjoint process in stochastic optimal control, *in* N. Christopeit, K. Helmes and M. Kohlmann (eds), *Stochastic Differential Systems*, Vol. 126 of *Lecture Notes in Control and Information Sciences*, Springer Berlin Heidelberg, pp. 115–127.

Elliott, R. J. and Kohlmann, M. (1989b). Integration by parts and densities for jump processes, *Stochastics* **27**(2): 83–97.

Elliott, R. J. and Kohlmann, M. (1989c). Martingale representation and the Malliavin calculus, *Applied Mathematics and Optimization* **20**(1): 105–112.

Elliott, R. J. and Kohlmann, M. (1989d). The variational principle for optimal control of diffusions with partial information, *Systems & Control Letters* **12**(1): 63 – 69.

Elliott, R. and Kopp, P. (2005). *Mathematics of Financial Markets*, Vol. 10 of *Springer finance*, Springer.

Engl, H. W., Hanke-Bourgeois, M. and Neubauer, A. (1996). *Regularization of Inverse Problems*, number 375 in *Mathematics and Its Applications*, Kluwer Academic, Dordrecht.

Fang, H. (2000). Option pricing implications of a stochastic jump rate, *Technical report*, Department of Economics. University of Virginia, Charlottesville, VA. Working Paper.

Fatone, L., Giacinti, M., Mariani, F., Recchioni, M. and Zirilli, F. (2012). Parallel option pricing on GPU: Barrier options and realized variance options, *The Journal of Supercomputing* **62**.

Feller, W. (1951). Two singular diffusion problems, *Annals of Mathematics* **54**(1): 173–182.

Föllmer, H. and Schied, A. (2004). *Stochastic Finance: An Introduction in Discrete Time*, de Gruyter Studies in Mathematics, 2nd edn, Walter de Gruyter, Berlin.

Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide*, John Wiley & Sons.

Gentle, J. E. (2004). *Random Number Generation and Monte Carlo Methods*, Statistics and Computing, 2nd edn, Springer.

Giannakoglou, K. C. and Papadimitriou, D. I. (2008). Adjoint methods for shape optimization, *in* D. Thevenin and G. Janiga (eds), *Optimization and Computational Fluid Dynamics*, Springer Berlin Heidelberg, pp. 79–108.

Giles, M. (1997). On adjoint equations for error analysis and optimal grid adaptation in CFD. `https://people.maths.ox.ac.uk/gilesm/files/NA-97-11.pdf`.

Giles, M. (2002). On the iterative solution of adjoint equations, *in* G. Corliss, C. Faure, A. Griewank, L. Hascoët and U. Naumann (eds), *Automatic Differentiation of Algorithms*, Springer New York, pp. 145–151.

Giles, M. (2015). Multilevel Monte Carlo methods, *Acta Numerica* **24**: 259–328.

Giles, M. B. and Glasserman, P. (2006). Smoking adjoints: fast Monte Carlo Greeks, *Risk* pp. 88–92.

Giles, M. B. and Pierce, N. A. (2000). An introduction to the adjoint approach to design, *Flow, Turbulence and Combustion* **65**(3-4): 393–415.

Giles, M., Debrabant, K. and Rößler, A. (2013). Numerical analysis of multilevel Monte Carlo path simulation using the Milstein discretisation, *arXiv preprint arXiv:1302.4676* .

Giles, M., Higham, D. and Mao, X. (2009). Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff, *Finance and Stochastics* **13**(3): 403–413.

Giles, M. and Süli, E. (2002). Adjoint methods for PDEs: a posteriori error analysis and postprocessing by duality, *Acta Numerica* **11**: 145–236.

Giles, M. and Szpruch, L. (2013). Antithetic multilevel Monte Carlo estimation for multidimensional SDEs, *in* J. Dick, F. Y. Kuo, G. W. Peters and I. H. Sloan (eds), *Monte Carlo and Quasi-Monte Carlo Methods 2012*, Vol. 65 of *Springer Proceedings in Mathematics & Statistics*, Springer Berlin Heidelberg, pp. 367–384.

Gill, P. E. and Murray, W. (1974). *Numerical Methods for Constrained Optimization*, Academic Press, London.

Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*, 1st edition edn, Springer-Verlag, New York.

Gobet, E. and Labart, C. (2007). Error expansion for the discretization of backward stochastic differential equations, *Stochastic Processes and their Applications* **117**(7): 803–829.

Golub, G. H. and van Loan, C. F. (1996). *Matrix Computations*, John Hopkins Studies in the Mathematical Sciences, 3rd edn, The John Hopkins University Press, Baltimore, Maryland, USA.

Griewank, A., Kulshreshtha, K. and Walther, A. (2012). On the numerical stability of algorithmic differentiation, *Computing* **94**(2-4): 125–149.

Griewank, A. and Walther, A. (2008). *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, 2nd edn, Society for Industrial and Applied Mathematics.

Groß, B. P. (2009). *Numerische Simulation von Lévy Modellen zur Optionspreisberechnung*, Diplomarbeit, Universität Trier, Germany.

Groß, B. P. and Sachs, E. W. (2012). Fast calibration of SDE-based financial models using an adjoint technique, *Preprint* .

Gutiérrez, R., Gutiérrez-Sanchez, R. and Nafidi, A. (2005). Forecasting total natural-gas consumption in Spain by using the stochastic Gompertz innovation diffusion model, *Applied Energy* **80**: 115–124.

Gutiérrez, R., Gutiérrez-Sanchez, R. and Nafidi, A. (2006). Electricity consumption in Morocco: Stochastic Gompertz diffusion analysis with exogenous factors, *Applied Energy* **83**: 1139–1151.

Hadamard, J. (1902). Sur les problèmes aux dérivés partielles et leur signification physique, *Princeton University Bulletin* **13**: 49–52.

Hakala, J. and Wystup, U. (2002). *Foreign Exchange Risk: Models, Instruments and Strategies*, Riskbooks.

Haslinger, J. and Mäkinen, R. (2003). *Introduction to Shape Optimization*, Society for Industrial and Applied Mathematics.

Haugh, M. B. and Kogan, L. (2004). Pricing American options: A duality approach, *Operations Research* **52**(2): 258–270.

Haussmann, U. (1981). On the adjoint process for optimal control of diffusion processes, *SIAM Journal on Control and Optimization* **19**(2): 221–243.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* **6**: 327–343.

Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations, *SIAM Review* **43**(3): 525–546.

Higham, D. J., Mao, X. and Stuart, A. (2002). Strong convergence of Euler-type methods for nonlinear stochastic differential equations, *SIAM Journal on Numerical Analysis* **40**(3): 1041–1063.

Hinze, M. (2009). Discrete concepts in PDE constrained optimization, *Optimization with PDE Constraints*, Vol. 23 of *Mathematical Modelling: Theory and Applications*, Springer Netherlands, pp. 157–232.

Hinze, M. and Rösch, A. (2012). Discretization of optimal control problems, *in* G. Leugering, S. Engell, A. Griewank, M. Hinze, R. Rannacher, V. Schulz, M. Ulbrich and S. Ulbrich (eds), *Constrained Optimization and Optimal Control for Partial Differential Equations*, Vol. 160 of *International Series of Numerical Mathematics*, Springer Basel, pp. 391–430.

Hinze, M. and Tröltzsch, F. (2010). Discrete concepts versus error analysis in PDE-constrained optimization, *GAMM-Mitteilungen* **33**(2): 148–162.

Horváth, A. and Manini, D. (2008). Parameter estimation of kinetic rates in stochastic reaction networks by the EM method, *Proceedings of the 2008 International Conference on BioMedical Engineering and Informatics - Volume 01*, BMEI '08, IEEE Computer Society, Washington, DC, USA, pp. 713–717.

Hull, J. (2008). *Options, Futures, and Other Derivatives*, 7th edn, Pearson Prentice Hall.

Hull, J. C. and White, A. D. (1987). The pricing of options on assets with stochastic volatilities, *Journal of Finance* **42**(2): 281–300.

Ikeda, N. and Watanabe, S. (1989). *Stochastic differential equations and diffusion processes*, North-Holland mathematical library, North-Holland Pub. Co. Tokyo, Amsterdam, New York.

Jäckel, P. (2002). *Monte Carlo Methods in Finance*, Wiley.

Jacod, J. and Protter, P. (2013). *Probability Essentials*, Wiley Series in probability and mathematical statistics, 2 edn, Springer-Verlag.

Jameson, A. (1988). Aerodynamic design via control theory, *Journal of Scientific Computing* **3**(3): 233–260.

Janczura, J. (2014). Pricing electricity derivatives within a markov regime-switching model: a risk premium approach, *Mathematical Methods of Operations Research* **79**(1): 1–30.

Joshi, M. S. (2010). Graphical Asian options, *Wilmott Journal* **2**(2): 97–107.

Käbe, C. (2010). *Feasibility and Efficiency of Monte Carlo Based Calibration of Financial Market Models*, PhD thesis, Universität Trier.

Käbe, C., Maruhn, J. H. and Sachs, E. W. (2009). Adjoint-based Monte Carlo calibration of financial market models, *Finance and Stochastics - Special Issue on Computational Methods in Finance (Part I)* **13**(3): 351–379.

Kahl, C. and Jäckel, P. (2006). Fast strong approximation Monte Carlo schemes for stochastic volatility models, *Quantitative Finance* **6**(6): 513–536.

Kantorovich, L. V. and Akilov, G. P. (1982). *Functional Analysis*, 2nd edn, Pergamon.

Karatzas, I. and Shreve, S. (1998). *Methods of Mathematical Finance*, number 39 in *Stochastic Modelling and Applied Probability*, Springer.

Karatzas, I. and Shreve, S. (2000). *Brownian Motion and Stochastic Calculus*, Graduate texts in mathematics, Springer.

Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*, 2nd edn, Academic Press, New York.

Karlin, S. and Taylor, H. M. (1981). *A Second Course in Stochastic Processes*, Academic Press, New York.

Kelley, C. (1995). *Iterative Methods for Linear and Nonlinear Equations*, Frontiers in applied mathematics, Society for Industrial and Applied Mathematics.

Kloeden, P. and Platen, E. (1999). *Numerical Solution of Stochastic Differential Equation*, number 23 in *Stochastic Modelling ans Applied Probability*, Springer.

Kohlmann, M. and Zhou, X. Y. (2000). Relationship between backward stochastic differential equations and stochastic controls: A linear-quadratic approach, *SIAM Journal on Control and Optimization* **38**(5): 1392–1407.

Kolb, C. and Pharr, M. (2005). Options pricing on the GPU, *GPU Gems 2: programming techniques for high-performance graphics and general-purpose computation* pp. 719–732. `https://developer.nvidia.com/gpugems/GPUGems2/gpugems2_chapter45.html`.

Komori, Y. (2007). Weak second-order stochastic Runge-Kutta methods for non-commutative stochastic differential equations, *Journal of Computational and Applied Mathematics* **206**(1): 158–173.

Komori, Y. (2008). Weak first- or second-order implicit Runge-Kutta methods for stochastic differential equations with a scalar Wiener process, *Journal of Computational and Applied Mathematics* **217**(1): 166–179.

Kosmol, P. and Müller-Wichards, D. (2014). *Optimization in Function Spaces*, Series in Nonlinear Analysis and Applications 13, De Gruyter, Berlin, Boston.

Kou, S. G. (2002). A jump-diffusion model for option pricing, *Management Science* **48**(8): 1086–1101.

Kunita, H. (1982). On backward stochastic differential equations, *Stochastics* **6**(3-4): 293–313.

Kushner, H. (1972). Necessary conditions for continuous parameter stochastic optimization problems, *SIAM Journal on Control* **10**(3): 550–565.

Lamberton, D. and Lapeyre, B. (1996). *Introduction to Stochastic Calculus Applied to Finance*, Chapman & Hall.

L'Ecuyer, P. (1996). Combined multiple recursive random number generators, *Operations Research* **44**(5): 816–822.

Lee, A., Yau, C., Giles, M. B., Doucet, A. and Holmes, C. C. (2010). On the utility of graphics cards to perform massively parallel simulation of advanced Monte Carlo methods, *Journal of Computational and Graphical Statistics* **19**(4): 769–789.

Lemor, J.-P., Gobet, E. and Warin, X. (2006). Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations, *Bernoulli* **12**(5): 889–916.

Li, H., Xiao, L. and Ye, J. (2013). Strong predictor-corrector Euler-Maruyama methods for stochastic differential equations with Markovian switching, *Journal of Computational and Applied Mathematics* **237**(1): 5–17.

Lipton, A. (2002). The vol smile problem, *Risk* **February**: 61–65.

Liu, X. and Zhai, Z. (2007). Inverse modeling methods for indoor airborne pollutant tracking: literature review and fundamentals, *Indoor Air* **17**(6): 419–438.

Longstaff, F. and Schwartz, E. (2001). Valuing American options by simulation: a simple least-squares approach, *Review of Financial Studies* **14**(1): 113–147.

Lörx, A. (2013). *Adjoint-Based Calibration of Local Volatility Models*, Berlin, Mensch & Buch.

Lörx, A. and Sachs, E. W. (2012). Model calibration in option pricing, *SQU Journal for Science* **17**(1): 84–102.

Luenberger, D. G. (1998). *Optimization by Vector Space Methods*, Wiley Interscience.

Lusternik, L. A. and Sobolev, V. I. (1965). *Elements of functional analysis.*

Ma, J., Protter, P., San Martin, J. and Torres, S. (2002). Numberical method for backward stochastic differential equations, *The Annals of Applied Probability* **12**(1): 302–316.

Ma, J. and Yong, J. (1999). *Forward-Backward Stochastic Differential Equations and Their Applications*, Vol. 170 of *Lecture Notes in Mathematics*, Springer, Berlin.

Marchuk, G. I., Agoshkov, V. I. and Shutyaev, V. P. (1996). *Adjoint Equations and Perturbation Algorithms in Nonlinear Problems*, CRC Press, Boca Raton, New York, London.

Marchuk, G., Shutyaev, V. and Bocharov, G. (2005). Adjoint equations and analysis of complex systems: Application to virus infection modelling, *Journal of Computational and Applied Mathematics* **184**(1): 177–204.

Maruhn, J. H. and Sachs, E. W. (2006). Robust static super-replication of barrier options in the Black-Scholes model, *in* A. J. Kurdila, P. M. Pardalos and M. Zabarankin (eds), *Robust Optimization-Directed Design*, Vol. 81 of *Nonconvex Optimization and Its Applications*, Springer US, pp. 135–155.

Meise, R. and Vogt, D. (1997). *Introduction to Functional Analysis*, Vol. 2 of *Oxford Graduate Texts in Mathematics*, Clarendon Press.

Merton, R. (1973). Theory of rational option pricing, *Bell Journal of Economics and Management Science* **4**: 141–183.

Merton, R. (1976). Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics* **3**(1-2): 125–144.

Metropolis, N. and Ulam, S. (1949). The Monte Carlo method, *Journal of the American Statistical Association* **44**(247): 335–341.

Michalak, A. M. and Kitanidis, P. K. (2004). Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling, *Water Resources Research* **40**(8).

Mikhailov, S. and Nögel, U. (2003). Heston's stochastic volatility model-implementation, calibration and some extensions, *Wilmott Magazine* .

Milstein, G. N. (1975). Approximate integration of stochastic differential equations, *Theory of Probability & Its Applications* **19**(3): 557–562.

Mou, L. and Yong, J. (2007). A variational formula for stochastic controls and some applications, *Pure and Applied Mathematics Quarterly* **3**(2): 539–567.

Moummou, E. K., Gutiérrez, R. and Gutiérrez-Sanchez, R. (2012). A stochastic Gompertz model with logarithmic therapy functions: Parameters estimation, *Applied Mathematics and Computation* **219**(8): 3729–3739.

Myneni, R. (1992). The pricing of the American option, *The Annals of Applied Probability* **2**(1): 1–23.

Niu, Y. and Zhang, C. (2012). Almost sure and moment exponential stability of predictor-corrector methods for stochastic differential equations, *Journal of Systems Science and Complexity* **25**(4): 736–743.

Noack, A. and Walther, A. (2007). Adjoint concepts for the optimal control of burgers equation, *Computational Optimization and Applications* **36**(1): 109–133.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*, Springer Series in Operations Research, Springer-Verlag, New York, Berlin, Heidelberg.

Nualart, D. (2006). *The Malliavin Calculus and Related Topics*, Probability and its Applications, Springer Berlin Heidelberg.

Øksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications*, 6th edn, Springer-Verlag, Berlin Heidelberg.

Pagès, G. and Wilbertz, B. (2010). Parallel implementation of quantization methods for the valuation of swing options on GPGPU, *Workshop on High Performance Computational Finance (WHPCF), 2010 IEEE*, IEEE Computer Society, pp. 1–5.

Pagès, G. and Wilbertz, B. (2012). GPGPUs in computational finance: massive parallel computing for American style options, *Concurrency and Computation: Practice and Experience* **24**(8): 837–848.

Pardoux, E. and Peng, S. (1990). Adapted solution of a backward stochastic differential equation, *Systems & Control Letters* **14**(1): 55–61.

Peng, S. (1992). Stochastic Hamilton-Jacobi-Bellman equations, *SIAM Journal on Control and Optimization* **30**(2): 284–304.

Peng, Y., Gong, B., Liu, H. and Dai, B. (2011). Option pricing on the GPU with Backward Stochastic Differential Equation, *Proceedings - 2011 4th International Symposium on Parallel Architectures, Algorithms and Programming, PAAP 2011* .

Pironneau, O. (1974). On optimum design in fluid mechanics, *Journal of Fluid Mechanics* **64**: 97–110.

Pironneau, O. (2007). Dupire identities for complex options, *Compte rendu de l'academie des sciences I* **344**: 127 – 133.

Platen, E. (1995). On weak implicit and predictor-corrector methods, *Mathematics and Computers in Simulation* **38**(1-3): 69–76.

Platen, E. and Shi, L. (2008). On the numerical stability of simulation methods for SDEs, *Technical Report 234*, Quantitative Finance Research Centre: University of Technology, Sidney.

Protter, P. (1991). *Stochastic integration and differential equations*, 1st edn, Springer-Verlag, New York.

Reinker, S., Altman, R. M. and Timmer, J. (2006). Parameter estimation in stochastic biochemical reactions, *Systems Biology, IEE Proceedings* **153**(4): 168–178.

Rogers, L. C. G. (2002). Monte Carlo valuation of American options, *Mathematical Finance* **12**(3): 271–286.

Rogers, L. C. G. (2007). Pathwise stochastic optimal control, *SIAM Journal on Control and Optimization* **46**(3): 1116–1132.

Rößler, A. (2009). Second order Runge-Kutta methods for Itô stochastic differential equations, *SIAM J. Numerical Analysis* **47**(3): 1713–1738.

Rößler, A. (2010). Runge-Kutta methods for the strong approximation of solutions of stochastic differential equations, *SIAM J. Numerical Analysis* **48**(3): 922–952.

Rößler, A., Seaïd, M. and Zahri, M. (2009). Numerical simulation of stochastic replicator models in catalyzed RNA-like polymers, *Mathematics and Computers in Simulation* **79**(12): 3577–3586.

Rudin, W. (1976). *Principles of Mathematical Analysis*, International Series in Pure and Applied Mathematic, 3rd edn, McGraw Hill.

Rudin, W. (1991). *Functional Analysis*, 2nd edn, McGraw Hill.

Russo, F. and Vallois, P. (1993). Forward, backward and symmetric stochastic integration, *Probability Theory and Related Fields* **97**(3): 403–421.

Russo, F. and Vallois, P. (1995). The generalized covariation process and Itô formula, *Stochastic Processes and their Applications* **59**(1): 81–104.

Sachs, E. W. and Schneider, M. (2014). Reduced order models for the implied variance under local volatility, *International Journal of Theoretical and Applied Finance* **in press**.

Sachs, E. W., Schneider, M. and Schu, M. (2014). Adaptive trust-region POD methods in PIDE-constrained optimization, *in* G. Leugering, P. Benner, S. Engell, A. Griewank, H. Harbrecht, M. Hinze, R. Rannacher and S. Ulbrich (eds), *Trends in PDE Constrained Optimization*, Birkhäuser, Basel, p. in press.

Sachs, E. W. and Schu, M. (2008). Reduced order models (POD) for calibration problems in finance, *in* K. Kunisch, G. Of and O. Steinbach (eds), *Numerical Mathematics and Advanced Applications, ENUMATH 2007*, Springer, pp. 735–742.

Sachs, E. W. and Schu, M. (2010). Reduced order models in PIDE constrained optimization, *Control and Cybernetics* **39**: 661–675.

Sachs, E. W. and Schu, M. (2013). Gradient computation for model calibration with pointwise observations, *in* K. Bredies, C. Clason, K. Kunisch and G. Winckel (eds), *Control and Optimization with PDE Constraints*, Vol. 164 of *International Series of Numerical Mathematics*, Springer Basel, pp. 117–136.

Saito, M. and Matsumoto, M. (2008). SIMD-oriented fast Mersenne twister: a 128-bit pseudorandom number generator, *in* A. Keller, S. Heinrich and H. Niederreiter (eds), *Monte Carlo and Quasi-Monte Carlo Methods 2006*, Springer Berlin Heidelberg, pp. 607–622.

Schachermayer, W. (1992). A Hilbert space proof of the fundamental theorem of asset pricing in finite discrete time, *Insurance: Mathematics and Economics* **11**(4): 249–257.

Schachermayer, W. (2008). The notion of arbitrage and free lunch in mathematical finance, *in* M. Yor (ed.), *Aspects of Mathematical Finance*, Springer Berlin Heidelberg, pp. 15–22.

Schillings, C., Schmidt, S. and Schulz, V. (2011). Efficient shape optimization for certain and uncertain aerodynamic design, *Computers & Fluids* **46**(1): 78–87. 10th ICFD Conference Series on Numerical Methods for Fluid Dynamics (ICFD 2010).

Schmidt, S., Ilic, C., Schulz, V. and Gauger, N. R. (2013). Three-dimensional large-scale aerodynamic shape optimization based on shape calculus, *American Institute of Aeronautics and Astronautics Journal* **51**(11): 2615–2627.

Schmitz Abe, K. E. (2010). θ scheme (orthogonal Milstein scheme), a better numerical approximation for multi-dimensional SDEs, *Global Journal of Computer Science and Technology* **9**(5).

Schmitz Abe, K. E. (2011). Pricing exotic options using MSL-MC, *Quantitative Finance* **11**(9): 1379–1392.

Schmitz Abe, K. E. and Giles, M. (2008). Pricing exotic options using strong convergence properties, *in* H.-G. Bock and et al. (eds), *Progress in Industrial Mathematics at ECMI 2006*, Vol. 12 of *Mathematics in Industry*, Springer Berlin Heidelberg, pp. 614–629.

Schmitz Abe, K. E. and Shaw, W. T. (2005). Measure order of convergence without an exact solution, Euler vs Milstein scheme, *International Journal of Pure and Applied Mathematics* **24**(3): 365 – 382.

Schoutens, W. (2003). *Levy-Processes in Finance*, Wiley Series in Probability and Statistics, 1st edn, Wiley, Chichester.

Schulze, M. (2002). *Parameter Identification for Underdetermined Systems Arising in Option Pricing Models and Neural Networks*, PhD thesis, Universität Trier.

Shiryaev, A. N. (1996). *Probability*, Graduate Textes in Mathematics, 2nd edn, Springer-Verlag, Berlin.

Shiryaev, A. N. (1999). *Essentials of Stochastic Finance*, Vol. 3 of *Advanced Series on Statistical Science & Applied Probability*, World Scientific, Singapore.

Shreve, S. E. (2004). *Stochastic calculus for finance II: Continuous-time models*, New York, NY: Springer.

Skiadas, C. H. (2010). Exact solutions of stochastic differential equations: Gompertz, generalized logistic and revised exponential, *Methodology and Computing in Applied Probability* **12**(2): 261–270.

Stein, E. M. and Stein, J. C. (1991). Stock price distributions with stochastic volatility: An analytic approach, *Review of Financial Studies* **4**(4): 727–52.

Stengel, R. F. and Ghigliazza, R. (2004). Stochastic optimal therapy for enhanced immune response, *Mathematical Biosciences* **191**(2): 123–142.

Stoll, H. R. (1969). The relationship between put and call option prices, *The Journal of Finance* **24**(5): 801–824.

Tikanmäki, H. (2013). Robust hedging and pathwise calculus, *Applied Mathematical Finance* **20**(3): 287–303.

Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of Ill-posed problems*, W.H. Winston.

Toivanen, J. (2008). Numerical valuation of European and American options under Kou's jump-diffusion model, *SIAM Journal on Scientific Computing* **30**(4): 1949–1970.

Vasicek, O. (1977). An equilibrium characterization of the term structure, *Journal of Financial Economics* **5**(2): 177–188.

Wai-Yuan, T. (2002). *Stochastic Models with Applications to Genetics, Cancers, AIDS and other Biomedical Systems*, Vol. 4 of *Series on Concrete and Applicable Mathematics*, World Scientific.

Wengenroth, J. (2008). *Wahrscheinlichkeitstheorie*, de Gryter.

Wilbertz, B. (2005). *Computational aspects of Functional Quantization for Gaussian measures and applications*, Master's thesis, Universität Trier.

Williams, D. (1991). *Probability with Martingales*, Cambridge mathematical textbooks, Cambridge University Press.

Wilmott, P., Dewynne, J. and Howison, J. (1993). *Option Pricing: Mathematical Models and Computation*, Oxford Financial Press.

Yong, J. (2010). Optimality variational principle for controlled forward-backward stochastic differential equations with mixed initial-terminal conditions, *SIAM Journal on Control and Optimization* **48**(6): 4119–4156.

Yong, J. and Zhou, X. (1999). *Stochastic Controls - Hamiltonian Systems and HJB Equations*, Vol. 43 of *Applications of Mathematics*, Springer.

Zeidler, E. (1995). *Applied Functional Analysis: Main Principles and Their Applications*, Vol. 109 of *Applied Mathematical Sciences*, 1st edn, Springer-Verlag.

Zhang, J. (2001). *Some fine properties of backward stochastic differential equations*, PhD thesis, Purdue University.