Quadrature of discontinuous SDE functionals using Malliavin integration by parts

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
der Universität Mannheim

vorgelegt von

Dipl. Math. Martin Altmayer
aus Ludwigshafen

Mannheim, März 2015

Dekan: Professor Dr. Heinz Jürgen Müller, Universität Mannheim
Referent: Professor Dr. Andreas Neuenkirch, Universität Mannheim
Korreferent: Professor Dr. Peter E. Kloeden, Goethe-Universität Frankfurt am Main

Tag der mündlichen Prüfung: 12. Mai 2015

Abstract

One of the major problems in mathematical finance is the pricing of options. This requires the computation of expectations of the form $E(f(S_T))$ with f being the payoff function of the option and S_T being the solution to a stochastic differential equation at a specific time T. A very popular choice for S is the Heston model.
While in the one-dimensional case $E(f(S_T))$ can often be computed using methods based on PDEs or the FFT, multidimensional models typically require the use of Monte-Carlo methods. Here, the multilevel Monte-Carlo algorithm provides considerably better performance — a benefit that is however reduced if the function f is discontinuous. This thesis introduces an approach based on the integration by parts formula from Malliavin calculus to overcome this problem: The original function f is replaced by a term including an antiderivative of f and a Malliavin weight term. We will prove that because the antiderivative is continuous, we can now apply multilevel Monte-Carlo to compute the value of the original expectation without performance reduction. This theoretical result is accompanied by numerical experiments which demonstrate that using the smoothed functional improves performance by a factor between 2 and 4.
Furthermore, the same integration by parts trick that was used to smooth the functional can be applied to derive a weak rate of convergence in the Heston model without making any smoothness assumptions on the payoff f — at the price of a rather strong condition on the model parameters.

Zusammenfassung

Eines der Hauptproblemfelder der Finanzmathematik ist die Berechnung korrekter Optionspreise. Zur Lösung muss ein Erwartungswert der Form $E(f(S_T))$ berechnet werden, wobei f das sogenannte Payoff-Funktional der Option ist und S_T die Lösung einer stochastischen Differentialgleichung zu einem festgelegten Zeitpunkt T ist. Sehr gerne wird zur Beschreibung des Preises S das Heston–Modell gewählt.
Im eindimensionalen kann $E(f(S_T))$ häufig mit Methoden berechnet werden, die auf partiellen Differentialgleichungen oder der FFT basieren. Dagegen müssen in mehrdimensionalen Modellen üblicherweise Monte-Carlo Methoden verwendet werden. Eine besonders schnelle solche Methode ist der Multilevel Monte-Carlo Algorithmus – allerdings reduziert sich ihr Geschwindigkeitsvorteil, falls die Funktion f unstetig ist. In dieser Dissertation stellen wir eine Methode vor, die die Partielle Integration aus dem Malliavin–Kalkül nutzt um dieses Problem zu lösen: Die ursprüngliche Funktion f wird durch eine Funktion ersetzt, die eine Stammfunktion von f und einen Gewichtsterm enthält. Wie wir zeigen werden, erlaubt die Stetigkeit der Stammfunktion uns, den ursprünglichen Erwartungswert mit dem Multilevel Monte-Carlo Algorithmus zu berechnen, ohne dass sich die Geschwindigkeit reduziert. Wir ergänzen dieses theoretische Resultat durch numerische Experimente, die belegen, dass der Wechsel zum geglätteten Funktional die Geschwindigkeit um einen Faktor zwischen 2 und 4 steigert.
Der gleiche Rechentrick mit der Partiellen Integration kann genutzt werden, um ohne jegliche Glattheitsannahmen an das Payoff-Funktional eine schwache Konvergenzrate im Heston–Modell zu beweisen – allerdings unter starken Voraussetzungen an die Modellparameter.

This thesis was written during my time as a researcher at the Technical University of Kaiserslautern and at the University of Mannheim. At both places my position was part of the DFG priority program SPP 1324 "Extraction of quantifiable information from complex systems".

First of all, I am deeply grateful to my supervisor Prof. Dr. A. Neuenkirch. I have greatly benefited from his enormous support and helpful advice. I would like to express my gratitude to Prof. Dr. P.E. Kloeden for being the co-referee of this thesis and to Prof. Dr. M. Schlather for being the third examiner.

I would also like to thank my colleagues of the Stochastic Numerics group at the University of Mannheim.

Last but not least I wish to thank my family for their never-ending support.

Contents

1 Introduction **1**
1.1 Notation and Calculation Tricks . . . 2

2 The Heston Model **3**
2.1 The Cox–Ingersoll–Ross Process (CIR) . . . 3
2.2 The Constant Elasticity of Volatility Process (CEV) . . . 6
2.3 The Heston Price . . . 7
2.4 An Almost Explicit Formula for European Call Options . . . 8

3 Malliavin Calculus **15**
3.1 Preliminaries . . . 15
3.2 The Malliavin Derivative . . . 16
3.3 Chain Rules . . . 19
3.4 The Skorohod Integral . . . 23
3.5 Partial Malliavin Derivatives . . . 24
3.6 Further Results . . . 26

4 Malliavin Calculus in the Heston Model **29**
4.1 The Derivative of the Volatility . . . 29
4.2 The Derivative of the Price Process . . . 32
4.3 Higher Derivatives . . . 34

5 Multidimensional Models **37**

6 Numerical Approximation **41**
6.1 The Euler Scheme for the CIR/CEV Process . . . 41
6.2 The Drift-Implicit Square Root Euler Scheme (DISE) . . . 42
6.3 The Drift-Implicit Milstein Scheme (DIMIL) . . . 45
6.4 Extension to the Heston Model . . . 46
6.5 Further Schemes . . . 47
6.6 Numerical Comparison of Approximation Schemes . . . 47

7 Weak Error **51**
7.1 Notation . . . 51
7.2 The differentiability of u . . . 52
7.3 Bounds on the Derivatives of u . . . 57
7.4 Weak Error Expansion . . . 62
7.5 Weak Error Rate . . . 64

8 Quadrature of Discontinuous Functionals **69**
8.1 An Integration by Parts Formula . . . 70
8.2 Order of Convergence Using Integration By Parts . . . 72
8.3 Payoff Splitting . . . 77

9 Monte-Carlo Algorithms **79**
9.1 Standard Monte-Carlo 79
9.2 Multilevel Monte-Carlo (MLMC) 80

10 Numerical Results **87**
10.1 The Algorithms 87
10.2 The Functional 88
10.3 The Models 88
10.4 Numerical Results 90
10.5 Conclusion 93

Appendix A **95**

Bibliography **97**

Chapter 1

Introduction

Three topics will be combined in this thesis: The stochastic volatility model of Steve L. Heston, [Hes93], the stochastic differential calculus developed by Paul Malliavin, [Mal76], and the multilevel Monte-Carlo algorithm by [Hei01] and [Gil08].
The Heston model was designed to solve a serious drawback of the famous Black–Scholes model: The latter assumes that the volatility remains constant, which contradicts real-world market data. To solve this problem, the Heston model replaces the constant volatility by a stochastic process on its own and is given by the SDEs

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{\nu_t} S_t dB_t \\ d\nu_t &= \kappa(\lambda - \nu_t)dt + \theta \nu_t^\gamma dW_t \end{aligned}$$

Among the many stochastic volatility models that are proposed in the literature, the Heston model stands out because it provides both flexibility and a semi-explicit representation of the price of call, put and digital options. However, in multidimensional models or with more complex options, there is usually no formula available and prices need to be computed using a Monte-Carlo approach. A major advance in this area was the introduction of the multilevel Monte-Carlo algorithm in the context of SDEs by Michael Giles, [Gil08]. For Lipschitz continuous option payoffs the cost to reach a certain root-mean-square error ε typically reduces from $\mathcal{O}(\varepsilon^{-3})$ for standard Monte-Carlo to $\mathcal{O}(\varepsilon^{-2}(\log \varepsilon)^2)$ for multilevel Monte-Carlo. However, it turns out that the benefit of multilevel Monte-Carlo depends on the rate of L^2-approximation of the underlying scheme. Unfortunately, for discontinuous option payoffs this rate is smaller. It is here that Malliavin calculus comes into play: One of the core rules of this calculus is an integration by parts rule, which can be used to replace a functional by its antiderivative. By replacing the discontinuous functional by its Lipschitz continuous antiderivative, we will be able to regain the computational complexity of $\mathcal{O}(\varepsilon^{-2}(\log \varepsilon)^2)$. This theoretical result (Theorems 8.9 and 9.2) is accompanied by our numerical experiments, which show that using this technique leads to a significantly faster algorithm. The integration by parts approach has been published first in [AN15].

The same integration by parts trick that accomplished to smooth a discontinuous payoff for use in multilevel Monte-Carlo is employed in two other places in this thesis: To prove that the multidimensional Heston model has a density; see Theorem 5.5. And to prove a weak error rate of 1 for the drift-implicit Milstein scheme in the Heston model, even if the payoff is not continuous; see Theorem 7.13.

This thesis is structured as follows: In the first chapter we will introduce the (generalized) Heston model as main object of this thesis and collect many results that will be needed later. In a small excursion in Section 2.4, we will prove the exact formulas for the prices of the call and digital option in a one-dimensional standard Heston model. The multidimensional model will be introduced later in Chapter 5.

Chapters 3 and 4 will be used to give an introduction to Malliavin calculus and apply it to the Heston model. While most results of Chapter 3 are standard and found in every book on the topic, we need a few extensions — particularly of the chain rule — which are not found in the literature, to the best of our knowledge.
The square root in the SDE of the volatility process and even more the fact that we will later require a numerical approximation that always remains positive, will force us to use non-standard approximation schemes. These schemes will be introduced and compared in Chapter 6.
In Chapter 7 we consider the function $u(t, x, v) := E(f(X_t)|X_0 = x, v_0 = v)$, where $X_t = \log S_t$. This function is typically used when studying the weak error rate of a scheme and we will be able to prove that one of the schemes of Chapter presented in 6 has a weak error rate of 1.
In Chapter 8 we will finally turn our attention to the strong convergence rate and show how it can be improved using Malliavin integration by parts.
In the last two chapters we will introduce the multilevel Monte-Carlo method and apply it to our problem. We will find that the improved strong convergence rate indeed leads to a faster algorithm for the quadrature of discontinuous options in the Heston model.

1.1 Notation and Calculation Tricks

Many results in this thesis will be concerned with asymptotic bounds on various quantities. Because we rarely care about the involved constants, we will use the constant c throughout this thesis to denote arbitrary constants. The value of c may change from term to term and — unless otherwise stated — may depend on the model parameters. Of course, it is always independent of the quantity that moves to the limit (typically the stepsize $\Delta \to 0$). When it is not clear from the context, we will give additional information. When using this notation, computations often use the following two rules, which we will use without further notice. For $p \geq 0$ we have

$$X \leq c \cdot (1 + Y) \Rightarrow X^p \leq c \cdot (1 + Y^p)$$
$$X_i \leq c \cdot (1 + Y_i), i = 1, \dots, N \Rightarrow \sum_{i=1}^{N} X_i \leq c \cdot (1 + \max_{i=1,\dots,N} Y_i) \leq c \cdot \left(1 + \sum_{i=1}^{N} Y_i\right)$$

Of course, the value of c will depend on p and in the last rule also on the number of summands.
We use the notation x^+ to denote the positive part: $x^+ := \max\{0, x\}$.
To save parentheses we will follow the rule that in case of doubt exponents are always included in expectations, e.g.

$$E\left(\int_0^t x_s ds\right)^p = E\left(\left(\int_0^t x_s ds\right)^p\right)$$

Because we will make extensive use of the Burkholder–Davis–Gundy inequality, we state it here. A proof can be found e.g. in Theorem 3.28 of [KS10].

Theorem 1.1 Let W_t be a Brownian motion and $(X_t)_{t\in[0,T]}$ be a square-integrable adapted process. Then for every $p \geq 0$ there exists a constant $C_p > 0$ such that

$$E \sup_{t\in[0,T]} \left(\int_0^t X_s dW_s\right)^p \leq C_p \cdot E\left(\int_0^T X_s^2 ds\right)^{p/2}$$

The constant only depends on p.

Chapter 2

The Heston Model

When the Black–Scholes formula was published in 1973, [BS73], it accounted for a major boost both in the theory of financial mathematics as well as in the finance industry. An important advantage of the model is that all parameters except one, the volatility, are observable from the market. Since the price as computed by the Black–Scholes formula is strictly increasing in the volatility, it is possible to invert the formula to compute the missing volatility parameter from real prices, the so-called *implied volatility*. This makes it easy to calibrate the model. However, it turns out that the implied volatility is not constant, violating a central assumption of the model.

Many approaches exist which try to overcome this problem by making the volatility stochastic; see e.g. [HW87], [Sco87], [SS91], [Hes93]. The first approach that provided a (semi)-analytical formula for the option price was [SS91]. However, it assumes that the price and the variance processes are independent which is too restrictive to fit the model to real market data. So it was Steven L. Heston's 1993 work "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options" that first provided both the necessary flexibility and (almost) explicit formulas for the price of European call and digital options.

More details on the development of option pricing theory can be found in [KK99] and in the introduction to [KJ06].

2.1 The Cox–Ingersoll–Ross Process (CIR)

Heston chose to use a Cox–Ingersoll–Ross process (CIR process; [CIR85]) as variance process for his model. For this reason we will collect some results on this process before proceeding to the Heston model. In fact, much of the complexity of the Heston model stems from the fact that the SDE for the CIR process contains a square root. Many standard results from the theory of SDEs require globally Lipschitz coefficient functions and cannot be used with the CIR process.

The CIR process is the solution to the SDE

$$dv_t = \kappa(\lambda - v_t)dt + \theta\sqrt{v_t}dW_t \tag{2.1}$$

with given start value $v_0 > 0$ and parameters $\kappa, \lambda, \theta > 0$. It is known that this SDE possesses a unique strong solution; see e.g. Theorem IV 3.2 in [IW89].

It turns out that the behavior of the CIR process — and thus of the Heston model — depends heavily on the fraction

$$\nu := \frac{2\kappa\lambda}{\theta^2}$$

A particularly important example is the question, whether zero is attainable. The Feller test (see [Fel51] or Chapter 15.6 in [Kar81]) can be used to prove

$$P(\nu_t > 0 \ \forall t \geq 0) = 1$$

if and only if $\nu \geq 1$.
The exact marginal distributions of the CIR process are known and can be written in terms of non-central χ^2-distributions; see [CIR85]. While this is of limited practical use, because the number of degrees of freedom of the χ^2-distribution — and consequently the cost of an algorithm — grows with ν, it can be used to compute moments exactly. A very general formula in this direction is given in Theorem 3.1 in [HK08]. In Chapter 7, it will be important to know not only the existence of various moments, but also their dependence on the start value ν_0.

Theorem 2.1 Fix $T \geq 0$ and $p \in \mathbb{R}$.

1. If $p \geq 0$, then
$$E \sup_{t\in[0,T]} \nu_t^p \leq c \cdot (1 + \nu_0^p)$$
2. If $p \leq -\nu$, then $E\nu_t^p = \infty$ for all $t > 0$.
3. If $p \in (-\nu, 0]$, then
$$\sup_{t\in[0,T]} E\nu_t^p \leq c \cdot \nu_0^p$$
4. If $1 \leq p < \nu - 1$, then
$$E \sup_{t\in[0,T]} \nu_t^{-p} \leq c \cdot (1 + \nu_0^{-p-1})$$

All constants depend on p, T, κ, λ, θ, but not on ν_0.

Proof. Theorem 3.1 in [HK08] gives an explicit formula for $E\nu_t^p$. If $p \leq -\nu$ the expectation is infinite as stated in the second claim. If $p > -\nu$, then

$$E\nu_t^p = \nu_0^p \cdot e^{-\kappa t p} \cdot z^{-p} \cdot \frac{\Gamma(\nu + p)}{\Gamma(\nu)} \cdot {}_1F_1(-p, \nu, -z)$$

with

$$z = \frac{\nu_0 \nu}{\lambda(e^{\kappa t} - 1)}$$

and the confluent hypergeometric function ${}_1F_1$. This function is defined in Chapter 13.1 of [AS64] where one can also find the following asymptotic behavior of this function for large z.

$${}_1F_1(a, b, -z) = \frac{\Gamma(b)}{\Gamma(b-a)} |z|^{-a} \cdot \mathcal{O}(|z|^{-1}) \tag{2.2}$$

In our setting we get for $z > 1$

$$\sup_{t\in[0,T]} E\nu_t^p \leq \nu_0^p \cdot \sup_{t\in[0,T]} e^{-\kappa t p} \cdot \mathcal{O}(1) \tag{2.3}$$

For $z \in [0, 1]$ we can use the fact that ${}_1F_1$ is continuous to bound $\sup_{t\in[0,T]} E\nu_t^p$ for $p \geq 0$ by

$$\sup_{t\in[0,T]} E\nu_t^p \leq \sup_{t\in[0,T]} \left(\frac{\lambda(1 - e^{-\kappa t})}{\nu}\right)^p \cdot \frac{\Gamma(\nu + p)}{\Gamma(\nu)} \cdot \sup_{z\in[0,1]} {}_1F_1(-p, \nu, -z) \in \mathcal{O}(1) \tag{2.4}$$

and for $-\nu < p < 0$ by

$$\sup_{t\in[0,T]} E\nu_t^p \leq \nu_0^p \cdot e^{\kappa T|p|} \cdot \frac{\Gamma(\nu + p)}{\Gamma(\nu)} \cdot \sup_{z\in[0,1]} {}_1F_1(-p, \nu, -z) \leq c \cdot \nu_0^p$$

Together with (2.3) this proves the third claim.
To handle the first claim, we use the Burkholder–Davis–Gundy inequality

$$\begin{aligned}\mathrm{E}\sup_{t\in[0,T]} v_t^p &= \mathrm{E}\sup_{t\in[0,T]}\left(v_0+\int_0^t \kappa(\lambda-v_s)ds+\int_0^t \theta\sqrt{v_s}dW_s\right)^p \\ &\le c\cdot\left(v_0^p+\mathrm{E}\left(\int_0^T|\kappa(\lambda-v_s)|ds\right)^p+\mathrm{E}\left(\int_0^T\theta^2 v_s ds\right)^{p/2}\right) \qquad (2.5)\end{aligned}$$

If $p\ge 1$, we apply Jensen's inequality to get

$$\mathrm{E}\left(\int_0^T v_s ds\right)^p \le c\cdot\mathrm{E}\int_0^T v_s^p ds \le cT\cdot\sup_{t\in[0,T]}\mathrm{E}v_s^p$$

For $p\in[0,1]$, the same result follows by moving the exponent out of the expectation instead of into the integral. Now (2.3) together with (2.4) prove that (2.5) is bounded by $c\cdot(1+v_0^p)$. It remains to prove the fourth claim. We use Lemma 2.12 from [NS14] to replace E sup by sup E at the cost of a higher inverse moment:

$$\mathrm{E}\sup_{t\in[0,T]} v_t^{-p}\le c\cdot\left(1+\sup_{t\in[0,T]}\mathrm{E}v_t^{-p-1}\right)$$

Now the assertion follows from the third claim. □

Surprisingly, the integral over v_s^{-1} is p-integrable for all $p\ge 0$, although integrability of v_s^{-p} itself is restricted to $p<\nu$ by the last theorem.

Theorem 2.2 Assume $\nu>1$ and $T>0$. Then for each $p\ge 0$

$$\sup_{t\in[0,T]}\mathrm{E}\left(\int_0^t\frac{1}{v_s}ds\right)^p<\infty$$

Proof. Because $\nu>1$ we can choose $\varepsilon<(\nu-1)^2\theta^2/8$ such that we can define

$$q:=\frac{1}{2}\left(\nu-1-\sqrt{(\nu-1)^2-\frac{8\varepsilon}{\theta^2}}\right)>0$$

Now we choose a $C\ge 0$ such that $x^p<Ce^{\varepsilon x}$ for all $x\ge 0$. Our choice of ε implies $\nu-q\ge\nu-(\nu-1)/2>0$, so we can use Theorem 3.1 in [HK08] (note that their ν_2 equals $-q$) to get

$$\begin{aligned}\mathrm{E}\left(\int_0^t\frac{1}{v_s}ds\right)^p &\le C\cdot\mathrm{E}\left(\exp\left(\varepsilon\cdot\int_0^t\frac{1}{v_s}ds\right)\right) \\ &= C\cdot e^{\kappa q t}\cdot v_0^{-q}\cdot\gamma_t^{-q}\cdot\frac{\Gamma(\nu-q)}{\Gamma(\nu-2q)}\cdot{}_1F_1\left(-q,\nu-2q,-\gamma_t v_0 e^{-\kappa t}\right)\end{aligned}$$

with

$$\gamma_t=\frac{2\kappa}{\theta^2}\cdot\frac{1}{1-e^{-\kappa t}}>\frac{2\kappa}{\theta^2}$$

and the confluent hypergeometric function ${}_1F_1(a,b,-z)$. Because this function is continuous, the right-hand side is bounded by $c\cdot e^{\kappa q T}v_0^{-q}$ for small $z=\gamma_t v_0 e^{-\kappa t}$. For large z the asymptotic expansion (2.2) gives

$$\mathrm{E}\left(\int_0^t\frac{1}{v_s}ds\right)^p\in\mathcal{O}((\gamma_t v_0 e^{-\kappa t})^{-1})=\mathcal{O}(1)$$

□

Because the volatility is strictly positive in the case $\nu \geq 1$, we can apply the Lamperti transformation to get a process with constant diffusion term; see e.g. [Iac08]. For a general SDE $dx_t = a(t, x_t)dt + b(t, x_t)dW_t$ with strictly positive diffusion term b, the Lamperti transformation is given as

$$F(t, x) := C \cdot \int_0^x \frac{1}{b(t, \xi)} d\xi$$

for an arbitrary constant C. If $F \in C^{1,2}$ then Itō's formula shows that the transformed process $F(t, x_t)$ has a constant diffusion term C:

$$dF(t, x_t) = \left(\partial_t F(t, x_t) + C \cdot \frac{a(t, x_t)}{b(t, x_t)} - \frac{C}{2} \cdot \partial_x b(t, x_t) \right) dt + C dW_t$$

In the setting of the CIR process we choose $C = \theta/2$ so that the transform is simply the square root function: $F(t, x) = \sqrt{x}$. The SDE of the resulting process, which will be denoted by σ_t, is

$$d\sigma_t = \left(\left(\frac{\kappa\lambda}{2} - \frac{\theta^2}{8} \right) \frac{1}{\sigma_t} - \frac{\kappa}{2} \sigma_t \right) dt + \frac{\theta}{2} dW_t, \qquad \sigma_0 = \sqrt{\nu_0} \tag{2.6}$$

Note that the factor $\kappa\lambda/2 - \theta^2/8$ is positive, because $\nu \geq 1$. Like (2.1) this equation has only one solution: Applying Itō's formula again shows that the square of each solution solves (2.1). Thus only $\sqrt{\nu_t}$ and $-\sqrt{\nu_t}$ can be solutions and the positive start value makes the latter impossible.
Finally, we state that the processes σ_t and ν_t are continuous in L^p.

Theorem 2.3 Let $T > 0$, $\alpha \geq 1$, $p \geq 0$ and assume $\nu > 1$. Then

$$\|\sigma_s^\alpha - \sigma_t^\alpha\|_p \leq c \cdot \sqrt{|t - s|}$$

for all $s, t \in [0, T]$. The constants are independent of s and t.

Proof. First assume $\alpha = 1$. Set $K := \kappa\lambda/2 - \theta^2/8$ and assume without loss of generality that $t \geq s$. The SDE of σ_t shows that

$$|\sigma_s - \sigma_t| \leq K \cdot \left| \int_s^t \sigma_u^{-1} du \right| + \frac{\kappa}{2} \cdot \left| \int_s^t \sigma_u du \right| + \frac{\theta}{2} \cdot |W_t - W_s|$$

The Cauchy–Schwarz inequality gives

$$|\sigma_s - \sigma_t| \leq K \cdot \sqrt{t - s} \cdot \left| \int_s^t \sigma_u^{-2} du \right|^{1/2} + \frac{\kappa}{2} \cdot \sqrt{t - s} \cdot \left| \int_s^t \sigma_u^2 du \right|^{1/2} + \frac{\theta}{2} \cdot |W_t - W_s|$$

After taking the L^p-norm, the result follows from $\|W_t - W_s\|_p \leq c \cdot \sqrt{t - s}$ and Theorems 2.1 and 2.2.
For general $\alpha \geq 1$ we use the mean value theorem and Hölder's inequality:

$$\|\sigma_s^\alpha - \sigma_t^\alpha\|_p = \alpha \cdot \|\Theta^{\alpha-1}\|_{2p} \cdot \|\sigma_s - \sigma_t\|_{2p}$$

for some random variable Θ with values between σ_s and σ_t. The result now follows from Theorem 2.1. □

2.2 The Constant Elasticity of Volatility Process (CEV)

Mean-reverting constant elasticity of volatility processes provide a simple generalization of the CIR process and can also be used as variance processes in the (generalized) Heston

model; see e.g. [LKvD10]. These processes result from replacing the square root of the CIR process by v^γ with $\gamma \in [1/2, 1)$. Thus the CEV process is defined by the SDE

$$dv_t^{(\gamma)} = \kappa(\lambda - v_t^{(\gamma)})dt + \theta(v_t^{(\gamma)})^\gamma dW_t \tag{2.7}$$

with given start value $v_0 > 0$. From now on, we will omit the superscript (γ). As for the CIR process, existence and uniqueness of a solution follows from Theorem IV 3.2 in [IW89].
Many properties of the CIR process carry over to the generalization. In fact, many desirable properties are always true for $\gamma > 1/2$ while they might be false if $\gamma = 1/2$ and ν is too small. As an example, for $\gamma > 1/2$ the CEV process is always strictly positive; see the boundary classification results in Chapter 15.6 of [Kar81]. This allows us to compute the Lamperti transformation as in the previous section. This time, the transformation takes the form $F(t, x_t) = x_t^{1-\gamma}$ (for $C = \theta(1-\gamma)$). The transformed process $\sigma_t^{(\gamma)} = \sigma_t$ is the solution of the SDE

$$\begin{aligned} d\sigma_t &= (1-\gamma) \cdot \left(\kappa\lambda\sigma_t^{-\frac{\gamma}{1-\gamma}} - \kappa\sigma_t - \frac{\gamma\theta^2}{2} \cdot \sigma_t^{-1} \right) dt + \theta(1-\gamma)dW_t \\ \sigma_0 &= v_0^{1-\gamma} \end{aligned} \tag{2.8}$$

The next theorem concerning the moments of the CEV process provides another example where the case $\gamma > 1/2$ behaves much better.

Theorem 2.4 Assume $\gamma > 1/2$ and let $p, T \geq 0$. Then

$$E \sup_{t\in[0,T]} v_t^p < \infty \qquad \text{and} \qquad \sup_{t\in[0,T]} Ev_t^{-p} < \infty$$

Proof. See Lemma 2.1 in [BBD07]. □

The L^p-continuity that was established in Theorem 2.3 extends to the case $\gamma > 1/2$.

Theorem 2.5 Let $T > 0$, $\alpha \geq 1$, $p \geq 0$ and assume $\gamma > 1/2$. Then

$$\|\sigma_s^\alpha - \sigma_t^\alpha\|_p \leq c \cdot \sqrt{|t-s|}$$

for all $s, t \in [0, T]$. The constants are independent of s and t.

Proof. The claim is shown exactly as Theorem 2.3, this time using Theorem 2.4 to bound the integrals. □

2.3 The Heston Price

The SDE for the Heston price differs from that of the Black–Scholes price only by using a CIR process as stochastic variance process. The driving Brownian motions of both processes are usually correlated. Let $(Z_t^1)_t$, $(Z_t^2)_t$ be independent Brownian motions. Assume $\rho \in [-1, 1]$ and set $\rho' := \sqrt{1-\rho^2}$. Then we define the driving Brownian motions of the price and variance process as $B_t := \rho Z_t^2 + \rho' Z_t^1$ and $W_t := Z_t^2$, respectively. Using three different symbols for these Brownian motions will make it easier to deal with multidimensional models later.
We consider the generalized Heston model, where the variance process is given by a CEV process. The model is given by the following two SDEs.

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dB_t \\ dv_t &= \kappa(\lambda - v_t)dt + \theta v_t^\gamma dW_t \end{aligned}$$

Table 2.1: The parameters of the Heston model.

$\mu \in \mathbb{R}$	Interest rate. It is often omitted, because Itō's formula shows that the transformation $\tilde{S}_t := e^{-\mu t} S_t$ leads to a Heston model with $\mu = 0$.
$\kappa > 0$	Strength of mean-reversal property of the variance.
$\lambda > 0$	Mean variance: If $v_0 = \lambda$ then $Ev_t = \lambda$ for all $t > 0$. Otherwise the expectation converges to λ as $t \to \infty$; see the formulas in the appendix of [And07].
$\theta > 0$	Volatility of the volatility.
$\rho \in [-1, 1]$	Correlation of Brownian motions. We abbreviate $\rho' := \sqrt{1-\rho^2}$.
$s_0 > 0$	Start price.
$v_0 > 0$	Start variance.
$\gamma \in [1/2, 1)$	Elasticity parameter for the CEV process in the generalized Heston model.

Table 2.1 provides an overview of the parameters, defining in particular the valid ranges. Similar to the treatment of the CIR/CEV process, we will often apply a transformation to the price and examine the transformed process. This time, the transformation consists of the logarithm, which has the advantage that Itō's formula yields an explicit representation for the log-price $X_t := \log(S_t)$ in terms of the volatility, because S_t cancels out on the right-hand side of the SDE:

$$dX_t = \left(\mu - \frac{v_t}{2}\right) dt + \sqrt{v_t} dB_t \tag{2.9}$$

Applying the Burkholder–Davis–Gundy inequality and Theorems 2.1 and 2.4 to the integral form of this representation gives

$$E \sup_{t \in T} X_t^p < \infty, \qquad \forall\, p \geq 0 \tag{2.10}$$

Unfortunately, this result does not hold for the price process itself.

Theorem 2.6 For $p \in (1, \infty)$ define

$$T^*(p) := \inf\{t \geq 0 : ES_t^p = \infty\}$$

Then in the case $\gamma = 1/2$ we have

$$T^*(p) = \infty \iff \rho \leq -\sqrt{\frac{p-1}{p}} + \frac{\kappa}{\theta p}$$

whereas in the case $\gamma > 1/2$ the following is true:

$$T^*(p) = \begin{cases} \infty & \text{if } \rho < -\sqrt{\frac{p-1}{p}} \\ 0 & \text{if } \rho > -\sqrt{\frac{p-1}{p}} \end{cases}$$

Proof. For $\gamma = 1/2$ see [AP06] and [FKR10]. For $\gamma > 1/2$ see [LM07]. □

2.4 An Almost Explicit Formula for European Call Options

In the finance industry a *European call option* is the right to buy one unit of an underlying stock at a future *maturity time* T for a price $K(S_T)$ which might depend on the future stock

price S_T. For classical European call options, $K(S_T)$ is simply a constant fixed in advance. Every finance company trading in options has to compute the correct price for the option at a time $t < T$. As usual in mathematical finance, the correct price is the one that avoids arbitrage; see e.g. [KK99].
In this section we will construct the price of a European call option in the standard Heston model as presented in Heston's original paper. As bonus we will get a formula for digital options. Heston uses the same approach that Black and Scholes used to discover their famous formula: build a risk-free portfolio containing the option. Being risk-free, the portfolio must always have the same price as the risk-free bond or otherwise arbitrage will arise. This reasoning can be captured in a PDE and the price is the solution to this PDE.
We will consider an option with payoff $g(S_T)$, depending on the price at maturity time T. For the call option with fixed strike price K we have $g(s) = (s - K)^+$, because when the maturity time T is reached, the owner of a call option will only use it if $K < S_T$. An option whose payoff is an indicator function, e.g. $g(s) = \mathbb{1}_{[0,K]}(s)$, is called *digital option*.

2.4.1 Deriving the Heston PDE

We denote the price function of a given option by $C(s, \nu, t)$. Inserting the current stock price S_t, the variance ν_t and time point t gives the current price of the option. Obviously, the function C must satisfy the terminal condition $C(s, \nu, T) = g(s)$.
We now start to construct a risk-free portfolio including the option. Due to the additional source of randomness in a stochastic volatility model, we need three building blocks to construct a risk-free portfolio (compared to two in the Black–Scholes model):

1. the option $C = C(S_t, \nu_t, t)$,
2. the underlying asset S_t,
3. an additional option $U = U(S_t, \nu_t, t)$. We will assume that we know the price of this option and that $\frac{\partial U}{\partial \nu} \neq 0$. The price of C will not depend on U but the risk-free portfolio will.

We build the portfolio as

$$\Pi = C + \alpha S_t + \beta U \tag{2.11}$$

where α and β are functions in S_t, V_t, t.
We assume that both prices C and U are twice continuously differentiable in s and ν and continuously differentiable in t, so that we can apply Itō's formula to both. For a function $f(s, \nu, t)$ we define the differential operator

$$\Gamma f = \frac{\partial f}{\partial t} + \frac{1}{2}\nu s^2 \frac{\partial^2 f}{\partial s^2} + \rho\theta\nu s \frac{\partial^2 f}{\partial s \partial \nu} + \frac{1}{2}\theta^2 \nu \frac{\partial^2 f}{\partial \nu^2}$$

Itō's formula leads to

$$\begin{aligned} d\Pi &= dC + \alpha dS_t + \beta dU \\ &= (\Gamma C + \beta \Gamma U)dt + \left(\frac{\partial C}{\partial s} + \beta \frac{\partial U}{\partial s} + \alpha\right) dS_t + \left(\frac{\partial C}{\partial \nu} + \beta \frac{\partial U}{\partial \nu}\right) d\nu_t \end{aligned} \tag{2.12}$$

If the portfolio is to be risk-free, the dS_t-term and the $d\nu_t$-term must vanish. This settles how to choose the portfolio:

$$\alpha = -\frac{\partial C}{\partial s} - \beta \frac{\partial U}{\partial s} \qquad \beta = -\frac{\partial C}{\partial \nu} \cdot \left(\frac{\partial U}{\partial \nu}\right)^{-1} \tag{2.13}$$

To avoid arbitrage opportunities, a risk-free portfolio must earn the risk-free rate μ of the bond, i.e. $d\Pi = \mu \Pi dt$. Together with (2.11) and (2.12) this leads to

$$\Gamma C + \beta \Gamma U = \mu(C + \alpha S_t + \beta U)$$

After inserting the specifications of α and β from (2.13), we can rearrange terms to get

$$\frac{\Gamma C - \mu C + \mu S_t \frac{\partial C}{\partial s}}{\frac{\partial C}{\partial v}} = \frac{\Gamma U - \mu U + \mu S_t \frac{\partial U}{\partial s}}{\frac{\partial U}{\partial v}} \tag{2.14}$$

This equation shows that the term on both sides is an invariant which is the same for all options in the model and depends only on S_t, V_t, t. We call this invariant $-f$:

$$f(s,v,t) = -\frac{\Gamma C - \mu C + \mu s \frac{\partial C}{\partial s}}{\frac{\partial C}{\partial v}} = -\frac{\Gamma U - \mu U + \mu s \frac{\partial U}{\partial s}}{\frac{\partial U}{\partial v}}$$

In particular, the function C must fulfill the PDE

$$\Gamma C + \mu s \frac{\partial C}{\partial s} + f(s,v,t)\frac{\partial C}{\partial v} - \mu C = 0 \tag{2.15}$$

with terminal condition $C(s,v,T) = g(s)$.
One might expect that f is some function inherent in the model but in fact we can choose f arbitrarily, as long as the PDE 2.15 admits a solution. Conditions under which such a solution exists can be found in [Fri64]; if it exists, the Feynman–Kac theorem (Section 4.4 in [KS10]) allows to write it as

$$C(s,v,t) := E\left(e^{-\mu(T-t)} \cdot g(S'_T) \middle| S'_t = s, v'_t = v \right)$$

where (S'_t, V'_t) are the solutions of the following SDEs:

$$\begin{aligned} dS'_t &= \mu S'_t dt + \sqrt{v'_t} S'_t dB_t \\ dv'_t &= f(S'_t, v'_t, t)dt + \theta\sqrt{v'_t} dW_t \end{aligned}$$

This modified model is called the *risk-neutral model*, because if (2.15) admits a solution, it can be computed as a simple expectation in the risk-neutral model.
Because the Heston model forms an incomplete market, it comes as no surprise that there is no unique option price without further assumptions on the model. Only after specifying f has the model enough information to allow unique option prices.
Following Heston, we specify the function f as

$$f(s,v,t) = \kappa(\lambda - v) - \gamma v$$

for some parameter $\gamma \in \mathbb{R}$ and call the additional term γv the "price of volatility risk". With this choice the risk-neutral model is again a Heston model (with modified parameters $\kappa' = \kappa + \gamma$ and $\lambda' = \kappa\lambda/(\kappa+\gamma)$). This has consequences for the numerical parts of this thesis: Because we can always assume that we are already given a risk-neutral model, the problem of computing option prices reduces to computing expectations in an Heston model. Note also that the fraction $\nu = 2\kappa\lambda/\theta^2$ is the same in both models.
Inserting our specification of f into (2.15) gives the final Heston PDE:

$$\frac{\partial C}{\partial t} + \frac{1}{2}vs^2\frac{\partial^2 C}{\partial s^2} + \rho\theta v s \frac{\partial^2 C}{\partial s \partial v} + \frac{1}{2}\theta^2 v \frac{\partial^2 C}{\partial v^2} + \mu s \frac{\partial C}{\partial s} + (\kappa(\lambda - v) - \gamma v)\frac{\partial C}{\partial v} - \mu C = 0 \tag{2.16}$$

with terminal condition $C(s,v,T) = g(s)$.
Equivalently, the price of an option can be given in terms of the log-price. Then $C'(x,v,t) := C(e^x, v, t)$ must satisfy the PDE

$$\frac{\partial C'}{\partial t} + \frac{1}{2}v\frac{\partial^2 C'}{\partial x^2} + \rho\theta v \frac{\partial^2 C'}{\partial x \partial v} + \frac{1}{2}\theta^2 v \frac{\partial^2 C'}{\partial v^2} + \left(\mu - \frac{1}{2}v\right)\frac{\partial C'}{\partial x} + (\kappa(\lambda - v) - \gamma v)\frac{\partial C'}{\partial v} - \mu C = 0 \tag{2.17}$$

with terminal condition $C'(x,v,T) = g(e^x)$.

2.4.2 Solving the Heston PDE

Inspired by the Black–Scholes formula Heston assumed that in the case of a call option the solution would have the form

$$C(S_t, \nu_t, t) = S_t \cdot P_1 - Ke^{-\mu(T-t)}P_2$$

where K is the strike price and P_1, P_2 are some probabilities depending on S_t, ν_t, t. He then derived PDEs for P_1 and P_2 (see (2.24) below) and found an interpretation as probabilities in slightly modified models (see X_j, ν_j below). Finally, he was able to invert the characteristic function of these models to compute a semi-explicit formula for the probabilities and thus for the price.
Define the processes X_1, X_2, ν_1, ν_2 by the SDEs

$$\begin{aligned} dX_j(t) &= (\mu + u_j\nu_j(t))dt + \sqrt{\nu_i(t)}dB_t \\ d\nu_j(t) &= (\kappa\lambda - b_j\nu_j(t))dt + \theta\sqrt{\nu_j(t)}dW_t \end{aligned}$$

with $u_1 = \frac{1}{2}$, $u_2 = -\frac{1}{2}$, $b_1 = \kappa + \gamma - \rho\theta$, $b_2 = \kappa + \gamma$. Note that for $j = 2$ this is exactly the risk-neutral model from the last section.
We will have to compute the characteristic function of the X_j first. It will be easier if we let the characteristic function depend on $\tau = T - t$ and thus we define for $j \in \{1, 2\}$

$$\varphi_j(x, \nu, \tau; y) := E(e^{iyX_j(T)}|X_j(T-\tau) = x, \nu_j(T-\tau) = \nu)$$

Theorem 2.7 The characteristic functions are given by

$$\varphi_j(x, \nu, \tau; y) = \exp\left(C(\tau, y) + D(\tau, y) \cdot \nu + iyx\right)$$

with

$$\begin{aligned} C_j(\tau, y) &= \mu y i\tau + \frac{\kappa\lambda}{\theta^2}\left((b_j - \rho\theta y i + d_j)\tau - 2 \cdot \log\left(\frac{1 - g_j e^{d_j\tau}}{1 - g_j}\right)\right) \qquad (2.18) \\ D_j(\tau, y) &= \frac{b_j - \rho\theta y i + d_j}{\theta^2} \cdot \frac{1 - e^{d_j\tau}}{1 - g_j e^{d_j\tau}} \\ g_j &= \frac{b_j - \rho\theta y i + d_j}{b_j - \rho\theta y i - d_j} \\ d_j &= \sqrt{(\rho\theta y i - b_j)^2 - \theta^2(2u_j y i - y^2)} \end{aligned}$$

Proof. For brevity we will omit the index j. Theorem 3.7 in [Fri64] shows that for each $y \in \mathbb{R}$ the following PDE has a unique solution which is twice differentiable:

$$(\mu + u\nu)\frac{\partial\varphi}{\partial x} + (\kappa\lambda - b\nu)\frac{\partial\varphi}{\partial\nu} + \frac{1}{2}\nu\left(\frac{\partial^2\varphi}{\partial x^2} + 2\rho\theta\frac{\partial^2\varphi}{\partial x\partial\nu} + \theta^2\frac{\partial^2\varphi}{\partial\nu^2}\right) - \frac{\partial\varphi}{\partial\tau} = 0 \qquad (2.19)$$

with initial condition $\varphi(x, \nu, 0) = e^{ixy}$. The Feynman–Kac theorem now shows that the solution of this PDE is exactly the characteristic function.
Assuming a solution of the form

$$\varphi(x, \nu, \tau; y) = \exp(C(\tau, y) + D(\tau, y) \cdot \nu + iyx) \qquad (2.20)$$

for some functions C and D with $C(0, y) = D(0, y) = 0$ transforms above PDE into

$$\begin{aligned} \nu\varphi \cdot \left[-\frac{1}{2}y^2 + \rho\theta i y D(\tau, y) + \frac{1}{2}\theta^2 D^2(\tau, y) + uiy - bD(\tau, y) - \frac{\partial D(\tau, y)}{\partial\tau}\right] \\ + \varphi \cdot \left[\mu i y + \kappa\lambda D(\tau, y) - \frac{\partial C(\tau, y)}{\partial\tau}\right] = 0 \end{aligned}$$

Thus if we can solve the system of ODEs

$$\frac{\partial D}{\partial \tau} = -\frac{1}{2}y^2 + \rho\theta i y D(\tau, y) + \frac{1}{2}\theta^2 D^2(\tau, y) + uiy - bD(\tau, y) \tag{2.21}$$

$$\frac{\partial C}{\partial \tau} = \mu i y + \kappa\lambda D(\tau, y) \tag{2.22}$$

we get a solution to the original PDE (2.19) by (2.20).
Equation (2.21) is a Riccati ODE with solution

$$D(\tau, y) = \frac{b - \rho\theta y i + d}{\theta^2} \cdot \frac{1 - e^{d\tau}}{1 - ge^{d\tau}}$$

(see e.g. [Heu06]). This allows us to solve the ODE for C (2.22):

$$C(\tau, y) = \mu i y \tau + \kappa\lambda \cdot \int_0^\tau D(s, y)\, ds = \mu i y \tau + \kappa\lambda \cdot \frac{b - \rho\theta y i + d}{\theta^2} \cdot \int_0^\tau \frac{1 - e^{ds}}{1 - ge^{ds}}\, ds$$

It remains to solve the integral:

$$\begin{aligned}
\int_0^\tau \frac{1 - e^{ds}}{1 - ge^{ds}} ds &= \int_0^\tau \frac{1 - e^{ds}}{(1 - ge^{ds})e^{ds} d} \cdot de^{ds}\, ds = \frac{1}{d} \cdot \int_1^{e^{d\tau}} \frac{1 - x}{(1 - gx)x} dx \\
&= \frac{1}{d} \cdot \int_1^{e^{d\tau}} \frac{1}{x} - \frac{1 - g}{1 - gx} dx = \frac{1}{d} \cdot \left[\log(x) + \frac{1 - g}{g} \cdot \log(1 - gx)\right]_1^{e^{d\tau}} \\
&= \frac{1}{d}\left(d\tau + \frac{1 - g}{g} \cdot \log\left(\frac{1 - ge^{d\tau}}{1 - g}\right)\right) = \tau + \frac{1 - g}{gd} \cdot \log\left(\frac{1 - ge^{d\tau}}{1 - g}\right)
\end{aligned}$$

The result now follows from the simple equation

$$(b - \rho\theta y i + d) \cdot \frac{1 - g}{gd} = \frac{1}{d} \cdot \big((b - \rho\theta i y - d) - (b - \rho\theta i y + d)\big) = -2$$

□

For $j \in \{1, 2\}$ define

$$P_j(x, v, t) := P(X_j(T) \geq \log K \mid X_j(t) = x, v_j(t) = v)$$

The formula of Gil-Pelaez ([GP51]) can be used to compute these probabilities from the characteristic function:

$$P_j(x, v, t) = \frac{1}{2} + \frac{1}{\pi} \cdot \int_0^\infty \frac{1}{y} \cdot \mathrm{Im}\left(e^{-iy \log K} \cdot \varphi_j(x, v, T - t; y)\right)\, dy \tag{2.23}$$

From (2.19) it follows that these probabilities satisfy the PDE

$$\frac{\partial P_i}{\partial t} + (\mu + u_i v)\frac{\partial P_i}{\partial x} + (\kappa\lambda - b_i v)\frac{\partial P_i}{\partial v} + \frac{1}{2}v\left(\frac{\partial^2 P_i}{\partial x^2} + 2\rho\theta\frac{\partial^2 P_i}{\partial x \partial v} + \theta^2\frac{\partial^2 P_i}{\partial v^2}\right) = 0 \tag{2.24}$$

with terminal condition $P_j(x, v, T) = \mathbb{1}_{\{x \geq \log K\}}$.

Theorem 2.8 The correct price function of the digital option with payoff $g(s) = \mathbb{1}_{[K,\infty)}$ is given by $e^{-\mu(T-t)}P_2$.

Proof. In case $j = 2$ equation (2.24) is exactly the price equation (2.17). As P_2 solves this PDE with the correct terminal condition, it must be the price. □

Using the probabilities P_j, we can finally compute the price of a call option with Theorem 2.9. The price of a put option can then be computed using the well-known put-call parity; see e.g. [KK99].

Theorem 2.9 The correct price function of the European call option with strike price K and time to maturity T is given by

$$C(s, v, t) := s \cdot P_1(\log s, v, t) - e^{-\mu(T-t)} K \cdot P_2(\log s, v, t)$$

Proof. By definition, C satisfies the terminal condition $C(s, v, T) = (s - K)^+$. Define $C'(x, v, t) := C(e^x, v, t) = e^x P_1(x, v, t) - e^{-\mu(T-t)} K P_2(x, v, t)$. Partial derivatives of C' can be written in terms of partial derivatives of the P_i:

$$\begin{aligned}
\frac{\partial C'}{\partial t} &= e^x \frac{\partial P_1}{\partial t} - \mu e^{-\mu(T-t)} K P_2 - e^{-\mu(T-t)} K \frac{\partial P_2}{\partial t} \\
\frac{\partial C'}{\partial x} &= e^x P_1 + e^x \frac{\partial P_1}{\partial x} - e^{-\mu(T-t)} K \frac{\partial P_2}{\partial x} \\
\frac{\partial^2 C'}{\partial x^2} &= e^x P_1 + 2e^x \frac{\partial P_1}{\partial x} + e^x \frac{\partial^2 P_1}{\partial x^2} - e^{-\mu(T-t)} K \frac{\partial^2 P_2}{\partial x^2} \\
\frac{\partial^2 C'}{\partial x \partial v} &= e^x \frac{\partial P_1}{\partial v} + e^x \frac{\partial^2 P_1}{\partial x \partial v} - e^{-\mu(T-t)} K \frac{\partial^2 P_2}{\partial x \partial v}
\end{aligned}$$

Inserting these formulas into the left-hand side of the Heston PDE for C' (2.17) gives

$$\begin{aligned}
e^x \Bigg[& \frac{\partial P_1}{\partial t} + \frac{1}{2} v \left(P_1 + 2 \frac{\partial P_1}{\partial x} + \frac{\partial^2 P_1}{\partial x^2} \right) + \rho \theta v \left(\frac{\partial P_1}{\partial v} + \frac{\partial^2 P_1}{\partial x \partial v} \right) + \frac{\theta^2 v}{2} \cdot \frac{\partial^2 P_1}{\partial v^2} \\
& + \left(\mu - \frac{1}{2} v \right) \cdot \left(P_1 + \frac{\partial P_1}{\partial x} \right) + (\kappa(\lambda - v) - \gamma v) \frac{\partial P_1}{\partial v} - \mu P_1 \Bigg] \\
- e^{-\mu(T-t)} K \Bigg[& \frac{\partial P_2}{\partial t} + \mu P_2 + \frac{1}{2} v \frac{\partial^2 P_2}{\partial x^2} + \rho \theta v \frac{\partial^2 P_2}{\partial x \partial v} + \frac{\theta^2 v}{2} \cdot \frac{\partial^2 P_2}{\partial v^2} \\
& + \left(\mu - \frac{v}{2} \right) \frac{\partial P_2}{\partial x} + (\kappa(\lambda - v) - \gamma v) \frac{\partial P_2}{\partial v} - \mu P_2 \Bigg]
\end{aligned}$$

Both terms in brackets are 0 due to (2.24) and thus the function C' satisfies the Heston PDE (2.17). Consequently, C must be the correct price function. □

2.4.3 Implementation of Complex Logarithm

Practical use of (2.23) seems straight-forward, in particular as the integrand turns out to be a rapidly decreasing function and thus the indefinite integral poses no problem. What makes problems, though, is the complex logarithm hidden in the formula of $C_j(\tau, y)$; see (2.18). The complex logarithm is a multibranched function with a branch cut at the negative real axis. For $z = re^{i\varphi} = re^{i\varphi + 2\pi i k}$, $r > 0$ and $\varphi \in (-\pi, \pi]$ the values of all branches are given as

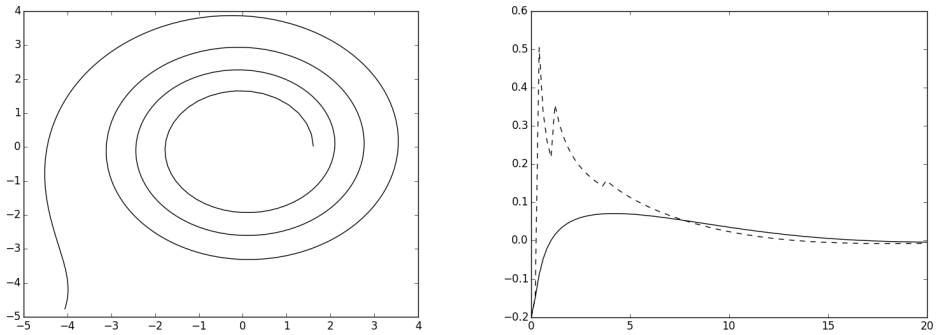
$$\log z = \log r + i\varphi + 2\pi i k, \qquad k \in \mathbb{Z}$$

In our case, the argument of the logarithm is

$$\psi(y) := \frac{1 - g_j(y) e^{d_j(y)\tau}}{1 - g_j(y)}$$

Figure 2.1 shows that the graph of this function starts as a spiral that rapidly moves away from zero. To compute the probabilities of (2.23) correctly, we must start at the default

Figure 2.1: Left: Argument of complex logarithm. Because ψ moves outward extremely fast, this plot shows $\psi(y)/|\psi(y)| \cdot (\log\log|\psi(y)|)$ with $y \in (0, 120]$. Right: The integrand for (2.23) with and without using the correct branch of the complex logarithm. In both cases parameters were taken from [And07]: $T = 10$, $\mu = 0$, $\kappa = 0.5$, $\lambda = 0.04$, $\theta = 1$, $\rho = -0.9$, $v_0 = \lambda$, $s_0 = 100$.



branch ($k = 0$) for y close to zero and then increase the value of k whenever ψ crosses the negative real axis. For a detailed discussion of this issue, see [KJ05].

The results of this section will be used to compute the reference values for the standard Heston model in all numerical examples in the rest of this thesis.

Chapter 3

Malliavin Calculus

Malliavin calculus introduces a stochastic derivative operator to stochastic calculus. Loosely speaking, the derivative allows to differentiate random variables "in direction" of the chance parameter ω. It was initially developed by Paul Malliavin in 1976 [Mal76] as a tool to find sufficient conditions for the existence and regularity of densities. The most important application in this direction is Malliavin's probabilistic proof of Hörmander's theorem, which gives a sufficient condition for the solution of a stochastic differential equation to possess an infinitely differentiable density. Besides regularity questions, Malliavin calculus is nowadays also used for anticipating stochastic differential equations and mathematical finance, where it is for example used to examine the sensitivity of option prices on various parameters; see e.g. [DNOP09], [Nua06], [FLL+99].
In this chapter we will develop a basic theory of Malliavin calculus including almost all proofs. In Sections 3.3 and 3.5 we will spend some effort to prove a very general chain rule and to construct a slightly extended version of the Malliavin derivative. Both are tools that we are going to use later and which are not found in the literature yet — to the best of our knowledge. Finally, in Section 3.6 we state some more advanced results which we are going to use, but whose proofs are well beyond the scope of this thesis.
Monographs on Malliavin calculus are for example [Nua06] and [DNOP09], a short introduction for beginners can be found in [Øk97] and [Alt11]. This chapter is mainly based on [Nua06] and [Alt11].

3.1 Preliminaries

There are several objects that can be used as fundamental objects upon which Malliavin calculus is constructed. We will use isonormal Gaussian processes. We assume that we are given a complete probability space $(\Omega, \mathcal{F}, P)$, a real separable Hilbert space H and a process $(W(h))_{h\in H}$ conforming with the following definition. We further assume that $\mathcal{F}$ is generated by the random variables $\{W(h) : h \in H\}$.

Definition 3.1 An *isonormal Gaussian process* on a Hilbert space H is a linear mapping $W\colon H \to L^2(\Omega, \mathcal{F}, P)$ such that

- For each $h \in H$ the random variable $W(h)$ is centered and Gaussian.
- W is an isometry: $\|W(h)\|_2 = \|h\|_H$ for all $h \in H$. ◁

The linearity of W implies that the random variables $(W(h))_{h\in H}$ are jointly Gaussian[1] The polarization identity proves $\langle W(h), W(g)\rangle_{L^2(\Omega)} = \langle h, g\rangle_H$. Put together we get that

[1] A set of random variables is called *jointly Gaussian* if each linear combination follows a Gaussian distribution. A subset of such random variables is independent if and only it is uncorrelated.

two random variables $W(h)$ and $W(g)$ are uncorrelated/independent if and only if $h \perp g$. Kolmogorov's consistency theorem (see e.g. Theorem 62.3 in [Bau74]) shows that on every Hilbert space there exists an isonormal Gaussian process.
The classical example of an isonormal Gaussian process is the Wiener integral with respect to a d-dimensional Brownian motion Z. It maps $h \in H := L^2([0,T];\mathbb{R}^d)$ to

$$W(h) := \int_0^T h(t)dZ_t = \sum_{i=1}^{d} \int_0^T h_i(t)dZ_t^i \tag{3.1}$$

(with a fixed endtime T). Although this so-called *Wiener setting* is the only isonormal Gaussian process that we are going to use, we believe that it is easier to introduce Malliavin calculus in the more abstract framework of above definition, because it provides exactly the structure necessary for Malliavin calculus.
We denote by $C^\infty_{pol}(\mathbb{R}^d,\mathbb{R})$ the set of functions $f\colon \mathbb{R}^d \to \mathbb{R}$ which are infinitely often differentiable such that each (partial) derivative of any order is bounded by some polynomial. $C^\infty_b(\mathbb{R}^d,\mathbb{R})$ is the subset of functions which are bounded and have bounded partial derivatives of all orders.

Definition 3.2 A random variable $X \in L^2(\Omega)$ is called *smooth*, if there exists $n \in \mathbb{N}$, a function $f \in C^\infty_{pol}(\mathbb{R}^n,\mathbb{R})$ and $h_1,\dots,h_n \in H$ such that

$$X = f(W(h_1),\dots,W(h_n))$$

The set of smooth random variables is denoted by $\mathcal{S}$. The set $\mathcal{S}_b$ consists of those smooth random variables which allow a representation using $f \in C^\infty_b(\mathbb{R}^n,\mathbb{R})$. ◁

The representation of a smooth random variable as a functional of several $W(h_i)$ is of course not unique. Often we will require the h_i to be orthonormal; because W is linear, this clearly is no restriction.

Proposition 3.3 The sets $\mathcal{S}$ and $\mathcal{S}_b$ are dense in $L^p(\Omega;\mathcal{F},P)$ for all $p \geq 1$.

Proof (Sketch). It suffices to consider the case of $\mathcal{S}_b$ and $p > 1$. Let q be the Hölder conjugate of p. Assume $Z \in L^q(\Omega)$ such that $E(XZ) = 0$ for all $X \in \mathcal{S}_b$. Using an approximation procedure via the exponential series , we then can show that even $E(e^{W(h)}Z) = 0$ for all $h \in H$ (for the details see Theorem A.2 in [Alt11]). Because the set $\{e^{W(h)} : h \in H\}$ is total in $L^p(\Omega)$, see Lemma 1.1.2 in [Nua06], this implies $Z = 0$ and thus $\mathcal{S}_b$ must be dense. □

3.2 The Malliavin Derivative

We will first define the derivative for smooth random variables and then extend it to a bigger class of random variables.

Definition 3.4 Let X be a smooth random variable that can be written as

$$X = f(W(h_1),\dots,W(h_n)) \tag{3.2}$$

◁

for $f \in C^\infty_{pol}(\mathbb{R}^n,\mathbb{R})$ and $h_1,\dots,h_n \in H$. Then we define the Malliavin derivative DX of X as the H-valued random variable

$$DX = \sum_{i=1}^{n} \partial_i f(W(h_1),\dots,W(h_n)) \cdot h_i$$

Note that the Malliavin derivative DX is an element of $L^p(\Omega; H)$. In the Wiener setting this is a random function in $H = L^2([0,T]; \mathbb{R}^d)$ and for $t \in [0,T]$ we write $D_t^i X$ instead of $((DX)(t))_i$. A simple consequence of above definition is $D(W(h)) = h$. In the Wiener setting this means for example that $D^i(Z_t^i - Z_s^i) = \mathbb{1}_{[s,t]}$ for $0 \le s \le t \le T$ and $D^j(Z_t^i - Z_s^i) = 0$ for $j \neq i$. This explains, why the i-th component of DX is regarded as derivative with respect to the i-th Brownian motion Z^i.
A random variable which is not differentiable is e.g. $\mathbb{1}_A$ for an event A with $P(A) \notin \{0,1\}$; see Proposition 1.2.6 in [Nua06].
Of course, we have to prove that above definition is well-defined:

Theorem 3.5 The definition of the Malliavin derivative of a smooth random variable X does not depend on the particular representation of X.

Proof (Sketch). For a detailed proof see Proposition 2.8 in [Alt11]. Given a random variable X with representation (3.2), choose orthonormal vectors $e_1, \dots, e_r$ and a matrix $A \in \mathbb{R}^{n \times r}$ such that $h_i = \sum_{j=1}^r a_{ij} e_j$, $i = 1, \dots, n$. Then for the linear map φ associated with A we have

$$\varphi(W(e_1), \dots, W(e_r)) = (W(h_1), \dots, W(h_n))$$

and thus

$$\begin{aligned}\sum_{i=1}^n \partial_i f(W(h_1), \dots, W(h_n)) \cdot h_i &= \sum_{j=1}^r \sum_{i=1}^n \partial_i f(\varphi(W(e_1), \dots, W(e_r)) \cdot a_{ij} e_j \\ &= \sum_{j=1}^r \partial_j (f \circ \varphi)(W(e_1), \dots, W(e_r)) \cdot e_j\end{aligned}$$

Given two representations of X, we can thus switch both to representations using the same orthonormal system without changing the derivatives. But because $(W(e_1), \dots, W(e_r))$ has an r-dimensional standard normal distribution, both representations must coincide. Thus both derivatives must be equal. □

The key to extending the derivative to a bigger domain is the integration by parts rule. Later, we will use the same rule to improve the efficiency of the quadrature of discontinuous payoffs; see Theorem 8.4.

Proposition 3.6 Let $X \in \mathcal{S}$ and $h \in H$. Then

$$E(\langle DX, h \rangle_H) = E(X \cdot W(h))$$

Proof. We can assume that X is given using a representation $X = f(W(e_1), \dots, W(e_n))$ with orthonormal vectors $e_i \in H$ and $e_1 = h/\|h\|_H^2$. We write ν^n for the n-dimensional standard normal distribution. Then

$$\begin{aligned}E(\langle DX, h \rangle_H) &= \sum_{i=1}^n E\big(\partial_i f(W(e_1), \dots, W(e_n))\big) \cdot \langle e_i, h \rangle_H \\ &= E\big(\partial_1 f(W(e_1), \dots, W(e_n))\big) \\ &= \int_{\mathbb{R}^n} \partial_1 f \, d\nu^n \\ &= \int_{\mathbb{R}^n} \partial_1 f(x) \cdot (2\pi)^{-\frac{n}{2}} e^{-\frac{\|x\|^2}{2}} \, dx\end{aligned}$$

Now we can use the integration by parts formula from ordinary calculus to replace $\partial_1 f$ by f. Because f is polynomially bounded, while the density of the normal distribution decreases exponentially, the boundary term vanishes.

$$\begin{aligned}
&= -\int_{\mathbb{R}^n} f(x) \cdot (2\pi)^{-\frac{n}{2}} e^{-\frac{\|x\|^2}{2}} \cdot (-x_1)\, dx \\
&= \int_{\mathbb{R}^n} f(x) x_1 d\nu^n \\
&= E(f(W(e_1), \dots, W(e_n)) \cdot W(e_1)) \\
&= E(X \cdot W(h))
\end{aligned}$$

□

If we apply the integration by parts rule to a product $X \cdot Y$ of smooth random variables we get

$$E(Y \cdot \langle DX, h\rangle_H) = E(X \cdot Y \cdot W(h)) - E(X \cdot \langle DY, h\rangle_H) \tag{3.3}$$

This small formula is the central tool in the extension of the Malliavin derivative to a closed operator.

Definition 3.7 Let V, U be normed vector spaces. An operator $A\colon \operatorname{dom} A \subset V \to U$ is called

1. *closed*, if for all sequences $(x_n) \subset V$ the existence of $\lim x_n$ and $\lim Ax_n$ imply $\lim x_n \in \operatorname{dom} A$ and $A(\lim x_n) = \lim Ax_n$.

2. *closable*, if for all sequences $(x_n), (y_n) \subset V$ the existence and equality of $\lim x_n = \lim y_n$ and the existence of $\lim Ax_n$ and $\lim Ay_n$ imply $\lim Ax_n = \lim Ay_n$. ◁

Clearly, the second property allows to extend the closable operator A to a closed operator by setting $A(\lim x_n) := \lim Ax_n$. To show that an operator is closable, it suffices to prove that whenever a sequence $(x_n) \subset V$ converges to 0 and $\lim Ax_n$ exists, then $\lim Ax_n = 0$ (because in this case the conditions of Definition 3.7 2. imply $x_n - y_n \to 0$ and $\lim A(x_n - y_n)$ exists, so $\lim A(x_n - y_n) = 0$).

Theorem 3.8 Let $p \geq 1$. The operator D as defined in Definition 3.4 is closable from $\mathcal{S} \subset L^p(\Omega)$ to $L^p(\Omega; H)$.

Proof. Let $(X_n)_{n\in\mathbb{N}} \subset \mathcal{S}$ converge to 0 in $L^p(\Omega)$ such that DX_n converges to some Z in $L^p(\Omega; H)$. Formula (3.3) gives for arbitrary $h \in H$ and $Y \in \mathcal{S}_b$ that

$$E(Y\langle Z, h\rangle_H) = \lim_{n\to\infty} E(Y\langle DX_n, h\rangle_H) = \lim_{n\to\infty} \big(E(YW(h) \cdot X_n) - E(\langle DY, h\rangle_H \cdot X_n)\big) \tag{3.4}$$

Because $Y \in \mathcal{S}_b$, hence DY is bounded, the second summand converges to 0. The first summand converges to 0, if we require that $YW(h)$ is bounded. So let us define

$$A(h) := \{Y \in \mathcal{S}_b : FW(h) \text{ is bounded}\}$$

For every $Y \in \mathcal{S}_b$ and $\varepsilon > 0$ the random variable $Ye^{-\varepsilon W(h)^2}$ is in $A(h)$ and

$$\lim_{\varepsilon\to 0} Ye^{-\varepsilon W(h)^2} = Y$$

in $L^p(\Omega)$ by dominated convergence. Thus $A(h)$ is dense in $\mathcal{S}_b$ for all $h \in H$. Therefore (3.4) implies $E(Y\langle Z, h\rangle_H) = 0$ for all $Y \in \mathcal{S}_b$ and $h \in H$. Because $\mathcal{S}_b$ is dense in $L^p(\Omega)$, see Proposition 3.3, this means that $\langle Z, h\rangle_H = 0$ almost surely for all h and thus $Z = 0$ almost surely. □

Definition 3.9 Let $p \geq 1$. We define the Malliavin derivative D as the closure of the operator D defined in Definition 3.4. The domain of D is denoted by $\mathbb{D}^{1,p}$ and given by

$$\mathbb{D}^{1,p} := \left\{ X \in L^p(\Omega) : \exists (X_n)_{n\in\mathbb{N}} \subset \mathcal{S}, Y \in L^p(\Omega; H) \text{ s.t. } \begin{matrix} X_n \to X \text{ in } L^p(\Omega) \\ DX_n \to Y \text{ in } L^p(\Omega; H) \end{matrix} \right\}$$

A random variable within $\mathbb{D}^{1,p}$ is called differentiable. ◁

While the domain of D depends on the parameter p, the actual derivative DX does not depend on p (as long as $X \in \mathbb{D}^{1,p}$). We define $\mathbb{D}^{1,\infty} = \bigcap_{p\geq 1} \mathbb{D}^{1,p}$. Note that as a consequence of the definition all differentiable random variables must be measurable with respect to the underlying isonormal Gaussian process W.

3.3 Chain Rules

An important tool in Malliavin calculus is the chain rule. Because the derivative — like most of stochastic calculus — is only defined up to sets of measure zero, the chain rule can be extended to functions that are not everywhere differentiable. In our application of Malliavin calculus to the quadrature of discontinuous functionals we will need a chain rule which is more general than what can be found in the literature. In this section we will develop this result in a series of increasingly general chain rules.
We say that the chain rule holds for a function $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ and a random vector $X = (X_1, \dots, X_d)$ with $X_i \in \mathbb{D}^{1,p}$, if $\varphi(X) \in \mathbb{D}^{1,p}$ and

$$D\varphi(X) = \sum_{i=1}^{d} \partial_i \varphi(X) \cdot DX_i \tag{3.5}$$

Because the Malliavin derivative is a closed operator, in order to prove the chain rule for φ and X, it clearly suffices to find functions $\varphi_k\colon \mathbb{R}^d \to \mathbb{R}$, $k \in \mathbb{N}$, with the following properties:

1. The chain rule holds for all φ_k.
2. $\varphi_k(X) \to \varphi(X)$ in $L^p(\Omega)$.
3. $\partial_i \varphi_k(X) \cdot DX_i \to \partial_i \varphi(X) \cdot DX_i$ in $L^p(\Omega; H)$ for all $i = 1, \dots, d$.

The proofs in this section will frequently use the following mollifier functions: For each $n \in \mathbb{N}$ choose $\psi_n \in C^\infty(\mathbb{R}^d, \mathbb{R})$ such that $\psi_n(x) = 0$ if $\|x\|_\infty \geq 1/n$ and $\int_{\mathbb{R}^d} \psi_n(x)dx = 1$. Note that

$$\int_{\mathbb{R}^d} \|y\| \cdot \psi_n(y)dy \leq \int_{\{\|y\|\leq 1/n\}} \|y\| \cdot \psi_n(y)dy \leq \frac{1}{n} \int_{\mathbb{R}^d} \psi_n(y)dy = \frac{1}{n} \tag{3.6}$$

Proposition 3.10 The chain rule holds if $X_1, \dots, X_d \in \mathcal{S}$ and $\varphi \in C^\infty_{pol}(\mathbb{R}^d, \mathbb{R})$.

Proof. We can represent all X_i using the same $h_1, \dots, h_n \in H$ so that

$$X_i = f_i(W(h_1), \dots, W(h_n))$$

Set $f = (f_1, \dots, f_d)$. Then $\varphi \circ f \in C^\infty_{pol}(\mathbb{R}^n, \mathbb{R})$ and thus $\varphi(X)$ is a smooth random variable. By definition, the derivative is

$$\begin{aligned} D(\varphi(X)) &= \sum_{j=1}^{n} \partial_j(\varphi \circ f)(W(h_1), \dots, W(h_n)) \cdot h_j \\ &= \sum_{j=1}^{n} \sum_{i=1}^{d} \partial_i \varphi(f(W(h_1), \dots, W(h_n))) \cdot \partial_j f_i(W(h_1), \dots, W(h_n)) \cdot h_j \\ &= \sum_{i=1}^{d} \partial_i \varphi(X) \cdot DX_i \end{aligned}$$

□

Proposition 3.11 The chain rule holds if $X_1, \dots, X_d \in \mathbb{D}^{1,p}$ and $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable with bounded partial derivatives.

Proof. $\varphi(X) \in L^p(\Omega)$ follows from the mean value theorem. First assume that all X_i are smooth random variables. Set

$$\tilde{\varphi}_n(x) := \begin{cases} \varphi(x) & \text{if } \|x\| \le n \\ 0 & \text{if } \|x\| > n \end{cases}$$

and define

$$\varphi_n := \tilde{\varphi}_n * \psi_n$$

φ_n is infinitely often differentiable and has bounded support and consequently $\varphi_n \in C^\infty_{\text{pol}}(\mathbb{R}^d, \mathbb{R})$ (note that this is not necessarily true for $\varphi * \psi_n$). Thus the previous theorem proves the chain rule for φ_n.
Moreover for each $\omega \in \Omega$ we have (we abbreviate $X = (X_1(\omega), \dots, X_d(\omega))$ and $Q_n := [-n, n]^d$)

$$\begin{aligned} |\varphi_n(X) - \varphi(X)| &= \left| \int_{Q_n} \varphi(y) \cdot \psi_n(X-y)dy - \int_{\mathbb{R}^d} \varphi(X) \cdot \psi_n(X-y)dy \right| \\ &\le \int_{Q_n} |\varphi(y) - \varphi(X)| \cdot \psi_n(X-y)dy + |\varphi(X)| \cdot \int_{Q_n^c} \psi_n(X-y)dy \\ &\le \sup \|\nabla\varphi\| \cdot \int_{Q_n} \|y - X\| \cdot \psi_n(X-y)dy + |\varphi(X)| \cdot \int_{Q_n^c} \psi_n(X-y)dy \\ &\le \sup \|\nabla\varphi\| \cdot \frac{1}{n} + |\varphi(X)| \cdot \int_{Q_n^c} \psi_n(X-y)dy \end{aligned}$$

This converges to 0 almost surely and in $L^p(\Omega)$ because it is bounded by $\sup \nabla\varphi + \varphi(X)$. Analogously

$$\begin{aligned} \|\partial_i\varphi_n(X) \cdot DX_i - \partial_i\varphi(X) \cdot DX_i\|_H \le \Bigg(&\int_{Q_n} |\partial_i\varphi(y) - \partial_i\varphi(X)| \cdot \psi_n(X-y)dy \\ &+ |\partial_i\varphi(X)| \cdot \int_{Q_n^c} \psi_n(X-y)dy \Bigg) \cdot \|DX_i\|_H \end{aligned}$$

This converges to 0 almost surely because for a given $\varepsilon > 0$ we can choose $\delta = \delta(\omega) > 0$ such that $|\partial_i\varphi(y) - \partial_i\varphi(X(\omega))| < \varepsilon$ for all y with $\|y - X(\omega)\| < \delta$. Thus for all $n > 1/\delta$ the first integral is at most ε. Because $\partial_i\varphi$ is bounded, convergence holds in $L^p(\Omega)$, too.
Now consider the general case. Choose $X^{(n)} = (X_1^{(n)}, \dots, X_d^{(n)})$ with smooth random variables $X_i^{(n)}$ such that $X_i^{(n)} \to X_i$ in $\mathbb{D}^{1,p}$ and almost surely as $n \to \infty$ uniformly in i. Then

$$|\varphi(X^{(n)}) - \varphi(X)| \le \sup \|\nabla\varphi\| \cdot \|X^{(n)} - X\|$$

and

$$\begin{aligned} \|\partial_i\varphi(X^{(n)}) \cdot DX_i^{(n)} - \partial_i\varphi(X) \cdot DX_i\| \le{}& |\partial_i\varphi(X^{(n)})| \cdot \|DX_i^{(n)} - DX_i\|_H \\ &+ |\partial_i\varphi(X^{(n)}) - \partial_i\varphi(X)| \cdot \|DX_i\|_H \end{aligned}$$

converge to 0 by bounded convergence. As above this suffices to prove the claim. □

Surprisingly the chain rule can be extended to some functions which are not differentiable (to make equation (3.5) meaningful, we set $\partial_i\varphi(x) := 0$ where this partial derivative does

not exist). Because we will continue to use the mollifier functions ψ_n, we need the following lemma. We call a function *Lipschitz in direction* e_i, if there exists a constant $L \geq 0$ such that

$$|\varphi(x + he_i) - \varphi(x)| \leq L \cdot h \qquad \forall x \in \mathbb{R}^d, h \in \mathbb{R}$$

Being *locally* Lipschitz in a direction e_i means that for each $x \in \mathbb{R}^d$ there exist $L(x) \geq 0$ and $\varepsilon(x) > 0$ such that

$$|\varphi(x + he_i) - \varphi(x)| \leq L(x) \cdot h \qquad \forall h \in \mathbb{R}, |h| \leq \varepsilon(x)$$

Lemma 3.12 Let $i \in \{1, \dots, d\}$ and assume $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ is Lipschitz in direction e_i. Then $\partial_i(\varphi * \psi_n)$ exists almost everywhere and equals $\partial_i \varphi * \psi_n$. If φ is only locally Lipschitz in direction e_i this is still true if n is large enough.

Proof. Rademacher's theorem (see e.g. Theorem 6 in §5.8.3 of [Eva10]) proves that $\partial_i \varphi$ exists for almost every $x \in \mathbb{R}^d$. For such x we have

$$\lim_{h\to 0} \frac{1}{h}(\varphi * \psi_n(x + he_i) - \varphi(x)) = \lim_{h\to 0} \int_{\mathbb{R}^d} \frac{1}{h}(\varphi(x + he_i - y) - \varphi(x - y)) \cdot \psi_n(y)dy$$

The integrand is bounded by $L \cdot \sup \psi_n$ and thus we can exchange integral and limit.

$$\begin{aligned} &= \int_{\mathbb{R}^d} \partial_i \varphi(x - y) \cdot \psi_n(y)dy \\ &= \partial_i \varphi * \psi_n(x) \end{aligned}$$

If φ is only locally Lipschitz, we must choose n large enough such that the bound $|\varphi(x + he_i - y) - \varphi(x - y)| \leq L(x) \cdot h$ holds for all $y \in \operatorname{supp} \psi_n$ for h small enough. □

Assumption 3.13 The set M of points where φ is not partially continuously differentiable is of the form $M = N \cup Z$ where $P(X \in N) = 0$ and Z is at most countable. ◁

Rademacher's theorem shows that for every locally Lipschitz continuous function φ the set M has Lebesgue measure zero. In particular, for these functions Assumption 3.13 is always fulfilled if P_X is absolutely continuous with respect to the Lebesgue measure.

Proposition 3.14 The chain rule holds if φ is Lipschitz continuous and Assumption 3.13 holds.

Proof. Let L be the Lipschitz constant of φ. $\varphi(X) \in L^p(\Omega)$ follows from

$$|\varphi(X)| \leq |\varphi(X) - \varphi(0)| + |\varphi(0)| \leq L|X| + |\varphi(0)|$$

Set $\varphi_n := \varphi * \psi_n$. Because φ is Lipschitz, all partial derivatives of φ_n are bounded by L so the chain rule holds for φ_n by the previous theorem. Furthermore

$$\begin{aligned} |\varphi_n(X) - \varphi(X)| &\leq \int_{\mathbb{R}^d} |\varphi(y) - \varphi(X)| \cdot \psi_n(X - y)dy \\ &\leq L \cdot \int_{\mathbb{R}^d} |y - X| \cdot \psi_n(X - y)dy \\ &\leq L \cdot \frac{1}{n} \end{aligned}$$

Thus $\varphi_n(X) \to \varphi(X)$ in $L^p(\Omega)$. For each $\omega \in \Omega$ we have

$$|\partial_i \varphi_n(X) - \partial_i \varphi(X)| \cdot \|DX_i\|_H \leq \int_{\mathbb{R}^d} |\partial_i \varphi(y) - \partial_i \varphi(X)| \cdot \psi_n(X - y)dy \cdot \|DX_i\|_H$$

If $X(\omega) \notin M$, i.e. $\partial_i \varphi$ is continuous at $X(\omega)$, the last integral converges to zero. On the other hand, by Proposition 1.3.16 in [Nua06], for each $x \in \mathbb{R}^d$ there exists a null set $N_x \subset \Omega$ such that $DX_i = 0$ on $\{X = x\} \setminus N_x$. Because Z is at most countable, above integral converges almost surely (except on $\bigcup_{z\in Z} N_z \cup N$). The sequence is dominated by $2L \cdot DX$ and thus converges to 0 also in L^p and the proof is complete. □

To extend the chain rule to functions which are not globally Lipschitz continuous we clearly need the following assumption.

Assumption 3.15 $\varphi(X) \in L^p(\Omega)$ and $\partial_i\varphi(X) \cdot DX_i \in L^p(\Omega; H)$ for all $i = 1, \dots, d$. ◁

Proposition 3.16 The chain rule holds if φ is locally Lipschitz continuous and bounded and Assumptions 3.13 and 3.15 hold.

Proof. Choose a sequence $(n_k)_k \subset \mathbb{R}$ such that $n_k \to \infty$ and $P(\|X\|_\infty = n_k) = 0$ for all $k \in \mathbb{N}$. Define $\varphi_k(x) := \varphi(-n_k \vee x \wedge n_k)$, component-wise. By construction $|\partial_i\varphi_k| \leq |\partial_i\varphi|$ and φ_k is Lipschitz continuous and Assumption 3.13 holds for φ_k. Thus the previous proposition shows that the chain rule holds for $\varphi_k(X)$. By dominated convergence $\varphi_k(X) \to \varphi(X)$ in $L^p(\Omega)$ and $\partial\varphi_k(X)DX_i \to \partial\varphi(X)DX_i$ in $L^p(\Omega; H)$ which implies the chain rule for φ. □

Theorem 3.17 (Chain rule) The chain rule holds if φ is locally Lipschitz continuous and Assumptions 3.13 and 3.15 hold.

Proof. Choose a sequence $(n_k)_k \subset \mathbb{R}$ such that $n_k \to \infty$ and $P(|\varphi(X)| = n_k) = 0$ for all $k \in \mathbb{N}$. Set $\varphi_k(x) := -n_k \vee \varphi(x) \wedge n_k$. By construction φ_k is bounded and Assumption 3.13 also holds for φ_k. Thus the chain rule holds for φ_k. By dominated convergence $\varphi_k(X) \to \varphi(X)$ and $\partial_i\varphi_n(X)DX_i \to \partial_i\varphi(X)DX_i$. □

Corollary 3.18 The chain rule holds for a continuously differentiable function if Assumption 3.15 holds.

The next step is to extend the chain rule to functions which may even be not continuous in some directions. This works as long as the Malliavin derivative of the random variables in this direction vanishes. For this we have to extend our definition of chain rule slightly by replacing (3.5) by

$$D\varphi(X) = \sum_{\substack{i=1,\dots,d \\ DX_i \neq 0}} \partial_i\varphi(X) \cdot DX_i \tag{3.7}$$

Let A denote the set of indices $i = 1, \dots, d$ with $DX_i \neq 0$.

Proposition 3.19 Let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ be continuous except on a P_X-zero set and assume that φ is locally Lipschitz continuous in all directions e_i, $i \in A$. If Assumption 3.15 holds, then the chain rule holds for φ.

Proof. First assume that φ is bounded and globally Lipschitz continuous in all directions e_i, $i \in A$. Because the chain rule holds for $\varphi * \psi_n$ we only have to prove $\varphi * \psi_n(X) \to \varphi(X)$ in $L^p(\Omega)$ and $\partial_i\varphi * \psi_n(X) \cdot DX_i \to \partial_i\varphi(X) \cdot DX_i$ in $L^p(\Omega; H)$ for $i \in A$. We have

$$|\varphi * \psi_n(X) - \varphi(X)| \leq \int_{\mathbb{R}^d} |\varphi(y) - \varphi(X)| \cdot \psi_n(X - y)dy$$

The integrand converges almost surely to 0 because φ is continuous P_X-almost everywhere. Because the integral is bounded by $2 \sup|\varphi|$, above expression converges to 0 in $L^p(\Omega)$. Furthermore

$$|(\partial_i(\varphi * \psi_n)(X) - \partial_i\varphi(X)) \cdot DX_i| \leq \int_{\mathbb{R}^d} |\partial_i\varphi(y) - \partial_i\varphi(X)| \cdot \psi_n(X - y) \cdot |DX_i|dy$$

The integral is dominated by $\frac{L_i}{n} \cdot DX_i$ due to (3.6) and thus this converges to 0 in $L^p(\Omega; H)$. The extension to unbounded and locally Lipschitz functions is the same as in Proposition 3.16 and Theorem 3.17. □

A direct consequence of the chain rule applied to $(x, y) \mapsto x \cdot y$ is the product rule.

Corollary 3.20 Assume $X, Y \in \mathbb{D}^{1,p}$ such that $X \cdot DY \in L^p(\Omega; H)$ and $Y \cdot DX \in L^p(\Omega; H)$. Then $X \cdot Y \in \mathbb{D}^{1,p}$ and

$$D(XY) = X \cdot DY + Y \cdot DX$$

3.4 The Skorohod Integral

The adjoint operator of the Malliavin derivative plays an important and somewhat surprising role: In the Wiener setting it is a generalization of the Itō integral to non-adapted processes.

Definition 3.21 Let δ denote the adjoint operator of $D\colon \mathbb{D}^{1,2} \subset L^2(\Omega) \to L^2(\Omega;H)$. δ is called the *divergence operator* or in the Wiener setting the *Skorohod integral*. ◁

Because D is an unbounded operator defined on a dense subset of $L^2(\Omega)$, the precise meaning of the previous definition is the following; see e.g. Definition VII.2.3 in [Wer07].

1. δ is defined on the subspace

$$\operatorname{dom}(\delta) := \{u \in L^2(\Omega;H) \,:\, \mathbb{D}^{1,2} \ni X \mapsto \langle DX, u\rangle_{L^2(\Omega;H)} \text{ is continuous}\}$$

   Because $\mathbb{D}^{1,2}$ is dense in $L^2(\Omega)$, for $u \in \operatorname{dom}\delta$ the functional $X \mapsto \langle DX, u\rangle_{L^2(\Omega;H)}$ can be uniquely extended to a linear functional φ on $L^2(\Omega)$.

2. If $u \in \operatorname{dom}(\delta)$, then $\delta(u)$ is the unique (by the Riesz representation theorem) element $Y \in L^2(\Omega)$ such that $\varphi = \langle \cdot, Y\rangle_{L^2(\Omega)}$.

As a direct consequence of the definition we have

$$E(\langle DX, u\rangle_H) = E(X\delta(u)) \tag{3.8}$$

for all $X \in \mathbb{D}^{1,2}$ and $u \in \operatorname{dom}(\delta)$. This relation is called the *integration by parts rule* and generalizes Proposition 3.6.

Remark 3.22 To prove that $u \in \operatorname{dom}(\delta)$ and $\delta(u) = Y$ for some $Y \in L^2(\Omega)$, it is sufficient to show $E(\langle DX, u\rangle_H) = E(XY)$ for all $X \in \mathcal{S}$ (or all $X \in \mathcal{S}_b$ or any other dense subset of $\mathbb{D}^{1,2}$). Because if we approximate an arbitrary $X \in \mathbb{D}^{1,2}$ by $X_n \in \mathcal{S}$ we have

$$E(\langle DX, u\rangle_H) = \lim_{n\to\infty} E(\langle DX_n, u\rangle_H) = \lim_{n\to\infty} E(X_n Y) = E(XY)$$

In particular the mapping $\mathbb{D}^{1,2} \ni X \mapsto \langle DX, u\rangle_{L^2(\Omega;H)}$ is continuous, hence $u \in \operatorname{dom}\delta$, and $\delta(u) = Y$. ◁

Using this remark, we can compute the divergence of simple processes of the form $u = \sum_{i=1}^n X_i h_i$ with $X_i \in \mathcal{S}$ and $h_i \in H$, $i = 1, \dots, n$: For every $Y \in \mathcal{S}$ we get from (3.3)

$$E(\langle DY, u\rangle_H) = \sum_{i=1}^n E(X_i \langle DY, h_i\rangle_H) = \sum_{i=1}^n E(X_i Y W(h_i) - Y\langle DX_i, h_i\rangle_H)$$

and thus

$$\delta(u) = \sum_{i=1}^n X_i W(h_i) - \langle DX_i, h_i\rangle_H \tag{3.9}$$

For more complex random processes we will often use the following generalization of (3.9).

Proposition 3.23 Let $X \in \mathbb{D}^{1,2}$, $u \in \operatorname{dom}(\delta)$ such that $Xu \in L^2(\Omega;H)$ and $X\delta(u) - \langle DX, u\rangle_H \in L^2(\Omega)$. Then $Xu \in \operatorname{dom}(\delta)$ and

$$\delta(Xu) = X\delta(u) - \langle DX, u\rangle_H$$

Proof. Define $A := X\delta(u) - \langle DX, u\rangle_H$. For every $Y \in \mathcal{S}_b$ the product rule (Corollary 3.20) implies

$$E(\langle DY, Xu\rangle_H) = E(\langle u, D(XY) - Y \cdot DX\rangle_H) = E(\delta(u)XY - \langle u, DX\rangle_H \cdot Y) = E(A \cdot Y)$$

By Remark 3.22, this is sufficient. □

In the Wiener setting, i.e. $H = L^2([0,T];\mathbb{R}^d)$ and W is the Wiener integral over a d-dimensional Brownian motion Z on $[0,T]$, the divergence operator is called the Skorohod integral and denoted by $\delta(u) = \int_0^T u_s \delta Z_s$. The resemblance to the notation of the Itō integral is justified by the following theorem.

Theorem 3.24 Consider the Wiener setting. If $u \in L^2(\Omega \times [0,T])$ is an adapted process, then $u \in \mathrm{dom}(\delta)$ and $\delta(u) = \int_0^T u(t)\delta Z_t = \int_0^T u(t)dZ_t = \sum_{i=1}^d \int_0^T u_i(t)dZ_t^i$.

Proof (sketch). The full proof is given in Proposition 1.3.11 of [Nua06]. The main idea (for simplicity assume $d = 1$) is that for an adapted step process $u = \sum_{i=1}^n X_i \mathbb{1}_{[t_i,t_{i+1})}$ with $0 \le t_1 < \cdots < t_n \le T$ and suitable X_i, we can compute the Skorohod integral using (3.9) or Proposition 3.23 as

$$\delta(u) = \sum_{i=1}^n X_i \cdot (Z_{t_{i+1}} - Z_{t_i}) - \langle DX_i, \mathbb{1}_{[t_i,t_{i+1})}\rangle_{L^2([0,T])}$$

Because X_i is measurable with respect to $\{Z_t : 0 \le t \le t_i\}$, the derivative $D_t X_i$ vanishes for $t > t_i$; see Proposition 1.2.8 in [Nua06]. Thus the second summand is zero, while the first one clearly is the Itō integral. □

3.5 Partial Malliavin Derivatives

When working with multidimensional Heston models later, it will be useful to consider models where only one price process is Malliavin differentiable. For this we introduce partial Malliavin derivatives which only differentiate with respect to a single Brownian motion. This is actually a special case of a much more general construction and because it makes the notation easier, we will consider this general construction and then introduce the special case that we will use later.
In the following let $\tilde{H}$ be a second Hilbert space and let $A\colon H \to \tilde{H}$ be a continuous linear operator. On the set of smooth random variables $\mathcal{S}$ define D^A as $A \circ D$, i.e.

$$D^A(X)(\omega) = A(DX(\omega)), \qquad X \in \mathcal{S}$$

As the following theorem shows, this is a closable operator, so that we can define the final D^A as the closure of $D^A = A \circ D$. The domain of the closed operator is denoted by $\mathbb{D}^{1,p,A}$.

Theorem 3.25 The operator D^A is closable from $\mathcal{S} \subset L^p(\Omega)$ to $L^p(\Omega;\tilde{H})$.

Proof. Let $h \in \tilde{H}$ and $X, Y \in \mathcal{S}$. From (3.3) we get

$$E(\langle ADX, h\rangle_{\tilde{H}} \cdot Y) = E(\langle DX, A^*h\rangle_H \cdot Y) = E(-X \cdot \langle ADY, h\rangle_{\tilde{H}} + XYW(A^*h))$$

Let X_n be a sequence of smooth random variables converging to 0 in $L^p(\Omega)$ such that ADX_n converges in $L^p(\Omega;\tilde{H})$ to some random variable u. For each $h \in \tilde{H}$ and $Y \in \mathcal{S}_b$ such that $YW(A^*h)$ is bounded, we have

$$\begin{aligned} E(\langle u, h\rangle_{\tilde{H}} Y) &= \lim_{n\to\infty} E(\langle ADX_n, h\rangle_{\tilde{H}} Y) \\ &= \lim_{n\to\infty} E(-X_n\langle ADY, h\rangle + X_n YW(A^*h)) \\ &= 0 \end{aligned}$$

If we can prove that the set of considered Y is dense in $\mathcal{S}_b$ for each $h \in \tilde{H}$, then this implies $u = 0$ and hence the assertion. In fact, for each $h \in \tilde{H}$ and $Y' \in \mathcal{S}_b$ we can choose $Y = Y' e^{-\varepsilon W(A^* h)^2}$ to approximate Y' arbitrarily well and such that $YW(A^*h))$ is bounded; see also the proof of Theorem 3.8. □

Because A is continuous, clearly $\mathbb{D}^{1,p} \subset \mathbb{D}^{1,p,A}$ and $D^A = A \circ D$ on $\mathbb{D}^{1,p}$.

Proposition 3.26 For each $p \geq 1$ we have $D^A X = 0$ for each random variable X in the $\mathbb{D}^{1,p,A}$-closure of the set

$$\mathcal{S}_{\ker} := \{X \in \mathcal{S} : X = f(W(h_1), \dots, W(h_n)),\ f \in C^\infty_{\text{pol}}(\mathbb{R}^n, \mathbb{R}),\ h_1, \dots, h_n \in \ker A\}$$

Proof. By our assumptions we can choose a sequence $(X_n) \subset \mathcal{S}_{\ker}$ such that $X_n \to X$ in $L^p(\Omega)$ and $D^A X_n \to D^A X$ in $L^p(\Omega; \tilde{H})$. Then $D^A X = \lim D^A X_n = 0$ by definition. □

The chain rules from Section 3.3 also hold for the operators D^A: The analogue to Proposition 3.10 can be shown by a similar calculation. All other chain rules use only the closability of the operator and thus the proofs remain valid when D is replaced by D^A.

We define the operator δ^A as adjoint operator of D^A, i.e.

$$\operatorname{dom} \delta^A := \{u \in L^2(\Omega; \tilde{H}) : X \mapsto \langle D^A X, u\rangle_{\tilde{H}} \text{ is continuous for } X \in \mathbb{D}^{1,2,A}\}$$

On this set $\delta^A(u)$ is defined as the unique element of $L^2(\Omega)$ such that

$$E(\langle D^A X, u\rangle_{\tilde{H}}) = E(X\delta^A(u)) \qquad \text{for all } X \in \mathbb{D}^{1,2,A}$$

Theorem 3.27 The operator δ^A equals $\delta \circ A^*$. In particular $\operatorname{dom} \delta^A = (A^*)^{-1}(\operatorname{dom} \delta)$.

Proof. For a smooth random variable X we have $\langle D^A X, u\rangle_{\tilde{H}} = \langle DX, A^* u\rangle_H$. Using Remark 3.22 we have $u \in \operatorname{dom} \delta^A$ if and only if $X \mapsto \langle DX, A^* u\rangle_H$ is continuous on $\mathcal{S}$, which is equivalent to $A^* u \in \operatorname{dom} \delta$. In this case $E(\langle D^A X, u\rangle_{\tilde{H}}) = E(X \cdot \delta(A^* u))$ which proves $\delta^A = \delta \circ A^*$. □

Theorem 3.28 Assume that $X \in \mathbb{D}^{1,2,A}$, $u \in \operatorname{dom} \delta^A$, $Xu \in L^2(\Omega; \tilde{H})$ and $X\delta^A(u) - \langle D^A X, u\rangle_{\tilde{H}} \in L^2(\Omega)$. Then $Xu \in \operatorname{dom} \delta^A$ and

$$\delta^A(Xu) = X \cdot \delta^A(u) - \langle D^A X, u\rangle_{\tilde{H}}$$

Proof. Let $Y \in \mathcal{S}_b$. Then $XY \in \mathbb{D}^{1,2,A}$ and we have

$$\begin{aligned} E(\langle D^A Y, Xu\rangle_{\tilde{H}}) &= E(\langle D^A(XY), u\rangle_{\tilde{H}}) - E(\langle Y \cdot D^A X, u\rangle_{\tilde{H}}) \\ &= E(XY\delta^A(u)) - E(\langle D^A X, u\rangle_{\tilde{H}} Y) \\ &= E((X\delta^A(u) - \langle D^A X, u\rangle_{\tilde{H}}) \cdot Y) \end{aligned}$$

which proves the claim; see Remark 3.22. □

We will now introduce the special case which will be later used in the multidimensional Heston model. Consider the Wiener setting, i.e. the isonormal Gaussian process is given by the Wiener integral over a d-dimensional Brownian motion Z on a fixed time interval $[0, T]$; see (3.1). Let $a \in \mathbb{R}^d$ with $\|a\| = 1$, so that $B := \langle a, Z\rangle$ defines a Brownian motion. Define $A\colon H = L^2([0,T]; \mathbb{R}^d) \to \tilde{H} := L^2([0,T])$ by

$$Af(x) = \langle a, f(x)\rangle_{\mathbb{R}^d} \qquad \text{for } x \in [0,T]$$

In this case we write D^B for the operator D^A and $\mathbb{D}^{1,p,B}$ for $\mathbb{D}^{1,p,A}$. D^B can be interpreted as Malliavin derivative with respect to the Brownian motion B. Random variables which are differentiable in the one-dimensional Malliavin calculus with respect to B (i.e. the isonormal Gaussian process in this calculus is the Wiener integral with respect to B), are also in $\mathbb{D}^{1,p,B}$ and both operators coincide. However, $\mathbb{D}^{1,2,B}$ contains much more elements. If for example $a = e_1$, i.e. $B = Z^1$, then only random variables measurable with respect to Z^1 will be differentiable in the one-dimensional Malliavin calculus with respect to Z^1. On the other hand $\mathbb{D}^{1,p,Z^1}$ contains all random variables X which are measurable with respect to $Z^2, \dots, Z^d$, their derivative being zero: Because $\mathcal{S}$ is dense, X can be approximated using only Wiener integrals with respect to $Z^2, \dots, Z^d$ and Proposition 3.26 shows $D^{Z^1}X = 0$. Also note that for $X \in \mathbb{D}^{1,p}$ the $L^2([0,T];\mathbb{R}^d)$-valued random variable DX can be interpreted as d-dimensional vector with the components $D^{Z^1}, \dots, D^{Z^d}$:

$$(D_r X)_i = D_r^{Z^i} X$$

The adjoint operator of D^B is $\delta^B = \delta \circ A^*$ with $A^*\colon L^2([0,T]) \to L^2([0,T];\mathbb{R}^d)$, $A(h)(t) = h(t) \cdot a$ for $t \in [0,T]$. If $a = e_i$, then the operator $\delta^A = \delta^{Z^i}$ is the Skorohod integral with respect to the i-th Brownian motion.
The following proposition computes the derivative of an Itō integral where the integrand is independent of the Brownian motion with respect to which we are differentiating. An analogous result using the normal Malliavin derivative instead of the partial derivative would require the integral to be differentiable with respect to all Z^i, $i = 1, \dots, d$ and thus require stronger assumptions; see e.g. Proposition 1.3.8 in [Nua06].

Proposition 3.29 Let $i, j \in \{1, \dots, d\}$. Assume that X_t be a process in $L^2(\Omega \times [0,T])$ which is adapted to the Brownian motion Z and independent of the Brownian motion Z^j. Then $\int_0^T X_t dZ_t^i \in \mathbb{D}^{1,p,Z^j}$ and for all $r \in [0,T]$

$$D_r^j \left(\int_0^T X_t dZ_t^i \right) = X_r \cdot \mathbb{1}_{i=j}$$

Proof. Choose simple processes X_t^n that converge against X_t in $L^2(\Omega \times [0,T])$. Assume $X_t^n = \sum_{k=1}^{K_n} A_k^n \mathbb{1}_{[t_k^n, t_{k+1}^n)}$. Then the product rule and the asserted independence give

$$D_r^j \left(\int_0^T X_t^n dZ_t^i \right) = \sum_{k=1}^{K_n} A_k^n \cdot D_r^j (Z_{t_{k+1}^n}^i - Z_{t_k^n}^i) = \sum_{k=1}^{K_n} A_k^n \cdot \mathbb{1}_{[t_k^n, t_{k+1}^n)}(r) \cdot \mathbb{1}_{i=j} = X_r^n \cdot \mathbb{1}_{i=j}$$

The assertion follows because the operator D^j is closed. □

Remark 3.30 Another instance of the general construction presented in this section is the operator D^h, $h \in H$, introduced in (1.33) of [Nua06]. On smooth random variables it is defined as $D^h X = \langle DX, h \rangle_H$. This differs from our operator D^B in that D^B only uses a scalar product over $\mathbb{R}^d$ in $D^B X = \langle DX, a \rangle_{\mathbb{R}^d}$ for $X \in \mathcal{S}$ and thus gives an $L^2([0,T])$-valued random variable, while D^h uses the scalar product of H and thus $D^h X$ is real valued. ◁

3.6 Further Results

This short section states further results from Malliavin calculus which will be used within this thesis. The proofs require much more of the theory of Malliavin calculus than what has been done in the this chapter and can be found in [Nua06]. The first theorem gives an SDE for the Malliavin derivative of an SDE solution — albeit under rather strong conditions on the coefficients.

Theorem 3.31 Let W_t, $t \in [0,T]$, be an m-dimensional Brownian motion and let X_t, $t \in [0,T]$, be the solution of the d-dimensional SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t$$

Assume that both $b\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are continuously differentiable with bounded derivatives (of first order). Then $X^i(t) \in \mathbb{D}^{1,\infty}$ for all $t \in [0,T]$, $i = 1,\dots,d$. The derivative $D_r^j X^i(t)$ satisfies almost everywhere the equation

$$D_r^j X^i(t) = \sigma_{ij}(X(r)) + \sum_{k=1}^{d}\sum_{l=1}^{m}\int_r^t \partial_k \sigma_{il}(X(s)) D_r^j X^k(s) dW^l(s) + \sum_{k=1}^{d}\int_r^t \partial_k b_i(X(s)) D_r^j X^k(s) ds$$

for $r \le t$ and the equation $D_r^j X^i(t) = 0$ for $r > t$.

Proof. This is a simplified version of Theorem 2.2.1 in [Nua06] together with the remark following that theorem. □

Finally, we need a result on the smoothness of densities.

Lemma 3.32 Let μ be a finite measure on $\mathbb{R}^d$. Assume that there exist constants c_i, $i = 1,\dots,d$ such that

$$\left|\int_{\mathbb{R}^d} \partial_i \varphi \, d\mu\right| \le c_i \|\varphi\|_\infty$$

for all $\varphi \in C_b^\infty(\mathbb{R}^d)$. Then μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$.

Proof. This is Lemma 2.1.1 in [Nua06]. □

Chapter 4

Malliavin Calculus in the Heston Model

At various occasions in this thesis we will apply results from Malliavin calculus to the Heston model. In this chapter we lay the foundations: We prove that both processes of the Heston model are (under some conditions) Malliavin differentiable and we calculate the derivatives.

4.1 The Derivative of the Volatility

The standard result to compute the Malliavin derivative of a process given by an SDE is Theorem 3.31. Unfortunately, the theorem requires the coefficients to be globally Lipschitz and thus cannot be applied to neither the volatility nor the price process in the (generalized) Heston model. This forces us to define a sequence of processes with globally Lipschitz coefficients which approximate the Heston model. In fact, we will approximate the transformed model (X_t, σ_t). This procedure was first carried out in [AE08] for the CIR process.
Define the function $f\colon \mathbb{R}_{>0} \to \mathbb{R}$ as

$$f(x) := \kappa\lambda x^{-\frac{\gamma}{1-\gamma}} - \frac{\gamma\theta^2}{2}x^{-1}$$

With this definition the SDE of the transformed process is given by

$$d\sigma_t = (1-\gamma)(f(\sigma_t) - \kappa\sigma_t)dt + \theta(1-\gamma)dW_t \tag{4.1}$$

While for $\gamma = 1/2$ the function f is always positive, it might become negative for $\gamma > 1/2$. However, the limit behaviour $\lim_{x\searrow 0} f(x) = \infty$ and $\lim_{x\to\infty} f(x) = 0$ ensures that there exists a constant $C_f \geq 0$ such that

$$f(x) \geq -C_f x, \qquad x > 0 \tag{4.2}$$

We will also need the following constant:

$$C_f' := \sup_{x>0} f'(x) = \sup_{x>0}\left(-\frac{\kappa\lambda\gamma}{1-\gamma}\cdot x^{-\frac{1}{1-\gamma}} + \frac{\gamma\theta^2}{2}\cdot x^{-2}\right) < \infty \tag{4.3}$$

For each $\varepsilon > 0$ choose a function $f_\varepsilon\colon \mathbb{R} \to \mathbb{R}$ satisfying the following conditions:

1. $f_\varepsilon \in C^\infty(\mathbb{R}, \mathbb{R})$,
2. $f_\varepsilon = f$ on $[\varepsilon, \infty)$,

3. $f_\varepsilon \leq f$ on $(0,\infty)$,
4. $f_\varepsilon(x) \geq -C_f x$ for all $x \geq 0$,
5. $f_\varepsilon'(x) \leq C_f'$ for all $x \in \mathbb{R}$,
6. $|f_\varepsilon'|$ is bounded on $\mathbb{R}$,
7. $f_\varepsilon'(x) \leq 0$ for all $x \leq 0$.

The approximating processes are now defined by replacing f by f_ε in the SDE (4.1): For $\varepsilon > 0$ let $(\sigma_t^\varepsilon)_t$ be the solution of the SDE

$$d\sigma_t^\varepsilon = (1-\gamma)(f_\varepsilon(\sigma_t^\varepsilon) - \kappa\sigma_t^\varepsilon)dt + \theta(1-\gamma)dW_t \tag{4.4}$$

This SDE satisfies the usual global Lipschitz and linear growth conditions and thus possesses a unique strong solution.
In fact, σ_t^ε not only approximates σ_t but eventually more and more paths coincide as $\varepsilon \to 0$.

Proposition 4.1 There exists a set $\Omega^* \subset \Omega$ with $P(\Omega^*) = 1$ and such that $\sigma_t^\varepsilon(\omega) = \sigma_t(\omega)$ for all $t \in [0,T]$, $\omega \in \Omega^*$, and $\varepsilon < \varepsilon^*(\omega) := \inf_{t\in[0,T]} \sigma_t(\omega)$. In particular $\sigma_t^\varepsilon \to \sigma_t$ almost surely as $\varepsilon \to 0$.

Proof. Let $\Omega^* \subset \Omega$ be a set of measure one such that for all $\omega \in \Omega^*$

- the integral form of the SDE (4.1) holds,
- $\sigma_t(\omega) > 0$ for all $t \in [0,T]$,
- and the integral form of (4.4) holds for all $\varepsilon > 0$.

To prove the existence of such a set one can adapt the proof of the usual existence and uniqueness theorem for SDEs (e.g. Theorem 2.9 in [KS10]) to the case of infinitely many SDEs with the same constant diffusion term.
Let $\omega \in \Omega^*$. From the integral form of the SDEs (4.1) and (4.4) we get

$$|\sigma_t^\varepsilon(\omega) - \sigma_t(\omega)| \leq (1-\gamma)\int_0^t |f_\varepsilon(\sigma_\tau^\varepsilon(\omega)) - f(\sigma_\tau(\omega))|d\tau + (1-\gamma)\kappa\int_0^t |\sigma_\tau^\varepsilon(\omega) - \sigma_\tau(\omega)|d\tau$$

For all $0 < \varepsilon < \varepsilon^*(\omega)$, $t \in [0,T]$ we have $f_\varepsilon(\sigma_t(\omega)) = f(\sigma_t(\omega))$ and thus

$$|\sigma_t^\varepsilon(\omega) - \sigma_t(\omega)| \leq \int_0^t (1-\gamma)(L_\varepsilon + \kappa)\,|\sigma_\tau^\varepsilon(\omega) - \sigma_\tau(\omega)|\,d\tau, \qquad t \in [0,T],$$

where L_ε is a Lipschitz constant for f_ε. By Gronwall's lemma it follows that $\sigma_t(\omega) = \sigma_t^\varepsilon(\omega)$ for all $t \in [0,T]$. □

Despite the last proposition, it is still possible for σ_t^ε to become negative. The proposition makes sure that this is only possible if σ_t falls below ε.
To prove L^p-convergence of σ_t^ε to σ_t we need the following pathwise bound:

Lemma 4.2 Let u_t be the Ornstein–Uhlenbeck process defined by the SDE

$$du_t = -(1-\gamma)(C_f + \kappa)u_t dt + \theta(1-\gamma)dW_t, \qquad u_0 = \sigma_0 \tag{4.5}$$

Then $P(u_t \leq \sigma_t^\varepsilon \leq \sigma_t \ \ \forall\, t \in [0,T]) = 1$ for all $\varepsilon > 0$.

Proof. The main tool in this proof is the Yamada–Watanabe comparison lemma, Theorem A.1, which allows to compare the solutions of two SDEs with the same diffusion coefficient but different drift coefficients. The first part of the claim, $P(u_t \leq \sigma_t^\varepsilon) = 1$ follows directly from the theorem because the drift coefficient of (4.5) is smaller than that of (4.4). Unfortunately, the comparison lemma requires the drift coefficients to be continuously defined on the whole of $\mathbb{R}$, which is not possible for the SDE of σ. We will still be able to prove the second part using the comparison lemma after transforming the SDEs back: Define the process $\nu_t^\varepsilon := (\sigma_t^\varepsilon)_+^{1/(1-\gamma)}$. Itō's formula shows that ν_t^ε follows the SDE

$$d\nu_t^\varepsilon = \left((\nu_t^\varepsilon)^\gamma f_\varepsilon((\nu_t^\varepsilon)^{1-\gamma}) - \kappa\nu_t^\varepsilon + \frac{\theta^2\gamma}{2}(\nu_t^\varepsilon)^{2\gamma-1} \right) dt + \theta(\nu_t^\varepsilon)^\gamma dW_t$$

Now the comparison lemma can be applied to ν_t^ε and ν_t: The diffusion coefficients coincide and satisfy condition (ii) in Theorem A.1 using $h(x) = \theta x^\gamma$. The drift coefficient of ν_t^ε is smaller than that of ν_t because $f_\varepsilon \leq f$:

$$x^\gamma \cdot f_\varepsilon(x^{1-\gamma}) - \kappa x + \frac{\theta^2\gamma}{2}x^{2\gamma-1} \leq x^\gamma \cdot f(x^{1-\gamma}) - \kappa x + \frac{\theta^2\gamma}{2}x^{2\gamma-1} = \kappa(\lambda - x)$$

We have established $P(\nu_t^\varepsilon \leq \nu_t) = 1$. By applying the transformation $x \mapsto x^{1-\gamma}$ again we get $P((\sigma_t^\varepsilon)_+ \leq \sigma_t)$. □

Proposition 4.3 Let $p \geq 1$. Then

$$\sup_{t\in[0,T]} |\sigma_t^\varepsilon - \sigma_t| \to 0$$

in $L^p(\Omega)$ as $\varepsilon \to 0$.

Proof. The claim follows from the almost-sure convergence established in Proposition 4.1, if we can provide an integrable dominating function. For this we use the previous lemma:

$$\sup_{t\in[0,T]} |\sigma_t^\varepsilon - \sigma_t| \leq \sup_{t\in[0,T]} |\sigma_t - u_t| \leq \sup_{t\in[0,T]} \sigma_t + \sup_{t\in[0,T]} u_t$$

$\sup \sigma_t \in L^p(\Omega)$ follows from Theorems 2.1 and 2.4. It is well-known that $\sup_{t\in[0,T]} u_t \in L^p(\Omega)$; see e.g. Example 6.8 in [KS10]. □

We are now ready to prove the Malliavin differentiability first of σ_t^ε, then of σ_t, finally of ν_t. Because all these processes depend only on one Brownian motion, one-dimensional Malliavin calculus over W_t (with $H := L^2([0,T])$) is sufficient.

Proposition 4.4 Let $t \in [0,T]$. Then $\sigma_t^\varepsilon \in \mathbb{D}^{1,\infty}$ with derivative

$$D_r\sigma_t^\varepsilon = \theta(1-\gamma) \cdot \exp\left(\int_r^t (1-\gamma)(f_\varepsilon'(\sigma_s^\varepsilon) - \kappa)\, ds \right) \cdot \mathbb{1}_{[0,t]}(r)$$

Moreover this derivative is bounded by

$$\sup_{r,t\in[0,T]} |D_r\sigma_t^\varepsilon| \leq \theta(1-\gamma) \cdot \exp(T(1-\gamma) \cdot C_f')$$

Proof. Thanks to the Lipschitz continuity of the SDE coefficients of σ_t^ε we can apply Theorem 3.31 and learn that its derivative follows the integral equation

$$D_r\sigma_t^\varepsilon = \theta(1-\gamma) + \int_r^t (f_\varepsilon'(\sigma_s^\varepsilon) - \kappa) \cdot D_r\sigma_s^\varepsilon ds$$

for $r \leq t$. Its solution is exactly the asserted expression for $D_r\sigma_t^\varepsilon$. Since one of the requirements for f_ε was that $f_\varepsilon'(x) \leq C_f'$, the bound follows trivially. □

Proposition 4.5 We have $\sigma_t \in \mathbb{D}^{1,\infty}$. The derivative is

$$D_r\sigma_t = \theta(1-\gamma) \cdot \exp\left(\int_r^t (1-\gamma)(f'(\sigma_s) - \kappa)ds\right) \cdot \mathbb{1}_{[0,t]}(r) \tag{4.6}$$

Furthermore

$$\sup_{r,t\in[0,T]} |D_r\sigma_t^\varepsilon - D_r\sigma_t| \to 0$$

as $\varepsilon \to 0$ almost surely and in $L^p(\Omega)$ for all $p \geq 1$.

Proof. For the moment, call the proposed derivative $\sigma'_{r,t}$:

$$\sigma'_{r,t} := \theta(1-\gamma) \cdot \exp\left(\int_r^t (1-\gamma)(f'(\sigma_s) - \kappa)ds\right) \cdot \mathbb{1}_{[0,t]}(r)$$

Proposition 4.1 proves that there exists a set $\Omega^* \subset \Omega$ with $P(\Omega^*) = 1$ and $\sigma_t^\varepsilon(\omega) = \sigma_t(\omega)$ for all $\omega \in \Omega^*$ and $\varepsilon \leq \varepsilon^*(\omega) := \inf_{s\in[0,T]} \sigma_s(\omega)$. But then also $D_r\sigma_t^\varepsilon(\omega) = \sigma'_{r,t}(\omega)$. This proves the asserted almost-sure convergence. The L^p-convergence follows from the uniform bound for $D_r\sigma_t^\varepsilon$ established in the previous proposition. Because the Malliavin derivative is a closable operator, this implies $\sigma_t \in \mathbb{D}^{1,\infty}$ with $D_r\sigma_t = \sigma'_{r,t}$. □

Note that in the standard Heston model the derivative is given by

$$D_r\sigma_t = \frac{\theta}{2} \cdot \exp\left(-\int_r^t \left(\frac{\kappa\lambda}{2} - \frac{\theta^2}{8}\right)\sigma_s^{-2} + \frac{\kappa}{2}\, ds\right) \cdot \mathbb{1}_{[0,t]}(r)$$

Theorem 4.6 We have $\nu_t \in \mathbb{D}^{1,\infty}$ with derivative

$$\begin{aligned} D_r\nu_t &= \frac{1}{1-\gamma} \cdot \sigma_t^{\frac{\gamma}{1-\gamma}} \cdot D_r\sigma_t \\ &= \theta\nu_t^\gamma \cdot \exp\left(\int_r^t \frac{\theta^2\gamma(1-\gamma)}{2} \cdot \nu_s^{-2(1-\gamma)} - \kappa\lambda\gamma\nu_s^{-1} + -\kappa(1-\gamma)\, ds\right) \cdot \mathbb{1}_{[0,t]}(r) \end{aligned}$$

Proof. This follows directly from the chain rule Corollary 3.18. □

4.2 The Derivative of the Price Process

Similar to the volatility process, the SDE of the log-price process,

$$dX_t = \left(\mu - \frac{1}{2}\sigma_t^{\frac{1}{1-\gamma}}\right) dt + \sigma_t^{\frac{1}{2(1-\gamma)}} dB_t$$

does not have globally Lipschitz coefficients. To compute the derivative, we again need to approximate X_t by the solutions of SDEs with suitable coefficient functions. To ensure that the whole system of SDEs has globally Lipschitz coefficients, we must in particular replace the process σ_t by the approximation σ_t^ε from the previous section. For $\varepsilon > 0$ choose a function $h_\varepsilon : \mathbb{R} \to \mathbb{R}$ satisfying the following properties:

1. h_ε is bounded and continuously differentiable,
2. $h_\varepsilon(x) = x$ on $[0, \varepsilon^{-1}]$,
3. $|h_\varepsilon(x)| \leq |x|$ for all $x \in \mathbb{R}$,
4. $|h'_\varepsilon(x)| \leq 1$ for all $x \in \mathbb{R}$.

Now define the process X_t^ε as

$$X_t^\varepsilon = x_0 + \int_0^t \mu - \frac{1}{2} h_\varepsilon^{\frac{1}{1-\gamma}}(\sigma_s^\varepsilon)\,ds + \int_0^t h_\varepsilon^{\frac{1}{2(1-\gamma)}}(\sigma_s^\varepsilon)\,dB_t$$

Lemma 4.7 For each $p, q \geq 1$ the suprema

$$\sup_{t\in[0,T]} |\sigma_t^q - h_\varepsilon(\sigma_t^\varepsilon)^q| \qquad \text{and} \qquad \sup_{t\in[0,T]} |\sigma_t^q - h_\varepsilon^q(\sigma_t^\varepsilon)\cdot h_\varepsilon'(\sigma_t^\varepsilon)|$$

converge to 0 almost surely and in $L^p(\Omega)$ as $\varepsilon \to 0$.

Proof. Almost sure convergence follows from Proposition 4.1: If ε is small enough such that

$$\varepsilon \leq \inf_{t\in[0,T]} \sigma_t(\omega) \leq \sup_{t\in[0,T]} \sigma_t(\omega) \leq \frac{1}{\varepsilon}$$

then $h_\varepsilon(\sigma_t^\varepsilon(\omega)) = \sigma_t^\varepsilon(\omega) = \sigma_t(\omega)$ and $h_\varepsilon'(\sigma_t^\varepsilon(\omega)) = 1$, $t \in [0,T]$, and thus both suprema are 0. L^p-convergence holds by dominated convergence because both suprema are bounded by $2\sup_{t\in[0,T]} |\sigma_t|^q \in L^p(\Omega)$; see Theorems 2.1 and 2.4. □

Proposition 4.8 $\sup_{t\in[0,T]} |X_t^\varepsilon - X_t| \to 0$ in $L^p(\Omega)$ for all $p \geq 1$.

Proof. The SDEs in integral form and the Burkholder–Davis–Gundy inequality yield

$$\begin{aligned}\mathbb{E} \sup_{t\in[0,T]} |X_t^\varepsilon - X_t|^p \leq c\cdot \mathbb{E} \sup_{t\in[0,T]} \left|\int_0^t \sigma_s^{\frac{1}{1-\gamma}} - h_\varepsilon^{\frac{1}{1-\gamma}}(\sigma_s^\varepsilon)\,ds\right|^p \\ + c\cdot \mathbb{E} \sup_{t\in[0,T]} \left|\int_0^t \left(\sigma_s^{\frac{1}{2(1-\gamma)}} - h_\varepsilon^{\frac{1}{2(1-\gamma)}}(\sigma_s^\varepsilon)\right)^2 ds\right)^{p/2}\end{aligned}$$

for some $c \geq 0$. Because $|x-y|^2 \leq |x^2 - y^2|$ for $x, y \geq 0$, the assertion now follows from the previous lemma. □

Theorem 4.9 The log-price X_t is in $\mathbb{D}^{1,p}$ for all $p \geq 1$. The derivative is given by

$$\begin{aligned}D_r^1 X_t &= \rho'\sqrt{v_r}\cdot \mathbb{1}_{[0,t]}(r)\\ D_r^2 X_t &= \left(\rho\sqrt{v_r} - \frac{1}{2(1-\gamma)}\cdot\int_0^t \sigma_s^{\frac{\gamma}{1-\gamma}} D_r\sigma_s ds \right.\\ &\qquad \left. + \frac{1}{2(1-\gamma)}\cdot\int_0^t \sigma_s^{\frac{1}{2(1-\gamma)}-1} D_r\sigma_s\,dB_s\right)\cdot \mathbb{1}_{[0,t]}(r)\end{aligned}$$

For each $r \in [0,T]$ and $i \in \{1,2\}$ the derivative converges uniformly:

$$\sup_{t\in[0,T]} |D_r^i X_t - D_r^i X_t^\varepsilon| \to 0$$

in $L^p(\Omega)$.

Proof. First we apply Theorem 3.31 to the system of SDEs with globally Lipschitz coefficients $(X_t^\varepsilon, \sigma_t^\varepsilon)$. For $r \leq t$ the result is

$$\begin{aligned}D_r^1 X_t^\varepsilon &= \rho' h_\varepsilon^{\frac{1}{2(1-\gamma)}}(\sigma_r^\varepsilon)\\ D_r^2 X_t^\varepsilon &= \rho h_\varepsilon^{\frac{1}{2(1-\gamma)}}(\sigma_r^\varepsilon) - \frac{1}{2(1-\gamma)}\cdot\int_r^t h_\varepsilon^{\frac{\gamma}{1-\gamma}}(\sigma_s^\varepsilon) h_\varepsilon'(\sigma_s^\varepsilon) D_r\sigma_s^\varepsilon\,ds\\ &\qquad + \frac{1}{2(1-\gamma)}\cdot\int_r^t h_\varepsilon^{\frac{1}{2(1-\gamma)}-1}(\sigma_s^\varepsilon) D_r\sigma_s^\varepsilon\,dB_s\end{aligned}$$

By Lemma 4.7 the first derivative as well as the first term of the second derivative converge to $\rho'\sqrt{v_r}$ and $\rho\sqrt{v_r}$, respectively. To show convergence of the second term of the second derivative we use the following inequality

$$\begin{aligned}
&\sup_{t\in[0,T]}\left|\int_r^t h_\varepsilon^{\frac{\gamma}{1-\gamma}}(\sigma_s^\varepsilon)h_\varepsilon'(\sigma_s^\varepsilon)D_r\sigma_s^\varepsilon - \sigma_s^{\frac{\gamma}{1-\gamma}}D_r\sigma_s\,ds\right| \\
&\le T\cdot\sup_{s\in[0,T]}\left|h_\varepsilon^{\frac{\gamma}{1-\gamma}}(\sigma_s^\varepsilon)h_\varepsilon'(\sigma_s^\varepsilon) - \sigma_s^{\frac{\gamma}{1-\gamma}}\right|\cdot\sup_{s\in[0,T]}|D_r\sigma_s^\varepsilon| \\
&\quad + T\cdot\sup_{s\in[0,T]}\left|\sigma_s^{\frac{\gamma}{1-\gamma}}\right|\cdot\sup_{s\in[0,T]}|D_r\sigma_s^\varepsilon - D_r\sigma_s|
\end{aligned}$$

Hölder's inequality, Lemma 4.7, Proposition 4.5 and Theorem 2.1 prove that this converges to 0 in $L^p(\Omega)$. After applying the Burkholder–Davis–Gundy inequality, the third term of the second derivative can be shown to converge to 0 analogously. □

The chain rule Corollary 3.18 can be used to compute the derivative of the actual price process as $D_r^i S_t = S_t\cdot D_r^i X_t$, provided both S_t and $S_t\cdot D_r^i X_t$ are L^p-integrable (see Theorem 2.6 for the moments). Because we will only need derivatives of the log-price later, we omit the details.

4.3 Higher Derivatives

For completeness, we will briefly show how to compute higher order Malliavin derivatives of the square root CIR process. Higher order derivatives are defined by an iterative procedure; see e.g. [Nua06]. Similar to the usual derivative, the domain of the N-the derivative depends on the chosen L^p-norm and is denoted by $\mathbb{D}^{N,p}$. The n-th derivative is a random variable with values in $L^p(\Omega;H^{\otimes n})$.
Because we will not need these derivatives later, we restrict ourselves to the standard Heston model (i.e. $\gamma = 1/2$) and the process σ_t.

Theorem 4.10 Let $2\le N\in\mathbb{N}$, $p\ge 1$ and $t\in[0,T]$ and assume

$$\nu > 3(N-1)\frac{p}{2}+1$$

Then $\sigma_t\in\mathbb{D}^{N,p}$.

Proof. We will define classes $A_n^p\subset L^p(\Omega;H^{\otimes n})$ of random variables such that $D\sigma_t\in A_1^p$ and

$$X\in A_n^p, n<N \Longrightarrow X\in\mathbb{D}^{1,p}(H^{\otimes n}) \text{ and } DX\in A_{n+1}^p \tag{4.7}$$

This implies the assertion.
Define A_1^p as the class of random variables $X\in L^p(\Omega;H)$ that can be written as

$$X\colon r\mapsto F_r\left(\int_{a(r)}^{b(r)}\sigma_u^{-2}du\right)$$

such that

- a and b are measurable with $0\le a(r)\le b(r)\le t$ for all $r\in[0,T]$,
- $[0,T]\times\mathbb{R}_+\ni(r,x)\mapsto F_r(x)$ is measurable,
- For each $r\in[0,T]$ the function F_r is infinitely often differentiable,
- $F_r(x)$ is bounded over all $(r,x)\in[0,T]\times\mathbb{R}_+$ and the same holds for all partial derivatives (w.r.t. x) of all orders of F_r.

For $n \geq 2$ define $\tilde{A}_n^p$ as the class of random variables $X \in L^p(\Omega; H^{\otimes n})$ that can be written as

$$X\colon (r_1, \dots, r_n) \mapsto \int_{[0,t]^d} F_{r,s} \left(\int_{a_1(r)}^{b_1(r)} \sigma_u^{-2} du, \dots, \int_{a_k(r)}^{b_k(r)} \sigma_u^{-2} du \right) \cdot \sigma_{s_1}^{-m_1} \cdots \sigma_{s_d}^{-m_d} ds$$

such that

- $d, k \in \mathbb{N}$,
- a_i and b_i are measurable with $0 \leq a_i(r) \leq b_i(r) \leq t$ for all $r \in [0,T]$ for all $i = 1, \dots, k$,
- $\sum_{i=1}^d m_i \leq 3(n-1)$,
- $[0,T]^d \times [0,t]^d \times \mathbb{R}_+^k \ni (r,s,x) \mapsto F_{r,s}(x)$ is measurable,
- for all $r \in [0,T]^n$, $s \in [0,t]^d$, the function $F_{r,s}$ is infinitely often differentiable,
- $F_{r,s}(x)$ is bounded over all $(r,s,x) \in [0,T]^n \times [0,T]^d \times \mathbb{R}_+^k$ and the same holds for all partial derivatives (w.r.t. x) of all orders of $F_{r,s}$.

Finally set $A_n^p := \mathrm{Lin}(\tilde{A}_n^p)$.
Abbreviate $C_\sigma = \kappa\lambda/2 - \theta^2/8$. In Proposition 4.5 we have proven that $\sigma_t \in \mathbb{D}^{1,p}$ and that

$$D_r\sigma_t = \frac{\theta}{2} \cdot \exp\left(\int_r^t -\frac{\kappa}{2} - \frac{C_\sigma}{\sigma_s^2} ds \right) \cdot \mathbb{1}_{[0,t]}(r) \tag{4.8}$$

$$= \frac{\theta}{2} \cdot \exp\left(-(t-r)\frac{\kappa}{2} \right) \cdot \exp\left(-C_\sigma \int_r^t \sigma_u^{-2} du \right) \tag{4.9}$$

Because $x \mapsto \theta/2 \cdot \exp(-(t-r)\kappa/2) \cdot \exp(-C_\sigma x)$ is bounded and infinitely often differentiable, we have $D\sigma_t \in A_1^p$ for all $p \geq 1$.
For the proof of (4.7) we will restrict ourselves to the case $n \geq 2$ as the case $n = 1$ is analogous and even easier. So assume $X \in \tilde{A}_n^p$ is given in the form specified above. Then for all $r' \in [0,T]$ we can compute the derivative using the chain rule. The use of the chain rule is justified because the result is in $L^p(\Omega; H^{\otimes n+1})$ because

$$\sum_{i=1}^n m_n + 3 \leq 3(n-1) + 3 \leq 3(N-1)$$

and thus thanks to the assumption enough inverse moments of σ exist; see Theorem 2.1. The result is

$$\begin{aligned} D_{r'}X &= \int_s \left(\sum_{i=1}^k \partial_i F_{r,s}(\cdots) \cdot \int_{a_i(r)}^{b_i(r)} \frac{-2}{\sigma_u^3} \cdot D_{r'}\sigma_u du + F_{r,s}(\cdots) \sum_{i=1}^n \frac{-m_i}{\sigma_{s_i}} \right) \cdot \sigma_{s_1}^{-m_1} \cdots \sigma_{s_n}^{-m_n} ds \\ &= \sum_{i=1}^k \int_{s \in [0,T]^d} \int_{a_i(r)}^{b_i(r)} -2\partial_i F_{r,s}(\cdots) \cdot D_r\sigma_u \cdot \sigma_{s_1}^{-m_1} \cdots \sigma_{s_n}^{-m_n} \cdot \sigma_u^{-3} du ds \\ &\quad + \sum_{i=1}^n \int_{s \in [0,T]^d} -m_i F_{r,s}(\cdots) \sigma_{s_1}^{-m_1} \cdots \sigma_{s_n}^{-m_n} \cdot \sigma_{s_i}^{-1} ds \end{aligned}$$

Using equation (4.9) it can be seen that this result is in A_{n+1}^p. □

Chapter 5

Multidimensional Models

In practice, one will often have to compute option prices in models containing several correlated stocks. This section introduces such a model. The exact formulas derived in Section 2.4 are only available for a single Heston model, so we will need to resort to Monte-Carlo methods to compute option prices in the multidimensional models.
The multidimensional model consists of d single models which are correlated only via the underlying Brownian motions. For $i = 1, \dots, d$ let (B_t^i) and (W_t^i) be Brownian motions on $[0, T]$ with correlation matrix

$$\Sigma^{(B,W)} = \begin{pmatrix} (E(B_1^i B_1^j))_{i,j=1,\dots,d} & (E(B_1^i W_1^j))_{i,j=1,\dots,d} \\ (E(W_1^i B_1^j))_{i,j=1,\dots,d} & (E(W_1^i W_1^j))_{i,j=1,\dots,d} \end{pmatrix}$$

Each model has its own group of parameters $\mu_i \in \mathbb{R}$, $\kappa_i, \lambda_i, \theta_i, s_0^i, v_0^i > 0$ and $\gamma_i \in [1/2, 1)$. Because the correlation is determined via $\Sigma^{(B,W)}$ there is no need for the parameter ρ anymore.

Definition 5.1 The d-dimensional Heston model consists of the $2d$ processes S^i, v^i that solve the following SDEs with start values s_0^i and v_0^i.

$$\begin{pmatrix} dS_t^i \\ dv_t^i \end{pmatrix} = \begin{pmatrix} \mu_i S_t^i \\ \kappa_i(\lambda_i - v_t^i) \end{pmatrix} dt + \begin{pmatrix} \sqrt{v_t^i} S_t^i & 0 \\ 0 & \theta_i (v_t^i)^{\gamma_i} \end{pmatrix} \begin{pmatrix} dB_t^i \\ dW_t^i \end{pmatrix}, \qquad i = 1, \dots, d \tag{5.1}$$

As in the one-dimensional model we define the transformed processes $X_t^i := \log(S_t^i)$ and $\sigma_t^i := (v_t^i)^{1-\gamma_i}$ and denote their start values by x_0^i and σ_0^i, respectively. ◁

This definition of the multidimensional model has the advantage that all properties of a one-dimensional model established earlier in this thesis are still true for the one-dimensional submodels of the multidimensional model.
We will always make the following assumption:

Assumption 5.2

- For each $i = 1, \dots, d$ we have either $\nu_i := 2\kappa_i\lambda_i/\theta_i^2 > 1$ or $\gamma_i > 1/2$.
- The matrix $\Sigma^{(B,W)}$ is positive definite. ◁

As discussed above, the first assumption guarantees positivity of v_t^i and the existence of some moments. The second assumption allows us to write

$$\begin{pmatrix} B \\ W \end{pmatrix} = R \cdot Z$$

with an upper triangular $2d \times 2d$-matrix R and a $2d$-dimensional standard Brownian motion Z:

Lemma 5.3 If (W_t^i) is an n-dimensional Brownian motion with positive definite correlation matrix $\Sigma^W := (E(W_1^i W_1^j))_{i,j=1,\dots,n}$, then there exists an upper triangular $n \times n$-matrix R and a n-dimensional standard Brownian motion (Z_t^i) such that $W = RZ$.

Proof. The following transformation reflects each component of a matrix through the central point of the matrix:

$$g\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}, \qquad (g(A))_{ij} := A_{n+1-i,n+1-j}, \qquad i,j = 1,\dots,n$$

Note that $g \circ g = \mathrm{id}$, $g(A^T) = g(A)^T$ and $g(A \cdot B) = g(A) \cdot g(B)$, because

$$g(AB)_{ij} = \sum_{k=1}^{n} A_{n+1-i,k} B_{k,n+1-j} = \sum_{k=1}^{n} A_{n+1-i,n+1-k} B_{n+1-k,n+1-j} = (g(A) \cdot g(B))_{ij}$$

Because $(g(\Sigma^W))_{ij} = (E(W_1^{n+1-i} W_1^{n+1-j}))$ the matrix $g(\Sigma^W)$ is the covariance matrix of the Brownian motion in reversed order $(W_t^n, \dots, W_t^1)$ and in particular also positive definite. Let $LL^T = g(\Sigma^W)$ be its Cholesky decomposition and define $R := g(L)$. Then R is an upper triangular matrix and

$$RR^T = g(L)g(L)^T = g(LL^T) = g(g(\Sigma^W)) = \Sigma^W$$

Thus the covariance matrix of $Z := R^{-1}W$ is

$$\Sigma^Z = R^{-1}\Sigma^W (R^{-1})^T = R^T (R^T)^{-1} \cdot R^{-1}\Sigma^W (R^{-1})^T = R^T (\Sigma^W)^{-1} \Sigma^W (R^T)^{-1} = \mathrm{Id}$$

and hence Z is a standard n-dimensional Brownian motion. □

This lemma has two important consequences: In practice, the covariance matrix is often fixed and one will need to construct the matrix R from the proof to generate the correlated Brownian motions from a standard Brownian motion that can be simulated directly. And in theory it allows to use multidimensional Malliavin calculus build on the Wiener integral with respect to the independent Brownian motions $Z^1, \dots, Z^{2d}$; see (3.1). Remember that for $i = 1, \dots, 2d$ we write D^i for the i-th component of the Malliavin derivative and that this derivative is interpreted as Malliavin derivative w.r.t. Z^i.
In order to prove the integration by parts rule Theorem 8.2, we need to find one derivative of the log-price which is particularly simple, because otherwise it will not be possible to compute the arising Skorohod integral. The trick will be to use the Malliavin derivative with respect to Z^1, because thanks to the triangular form of R, this Brownian motion appears only in S^1 and is independent of the other price processes and all volatility processes.
The following lemma makes this precise. It uses the partial Malliavin derivatives introduced in Section 3.5.

Lemma 5.4 For each log-price process X^i there exists a Brownian motion $\tilde{Z}^i$ and a constant $c_i \in \mathbb{R} \setminus \{0\}$ such that

- $D_r^{\tilde{Z}^i} X^i = c_i \cdot \sqrt{v_r^i} \cdot \mathbb{1}_{[0,t]}(r)$,
- $\tilde{Z}^i$ is independent from all other log-price processes X^j, $i \neq j$, and all volatility processes v^i, $i = 1, \dots, 2d$ (of course, this implies that $D^{\tilde{Z}^i}$ is zero for all these processes).

For $i = 1$ we can choose $\tilde{Z}^1 = Z^1$ and $c_i = R_{11}$.

Proof. First consider the case $i = 1$. For $j = 1, \dots, d$ we can compute the derivative of X_t^j using that Z^1 is independent from the volatility processes and Proposition 3.29: For $0 \le r \le t$ we have

$$D_r^{Z^1} X_t^j = D_r^{Z^1} \left(x_0^j + \int_0^t \mu_j - \frac{v_s^j}{2} ds + \sum_{k=1}^{2d} R_{j,k} \int_0^t \sqrt{v_s^j} dZ_s^k \right) = R_{j,1} \sqrt{v_r^j}$$

Now the result follows because $R_{j,1}$ is zero for $j > 1$.
The case $i > 1$ can be proven analogously if we can find a decomposition $(B, W)^\top = \tilde{R}\tilde{Z}$ with a 2d-standard Brownian motion $\tilde{Z}$ and a matrix $\tilde{R}$ whose first column has only a single entry c_i in the i-th column. To get this decomposition we reorder the Brownian motions B and W so that B^i is first, apply Lemma 5.3 to get an upper triangular matrix and return to the original order. □

A major tool in the derivation of the integration by parts formula for the Heston model (Theorem 8.2) will be the Malliavin chain rule. For discontinuous functionals — which are our main interest — the involved functions might only be Lipschitz continuous instead of being continuously differentiable. Before we can apply the chain rule, we thus need to verify that the multidimensional model has an absolutely continuous distribution; see Assumption 3.13.

Theorem 5.5 Let $t \in [0, T]$. Then the law of the random vector X_t is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$.

Proof. By Lemma 3.32 a sufficient condition for the assertion is that there exists a constant c such that

$$|E(\partial_i \varphi(X_t))| \le c \cdot \frac{1}{t} \cdot \|\varphi\|_\infty$$

for all $\varphi \in C_b^\infty(\mathbb{R}^d, \mathbb{R})$ and each $i = 1, \dots, d$.
Let $i \in \{1, \dots, d\}$. Use Lemma 5.4 to choose a Brownian motion $\tilde{Z}^i$ that is independent of all log-price and volatility processes except X^i and a constant c_i such that $D_r^{\tilde{Z}^i} X_t^i = c_i \cdot \sqrt{v_r^i} \cdot \mathbb{1}_{[0,t]}(r)$. For the rest of the proof we write D for the derivative $D^{\tilde{Z}^i}$. Applying the chain rule yields

$$D_r \varphi(X_t) = \partial_i \varphi(X_t) \cdot D_r X_t^i$$

because $D_r X_t^j = 0$ for $j \neq i$. We use this in the following calculation.

$$\begin{aligned} |E(\partial_i \varphi(X_t))| &= \frac{1}{t} \cdot \left| E\left(\int_0^t \partial_i \varphi(X_t) \cdot D_r X_t^i \cdot \frac{1}{D_r X_t^i} dr \right) \right| \\ &= \frac{1}{t c_i} \cdot \left| E\left(\int_0^t D_r \varphi(X_t) \cdot \frac{1}{\sqrt{v_r^i}} dr \right) \right| \end{aligned}$$

Now we apply the integration by parts rule. Because the integrand of the arising Skorohod integral is adapted, we can replace it by an Itō integral.

$$\begin{aligned} &= \frac{1}{t c_i} \cdot \left| E\left(\varphi(X_t) \int_0^t \frac{1}{\sqrt{v_r^i}} d\tilde{Z}_r^i \right) \right| \\ &\le \frac{1}{t c_i} \cdot \|\varphi\|_\infty \cdot E\left| \int_0^t \frac{1}{\sqrt{v_r^i}} d\tilde{Z}_r^i \right| \end{aligned}$$

The expectation is finite by Theorems 2.1 and 2.4. □

Chapter 6

Numerical Approximation

The literature presents a great number of approximation schemes for SDE solutions; see e.g. [KP92]. Most of these schemes do not preserve the positivity of the CIR process. Due to the square root in the SDE of this process, they cannot be applied to the Heston model without modification. The price process, on the other hand, presents no difficulties, because in most cases it suffices to approximate the log-price which can be given explicitly in terms of the variance process. Therefore most approximation schemes that are proposed for the Heston model are mainly schemes for the CIR process, complemented by a simple Euler method for the price process.
When discussing schemes we will always denote a scheme by the name of the process which is approximated and a hat to distinguish the two, i.e. $\hat{x}$ approximates $(x_t)_t$. We will assume to have a predefined endtime T and stepsize Δ, so that the scheme is indexed by $n = 0, \dots, N := \lceil T/\Delta \rceil$ and the n-th step approximates the process at time $n\Delta$. In order to simplify comparisons between the actual process and the approximation we will sometimes abuse notation and use time indexes: We set $\hat{x}_{k\Delta} = \hat{x}_k$ for $k = 0, \dots, N$ and $\hat{x}_T = \hat{x}_N$. Finally, we abbreviate the necessary Brownian increments by $\Delta_n W = W_{(n+1)\Delta} - W_{n\Delta}$, for an arbitrary Brownian motion W.
We will restrict ourselves to a single (generalized) Heston model. Because the multidimensional model introduced in Chapter 5 couples the model only via the correlation of the underlying Brownian motions, all schemes presented in this chapter can trivially be extended to the multidimensional case.

6.1 The Euler Scheme for the CIR/CEV Process

A first difficulty when approximating the CIR/CEV process is the γ-th power in the SDE: A numerical scheme must either make sure to never fall below zero or replace the power function by a function defined on the whole of $\mathbb{R}$. The obvious choices in the latter case lead to the numerical schemes:

$$\hat{v}_{n+1} = \hat{v}_n + \kappa(\lambda - \hat{v}_n)\Delta + \theta(\hat{v}_n^+)^\gamma \Delta_n W$$
$$\hat{v}_{n+1} = \hat{v}_n + \kappa(\lambda - \hat{v}_n)\Delta + \theta|\hat{v}_n|^\gamma \Delta_n W$$

Both schemes allow the approximation to become negative but nevertheless converge strongly in L^2, i.e.

$$E \max_{n=0,\dots,\lceil T/\Delta \rceil} |\hat{v}_n - v_{n\Delta}|^2 \to 0$$

as $\Delta \to 0$; see [DD98] and [HM05]. A pathwise convergence rate was established in [Gy98]: If $\nu \geq 1$ or $\gamma > 1/2$, then for each $\varepsilon > 0$ the convergence

$$\frac{1}{\Delta^{1/2-\varepsilon}} \cdot |\hat{v}_n - v_{n\Delta}| \to 0 \tag{6.1}$$

holds almost surely for $\Delta \to 0$.
Because the integration by parts formula of Theorem 8.2 requires to divide by the volatility, we will be only interested in schemes that never reach zero. A simple candidate is the symmetrized Euler scheme

$$\hat{v}_{n+1} = \left|\hat{v}_n + \kappa(\lambda - \hat{v}_n)\Delta + \theta\hat{v}_n^{\gamma}\Delta_n W\right| \tag{6.2}$$

This scheme satisfies the same pathwise convergence result (6.1). Unfortunately, L^p-convergence results for this scheme are only known under rather restrictive assumptions on ν, see [BBD07].
To find a scheme that preserves positivity, converges strongly in L^2 and whose rate of convergence is known, we have to examine more complex schemes.

6.2 The Drift-Implicit Square Root Euler Scheme (DISE)

This scheme was first suggested in [Alf05]. It preserves positivity and for the case $\gamma = 1/2$ a strong L^2-convergence rate of $1/2$ was proven in [DNS12]. Instead of using the original SDE, the scheme uses the SDE after the Lamperti transformation $\sigma_t := v_t^{1-\gamma}$,

$$d\sigma_t = f(\sigma_t)dt + \theta(1-\gamma)dW_t$$

with

$$f(x) = (1-\gamma)\cdot\left(\kappa\lambda x^{-\frac{\gamma}{1-\gamma}} - \kappa x - \frac{\gamma\theta^2}{2}\cdot x^{-1}\right)$$

Then the drift-implicit Euler method is used to approximate this process, and the approximation of the CIR process is obtained via back-transformation.

$$\begin{aligned}\hat{\sigma}_{n+1} &= \hat{\sigma}_n + f(\hat{\sigma}_{n+1})\cdot\Delta + \theta(1-\gamma)\cdot\Delta_n W \\ \hat{v}_{n+1} &= \hat{\sigma}_{n+1}^{1/(1-\gamma)}\end{aligned} \tag{6.3}$$

To solve the first equation for $\hat{\sigma}_{n+1}$ we have to find a positive solution to $x - f(x)\Delta = c$ for positive x and arbitrary right-hand sides c. This is always possible, if either $\gamma > 1/2$ or $\nu > 1/2$, because in these cases $x - f(x)\Delta \to -\infty$ for $x \to 0$ and $x - f(x)\Delta \to \infty$ for $x \to \infty$. Thus we have proven the following result.

Proposition 6.1 The drift-implicit square root Euler scheme is well-defined and preserves positivity if either $\gamma > 1/2$ or $\nu > 1/2$. ◁

While in general we will need root-finding algorithms to solve equation (6.3), the particularly important case of $\gamma = 1/2$ allows for an explicit solution. In this case f is given by

$$f(x) = \left(\frac{\kappa\lambda}{2} - \frac{\theta^2}{8}\right)\cdot x^{-1} - \frac{\kappa}{2}\cdot x \tag{6.4}$$

and leads to the equation

$$\hat{\sigma}_{n+1} - \left(\frac{\kappa\lambda}{2} - \frac{\theta^2}{8}\right)\Delta\cdot\hat{\sigma}_{n+1}^{-1} + \frac{\kappa\Delta}{2}\cdot\hat{\sigma}_{n+1} = \hat{\sigma}_n + \frac{\theta}{2}\cdot\Delta_n W$$

Multiplying this equation by $2\hat{\sigma}_{n+1}$ gives a quadratic equation which for $\kappa\lambda - \theta^2/4 > 0$, i.e. $\nu > 1/2$, always possesses a unique positive solution

$$\hat{\sigma}_{n+1} = \frac{\hat{\sigma}_n + \frac{\theta}{2}\Delta_n W}{\beta} + \sqrt{\frac{(\hat{\sigma}_n + \frac{\theta}{2}\Delta_n W)^2}{\beta^2} + \frac{\left(\kappa\lambda - \frac{\theta^2}{4}\right)\cdot\Delta}{\beta}} \tag{6.5}$$

with $\beta = 2 + \kappa\Delta$.
In the case $\gamma = 1/2$ the drift-implicit square root Euler scheme is very similar to the drift-implicit Milstein scheme, which is discussed in the next section. This can be seen by squaring the implicit equation (6.3):

$$\begin{aligned}
\hat{v}_{n+1} &= \hat{\sigma}_n^2 + \hat{\sigma}_n\theta \cdot \Delta_n W + \frac{\theta^2}{4}(\Delta_n W)^2 - f(\hat{\sigma}_{n+1})^2\Delta^2 \\
&\qquad + 2f(\hat{\sigma}_{n+1})\Delta \cdot \left(f(\hat{\sigma}_{n+1})\Delta + \hat{\sigma}_n + \frac{\theta}{2}\Delta_n W \right) \\
&= \hat{\sigma}_n^2 + \hat{\sigma}_n\theta \cdot \Delta_n W + \frac{\theta^2}{4}(\Delta_n W)^2 - f(\hat{\sigma}_{n+1})^2\Delta^2 \\
&\qquad + \left(\left(\kappa\lambda - \frac{\theta^2}{4} \right) \cdot \frac{1}{\hat{\sigma}_{n+1}} - \kappa\hat{\sigma}_{n+1} \right) \cdot \Delta \cdot \hat{\sigma}_{n+1} \\
&= \hat{v}_n + \kappa(\lambda - \hat{v}_{n+1}) \cdot \Delta + \theta\sqrt{\hat{v}_n} \cdot \Delta_n W + \frac{\theta^2}{4}\left((\Delta_n W)^2 - \Delta \right) - f(\hat{\sigma}_{n+1})^2\Delta^2 \qquad (6.6)
\end{aligned}$$

Later we will make use of the following moment bounds and convergence rates for the DISE scheme.

Theorem 6.2 1. Assume $\gamma = 1/2$ and $\nu > 2$.

$$\mathbb{E} \sup_{k=0,\dots,\lceil T/\Delta \rceil} |\hat{\sigma}_{k\Delta}|^p < \infty \qquad \text{for all } p > -\frac{2}{3}\nu$$

and there exist constants C_p such that

$$\left(\mathbb{E} \sup_{k=0,\dots,\lceil T/\Delta \rceil} |\sigma_{k\Delta} - \hat{\sigma}_{k\Delta}|^p \right)^{1/p} \leq C_p \cdot \Delta \qquad \text{for all } 1 \leq p < \frac{2}{3}\nu$$

$$\left(\mathbb{E} \sup_{k=0,\dots,\lceil T/\Delta \rceil} |\sigma_{k\Delta} - \hat{\sigma}_{k\Delta}|^p \right)^{1/p} \leq C_p \cdot \sqrt{\Delta} \qquad \text{for all } 1 \leq p < \frac{4}{3}\nu$$

Finally, if $1 \leq p < \nu$, then the linearly interpolated scheme, i.e.

$$\bar{\sigma}_t := \left(k + 1 - \frac{t}{\Delta} \right) \hat{\sigma}_{k\Delta} + \left(\frac{t}{\Delta} - k \right) \hat{\sigma}_{(k+1)\Delta} \qquad t \in [k\Delta, (k+1)\Delta]$$

satisfies

$$\left(\mathbb{E} \sup_{t\in[0,T]} |\sigma_t - \bar{\sigma}_t|^p \right)^{1/p} \leq c_p \cdot \sqrt{\Delta} \cdot \sqrt{|\log(\Delta)|}$$

for some constant c_p.

2. On the other hand, if $\gamma > 1/2$, then

$$\mathbb{E} \sup_{k=0,\dots,\lceil T/\Delta \rceil} |\hat{\sigma}_{k\Delta}|^p < \infty \qquad \text{for all } p \in \mathbb{R}$$

and there exist constants C_p such that

$$\mathbb{E} \sup_{k=0,\dots,\lceil T/\Delta \rceil} |\sigma_{k\Delta} - \hat{\sigma}_{k\Delta}|^p \leq C_p \cdot \Delta^p \qquad \text{for all } p \geq 1$$

Proof. These results were established in [DNS12] and [NS14]. □

Note that the convergence rate results extend to $\hat{\sigma}_t^\alpha$ for $\alpha \geq 1$, in particular to $\hat{v}_t = \hat{\sigma}_t^{1/(1-\gamma)}$ and $\sqrt{\hat{v}_t} = \hat{\sigma}_t^{1/2(1-\gamma)}$. Indeed, Taylor's formula shows

$$|x^\alpha - y^\alpha| \leq \alpha \sup_{z \in [x,y]} z^{\alpha-1} \cdot |x-y| \leq \alpha \cdot (x^{\alpha-1} + y^{\alpha-1}) \cdot |x-y| \tag{6.7}$$

and thus if $p' \in (p, 2\nu/3)$ and $q := pp'/(p'-p)$, then Hölder's inequality and the previous theorem give

$$\left(E \sup_{k=0,\dots,\lceil T/\Delta \rceil} |\sigma_{k\Delta}^\alpha - \hat{\sigma}_{k\Delta}^\alpha|^p \right)^{1/p} \leq c \cdot \left(E \sup_{k=0,\dots,\lceil T/\Delta \rceil} (\sigma_{k\Delta}^{\alpha-1} + \hat{\sigma}_{k\Delta}^{\alpha-1})^q \right)^{1/q} \cdot \left(E \sup_{k=0,\dots,\lceil T/\Delta \rceil} |\sigma_{k\Delta} - \hat{\sigma}_{k\Delta}|^{p'} \right)^{1/p'} \leq c \cdot \Delta$$

for all $k = 0, \dots, \lceil T/\Delta \rceil$ and $\alpha \geq 1$. Analogously for $p' \in (p, 4\nu/3)$

$$\left(E \sup_{k=0,\dots,\lceil T/\Delta \rceil} |\sigma_{k\Delta}^\alpha - \hat{\sigma}_{k\Delta}^\alpha|^p \right)^{1/p} \leq c \cdot \sqrt{\Delta}$$

In the case $\gamma = 1/2$, we can sometimes weaken the condition on ν if we use the following lemma to remove an inverse moment of $\hat{\sigma}_k$ at the cost of one $\Delta^{-1/2}$.

Lemma 6.3 Let $\gamma = 1/2$ and $p \geq 1$. There exists a constant C_p such that

$$\sup_{k=0,\dots,N} E\hat{\sigma}_k^{-p} \leq C_p \cdot \Delta^{-p/2}$$

Proof. Set

$$u_k := \hat{\sigma}_k + \frac{\theta}{2}\Delta_k W \qquad \text{and} \qquad C_\Delta = \left(\kappa\lambda - \frac{\theta^2}{4}\right) \cdot (2 + \kappa\Delta)$$

With these abbreviations we can write (6.5) as

$$\hat{\sigma}_{k+1} = \frac{u_k + \sqrt{u_k^2 + C_\Delta \Delta}}{2 + \kappa\Delta}$$

If $u_k \geq 0$, this gives immediately

$$\frac{1}{\hat{\sigma}_{k+1}} \leq \frac{2 + \kappa\Delta}{\sqrt{C_\Delta \Delta}}$$

Otherwise, if $u_k < 0$, first observe that $|u_k| < \frac{\theta}{2}|\Delta_k W|$ because $\hat{\sigma}_k > 0$. We use the mean value theorem to get

$$(2 + \kappa\Delta)\hat{\sigma}_{k+1} = \sqrt{u_k^2 + C_\Delta \Delta} - \sqrt{u_k^2} \geq \frac{C_\Delta \Delta}{2\sqrt{u_k^2 + C_\Delta \Delta}}$$

and thus

$$\frac{1}{\hat{\sigma}_{k+1}} \leq \frac{2(2 + \kappa\Delta)}{C_\Delta \Delta} \cdot \sqrt{u_k^2 + C_\Delta \Delta} \leq \frac{2(2 + \kappa\Delta)}{C_\Delta} \cdot \frac{\sqrt{\frac{\theta^2}{4}|\Delta_k W|^2 + C_\Delta \Delta}}{\Delta}$$

After taking the p-th power and expectation, the right-hand side is of order $\Delta^{-p/2}$ and the proof is complete. □

6.3 The Drift-Implicit Milstein Scheme (DIMIL)

The drift-implicit Milstein scheme for the CIR/CEV process reads

$$\hat{v}_{n+1} = \hat{v}_n + \kappa(\lambda - \hat{v}_{n+1}) \cdot \Delta + \theta \hat{v}_n^{\gamma} \cdot \Delta_n W + \frac{\theta^2 \gamma}{2} \hat{v}_n^{2\gamma-1} \cdot \left((\Delta_n W)^2 - \Delta \right) \tag{6.8}$$

and can be easily solved explicitly

$$\hat{v}_{n+1} = \frac{1}{1+\kappa\Delta} \left(\hat{v}_n + \kappa\lambda\Delta + \theta \hat{v}_n^{\gamma} \cdot \Delta_n W + \frac{\theta^2 \gamma}{2} \hat{v}_n^{2\gamma-1} \cdot \left((\Delta_n W)^2 - \Delta \right) \right) \tag{6.9}$$

Note that for $\gamma = 1/2$ the scheme is particularly simple and differs from the DISE scheme as given in (6.6) only by omitting the last summand $f(\hat{\sigma}_{n+1})^2 \Delta^2$.

Theorem 6.4 For $\gamma = 1/2$ and $\nu > 1/2$ the scheme DIMIL dominates the scheme DISE: Denoting the approximation obtained by DIMIL and DISE by $\hat{v}^{\mathrm{MIL}}$ and $\hat{v}^{\mathrm{SE}}$, respectively, we have $\hat{v}_n^{\mathrm{MIL}} \geq \hat{v}_n^{\mathrm{SE}}$ for all $n = 0, \ldots, \lceil T/\Delta \rceil$. In particular, the approximation remains strictly positive.

Proof. Let $h(\sigma, w)$ denote the value of the next step of the drift-implicit Euler scheme for (σ_t) as a function of the current step value $\hat{\sigma}_n = \sigma > 0$ and the increment $\Delta_n W = w$, i.e. $h(\sigma, w)$ is the unique positive solution to the equation

$$h(\sigma, w) = \sigma + f(h(\sigma, w))\Delta + \frac{\theta}{2} w$$

with the function f being defined in (6.4). From (6.5) we learn that

$$\partial_\sigma h(\sigma, w) = \frac{1}{\beta} \cdot \left(1 + \frac{\sigma + \frac{\theta}{2} w}{\sqrt{\left(\sigma + \frac{\theta}{2} w\right)^2 + C}} \right) \geq \frac{1}{\beta}$$

with $\beta = 2 + \kappa\Delta$ and $C = (\kappa\lambda - \theta^2/4)\Delta\beta > 0$ (thanks to $\nu > 1/2$). Thus $h(\sigma, w)$ is increasing in the first component. From (6.6) we can see that h fulfills the equation

$$h^2(\sigma, w) = \frac{1}{1+\kappa\Delta} \cdot \left(\sigma^2 + \kappa\lambda\Delta + \theta\sigma w + \frac{\theta^2}{4} \cdot \left(w^2 - \Delta \right) - f(h(\sigma, w))\Delta^2 \right)$$

Now assume that $\hat{v}_n^{\mathrm{MIL}} \geq \hat{v}_n^{\mathrm{SE}}$ for some $n \geq 0$ and set $\hat{\sigma}_n^{\mathrm{MIL}} := \sqrt{\hat{v}_n^{\mathrm{MIL}}}$. Putting our results together gives

$$\begin{aligned}
\hat{v}_{n+1}^{\mathrm{SE}} &= h^2(\hat{\sigma}_n^{\mathrm{SE}}, \Delta_n W) \\
&\leq h^2(\hat{\sigma}_n^{\mathrm{MIL}}, \Delta_n W) \\
&= \frac{1}{1+\kappa\Delta} \left(\hat{v}_n^{\mathrm{MIL}} + \kappa\lambda\Delta + \theta\hat{\sigma}_n^{\mathrm{MIL}} \cdot \Delta_n W + \frac{\theta^2}{4} \left(\Delta_n W^2 - \Delta \right) - f(h(\hat{\sigma}_n^{\mathrm{MIL}}, \Delta_n W))^2 \Delta^2 \right) \\
&\leq \frac{1}{1+\kappa\Delta} \cdot \left(\hat{v}_n^{\mathrm{MIL}} + \kappa\lambda\Delta + \theta\sqrt{\hat{v}_n^{\mathrm{MIL}}} \cdot \Delta_n W + \frac{\theta^2}{4} \cdot \left((\Delta_n W)^2 - \Delta \right) \right) \\
&= \hat{v}_{n+1}^{\mathrm{MIL}}
\end{aligned}$$

The result now follows by induction. □

Theorem 6.5 Assume $\gamma = 1/2$ and $\nu > 1$. Then for all $p > -\frac{2}{3}\nu$

$$\mathbb{E} \sup_{k=0,\ldots,\lceil T/\Delta \rceil} \hat{v}_{k\Delta}^p < \infty$$

Moreover, if $\nu > 3$, then

$$\left(E \sup_{k=0,\dots,\lceil T/\Delta \rceil} |\nu_{k\Delta} - \hat{\nu}_{k\Delta}|^2 \right)^{1/2} < c \cdot \sqrt{\Delta}$$

Proof. The convergence order was proven in Proposition 4.2 (ii) of [NS14]. For negative moments the first claim follows from the last result and Theorem 6.2. For positive moments we use (6.9) to get

$$\hat{\nu}_{n+1} \le \hat{\nu}_n + \kappa\lambda\Delta + \theta\sqrt{\hat{\nu}_n} \cdot \Delta_n W + \frac{\theta^2}{4} \cdot \left((\Delta_n W)^2 - \Delta \right)$$

and by induction

$$\hat{\nu}_n \le \nu_0 + n\kappa\lambda\Delta + \theta \int_0^{n\Delta} \sqrt{\hat{\nu}_{\Delta\lfloor t/\Delta \rfloor}} dW_t + \frac{\theta^2}{4} \cdot \sum_{k=0}^{n-1} ((\Delta_k W)^2 - \Delta)$$

Now the Burkholder–Davis–Gundy inequality shows for $p \ge 2$ that

$$E \sup_{k=0,\dots,\lceil T/\Delta \rceil} \hat{\nu}_n^p \le c \cdot \left(1 + E \sup_{k=0,\dots,\lceil T/\Delta \rceil} \hat{\nu}_n^{p/2} \right)$$

The result follows by induction; the induction start for $p = 1$ is trivial. □

6.4 Extension to the Heston Model

All schemes presented in the previous sections can easily be extended to a scheme for the full Heston model by using the standard Euler scheme for the price process. For our purposes it is usually easier to approximate the log-price and compute the actual price only when necessary. We call this scheme the log-Euler scheme:

$$\begin{aligned} \hat{X}_{n+1} &= \hat{X}_n + \left(\mu - \frac{1}{2}\hat{\nu}_n \right) \Delta + \sqrt{\hat{\nu}_n^+} \Delta_n B \\ \hat{S}_{n+1} &= \exp(\hat{X}_n) \end{aligned}$$

Sometimes it is possible to transfer the convergence rate of the scheme for the CIR process to the log-price scheme:

Theorem 6.6 Let $p \ge 2$. Assume that ν_t is approximated using a scheme $\hat{\nu}_k$ that satisfies $E|\hat{\nu}_{k\Delta}|^q < \infty$ for all $k = 1, \dots, \lceil T/\Delta \rceil$ and all $q \ge 0$ and

$$\sup_{k=0,\dots,\lceil T/\Delta \rceil} \left\| \sqrt{\nu_{k\Delta}} - \sqrt{\hat{\nu}_{k\Delta}} \right\|_{p'} \le c \cdot \sqrt{\Delta} \tag{6.10}$$

for some $p' > p$. If X is approximated by the log-Euler scheme, then

$$\sup_{k=0,\dots,\lceil T/\Delta \rceil} \|X_{k\Delta} - \hat{X}_{k\Delta}\|_p \le c \cdot \sqrt{\Delta}$$

Proof. First note that Hölder's inequality and the conditions on $\hat{\nu}$ imply for $q = pp'/(p'-p)$

$$\begin{aligned} \sup_{k=0,\dots,\lceil T/\Delta \rceil} \|\nu_{k\Delta} - \hat{\nu}_{k\Delta}\|_p \le \sup_{k=0,\dots,\lceil T/\Delta \rceil} & \left\| \sqrt{\nu_{k\Delta}} - \sqrt{\hat{\nu}_{k\Delta}} \right\|_{p'} \\ & \cdot \left\| \sqrt{\nu_{k\Delta}} + \sqrt{\hat{\nu}_{k\Delta}} \right\|_q \le c \cdot \sqrt{\Delta} \end{aligned} \tag{6.11}$$

Now we can prove the assertion:

$$E|X_{k\Delta} - \hat{X}_{k\Delta}|^p \le c \cdot \left(E\left| \int_0^{k\Delta} v_s - \hat{v}_{\eta(s)} ds \right|^p + E\left| \int_0^{k\Delta} \sqrt{v_s} - \sqrt{\hat{v}_{\eta(s)}} dW_s \right|^p \right)$$

We apply the Burkholder–Davis–Gundy inequality to the second summand.

$$\begin{aligned} &\le c \cdot E\left(\int_0^T |v_s - v_{\eta(s)}|^p + |v_{\eta(s)} - \hat{v}_{\eta(s)}|^p \right. \\ &\qquad \left. + |\sqrt{v_s} - \sqrt{v_{\eta(s)}}|^p + \left|\sqrt{v_{\eta(s)}} - \sqrt{\hat{v}_{\eta(s)}}\right|^p ds \right) \\ &\le c \cdot \Delta^{p/2} \end{aligned}$$

Now the claim follows from (6.10), (6.11) and the continuity in L^p of v_t; see Theorems 2.3 and 2.5. □

Unfortunately, so far the scheme DISE is the only scheme where the conditions of this theorem could be proven, see Theorem 6.2 and the remark thereafter.

Corollary 6.7 Let $2 \le p < \frac{4}{3}\nu$. Assume that v_t is approximated using the DISE scheme and X is approximated using the log-Euler scheme. Then

$$\sup_{k=0,\dots,\lceil T/\Delta \rceil} \|X_{k\Delta} - \hat{X}_{k\Delta}\|_p \le c \cdot \sqrt{\Delta}$$

6.5 Further Schemes

Many more schemes to approximate the Heston model — typically more complex than the ones presented here — are suggested in the literature, see e.g. [And07], [BK06], [LKvD10]. Somewhat outstanding among the schemes is the algorithm presented in work of Broadie and Kaya, [BK06], because it simulates the Heston model exactly. Unfortunately, the algorithm is complex and expensive so that its practical use is limited. This is discussed e.g. in [And07]. This work also presents two schemes which approximate the Broadie–Kaya scheme, sacrificing exactness for computational efficiency.
Most works try to improve the approximation of the volatility, which seems to be the main problem. [KJ06] are an exception and propose a scheme based on the trapezoidal rule for the log-price process.

6.6 Numerical Comparison of Approximation Schemes

We now perform a short comparison of the different schemes presented in this chapter. For simplicity and because the Monte-Carlo algorithms which we use later assume a bias of order 1, we will compare the so-called *weak error*, i.e. the bias achieved with a certain stepsize. To highlight the dependence of the approximation on the stepsize, we will write $\hat{S}_t^\Delta$ etc. in this section. The weak error is defined as

$$e_{\text{weak}}(\Delta) := |E(f(\hat{S}_T^\Delta)) - E(f(S_T))|$$

We say that a scheme achieves a weak order of convergence of α if

$$e_{\text{weak}}(\Delta) \le c \cdot \Delta^\alpha$$

The literature contains various results which prove an order of convergence of $\alpha = 1$ for certain schemes if the function f is smooth enough and the coefficients of the underlying

SDEs are nice enough; see e.g. [KP92]. So far none of those results cover the digital option in the Heston model; see Theorem 7.13 for a first result in this direction. Nevertheless, numerical practice has shown that a weak order of 1 is often achieved under much weaker assumptions; this is also confirmed in the following experiments. However, there are also rather simple models, where a weak order of 1 is provably not achieved by the Euler scheme; see e.g. Proposition 2.3 in [Keb05].
We will use two different models, both of them being standard Heston models, but with widely different values of ν.

M1) The parameters were taken from [ASK07]. ν is comfortably high at $\nu \approx 2.011$.

$$T = 2,\ \mu = 0,\ \kappa = 5.07,\ \lambda = 0.0457,\ \theta = 0.48,\ \rho = -0.767,\ \nu_0 = \lambda,\ s_0 = 100$$

M2) This model stems from [BK06]. The parameters make $\nu \approx 0.634$ and thus ν_t may hit zero.

$$T = 1,\ \mu = 0.0319,\ \kappa = 6.21,\ \lambda = 0.019,\ \theta = 0.61,\ \rho = -0.7,\ \nu_0 = 0.010201,\ s_0 = 100$$

Because we are mainly interested in discontinuous options, we will use the payoff function $f = \mathbb{1}_{[0,s_0]}$.
In Chapters 8 and 10 it will be vital to use schemes which are guaranteed to always remain positive (for $\nu > 1/2$). We thus chose to compare the three presented schemes with this property: The drift-implicit square root Euler scheme of Section 6.2, the drift-implicit Milstein scheme of Section 6.3 and the symmetrized Euler scheme from (6.2).

The numerical experiment consists of the following steps:

1. First, for each model a reference value x^* is computed using Theorem 2.8.
2. Now for each model and scheme and various step sizes of the form $T/2^k$, $k \geq 2$ we compute an estimate of the weak error as

$$\hat{e}_{\text{weak}} = \left| \frac{1}{M} \cdot \sum_{i=1}^{M} f(\hat{S}_T^{\Delta,i}) - x^* \right|$$

   with M being at least 10^7 and $\hat{S}_T^{\Delta,i}$ denoting independent repetitions of $\hat{S}_T^{\Delta}$.
3. The resulting figure plots the $\log_2$ of the number of steps (i.e. k) on the x-axis against the $\log_2$ of the error $\hat{e}_{\text{weak}}$ on the y-axis. The plot is complemented by a least-squares fitted affinely linear function visualizing the rate of convergence: Assuming that $e_{\text{weak}} \approx c \cdot \Delta^\alpha = c \cdot (T/N)^\alpha$ we have in $\log_2$-coordinates

$$\log_2 e_{\text{weak}} \approx \log(c \cdot T^\alpha) - \alpha \cdot \log_2 N$$

   Thus the (negative) slope of the fitted line gives the measured rate of weak convergence.

The plots show that the DIMIL scheme achieves a weak error rate of 1 under far weaker assumptions than required by Theorem 7.13. The two other schemes fail to reach weak error rate in model M2 ($\nu \approx 0.634$) but achieve a higher rate in model M1 ($\nu \approx 2.011$). However, the DISE scheme seems to reach a maximum accuracy of approximately 2^{12}. Neither computing more samples nor using more discretization steps nor changing the reference value to a value computed by DISE itself (using 2^{12} discretization steps) removes this behavior.

Figure 6.1: Weak error of different schemes in model M1. The least-squares fits have the convergence rates (i.e. negative slopes) 1.12 *(DISE),* 1.01 *(DIMIL) and* 1.49 *(symmetrized Euler).*

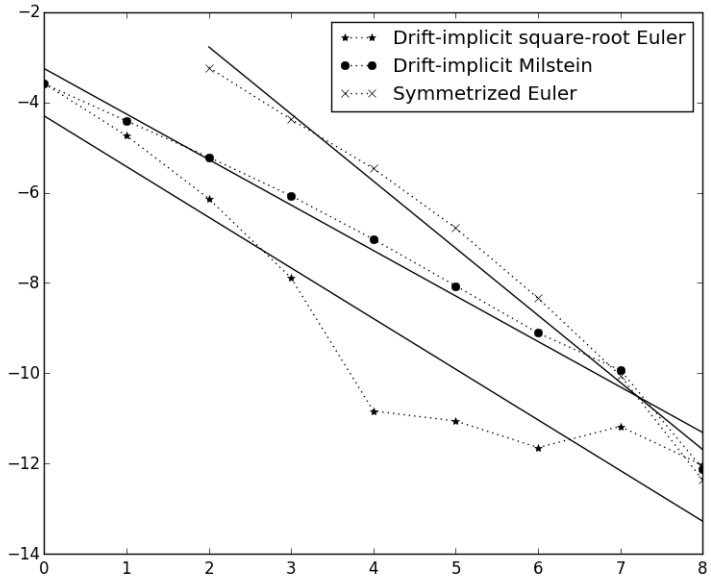
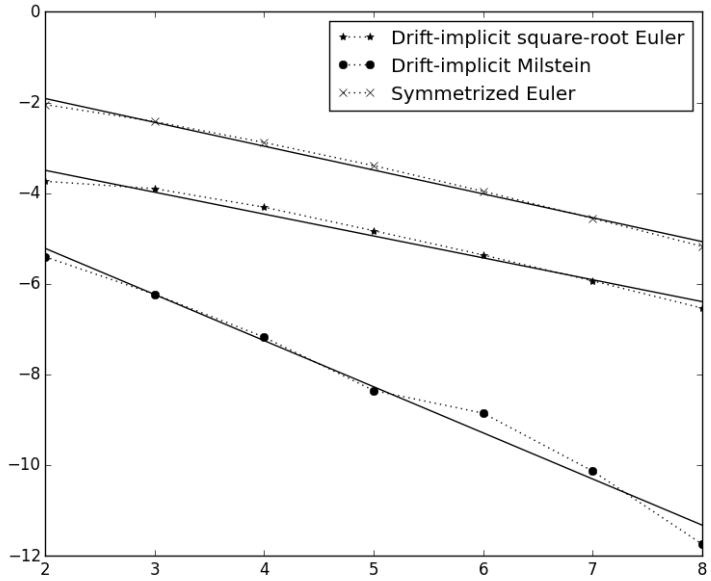


Figure 6.2: Weak error of different schemes in model M2. The least-squares fits have the convergence rates 0.48 *(DISE),* 1.02 *(DIMIL) and* 0.53 *(symmetrized Euler).*



Chapter 7

Weak Error

This chapter is devoted to the function

$$u(t, x, v) := \mathbb{E}(f(X_{T-t}^{x,v}))$$

where $X_t^{x,v}$ denotes the usual log-price process, started at $x_0 = x$ and $v_0 = v$. We will have to restrict ourselves to the one-dimensional standard Heston model, i.e. $\gamma = 1/2$. First, in Section 7.2, we will prove that u is infinitely often differentiable. Secondly, in Section 7.3, we will derive bounds on the derivatives. Unfortunately, both parts require different approaches and lengthy proofs. As an application of the bounds on the derivatives of u, we prove that the drift-implicit Milstein scheme has a weak error rate of 1 in Heston models with sufficiently high ν.

7.1 Notation

For $x \in \mathbb{R}$ we write $\mathbb{C}_{\geq x}$ for $\{z \in \mathbb{C} : \mathrm{Re}(z) \geq x\}$. For $\Omega \subset \mathbb{C}^d$ we denote by $C_{\mathrm{pol}}^\infty(\Omega)$ the set of functions $f\colon \Omega \to \mathbb{C}$ which are infinitely often differentiable with the function and all derivatives bounded by polynomials. $C_b^\infty(\Omega)$ denotes the subset where the function and all derivatives are bounded by constants. For $f \in C^\infty(\mathbb{R})$ and $n, m \in \mathbb{N}_0$ we define seminorms

$$p_{m,n}(f) := \sup_{x \in \mathbb{R}} (1 + |x|^m) \cdot |f^{(n)}(x)| \tag{7.1}$$

Finally, $\mathcal{S}$ denotes the Schwartz space of rapidly decreasing functions; see e.g. [Wer07].

$$\mathcal{S} = \left\{ f\colon \mathbb{R} \to \mathbb{C} \;\middle|\; p_{m,n}(f) < \infty \;\; \forall\, m, n \in \mathbb{N}_0 \right\}$$

The following lemma shows that the set $C_{\mathrm{pol}}^\infty(\Omega)$ for arbitrary Ω is stable under common operations.

Lemma 7.1 Let $f, g \in C_{\mathrm{pol}}^\infty(\Omega)$. Then $f \cdot g \in C_{\mathrm{pol}}^\infty(\Omega)$ and $f \circ g \in C_{\mathrm{pol}}^\infty(\Omega)$. If g is bounded away from 0, then also $f/g \in C_{\mathrm{pol}}^\infty(\Omega)$. Finally, if both f and g take only values in $\{x \in \mathbb{R} : x \geq \varepsilon\}$ for some $\varepsilon > 0$, then $(f/g)^p \in C_{\mathrm{pol}}^\infty(\Omega)$ for all $p \in \mathbb{R}$.

Proof. The first claim follows from the generalizations of the product and chain rule to higher order derivatives (the latter is also known as Faà di Bruno's formula; see [Bru57]). The second claim holds because $x \mapsto x^{-1}$ is $C_{\mathrm{pol}}^\infty(\Omega)$ on $\{z \in \mathbb{C} : |z| > \varepsilon\}$ for any $\varepsilon > 0$. By the generalized chain rule, the n-th derivative of h^p is a linear combination of terms

$$h^{p-(k_1+\cdots+k_n)} \cdot \prod_{m=1}^{n} (D^m(h))^{k_m}$$

for some $k_i \in \mathbb{N}$. If h and $1/h$ are in $C^\infty_{\text{pol}}(\Omega)$, then all these terms are bounded by polynomials. Setting $h = f/g$ proves the claim. □

7.2 The differentiability of u

In this section we will prove that the function

$$u(t, x, v) := \mathbb{E}(f(X_{T-t}^{x,v}))$$

is infinitely often differentiable; see Theorem 7.6.
In [BRFCU10] it is shown that for $t > 0$ the log-price $X_t^{x,v}$ has a C^∞-density $\varphi_{t,x,v}$ and that its characteristic function $\psi_{t,x,v}$ can be written as follows. Of course, this is just a different formulation for the characteristic function φ_2 of Section 2.4.

$$\psi_{t,x,v}(y) = \mathbb{E}(e^{iyX_t}) = e^{i(x_0+\mu t)y} \cdot g_1(y) \cdot \exp(-v_0 g_2(y))$$

with

$$\begin{aligned}
g_1(y) &= \left(\frac{e^{\xi t/2} \cdot d}{h(y)}\right)^\nu \\
g_2(y) &= \frac{(iy + y^2) \cdot \sinh(d \cdot t/2)}{h(y)} \\
h(y) &= d\cosh(d \cdot t/2) + \xi \sinh(d \cdot t/2) \\
D = D(y) &= (\kappa - i\theta\rho y)^2 + \theta^2(iy + y^2) \\
d = d(y) &= \sqrt{D} \\
\xi = \xi(y) &= \kappa - i\theta\rho y
\end{aligned}$$

Similar to the complex logarithm of Section 2.4.3, the complex square root and the exponentiation of a complex number are functions with multiple branches. The complex square root in the definition of d poses no problem, because $\operatorname{Re}(D) = \kappa^2 + (1-\rho^2)\theta^2 y^2 > 0$. The exponentiation in the definition g_1 is well-defined by the requirement that g_1 is continuous. This will be made more precise in the proof of Lemma 7.3 which establishes some properties of above functions.

Lemma 7.2 The real value of g_2 is always non-negative.

Proof. From Theorem 5.9 in [BRFCU10] we get the following representation of g_2 (note that in their notation we have $g_2(y) = iy \cdot G(iy)$):

$$g_2(y) = iy \cdot \frac{1 - e^{-\kappa t}}{2\kappa} + y^2 \sum_{n=1}^{\infty} \frac{b_n}{a_n^2 \cdot (1 - iy/a_n)}$$

with $b_n \in \mathbb{R}_+$, $a_n \in \mathbb{R} \setminus \{0\}$ such that $\sum_{n=1}^\infty b_n a_n^{-2}$ converges. The claim now follows from $\operatorname{Re}(1/z) = \operatorname{Re}(z)/|z|^2$ with $z = 1 - iy/a_n$. □

Lemma 7.3 We have

1. $g_1 \in \mathcal{S}$,
2. $g_2 \in C^\infty_{\text{pol}}(\mathbb{R})$,
3. There exist constants $c_1, c_2, y_0 > 0$ such that

$$\operatorname{Re}(g_2(y)) \geq c_1|y| \quad \forall y \in \mathbb{R}, |y| \geq y_0 \qquad \text{and} \qquad |g_2(y)| \leq c_2(1 + |y|), \quad \forall y \in \mathbb{R}$$

As a consequence

$$\frac{\partial^{k+l}}{\partial x^k \partial \nu^l}\psi_{t,x,\nu}(y) = e^{i(x+\mu t)y} \cdot g_1(y) \cdot \exp(-\nu g_2(y)) \cdot (iy)^k \cdot (-g_2(y))^l$$

as a function in y is in $\mathcal{S}$ for all $k, l \in \mathbb{N}_0$.

Proof. For simplicity we assume that $t = 2$. The general case follows by replacing every occurrence of e in the proof by $\tilde{e} := e^{t/2}$ so that $\cosh(dt/2) = (\tilde{e}^d + \tilde{e}^{-d})/2$ and $\sinh(dt/2) = (\tilde{e}^d - \tilde{e}^{-d})/2$. We prove the result in five steps, which prove necessary properties of the functions d, e^{-d}, h^{-1}, g_1, and g_2, respectively.
1. Because $d = \sqrt{D}$ with $\mathrm{Re}(D) \geq \kappa^2 > 0$ and because the complex square root is in $C^\infty_{\mathrm{pol}}(\mathbb{C}_{\geq \kappa^2})$ we have $d \in C^\infty_{\mathrm{pol}}(\mathbb{R})$. Also $\mathrm{Re}(d)$ grows at least linearly: $\mathrm{Re}(z)^2 \geq \mathrm{Re}(z^2)$ is true for all complex numbers z. In particular

$$\mathrm{Re}(d) \geq \sqrt{\mathrm{Re}(D)} \geq \sqrt{1-\rho^2}\theta|y| \tag{7.2}$$

for all $y \in \mathbb{R}$. On the other hand, $|d|$ grows at most linearly, because D grows clearly quadratically.
2. Therefore $e^{-d} \in \mathcal{S}$ decreases faster than every polynomial. Now all derivatives of e^{-d} are sums of products of d, derivatives of d and always one factor e^{-d} and thus bounded by a polynomial times e^{-d}. Consequently $e^{-d} \in \mathcal{S}$.
3. h can be rewritten as

$$h = \frac{1}{2} \cdot (d + \xi + (d - \xi) \cdot e^{-2d}) \cdot e^d$$

Given the linear growth of d, $\mathrm{Re}(d)$, and ξ, we know that there exists an $y_0 \geq 0$ such that for all $|y| \geq y_0$

$$|d - \xi| \cdot e^{-2\mathrm{Re}(d)} \leq \kappa \leq \frac{1}{2}\mathrm{Re}(d + \xi) \tag{7.3}$$

and thus

$$\left|d + \xi + (d - \xi) \cdot e^{-2d}\right| \geq |d + \xi| - |d - \xi| \cdot e^{-2\mathrm{Re}(d)} \geq \frac{1}{2}\mathrm{Re}(d + \xi) > \kappa > 0, \qquad |y| > y_0 \tag{7.4}$$

On the other hand, from Proposition 5.4 in [BRFCU10] it follows that h does not have real zeros (note that their function F relates to h via $h(y) = d \cdot F(iy)$). Because it is continuous, it must be bounded away from 0 on compact intervals. Together with (7.4) we can thus conclude that there exists $c > 0$ such that

$$\left|d + \xi + (d - \xi) \cdot e^{-2d}\right| \geq |d + \xi| - |d - \xi| \cdot e^{-2\mathrm{Re}(d)} \geq c > 0 \tag{7.5}$$

for all $y \in \mathbb{R}$. This implies $1/h \in \mathcal{S}$.
4. To prove that $g_1 \in \mathcal{S}$ we need to define the exponentiation of $\bar{g}_1 := e^\xi d/h$ by $\nu = 2\kappa\lambda/\theta^2$ more precisely. For $z \in \mathbb{C}$ the power z^ν can be any number of the form

$$|z|^\nu \cdot e^{i\nu \cdot (\arg z + 2\pi k)} \qquad k \in \mathbb{Z}$$

If $\bar{g}_1$ crosses the negative real line, $\bar{g}_1^\nu$ will be discontinuous for each fixed choice of k (in fact, $\bar{g}_1$ forms a spiral towards 0 and crosses the negative real line infinitely often). Instead, the only choice which makes g_1 continuous and satisfies $g_1(0) = 1$ is

$$g_1(y) := |\bar{g}_1(y)|^\nu \cdot e^{i\nu\alpha(y)}$$

where the "rotation count" $\alpha\colon \mathbb{R} \to \mathbb{R}$ is defined by $\alpha(0) = 0$, α is continuous and

$$\alpha \equiv \arg \bar{g}_1(y) \mod 2\pi$$

Therefore

$$\begin{aligned} g_1(y) &= \left|\frac{e^{\xi} d}{h(y)}\right|^{\nu} \cdot e^{i\nu\alpha(y)} \\ &= \left|e^{\xi-d} \cdot \frac{2d}{d+\xi+(d-\xi)e^{-2d}}\right|^{\nu} \cdot e^{i\nu\alpha(y)} \\ &= e^{\nu(\kappa-\mathrm{Re}(d))} \cdot \left(\frac{|2d|}{|d+\xi+(d-\xi)e^{-2d}|}\right)^{\nu} \cdot e^{i\nu\alpha(y)} \end{aligned} \tag{7.6}$$

The first factor is in $\mathcal{S}$ by 2.; using (7.5) and Lemma 7.1 the second factor is in $C^{\infty}_{\mathrm{pol}}(\mathbb{R})$. Now consider the last factor $e^{i\nu\alpha(y)}$ of (7.6). Note that $\alpha(y) \equiv \arg(\bar{g}_1(y)) = \arg(\bar{g}_1(y)\cdot e^{\mathrm{Re}(d)})$ and that

$$\bar{g}_1(y) \cdot e^{\mathrm{Re}(d)} = \frac{e^{\xi} d \cdot 2e^{-d}}{d+\xi+(d-\xi)e^{-2d}} \cdot e^{\mathrm{Re}(d)} = \frac{2e^{\xi-i\cdot\mathrm{Im}(d)} d}{d+\xi+(d-\xi)e^{-2d}}$$

This function is in $C^{\infty}_{\mathrm{pol}}(\mathbb{R})$ due to (7.5) and Lemma 7.1. As consequence of (7.3) we have

$$\left|\bar{g}_1(y) \cdot e^{\mathrm{Re}(d)}\right| \geq c \cdot \frac{|d|}{|d+\xi|}$$

for some constant c. Due to the linear growth of d and $d+\xi$, this is bounded away from 0. On $\{z \in \mathbb{C} : |z| \geq \varepsilon\}$ the function arg is in C^{∞}_{b} as long as the branch cut on the negative real axis is avoided by switching to other branches in a continuous way. α does so by definition and thus $\alpha \in C^{\infty}_{\mathrm{pol}}(\mathbb{R})$. We have proven that the last factor of (7.6) is in $C^{\infty}_{\mathrm{pol}}(\mathbb{R})$. Because the first factor was in $\mathcal{S}$ we have finally proven that $g_1 \in \mathcal{S}$.
5. g_2 can be rewritten as

$$\begin{aligned} g_2(y) &= \frac{iy+y^2}{2h(y)}e^{d} - \frac{iy+y^2}{2h(y)}e^{-d} \\ &= \frac{iy+y^2}{d+\xi} + \frac{iy+y^2}{2}\left(\frac{e^{d}}{h(y)} - \frac{2}{d+\xi}\right) - \frac{iy+y^2}{2h(y)}e^{-d} \end{aligned} \tag{7.7}$$

The first part can be simplified to

$$\frac{iy+y^2}{d+\xi} = \frac{(iy+y^2)(d-\xi)}{\theta^2(iy+y^2)} = \frac{d-\xi}{\theta^2} \tag{7.8}$$

Using $h = ((d+\xi)e^{d} + (d-\xi)e^{-d})/2$ we can simplify the central part of (7.7):

$$\frac{e^{d}}{h(y)} - \frac{2}{d+\xi} = \frac{(d+\xi)e^{d} - 2h(y)}{h(y)(d+\xi)} = -\frac{d-\xi}{d+\xi}\cdot\frac{e^{-d}}{h(y)} \tag{7.9}$$

Combining (7.7), (7.8) and (7.9) gives

$$\begin{aligned} g_2(y) &= \frac{d-\xi}{\theta^2} - \left(\frac{d-\xi}{d+\xi}+1\right)\cdot\frac{iy+y^2}{2}\cdot\frac{e^{-d}}{h(y)} \\ &= \frac{d-\xi}{\theta^2} - \frac{d(iy+y^2)}{d+\xi}\cdot\frac{e^{-d}}{h(y)} \end{aligned}$$

This proves that $g_2 \in C^{\infty}_{\mathrm{pol}}(\mathbb{R})$ because d and ξ are linearly bounded, $\mathrm{Re}(d+\xi) \geq 2\kappa$ and $e^{-d}/h \in \mathcal{S}$. The latter fact also proves that the asymptotic behaviour of this function is determined by the first summand $(d-\xi)/\theta^2$. Because $\mathrm{Re}(d)$ grows linearly, $\mathrm{Re}(\xi) = \kappa$, and $|d|, |\xi|$ are bounded linearly, there must exist constants c_1 and c_2 as required.
The final claim about the derivative $\frac{\partial^{k+l}}{\partial x^k \partial \nu^l}\psi_{t,x,\nu}$ is now obvious, because $g_1 \in \mathcal{S}$. □

We need some further lemmas before we can prove the differentiability of u:

Lemma 7.4 For $\nu > 0$ and $h \in \mathbb{R} \setminus \{0\}$ define the function $f_h \colon \mathbb{C}_{\geq 0} \to \mathbb{C}$

$$f_h(x) := e^{-\nu x} \cdot h^{-2} \cdot \left(e^{-hx} - 1 + hx\right)$$

Then there exists $h_0 > 0$ such that for all $h \in \mathbb{R} \setminus \{0\}$ with $|h| \leq h_0$ the function $f_h \circ g_2$ is in $C^\infty_{\text{pol}}(\mathbb{R})$ and the bounds of $f \circ g_2$ and its derivatives can be chosen to hold for all $h \in [-h_0, h_0] \setminus \{0\}$.

Proof. $f_h \circ g_2 \in C^\infty(\mathbb{R})$ is obvious from Lemmas 7.1-7.3. Assume $h_0 > 0$. Using Taylor's formula on the term in parentheses we get that there exists a $\Theta(x) \in \mathbb{C}$ with $|\Theta(x)| \leq |hx|$ and

$$|f_h(x)| = \left|e^{-\nu x} h^{-2} \cdot (xh)^2 e^{-\Theta(x)}\right| \leq |x|^2 e^{-\nu \mathrm{Re} x + h_0 |x|}$$

for all $x \in \mathbb{C}$ and $h \in [-h_0, h_0] \setminus \{0\}$. Now we need the constants c_1, c_2, y_0 from Lemma 7.3. If $y \in \mathbb{R}$, $|y| \leq y_0$, then

$$|f_h(g_2(y))| \leq |g_2(y)|^2 \cdot e^{h_0 c_2 (1 + y_0)}$$

Otherwise, if $|y| \geq y_0$, then

$$|f_h(g_2(y))| \leq |g_2(y)|^2 \cdot e^{-\nu c_1 |y| + h_0 c_2 (1 + |y|)} \leq (g_2(y))^2 \cdot e^{h_0 c_2} \cdot e^{-(\nu c_1 - h_0 c_2)|y|}$$

If h_0 is chosen small enough such that $\nu c_1 - h_0 c_2 \geq 0$, the last term is bounded.
Analogously, it can be shown that the derivatives of $f_h \circ g_2$ are polynomially bounded, with the bounds being independent of h (for $|h| \leq h_0$). □

Lemma 7.5 For all $k, l \in \mathbb{N}_0$ and all $x \in \mathbb{R}$, $\nu \in \mathbb{R}_+$ the following limits hold in the Schwartz space topology (see e.g. [Wer07]):

$$\lim_{h \to 0} \frac{1}{h} \cdot \frac{\partial^{k+l}}{\partial x^k \partial \nu^l} (\psi_{t,x+h,\nu} - \psi_{t,x,\nu}) = \frac{\partial^{k+l+1}}{\partial x^{k+1} \partial \nu^l} \psi_{t,x,\nu}$$

and

$$\lim_{h \to 0} \frac{1}{h} \cdot \frac{\partial^{k+l}}{\partial x^k \partial \nu^l} (\psi_{t,x,\nu+h} - \psi_{t,x,\nu}) = \frac{\partial^{k+l+1}}{\partial x^k \partial \nu^{l+1}} \psi_{t,x,\nu}$$

Proof. We start with the second claim. Convergence in $\mathcal{S}$ is equivalent to convergence in all of the seminorms $p_{m,n}$ of this locally convex space; see (7.1). Thus we have to show that for all $k, l, n, m \in \mathbb{N}_0$ the following converges holds

$$\sup_{y \in \mathbb{R}} (1 + |y|^m) \left| \frac{\partial^n}{\partial y^n} \left[\frac{1}{h} \cdot \frac{\partial^{k+l}}{\partial x^k \partial \nu^l} (\psi_{t,x,\nu+h} - \psi_{t,x,\nu})(y) - \frac{\partial^{k+l+1}}{\partial x^k \partial \nu^{l+1}} \psi_{t,x,\nu}(y) \right] \right| \longrightarrow 0$$

as $h \to 0$. With the derivatives from Lemma 7.3 the term in brackets can be written as

$$\begin{aligned} A_h(y) :&= e^{i(x+\mu t)y} g_1(y) e^{-\nu g_2(y)} (iy)^k (-g_2(y))^l \cdot \left(h^{-1} (e^{-h g_2(y)} - 1) + g_2(y) \right) \\ &= h \cdot e^{i(x+\mu t)y} g_1(y) (iy)^k (-g_2(y))^l \cdot \left[e^{-\nu g_2(y)} h^{-2} \left(e^{-h g_2(y)} - 1 + h g_2(y) \right) \right] \end{aligned}$$

By the previous lemma there exists $h_0 > 0$ such that for $h \in \mathbb{R} \setminus \{0\}$, $|h| \leq h_0$, the last term in brackets is in $C^\infty_{\text{pol}}(\mathbb{R})$ with the bounds being independent of h. Clearly all other factors are in the same set with g_1 being even a Schwartz–function. Hence $A_h(y) = h \cdot B_h(y)$ for some $B_h \in \mathcal{S}$ and thus

$$\sup_{y \in \mathbb{R}} (1 + |y|^m) \left| \frac{\partial^n}{\partial y^n} A_h(y) \right| \leq |h| \cdot \sup_{y \in \mathbb{R}} (1 + |y|^m) \left| \frac{\partial^n}{\partial y^n} B_h(y) \right|$$

which converges to 0. This proves the second claim.
The first assertion can be proven in the same way. Here the function A_h takes the form

$$\begin{aligned} A_h(y) &:= e^{i(x+\mu t)y} g_1(y) e^{-\nu g_2(y)} (iy)^k (-g_2(y))^l \cdot \big(h^{-1}(e^{ixh} - 1) - iy\big) \\ &= h \cdot e^{i(x+\mu t)y} g_1(y) e^{-\nu g_2(y)} (iy)^k (-g_2(y))^l \cdot \big[h^{-2}(e^{ixh} - 1 - ihy)\big] \end{aligned}$$

The term in brackets is

$$f_h(y) := h^{-2}\big(\cos(hy) - 1 + i(\sin(hy) - hy)\big)$$

With Taylor's theorem we can show that f_h is polynomially bounded: There exists $\Theta_1(hy), \Theta_2(hy) \in \mathbb{R}$ such that

$$|f_h(y)| = \big|h^{-2}\big(-(hy)^2 \cos(\Theta_1(hy)) - i(hy)^2 \sin(\Theta_2(hy))\big)\big| \le 2y^2$$

Analogously one can prove that all derivatives of f_h with respect to y are polynomially bounded, the bounds being independent of h. The proof concludes as above. □

Remember that the space of tempered distributions is defined as the dual $\mathcal{S}'$ of the Schwartz space; see e.g. [Wer07]. Each measurable and polynomially bounded function $g\colon \mathbb{R} \to \mathbb{C}$ yields a tempered distribution T_g defined as

$$T_g(\varphi) := \langle T_g, \varphi \rangle := \int_{\mathbb{R}} g(y)\varphi(y)dy, \qquad \varphi \in \mathcal{S}$$

Our function u is the application of the tempered distribution defined by the payoff f to the log-price density:

$$u(t, x, \nu) = E(f(X_{T-t}^{x,\nu})) = \int_{\mathbb{R}} f(y)\varphi_{T-t,x,\nu}(y)dy = \langle T_f, \varphi_{T-t,x,\nu} \rangle$$

The final theorem of this section proves that the application of an arbitrary tempered distribution to $\varphi_{T-t,x,\nu}$ yields a smooth function in x, ν.

Theorem 7.6 Let $0 \le t < T$. For each tempered distribution $F \in \mathcal{S}'$ the function

$$A\colon x, \nu \mapsto \langle F, \varphi_{T-t,x,\nu} \rangle$$

is in $C^\infty(\mathbb{R} \times \mathbb{R}_+)$. In particular, this holds for the function $u(t, \cdot, \cdot)$.

Proof. Let $\mathcal{F}$ denote the Fourier transformation $f \mapsto \int_{-\infty}^{\infty} e^{-isy} f(y)dy$. Then $\varphi_{T-t,x,\nu} = \mathcal{F}(\psi_{T-t,x,\nu})$ and $A(x, \nu) = \langle F \circ \mathcal{F}, \psi_{T-t,x,\nu} \rangle$. We claim that for every $k, l \in \mathbb{N}_0$ we have

$$\frac{\partial^{k+l}}{\partial x^k \partial \nu^l} A(x, \nu) = \left\langle F \circ \mathcal{F}, \frac{\partial^{k+l}}{\partial x^k \partial \nu^l} \psi_{T-t,x,\nu} \right\rangle$$

Assume this claim holds for some $k, l \in \mathbb{N}_0$. Then

$$\begin{aligned} &\lim_{h \to 0} \frac{1}{h} \cdot \left(\frac{\partial^{k+l}}{\partial x^k \partial \nu^l} A(x, \nu + h) - \frac{\partial^{k+l}}{\partial x^k \partial \nu^l} A(x, \nu) \right) \\ &= \lim_{h \to 0} \left\langle F \circ \mathcal{F}, \frac{1}{h} \left(\frac{\partial^{k+l}}{\partial x^k \partial \nu^l} (\psi_{T-t,x,\nu+h} - \psi_{T-t,x,\nu}) \right) \right\rangle \\ &= \left\langle F \circ \mathcal{F}, \frac{\partial^{k+l+1}}{\partial x^k \partial \nu^{l+1}} \psi_{T-t,x,\nu} \right\rangle \end{aligned}$$

by the continuity of $F \circ \mathcal{F}$ from $\mathcal{S}$ to $\mathbb{R}$ and Lemma 7.5. This was the induction step for $l \to l + 1$. The step for $k \to k + 1$ can be done in the same way. □

7.3 Bounds on the Derivatives of u

In this section we will prove the following bound on the derivatives of u. In the whole section we will assume that $f\colon \mathbb{R} \to \mathbb{R}$ is measurable and bounded and $\rho \in (-1, 1)$.

Theorem 7.7 Let $k, l \in \mathbb{N}_0$ with $l \leq 2$ and set $a_0 = 0$, $a_1 = 1$, and $a_2 = 3$. Assume $\nu > k/2 + l$. Then the partial derivatives of u are bounded in the following way

$$\frac{\partial^{k+l}}{\partial x_0^k \partial v_0^l} u(t, x_0, v_0) \leq c \cdot \left(1 + (T-t)^{-\frac{k}{2}} \left(1 + {v_0}^{-\frac{k}{2} - a_l}\right)\right)$$

We start with a lemma on the derivatives of σ_t. We write $\sigma_t(\sigma_0)$ to denote the square-root volatility process in dependence of its start value. We will denote derivatives with respect to σ_0 by ∂_σ and $\partial_{\sigma\sigma}$.

Lemma 7.8 For each $t \in [0, T]$ the process σ_t is twice differentiable with respect to the start value σ and

$$0 < \partial_\sigma \sigma_t \leq 1, \qquad |\partial_{\sigma\sigma} \sigma_t| \leq \frac{1}{2e} \sup_{s \in [0,t]} \sigma_s^{-1}$$

In particular, $|\sigma_t(\sigma) - \sigma_t(\tilde{\sigma})| \leq |\sigma - \tilde{\sigma}|$ for all $\sigma, \tilde{\sigma} > 0$.

Proof. Set $C := \kappa\lambda/2 - \theta^2/8$. From the integral equation of σ_t, see (2.6), we get

$$\partial_\sigma \sigma_t = 1 - \int_0^t \left(\frac{C}{\sigma_s^2} + \frac{\kappa}{2}\right) \partial_\sigma \sigma_s \, ds$$

and hence

$$\partial_\sigma \sigma_t = \exp\left(-\int_0^t \frac{C}{\sigma_s^2} + \frac{\kappa}{2} \, ds\right) \leq 1$$

From this formula we can easily compute the second derivative as

$$\begin{aligned}
\partial_{\sigma\sigma} \sigma_t &= \exp\left(-\int_0^t \frac{C}{\sigma_s^2} + \frac{\kappa}{2} ds\right) \cdot \int_0^t \frac{C}{2\sigma_s^3} ds \\
&\leq \frac{1}{2} e^{-\frac{\kappa t}{2}} \cdot \exp\left(-\int_0^t \frac{C}{\sigma_s^2} \, ds\right) \cdot \int_0^t \frac{C}{\sigma_s^2} \, ds \cdot \sup_{s \in [0,t]} \frac{1}{\sigma_s} \\
&\leq \frac{1}{2} e^{-\frac{\kappa t}{2} - 1} \cdot \sup_{s \in [0,t]} \frac{1}{\sigma_s}
\end{aligned}$$

because $x e^{-x} \leq e^{-1}$ for all $x \geq 0$. □

To prove Theorem 7.7 we rewrite u as composition

$$u(T-t, x_0, v_0) = E(f(X_t^{x_0, v_0})) = E(H(x_0, Y(\sqrt{v_0}), Z(\sqrt{v_0}))$$

with $Y = Y_t(\sigma_0) := \int_0^t \sigma_s^2(\sigma_0) ds$, $Z = Z_t(\sigma_0) := \int_0^t \sigma_s(\sigma_0) dW_s$ and

$$H(x_0, y, z) := E\big(f(X_t^{x_0, v_0}) \,\big|\, Y = y, Z = z\big)$$

Note that given Y and Z, the process $X_t^{x_0, v_0}$ does not depend on v_0 anymore and thus H does not depend on v_0; we will see this explicitly in (7.12).
Y and Z are both differentiable: For Y this is a trivial consequence of Lemma 7.8 and for Z it follows from the SDE of $v_t = \sigma_t^2$:

$$Z = \sigma_t^2 - \sigma_0^2 - \kappa\lambda t + \kappa \int_0^t \sigma_s^2 ds = \sigma_t^2 - \sigma_0^2 - \kappa\lambda t + \kappa Y \tag{7.10}$$

In the sequel we will need various bounds on the derivatives of Y and Z which are collected in the next lemma.

Lemma 7.9 1. Let $0 \le p < \nu$. Then $EY^{-p} \le c \cdot t^{-p} \cdot \sigma_0^{-2p}$.

2. For $p \ge 1$ we have $E|\partial_\sigma Y|^p \le c \cdot t^p \cdot (1 + \sigma_0^p)$.

3. For $p \ge 1$ we have $E|\partial_\sigma Z|^p \le c \cdot t^{p/2}$.

4. Let $p \ge 1$ such that $p/2 < \nu - 1$. Then

$$\|\partial_{\sigma\sigma} Y\|_p \le c \cdot t \cdot (1 + \sigma_0) \cdot \left(1 + \sigma_0^{-\frac{p+2}{p}}\right)$$

5. Let $p \ge 1$ such that $p/2 < \nu - 1$. Then

$$\|\partial_{\sigma\sigma} Z\|_p \le c \cdot \sqrt{t} \cdot \left(1 + \sigma_0^{-\frac{p+2}{p}}\right)$$

Proof. 1) For $0 \le p < \nu$ Jensen's inequality proves

$$EY^{-p} = t^{-p} \cdot E\left(\left(\frac{1}{t} \cdot \int_0^t \sigma_s^2 ds\right)^{-p}\right) \le t^{-p-1} \cdot E\int_0^t \sigma_s^{-2p} ds$$
$$\le t^{-p} \cdot \sup_{s\in[0,t]} E\sigma_s^{-2p} \le c \cdot t^{-p} \cdot \sigma_0^{-2p} \quad (7.11)$$

where the last inequality follows from Theorem 2.1.
2) By Jensen's inequality and Lemma 7.8 we have

$$E|\partial_\sigma Y|^p = E\left|\int_0^t 2\sigma_s \partial_\sigma \sigma_s ds\right|^p \le t^{p-1} \cdot \int_0^t 2E|\sigma_s \cdot \partial_\sigma \sigma_s|^p ds \le 2t^p \cdot \sup_{s\in[0,t]} E|\sigma_s|^p$$

and now the claim follows from Theorem 2.1.
3) An easy bound can be proven using (7.10): $\partial_\sigma Z = 2\sigma_t \partial_\sigma \sigma_t - 2\sigma_0 + \kappa \partial_\sigma Y$. As a consequence $\sup_{h\in[-\varepsilon,\varepsilon]} |\partial_\sigma Z(\sigma_0 + h)|^p$ is integrable for all $\varepsilon \in (0, \sigma_0)$ and $p \ge 0$. This allows us to exchange limit and expectation in the following equation:

$$E|\partial_\sigma Z|^p = E\left(\lim_{h\to 0}\left|\frac{Z(\sigma_0 + h) - Z(\sigma_0)}{h}\right|^p\right)$$
$$\le \sup_{|h|<\varepsilon} E\left|\frac{Z(\sigma_0 + h) - Z(\sigma_0)}{h}\right|^p$$

Now we apply the Burkholder–Davis–Gundy inequality to bound the difference quotient and derive a better bound.

$$\le c \cdot \sup_{|h|<\varepsilon} E\left(\int_0^t \left(\frac{\sigma_s(\sigma_0 + h) - \sigma_s(\sigma_0)}{h}\right)^2 ds\right)^{p/2}$$

Because $|\sigma_s(\sigma_0 + h) - \sigma_s(\sigma_0)| \le h$ by Lemma 7.8, this is bounded by $ct^{p/2}$.
4) The derivative is given by

$$\partial_{\sigma\sigma} Y = \int_0^t (\partial_\sigma \sigma_s)^2 + \sigma_s \cdot \partial_{\sigma\sigma} \sigma_s ds$$

Together with Lemma 7.8 this yields

$$E|\partial_{\sigma\sigma} Y|^p \le c \cdot t \cdot \left(1 + E\left(\sup_{s\in[0,t]} \sigma_s^p \cdot \sup_{s\in[0,t]} \sigma_s^{-p}\right)\right)$$

Now choose $\varepsilon > 0$ such that $p(1+\varepsilon)/2 < \nu - 1$ and set $q = (1+\varepsilon)/\varepsilon$. Thus

$$\|\partial_{\sigma\sigma} Y\|_p \le c \cdot t \cdot \left(1 + \left(\mathbb{E} \sup_{s\in[0,t]} \sigma_s^{pq}\right)^{\frac{1}{pq}} \cdot \left(\mathbb{E} \sup_{s\in[0,t]} \sigma_s^{-p(1+\varepsilon)}\right)^{\frac{1}{p(1+\varepsilon)}}\right)$$

By Theorem 2.1, these expectations are bounded respectively by $c(1+\sigma_0)$ and

$$c\left(1 + \sigma_0^{-\frac{p(1+\varepsilon)+2}{p(1+\varepsilon)}}\right) \le c \cdot \left(1 + \sigma_0^{-\frac{p+2}{p}}\right)$$

5) From (7.10) we get

$$\partial_{\sigma\sigma} Z = 2(\partial_\sigma \sigma_t)^2 + 2\sigma_t \cdot \partial_{\sigma\sigma}\sigma_t - 2 + \kappa \partial_{\sigma\sigma} Y$$

The terms $\sigma_t \cdot \partial_{\sigma\sigma}\sigma_t$ and $\partial_{\sigma\sigma} Y$ were bounded by $ct(1+\sigma_0)(1+\sigma_0^{-(p+2)/p})$ in part 4) and thus

$$\|\partial_{\sigma\sigma} Z\|_p \le c \cdot \left(1 + t \cdot (1+\sigma_0) \cdot \left(1 + \sigma_0^{-\frac{p+2}{p}}\right)\right)$$

Like in part 3) this allows to exchange limit and expectation when we derive a stronger bound using difference quotients. From Taylor's theorem we know that there exists $\Theta_1, \Theta_2 \in [0, h]$ such that

$$\begin{aligned}\sigma_s(\sigma_0 + h) - \sigma_s(\sigma_0) &= \partial_\sigma \sigma_s(\sigma_0) \cdot h + \partial_{\sigma\sigma}\sigma_s(\sigma_0 + \Theta_1) \cdot h^2 \\ \sigma_s(\sigma_0 - h) - \sigma_s(\sigma_0) &= -\partial_\sigma \sigma_s(\sigma_0) \cdot h + \partial_{\sigma\sigma}\sigma_s(\sigma_0 + \Theta_2) \cdot h^2\end{aligned}$$

Adding both rows and using the bound from Lemma 7.8 we get

$$|\sigma_s(\sigma_0+h) - 2\sigma_s(\sigma_0) + \sigma_s(\sigma_0-h)| = \left|\partial_{\sigma\sigma}\sigma_s(\sigma_0+\Theta) + \partial_{\sigma\sigma}\sigma_s(\sigma_0-\Theta)\right| \cdot h^2 \le c \cdot h^2 \cdot \sup_{s'\in[0,s]} \sigma_{s'}^{-1}$$

With the Burkhold–Davis–Gundy inequality we get

$$\begin{aligned}\|\partial_{\sigma\sigma} Z\|_p &= \lim_{h\to 0} \frac{1}{h^2} \|Z(\sigma_0 + h) - 2Z(\sigma_0) + Z(\sigma_0 - h)\|_p \\ &\le \lim_{h\to 0} c \cdot \left(\mathbb{E}\left(\int_0^t \frac{1}{h^4} \cdot |\sigma_s(\sigma_0 + h) - 2\sigma_s(\sigma_0) + \sigma_s(\sigma_0 - h)|^2 ds\right)^{p/2}\right)^{1/p} \\ &\le c \cdot \left(t^{p/2} \cdot \mathbb{E} \sup_{s\in[0,t]} \sigma_s^{-p}\right)^{1/p}\end{aligned}$$

This is bounded as required due to Theorem 2.1. □

It turns out that $\partial_\sigma Y$ can be used to absorb one term of $Y^{-1/2}$, decreasing the dependence on ν_0^{-1} by half a power. Additionally, the exploding behaviour of Y^{-1} for $t \to 0$ can be cancelled by sufficiently high powers of $\partial_\sigma Y$ and $\partial_\sigma Z$ (namely for $k + l/2 = p$ in the next lemma).

Lemma 7.10 Assume $k, l \in \mathbb{N}_0$ and $p \in \mathbb{R}$ with $k/2 \le p < \nu + k/2$. Then

$$\mathbb{E}|(1 + Y^{-p}) \cdot (\partial_\sigma Y)^k \cdot (\partial_\sigma Z)^l| \le c \cdot \left(1 + \nu_0^{\frac{k}{2}} + \nu_0^{-\left(p - \frac{k}{2}\right)}\right) \cdot t^{-\left(p - k - \frac{l}{2}\right)}$$

Proof. Using $0 \le |\partial_\sigma \sigma_s| \le 1$ and the Cauchy–Schwarz inequality we get

$$\partial_\sigma Y = \int_0^t \sigma_s \cdot \partial_\sigma \sigma_s ds \le \int_0^t \sigma_s ds \le \sqrt{t} \cdot \left(\int_0^t \sigma_s^2 ds\right)^{1/2} = \sqrt{tY}$$

Choose ε small enough such that $(1+\varepsilon)(p-k/2) < \nu$ and set $q = (1+\varepsilon)/\varepsilon$. Now we apply the previous line, Hölder's inequality and the bounds from the previous Lemma to get

$$\begin{aligned}
E|Y^{-p} \cdot (\partial_\sigma Y)^k \cdot (\partial_\sigma Z)^l| &\le c \cdot E|Y^{-(p-k/2)} \cdot t^{k/2} \cdot (\partial_\sigma Z)^l| \\
&\le c \cdot t^{k/2} \cdot \left(EY^{-(1+\varepsilon)(p-k/2)}\right)^{1/(1+\varepsilon)} \cdot \left(E(\partial_\sigma Z)^{ql}\right)^{1/q} \\
&\le c \cdot t^{k/2} \cdot t^{-(p-k/2)} \cdot (1+\nu_0^{-(p-k/2)}) \cdot t^{l/2} \\
&= c \cdot t^{-(p-k-l/2)} \cdot (1+\nu_0^{-(p-k/2)})
\end{aligned}$$

Furthermore

$$E|(\partial_\sigma Y)^k \cdot (\partial_\sigma Z)^l| \le \|(\partial_\sigma Y)^k\|_2 \cdot \|(\partial_\sigma Z)^l\|_2 \le c \cdot t^{k+\frac{l}{2}} \cdot (1+\sigma_0^k)$$

Adding both parts concludes the proof. □

Because the log-price has a very simple SDE we can compute an explicit formula for H: The conditional distribution of $\int_0^t \sigma_s dB_s$ given Y is $\mathcal{N}(0, Y)$. Thus if $N \sim \mathcal{N}(x_0 + \mu t - y/2 + \rho z, (1-\rho^2)y)$ is independent of $(W_s)_{s\in[0,T]}$, then

$$\begin{aligned}
H(x_0, y, z) &= E\left(f\left(x_0 + \mu t - \frac{1}{2}y + \rho z + \sqrt{1-\rho^2} \cdot \int_0^t \sigma_s dB_s\right)\right) \\
&= E(f(N)) = \int_{\mathbb{R}} (2\pi(1-\rho^2)y)^{-1/2} \cdot \exp\left(-\frac{\theta^2}{2(1-\rho^2)y}\right) \cdot f(x)dx \quad (7.12)
\end{aligned}$$

with $\theta = \theta(x_0, x, y, z) = x - x_0 - \mu t + y/2 - \rho z$.

Lemma 7.11 Let $k, l, m \in \mathbb{N}_0$. Then there exist $n \in \mathbb{N}$ and $c_i \in \mathbb{R}$ and $k_i, l_i \in \mathbb{N}_0$, $i = 1, \dots, n$ such that

$$\frac{\partial^{k+l+m}}{\partial x_0^k \partial y^l \partial z^m} H(x_0, y, z) = \int_{\mathbb{R}} (2\pi(1-\rho^2)y)^{-1/2} \cdot \sum_{i=1}^n c_i \frac{\theta^{k_i}}{y^{l_i}} \cdot \exp\left(-\frac{\theta^2}{2(1-\rho^2)y}\right) \cdot f(x)dx$$

and $k_i \le l_i$, $i = 1, \dots, n$, and

$$\min_{i=1,\dots,n} \frac{k_i}{2} - l_i = -\frac{k}{2} - l - \frac{m}{2}$$

Proof. We prove this result by induction on $k+l+m$. The induction start for $k = l = m = 0$ is obvious from (7.12). So suppose the lemma holds for k, l, m and consider the induction step $k \mapsto k+1$.

$$\frac{\partial^{k+1+l+m}}{\partial x_0^{k+1} \partial y^l \partial z^m} H = \partial_{x_0} \int_{\mathbb{R}} (2\pi(1-\rho^2)y)^{-1/2} \cdot \sum_{i=1}^n c_i \frac{\theta^{k_i}}{y^{l_i}} \cdot \exp\left(-\frac{\theta^2}{2(1-\rho^2)y}\right) \cdot f(x)dx$$

Due to the exponential decay of the exp-term we can move the additional derivative into the integral.[1] Thus

$$\begin{aligned}
\frac{\partial^{k+1+l+m}}{\partial x_0^{k+1} \partial y^l \partial z^m} H = \int_{\mathbb{R}} (2\pi(1-\rho^2)y)^{-1/2} \cdot \sum_{i=1}^n c_i \left[\frac{-k_i \theta^{k_i-1}}{y^{l_i}} + \frac{\theta^{k_i}}{y^{l_i}} \cdot \frac{\theta}{(1-\rho^2)y}\right] \\
\cdot \exp\left(-\frac{\theta^2}{2(1-\rho^2)y}\right) \cdot f(x)dx
\end{aligned}$$

[1] As a consequence of the dominated convergence theorem, given a two-parameter function g we can exchange differentiation and integration ($\partial_y \int_{\mathbb{R}} g(x,y)dx = \int_{\mathbb{R}} \partial_y g(x,y)dx$), if there exists an $\varepsilon > 0$ such that $\sup_{|h|<\varepsilon} |\partial_y g(x, y+h)| \in L^1(\mathbb{R})$.

From this expression it can be seen that the assertion holds for $k+1, l, m$.
Now consider the induction step $l \mapsto l+1$. As above we can move the additional derivative inside. With

$$\partial_y (2\pi(1-\rho^2)y)^{-1/2} = (2\pi(1-\rho^2)y)^{-1/2} \cdot \frac{-1}{2y}$$
$$\partial_y \frac{\theta^{k_i}}{y^{l_i}} = \frac{k_i \theta^{k_i-1}}{2y^{l_i}} - \frac{\theta^{k_i} l_i}{y^{l_i+1}}$$
$$\partial_y \exp\left(-\frac{\theta^2}{2(1-\rho^2)y}\right) = \frac{1}{2(1-\rho^2)}\left(\frac{-\theta}{y} + \frac{\theta^2}{y^2}\right) \cdot \exp\left(-\frac{\theta^2}{2(1-\rho^2)y}\right)$$

we can compute the derivative of the integrand and get

$$\frac{\partial^{k+l+1+m}}{\partial x_0^k \partial y^{l+1} \partial z^m} H = \int_{\mathbb{R}} (2\pi(1-\rho^2)y)^{-1/2}$$
$$\cdot \sum_{i=1}^{n} c_i \left[\frac{-\theta^{k_i}}{2y^{l_i+1}} + \frac{k_i \theta^{k_i-1}}{2y^{l_i}} - \frac{\theta^{k_i} l_i}{y^{l_i+1}} + \frac{1}{2(1-\rho^2)}\left(\frac{-\theta^{k_i+1}}{y^{l_i+1}} + \frac{\theta^{k_i+2}}{y^{l_i+2}}\right)\right]$$
$$\cdot \exp\left(-\frac{\theta^2}{2(1-\rho^2)y}\right) \cdot f(x)dx$$

This proves that the assertion holds for $k, l+1, m$.
Because x_0 and z both appear only in θ and $\partial_z \theta = \rho \partial_{x_0} \theta$, the induction step $m \mapsto m+1$ is the same as $k \mapsto k+1$ and the proof is complete. □

As a final preparation we need a bound for the derivatives of H.

Lemma 7.12 Let $k, l, m \in \mathbb{N}_0$. Then there exists a constant C such that for all $x_0, z \in \mathbb{R}$ and $y > 0$

$$\left|\frac{\partial^{k+l+m}}{\partial x_0^k \partial y^l \partial z^m} H(x_0, y, z)\right| \leq C \cdot \sup|f| \cdot \left(1 + y^{-(\frac{k}{2}+l+\frac{m}{2})}\right)$$

Proof. Let X be an $\mathcal{N}(0, (1-\rho^2)y)$-distributed random variable. Then

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(1-\rho^2)y}} \cdot \frac{|\theta|^{k_i}}{y^{l_i}} \cdot \exp\left(-\frac{\theta^2}{2(1-\rho^2)y}\right) dx = \frac{E(|X|^{k_i})}{y^{l_i}} = c y^{k_i/2 - l_i} \tag{7.13}$$

for some constant c depending on the k_i and ρ. The result now follows from Lemma 7.11.□

Proof (of Theorem 7.7). We will consider derivatives with respect to $\sigma_0 := \sqrt{v_0}$ instead of derivatives w.r.t. v_0 and show that the function $\tilde{u}(t, x_0, \sigma_0) := u(t, x_0, \sigma_0^2)$ satisfies

$$\frac{\partial^{k+l}}{\partial {x_0}^k \partial \sigma_0^l} \tilde{u}(t, x_0, \sigma_0) \leq c \cdot \left(1 + (T-t)^{-\frac{k}{2}} \left(1 + v_0^{\frac{1}{2}} + {v_0}^{-\frac{k}{2} - \tilde{a}_l}\right)\right)$$

with $\tilde{a}_0 = 0$, $\tilde{a}_1 = 1/2$, $\tilde{a}_2 = 2$. The result then follows from $\partial_{v_0} u = \partial_\sigma \tilde{u}/(2\sqrt{v_0})$ and $\partial_{v_0 v_0} u = (\partial_{\sigma\sigma}\tilde{u} \cdot {v_0}^{-1} - \partial_\sigma \tilde{u} \cdot {v_0}^{-3/2})/4$.
Because Y and Z do not depend on x_0, the case $l = 0$ follows from the previous lemma and Lemma 7.9:

$$\left|\frac{\partial^k}{\partial x_0^k} \tilde{u}(T-t, x_0, \sigma_0)\right| = E\left|\frac{\partial^k}{\partial x_0^k} H(x_0, Y, Z)\right|$$
$$\leq c \cdot \sup|f| \cdot E\left(1 + Y^{-k/2}\right) \leq c \cdot \sup|f| \cdot t^{-k/2} \cdot (1 + v_0^{-k/2})$$

Note that here and in the following, the assumption $\nu > k/2 + l$ makes sure that the arising inverse moments of Y can be bounded; see Lemma 7.9.

To bound the other derivatives we need the chain rule. First assume that $k = 0$, $l = 1$.

$$\begin{aligned}|\partial_\sigma \tilde{u}(T-t, x_0, \sigma_0)| &\le E|\partial_y H(x_0, Y, Z) \cdot \partial_\sigma Y| + E|\partial_z H(x_0, Y, Z) \cdot \partial_\sigma Z| \\ &\le c \cdot \sup|f| \cdot \Big(E\left|(1+Y^{-1}) \cdot \partial_\sigma Y\right| + E|(1+Y^{-1/2}) \cdot \partial_\sigma Z| \Big)\end{aligned}$$

Lemma 7.10 allows us to bound both expectations by $c \cdot (1 + v_0^{1/2} + v_0^{-1/2})$ which proves the assertion for $\partial_\sigma \tilde{u}$. The second derivative $\partial_{\sigma\sigma}\tilde{u}$ can be bounded as follows:

$$\begin{aligned}|\partial_{\sigma\sigma}\tilde{u}(T-t, x_0, \sigma_0)| &\le E|\partial_{yy} H \cdot (\partial_\sigma Y)^2| + E|\partial_{zz} H \cdot (\partial_\sigma Z)^2| + 2E|\partial_{yz} H \cdot \partial_\sigma Y \cdot \partial_\sigma Z| \\ &\quad + E|\partial_y H \cdot \partial_{\sigma\sigma} Y| + E|\partial_z H \cdot \partial_{\sigma\sigma} Z| \\ &\le c \cdot \sup|f| \cdot \Big(E\left|(1+Y^{-2}) \cdot (\partial_\sigma Y)^2\right| + E\left|(1+Y^{-1}) \cdot (\partial_\sigma Z)^2\right| \\ &\quad + E\left|(1+Y^{-3/2}) \cdot \partial_\sigma Y \cdot \partial_\sigma Z\right| \\ &\quad + E\left|(1+Y^{-1}) \cdot \partial_{\sigma\sigma} Y\right| + E\left|(1+Y^{-1/2}) \cdot \partial_{\sigma\sigma} Z\right| \Big)\end{aligned}$$

For the first three expectations we can again apply Lemma 7.10 to get a bound of $c(1+v_0+v_0^{-1})$. For the fourth expectation we use Hölder's inequality and Lemma 7.9:

$$\begin{aligned}E|(1+Y^{-1}) \cdot \partial_{\sigma\sigma} Y| &\le c \cdot \left(1 + E(Y^{-2})\right)^{1/2} \cdot \left(E(\partial_{\sigma\sigma} Y)^2\right)^{1/2} \\ &\le c \cdot (1 + t^{-1} v_0^{-1}) \cdot t \cdot (1 + v_0^{1/2} + v_0^{-1}) \\ &\le c \cdot (1 + v_0^{1/2} + v_0^{-2})\end{aligned}$$

Analogously we get a $c \cdot (1 + v_0^{-3/2})$-bound for the last expectation $E|(1+Y^{-1/2}) \cdot \partial_{\sigma\sigma} Z|$. Adding all terms together gives a $c(1 + v_0 + v_0^{-2})$-bound and the proof for the case $k = 0$, $l = 2$ is complete.
The cases with $k \ge 1$ follow analogously; one simply needs to replace $\tilde{u}$ by $\frac{\partial^k}{\partial x_0^k}\tilde{u}$ and H by $\frac{\partial^k}{\partial x_0^k} H$ and Y^{-p} by $Y^{-p-k/2}$ and v_0^{-p} by $v_0^{-p-k/2}$. □

7.4 Weak Error Expansion

One possible application of bounds on the function u is deriving the weak error rate. That is, we search for a number $\alpha \ge 0$ such that

$$|E(f(\hat{X}_T) - E(f(X_T))| \le c \cdot \Delta^\alpha$$

where Δ is the stepsize of the discretization scheme. This is connected to the function u via

$$|E(f(\hat{X}_T) - E(f(X_T))| = |E(u(\hat{X}_T, \hat{v}_T, T)) - u(x_0, v_0, 0)|$$

Remember that we only consider the one-dimensional standard Heston model. Because we want to use Kolmogorov's backward equation, we have to assume that $f \in C_0^2(\mathbb{R})$; see e.g. Theorem 8.1.1 in [Øk07]. This restriction will be removed, when we prove the weak rate.
We will use the drift-implicit Milstein scheme introduced in (6.8). For simplicity we only considered equidistant time discretizations in Chapter 6. We will allow an arbitrary discretization $0 = t_0 < \cdots < t_N = T$ in this chapter. Then the scheme is given by

$$\begin{aligned}\hat{v}_{n+1} = \hat{v}_n + \kappa(\lambda - \hat{v}_{n+1}) \cdot (t_{n+1} - t_n) + \theta\sqrt{\hat{v}_n} \cdot (W_{t_{n+1}} - W_{t_n}) \\ + \frac{\theta^2}{4}\left((W_{t_{n+1}} - W_{t_n})^2 - \Delta\right)\end{aligned}$$

We use the notation $\Delta := \max_{i=0,\dots,N-1}(t_{i+1} - t_i)$, $n(t) := \max\{n \in \{0,\dots,N\} : t_n \le t\}$, $\eta(t) := t_{n(t)}$, $\Delta_t = t - \eta(t)$ and $\tilde{W}_t := W_t - W_{\eta(t)}$ for an arbitrary Brownian motion W. Let us define the following processes

$$\begin{aligned}
\hat{X}_t &:= \hat{X}_{n(t)} + \Big(\mu - \frac{1}{2}\hat{v}_{n(t)}\Big)\Delta_t + \sqrt{\hat{v}_{n(t)}}\tilde{B}_t \\
\tilde{v}_t &:= \hat{v}_{n(t)} + \kappa\lambda\Delta_t + \theta\sqrt{\hat{v}_{n(t)}}\tilde{W}_t + \frac{\theta^2}{4}(\tilde{W}_t^2 - \Delta_t) \\
\hat{v}_t &:= \frac{1}{1+\kappa\Delta_t}\tilde{v}_t
\end{aligned} \tag{7.14}$$

By definition $\hat{X}_{t_n} = \hat{X}_n$, $\lim_{t \nearrow t_n} \hat{X}_t = \hat{X}_{t_n}$ and the same holds for $\hat{v}_t$. Furthermore, when restricted to an interval $[t_i, t_{i+1})$ the processes $\hat{X}_t$ and $\tilde{v}_t$ are Itō processes, as can be seen from

$$\begin{aligned}
\hat{X}_t &:= \hat{X}_{n(t)} + \int_{\eta(t)}^t \Big(\mu - \frac{1}{2}\hat{v}_{n(t)}\Big)ds + \int_{\eta(t)}^t \sqrt{\hat{v}_{n(t)}}\ d\tilde{B}_s \\
\tilde{v}_t &:= \hat{v}_{n(t)} + \int_{\eta(t)}^t \kappa\lambda\ ds + \int_{\eta(t)}^t \theta\left(\sqrt{\hat{v}_{n(t)}} + \frac{\theta}{2}\tilde{W}_s\right)d\tilde{W}_s
\end{aligned}$$

After these preparations we can expand the weak error into local errors, following [TT90].

$$|E(f(\hat{X}_N)) - E(f(X_T))| = \left|\sum_{n=1}^N E\big(u(t_n, \hat{X}_n, \hat{v}_n) - u(t_{n-1}, \hat{X}_{n-1}, \hat{v}_{n-1})\big)\right|$$

With the help of the function $\tilde{u}(t,x,v) := u(t,x,v/(1+\kappa\Delta_t))$ and above Itō processes we can rewrite the local error in a form that allows to use Itō's formula. To make notation simpler we will usually omit the arguments of $\tilde{u}(t,\hat{X}_t,\tilde{v}_t)$ and $u(t,\hat{X}_t,\hat{v}_t)$.

$$\begin{aligned}
e_n &:= E\big(u(t_{n+1},\hat{X}_{n+1},\hat{v}_{n+1}) - u(t_n,\hat{X}_n,\hat{v}_n)\big) \\
&= E\big(\tilde{u}(t_{n+1},\hat{X}_{t_{n+1}},\tilde{v}_{t_{n+1}}) - \tilde{u}(t_n,\hat{X}_{t_n},\tilde{v}_{t_n})\big) \\
&= \int_{t_n}^{t_{n+1}} E\left[\partial_t\tilde{u}(t,\hat{X}_t,\tilde{v}_t) + \left(\mu - \frac{\hat{v}_n}{2}\right)\partial_x\tilde{u} + \kappa\lambda\partial_v\tilde{u} + \frac{\hat{v}_n}{2}\partial_{xx}\tilde{u}\right. \\
&\qquad \left. + \sqrt{\hat{v}_n}\rho\theta\left(\sqrt{\hat{v}_n} + \frac{\theta}{2}\tilde{W}_t\right)\partial_{xv}\tilde{u} + \frac{\theta^2}{2}\left(\sqrt{\hat{v}_n} + \frac{\theta}{2}\tilde{W}_t\right)^2\partial_{vv}\tilde{u}\right]\ dt
\end{aligned}$$

Now we replace the derivatives of $\tilde{u}$ by the corresponding derivatives of u,

$$\begin{aligned}
\partial_t\tilde{u}(t,\hat{X}_t,\tilde{v}_t) &= \partial_t u(t,\hat{X}_t,\hat{v}_t) - \frac{\kappa\hat{v}_t}{1+\kappa\Delta_t}\cdot\partial_v u(t,\hat{X}_t,\hat{v}_t) \\
\frac{\partial^{k+l}}{\partial x^k\partial v^l}\tilde{u}(t,\hat{X}_t,\tilde{v}_t) &= \frac{1}{(1+\kappa\Delta_t)^l}\cdot\frac{\partial^{k+l}}{\partial x^k\partial v^l}u(t,\hat{X}_t,\hat{v}_t)
\end{aligned}$$

to get

$$\begin{aligned}
e_n = \int_{t_n}^{t_{n+1}} E\Bigg[\ &\partial_t u(t,\hat{X}_t,\hat{v}_t) + \left(\mu - \frac{\hat{v}_n}{2}\right)\partial_x u + \frac{\kappa(\lambda - \hat{v}_t)}{1+\kappa\Delta_t}\partial_v u + \frac{\hat{v}_n}{2}\partial_{xx}u \\
&+ \frac{\rho\theta\sqrt{\hat{v}_n}}{1+\kappa\Delta_t}\left(\sqrt{\hat{v}_n} + \frac{\theta}{2}\tilde{W}_t\right)\partial_{xv}u + \frac{\theta^2}{2(1+\kappa\Delta_t)^2}\left(\sqrt{\hat{v}_n} + \frac{\theta}{2}\tilde{W}_t\right)^2\partial_{vv}\Bigg]\ dt
\end{aligned}$$

To remove $\partial_t u$ we use Kolmogorov's backward equation.

$$\partial_t u = -\left(\mu - \frac{v}{2}\right)\partial_x u - \kappa(\lambda - v)\partial_v u - \frac{1}{2}v\partial_{xx}u - \rho\theta v\partial_{xv}u - \frac{\theta^2}{2}v\partial_{vv}u$$

Using additionally $(\sqrt{\hat{v}_n} + \frac{\theta}{2}\tilde{W}_t)^2 = \tilde{v}_t - (\kappa\lambda - \theta^2/4)\Delta_t$ we come up with

$$\begin{aligned} e_n = \int_{t_n}^{t_{n+1}} E\Bigg[& \frac{1}{2}(\hat{v}_t - \hat{v}_n)\partial_x u + \kappa(\lambda - \hat{v}_t)\left(\frac{1}{1+\kappa\Delta_t} - 1\right)\partial_v u \\ & + \frac{1}{2}(\hat{v}_n - \hat{v}_t)\,\partial_{xx} u + \rho\theta\left(\frac{\hat{v}_n}{1+\kappa\Delta_t} - \hat{v}_t + \frac{\theta\sqrt{\hat{v}_n}\tilde{W}_t}{2(1+\kappa\Delta_t)}\right)\partial_{xv} u \\ & + \frac{\theta^2}{2}\left(\left(\frac{1}{1+\kappa\Delta_t} - 1\right)\hat{v}_t - \frac{4\kappa\lambda - \theta^2}{4(1+\kappa\Delta_t)^2}\Delta_t\right)\partial_{vv} u\Bigg] dt \end{aligned}$$

Finally, we use $1 - 1/(1+\kappa\Delta_t) = \kappa\Delta_t/(1+\kappa\Delta_t)$ and

$$\hat{v}_t - \hat{v}_n = (1+\kappa\Delta_t)\hat{v}_t - \hat{v}_n - \kappa\Delta_t\hat{v}_t = \kappa\Delta_t(\lambda - \hat{v}_t) + \theta\sqrt{\hat{v}_n}\tilde{W}_t + \frac{\theta^2}{4}(\tilde{W}_t^2 - \Delta_t)$$

We sort terms in groups containing either Δ_t or $\sqrt{\hat{v}_n}\tilde{W}_t$ or $\tilde{W}_t^2 - \Delta_t$ and end up with

$$e_n = e_n^{(1)} + e_n^{(2)} + e_n^{(3)}$$

with

$$\begin{aligned} e_n^{(1)} &= \int_{t_n}^{t_{n+1}} \Delta_t \cdot E\left[\frac{\kappa^2}{1+\kappa\Delta_t}(\hat{v}_t - \lambda)\partial_v u - \frac{\theta^2}{2(1+\kappa\Delta_t)}\left(\kappa\hat{v}_t + \frac{4\kappa\lambda - \theta^2}{4(1+\kappa\Delta_t)}\right)\partial_{vv} u \right. \\ &\qquad \left. + \frac{\kappa}{2}(\lambda - \hat{v}_t)(\partial_x u - \partial_{xx} u) - \frac{\rho\theta\kappa\lambda}{1+\kappa\Delta_t}\partial_{xv} u\right] dt \\ e_n^{(2)} &= \int_{t_n}^{t_{n+1}} E\left[\sqrt{\hat{v}_n}\tilde{W}_t \cdot \left(\frac{\theta}{2}\partial_x u - \frac{\theta}{2}\partial_{xx} u - \frac{\rho\theta^2}{2(1+\kappa\Delta_t)}\partial_{xv} u\right)\right] dt \\ e_n^{(3)} &= \int_{t_n}^{t_{n+1}} E\left[(\tilde{W}_t^2 - \Delta_t) \cdot \left(\frac{\theta^2}{8}\partial_x u - \frac{\theta^2}{8}\partial_{xx} u - \frac{\theta^3\rho}{4(1+\kappa\Delta_t)}\partial_{xv} u\right)\right] dt \end{aligned}$$

7.5 Weak Error Rate

Using the error expansion derived in the last section we can now prove a rate of 1 for the weak error.

Theorem 7.13 Assume that $\nu > 9/2$. Then there exists a constant C such that for all measurable and bounded $f\colon \mathbb{R} \to \mathbb{R}$ the weak error is bounded by

$$|Ef(\hat{X}_N) - Ef(X_T)| \le C \cdot \sup|f| \cdot \Delta$$

Proof. First assume that $f \in C_0^2(\mathbb{R})$, so that we can use the error expansion from the last section. Using Lemmas 7.17 to 7.19 below, we can bound $e_n^{(1)}, e_n^{(2)}, e_n^{(3)}$ by

$$e_n^{(i)} \le \int_{t_n}^{t_{n+1}} c \cdot \sup|f| \cdot (t(T-t))^{-1/2} \cdot \Delta\, dt$$

Thus

$$|Ef(\hat{X}_N) - Ef(X_T)| = \left|\sum_{n=0}^{N-1} e_n\right| \le c \cdot \sup|f| \cdot \int_0^T (t(T-t))^{-1/2} dt \cdot \Delta = c \cdot \sup|f| \cdot \Delta$$

The result for general f follows now from a simple approximation procedure. □

Three lemmas are missing to complete the preceding proof. Common to them is the difficulty that for irregular f the derivatives $\partial_x u$, $\partial_{xx} u$ and $\partial_{xv} u$ become infinitely large when $t \to T$. Lemma 7.16 uses the integration by parts rule from Malliavin calculus to remove the derivatives w.r.t. x and will be the key to complete the proof.
First we need a bound on the inverse moments of $\hat{v}_t$. Note that for $\nu > 3$ this result is stronger than the corresponding statement in Theorem 6.5.

Theorem 7.14 Assume $0 \le p \le \nu - 1$. Then

$$\sup_{t\in[0,T]} E\hat{v}_t^{-p} \le c \cdot v_0^{-p}$$

Proof. To use Itō's formula for the function $x \mapsto x^{-p}$ we need to localize the process. For $\varepsilon > 0$ set $\tau_\varepsilon := \inf\{t \ge 0 : \tilde{v}_t < \varepsilon\}$. First note that for $u \in [0, T]$

$$\left(\sqrt{\hat{v}_{\eta(u)}} + \frac{\theta}{2}\tilde{W}_u\right)^2 = \tilde{v}_u - \left(\kappa\lambda - \frac{\theta^2}{4}\right)\Delta_u \le \tilde{v}_u$$

Using this equation and the Itō formula we get

$$\begin{aligned}
E\tilde{v}_{t\wedge\tau_\varepsilon}^{-p} &= E\tilde{v}_{\eta(t\wedge\tau_\varepsilon)}^{-p} - p\kappa\lambda \cdot E\int_{\eta(t\wedge\tau_\varepsilon)}^{t\wedge\tau_\varepsilon} \tilde{v}_u^{-p-1}du \\
&\qquad + \frac{p(p+1)\theta^2}{2} \cdot E\int_{\eta(t\wedge\tau_\varepsilon)}^{t\wedge\tau_\varepsilon} \tilde{v}_u^{-p-2} \cdot \left(\sqrt{\hat{v}_{\eta(t\wedge\tau_\varepsilon)}} + \frac{\theta}{2}\tilde{W}_u\right)^2 du \\
&\le E\tilde{v}_{\eta(t\wedge\tau_\varepsilon)}^{-p} + p\left((p+1)\frac{\theta^2}{2} - \kappa\lambda\right) \cdot E\int_{\eta(t\wedge\tau_\varepsilon)}^{t\wedge\tau_\varepsilon} \tilde{v}_u^{-p-1}du \\
&\le E\tilde{v}_{\eta(t\wedge\tau_\varepsilon)}^{-p}
\end{aligned}$$

because the assumption $p + 1 \le \nu$ implies that the second summand is negative. Thus

$$E\hat{v}_{t\wedge\tau_\varepsilon}^{-p} = (1 + \kappa\Delta_t)^p E\tilde{v}_{t\wedge\tau_\varepsilon}^{-p} \le e^{p\kappa\Delta_t} \cdot E\tilde{v}_{\eta(t\wedge\tau_\varepsilon)}^{-p}$$

Now induction over the discretization subintervals shows

$$\sup_{t\in[0,T]} E\hat{v}_{t\wedge\tau_\varepsilon}^{-p} \le e^{p\kappa T} v_0^{-p}$$

With Fatou's lemma the claim follows for $E\hat{v}_t^{-p}$. □

As described in Chapter 4, Malliavin calculus in the Heston model is based on two independent Brownian motions Z^1, Z^2. The SDEs of the model are driven by $W = Z^2$ and $B = \rho Z^2 + \rho' Z^1$. In the following D_r will always denote the Malliavin derivative with respect to Z^1. Because Z^1 is independent of v_t, the derivative $D_r\hat{v}_t$ is 0, while the derivative of $\hat{X}_t$ can easily be computed from (7.14) as

$$D_r\hat{X}_t = \rho'\sqrt{\hat{v}_{\eta(r)}} \cdot \mathbb{1}_{[0,t]}(r) \tag{7.15}$$

For $t \in [0, T]$ define $I_0(t) := 1$,

$$\begin{aligned}
I_1(t) &:= \int_0^t \hat{v}_{\eta(r)}^{-1/2} dZ_r^1 \\
I_2(t) &:= \left(\int_0^t \hat{v}_{\eta(r)}^{-1/2} dZ_r^1\right)^2 - \int_0^t \hat{v}_{\eta(r)}^{-1} dr
\end{aligned}$$

The previous theorem shows that I_1 and I_2 are well-defined if $\nu \ge 2$.

Proposition 7.15 Let $k \in \{0,1,2\}$, $l \in \mathbb{N}_0$, $p \geq 1$ and $t, t' \in [0,T]$. Further assume that

$$\nu \geq \max\{2, p(k/2 + l) + 1\}$$

Then we have

$$\mathbb{E}\left|\frac{I_k(t)}{\hat{v}_{t'}^l}\right|^p \leq c \cdot t^{\frac{pk}{2}}$$

Proof. The case $k = 0$ follows directly from Theorem 7.14. The same is true for the case $l = 0$, using additionally the Burkholder–Davis–Gundy inequality.
Now consider the case $k = 1$, $l > 0$. Setting $q_1 = 1 + 2l$, $q_2 = (1 + 2l)/(l2)$, Hölder's inequality and the Burkholder–Davis–Gundy inequality give

$$\mathbb{E}\left|\int_0^t \hat{v}_{\eta(r)}^{-1/2} dZ_r^1 \cdot \hat{v}_{t'}^{-l}\right|^p \leq \left(\mathbb{E}\left(\int_0^t \hat{v}_{\eta(r)}^{-1} dr\right)^{\frac{pq_1}{2}}\right)^{\frac{1}{q_1}} \cdot \left(\mathbb{E}\hat{v}_{t'}^{-lpq_2}\right)^{\frac{1}{q_2}}$$

If $pq_1/2 \geq 1$ we use Jensen's inequality and get

$$\leq t^{p/2} \sup_{r\in[0,t]} \left(\mathbb{E}\hat{v}_{\eta(r)}^{-p(\frac{1}{2}+l)}\right)^{\frac{1}{q_1}} \cdot \left(\mathbb{E}\hat{v}_{t'}^{-p(\frac{1}{2}+l)}\right)^{\frac{1}{q_2}}$$

This is bounded by our assumption and Theorem 7.14. If $pq_1/2 < 1$ we use also Jensen's inequality, this time to move the exponent out of the expectation. To bound the arising $\mathbb{E}\hat{v}_t^{-1}$ we need the assumption that $\nu > 2$.
The case $k = 2$, $l > 0$ is handled analogously. □

Lemma 7.16 Let $k \in \{1,2\}$ and let G be a random variable independent from $(Z_t^1)_{t\in[0,T]}$ such that $G \in L^p(\Omega)$ for all $p \geq 1$. Let $h\colon \mathbb{R} \to \mathbb{R}$ be k-times continuously differentiable. Then

$$\mathbb{E}\left(\frac{\partial^k}{\partial x^k} h(\hat{X}_{\eta(r)}) \cdot G\right) = (t\rho')^{-k} \cdot \mathbb{E}\left(h(\hat{X}_{\eta(r)}) \cdot G \cdot I_k\right)$$

Proof. The chain rule Theorem 3.17 and (7.15) allow us to rewrite the expectation on the left-hand side

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^k}{\partial x^k} h(\hat{X}_{\eta(r)}) \cdot G\right) &= t^{-1} \cdot \mathbb{E}\left(\int_0^t \frac{\partial^k}{\partial x^k} h(\hat{X}_{\eta(r)}) \cdot G \cdot D_r \hat{X}_{\eta(r)} \cdot \frac{1}{D_r \hat{X}_{\eta(r)}} dr\right) \\ &= (t\rho')^{-1} \cdot \mathbb{E}\left(\int_0^t D_r\left(\frac{\partial^{k-1}}{\partial x^{k-1}} h(\hat{X}_{\eta(r)})\right) \cdot G \cdot \hat{v}_{\eta(r)}^{-1/2} dr\right)\end{aligned}$$

Now we can apply the integration by parts rule from Malliavin calculus; see (3.8).

$$= (t\rho')^{-1} \cdot \mathbb{E}\left(\frac{\partial^{k-1}}{\partial x^{k-1}} h(\hat{X}_{\eta(r)}) \cdot \int_0^t G \cdot \hat{v}_{\eta(r)}^{-1/2} \delta Z_r^1\right)$$

Using Proposition 3.23, the Skorohod integral can be computed as GI_1. The conditions of the proposition are satisfied by the previous lemma.
If $k = 2$, we can repeat this procedure to get

$$\mathbb{E}\left(\partial_x h(\hat{X}_{\eta(r)}) \cdot GI_1\right) = (t\rho')^{-1} \cdot \mathbb{E}\left(h(\hat{X}_{\eta(r)}) \cdot \int_0^t GI_1 \cdot \hat{v}_{\eta(r)}^{-1/2} \delta Z_r^1\right)$$

Proposition 3.23 again shows that the Skorohod integral is exactly GI_2. Put together, this shows the claim. □

Lemma 7.17 Assume $\nu \geq 4$. For $k + l \leq 2$ and $G \in \{1, \hat{v}_t\}$ there is a constant C, independent of t, such that

$$\left| E\ G \cdot \frac{\partial^{k+l}}{\partial x^k \partial v^l} u(t, \hat{X}_t, \hat{v}_t) \right| \leq C \cdot \sup |f|$$

Proof. First assume $k = 0$ or $t \leq T/2$. By Theorem 7.7 we have

$$\left| E\ G \cdot \frac{\partial^{k+l}}{\partial x^k \partial v^l} u(t, \hat{X}_t, \hat{v}_t) \right| \leq c \cdot \sup |f| \cdot \left(1 + (T-t)^{-\frac{k}{2}} \cdot E \left| G \left(1 + \hat{v}_t^{-\frac{k}{2} - a_l}\right)\right|\right)$$

The expectation is finite, if $\nu \geq 4$; see Theorem 7.14.
In the case $k > 0$ and $t > T/2$ we first use the previous lemma to remove the derivatives w.r.t. x (which would cause a $(T-t)^{-k/2}$-term) and then apply Theorem 7.7:

$$\begin{aligned} \left| E\ G \cdot \frac{\partial^{k+l}}{\partial x^k \partial v^l} u(t, \hat{X}_t, \hat{v}_t) \right| &= (t\rho')^{-k} \cdot \left| E\ G \cdot I_k \cdot \frac{\partial^l}{\partial v^l} u(t, \hat{X}_t, \hat{v}_t) \right| \\ &\leq c \cdot \sup |f| \cdot E \left| G \cdot I_k \cdot (1 + \hat{v}_t^{-a_l}) \right| \end{aligned}$$

This expectation is finite by Proposition 7.15. □

Lemma 7.18 Assume $\nu > 9/2$. For $* \in \{x, xx, xv\}$ and $G \in \{1, \hat{v}_t\}$ there is a constant C, independent of t, such that

$$\left| E\ \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot \partial_* u(t, \hat{X}_t, \hat{v}_t) \right| \leq C \cdot \sup |f| \cdot \Delta \cdot (t(T-t))^{-1/2}$$

Proof. 1.) The main trick will be to use the mean value Theorem. However, this will introduce an additional derivative in both directions. Because each derivative w.r.t. x implies a $(T-t)^{-1/2}$-term in Theorem 7.7, we will first remove all of them using Lemma 7.16.

$$\left| E\ \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot \frac{\partial^{k+l}}{\partial x^k \partial v^l} u(t, \hat{X}_t, \hat{v}_t) \right| = (t\rho')^{-k} \cdot \left| E\ \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot I_k \cdot \frac{\partial^l}{\partial v^l} u(t, \hat{X}_t, \hat{v}_t) \right|$$

Now we apply Taylor's theorem

$$\begin{aligned} &\leq (t\rho')^{-k} \cdot \Bigg[\left| E\ \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot I_k \cdot \frac{\partial^l}{\partial v^l} u(t, \hat{X}_n, \hat{v}_n) \right| \\ &\quad + \left| E\ \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot I_k \cdot \frac{\partial^{l+1}}{\partial x \partial v^l} u(t, \Theta_1, \Theta_2) \cdot (\hat{X}_t - \hat{X}_n) \right| \\ &\quad + \left| E\ \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot I_k \cdot \frac{\partial^{l+1}}{\partial v^{l+1}} u(t, \Theta_1, \Theta_2) \cdot (\hat{v}_t - \hat{v}_n) \right| \Bigg] \end{aligned}$$

for some random variables Θ_1, Θ_2 with $0 < \min\{\hat{v}_{\eta(t)}, \hat{v}_t\} \leq \Theta_2$. The first expectation is zero because $\tilde{W}_t$ is independent of the remaining terms. We will skip the second expectation because it is similar to the third one. By Theorem 7.7 and Hölder's inequality the third expectation is smaller than

$$c \cdot \sup |f| \cdot (t\rho')^{-k} \cdot \left\| I_k \cdot \left(1 + \Theta_2^{-a_{l+1}}\right) \right\|_{1+\varepsilon} \cdot \left\| \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot (\hat{v}_t - \hat{v}_n) \right\|_{(1+\varepsilon)/\varepsilon}$$

The worst case for the first norm is $k = l = 1$, which makes $a_{l+1} = 3$. To apply Proposition 7.15 we thus need $\nu > 9/2$ and ε small enough. We get a bound of $c \cdot t^{k/2}$. For the second

norm we use Hölder's inequality and $\|\hat{v}_t - \hat{v}_{\eta(t)}\|_q \le c \cdot \sqrt{\Delta}$ for all $q \ge 1$; this follows from the definition of $\hat{v}_t$ and the Burkholder–Davis–Gundy inequality. We get for arbitrary $q \ge 1$

$$\left\| \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot (\hat{v}_t - \hat{v}_n) \right\|_q \le \|\hat{v}_n\|_{3q} \cdot \|\tilde{W}_t\|_{3q} \cdot \|\hat{v}_t - \hat{v}_{\eta(t)}\|_{3q} \le c\Delta$$

Summarizing, we have shown

$$\left| E\, \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot \frac{\partial^{k+l}}{\partial x^k \partial v^l} u(t, \hat{X}_t, \hat{v}_t) \right| \le c \cdot \sup|f| \cdot t^{-k/2} \cdot \Delta$$

2.) Only in case $k = 2$, $t < T/2$ this does not suffice to prove the assertion. The solution is to remove only a single x-derivative, instead of removing both:

$$\begin{aligned}
\left| E\, \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot \partial_{xx} u(t, \hat{X}_t, \hat{v}_t) \right| &= (t\rho')^{-1} \cdot \left| E\, \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot I_1 \cdot \partial_x u(t, \hat{X}_t, \hat{v}_t) \right| \\
&\le (t\rho')^{-1} \cdot \Bigg[\left| E\, \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot I_1 \cdot \partial_x u(t, \hat{X}_n, \hat{v}_n) \right| \\
&\quad + \left| E\, \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot I_1 \cdot \partial_{xx} u(t, \Theta_1, \Theta_2) \cdot (\hat{X}_t - \hat{X}_n) \right| \\
&\quad + \left| E\, \sqrt{\hat{v}_n} \cdot \tilde{W}_t \cdot I_1 \cdot \partial_{xv} u(t, \Theta_1, \Theta_2) \cdot (\hat{v}_t - \hat{v}_n) \right| \Bigg]
\end{aligned}$$

This can be bounded by $c \cdot \sup|f| \cdot t^{-1/2} \cdot \Delta$ similarly to the first part. Note that the ∂_{xx}-derivative introduces a term $(T-t)^{-1}$. Thus, in case $k = 2$ and $t > T/2$ we really need to remove all x-derivatives as was done in the first part. □

Lemma 7.19 Assume $\nu > 5/2$. For $* \in \{x, xx, x\nu\}$ and $G \in \{1, \hat{v}_t\}$ there is a constant C, independent of t, such that

$$\left| E\, (\tilde{W}_t^2 - \Delta_t) \cdot \frac{\partial^{k+l}}{\partial x^k \partial v^l} u(t, \hat{X}_t, \hat{v}_t) \right| \le C \cdot \sup|f| \cdot \Delta$$

Proof. First consider the case $k = 0$ or $t \le T/2$. By Hölder's inequality and Theorem 7.7 we get for every $\varepsilon > 0$

$$\begin{aligned}
&\left| E\, (\tilde{W}_t^2 - \Delta_t) \cdot \frac{\partial^{k+l}}{\partial x^k \partial v^l} u(t, \hat{X}_t, \hat{v}_t) \right| \\
&\qquad \le c \cdot \sup|f| \cdot \|\tilde{W}_t^2 - \Delta_t\|_{\frac{1+\varepsilon}{\varepsilon}} \cdot \left(1 + (T-t)^{-\frac{k}{2}} \cdot \left(1 + E\hat{v}_t^{-(1+\varepsilon)\left(\frac{k}{2} - a_l\right)} \right)^{\frac{1}{1+\varepsilon}} \right)
\end{aligned}$$

If ε is chosen small enough, the last expectation is finite by Theorem 7.14 (because $a_l \le 1$, the assumption $\nu > 5/2$ is sufficient). Thus we get

$$\le c \cdot \sup|f| \cdot \Delta \cdot \left(1 + (T-t)^{-k/2} \right)$$

In case $k > 0$ and $t > T/2$ we need to remove the derivatives with respect to x first using Lemma 7.16.

$$\left| E(\tilde{W}_t^2 - \Delta_t) \cdot \frac{\partial^k + l}{\partial x^k \partial v^l} u(t, \hat{X}_t, \hat{v}_t) \right| = (t\rho')^{-k/2} \cdot E \left| (\tilde{W}_t^2 - \Delta_t) \cdot I_k \cdot \frac{\partial^l}{\partial v^l} u(t, \hat{X}_t, \hat{v}_t) \right|$$

This can be bounded analogously to the first part. □

The last three lemmas complete the proof of Theorem 7.13.

Chapter 8

Quadrature of Discontinuous Functionals

In practice one will try to approximate functionals like $E(f(S_T))$ using multilevel Monte-Carlo techniques. Assuming a bias of order 1, the convergence order can reach up to $1/2$ instead of $1/3$ as with standard Monte-Carlo. The multilevel algorithm will be presented in the next chapter where we will also learn that a good rate of L^2-convergence is important to benefit from the algorithm. In this chapter we will first prove that this L^2-convergence order is very low if the functional f is discontinuous. Then we will discuss an approach based on Malliavin calculus to replace f by a smooth functional and thus regain the order that is achieved for smooth functionals. We remark that [GNR14] study a similar smoothing technique — without Malliavin calculus — in the case of distribution functions and densities. Because it is the only scheme with a guaranteed strong convergence order of $1/2$ for a wide range of parameters, we will always use the scheme DISE to approximate ν_t.

Theorem 8.1 Let $K > 0$. Assume either $\gamma > 1/2$ or $\nu > 1$. Define

$$r_0 := \begin{cases} 1/4 & \text{if } \gamma > 1/2 \\ \frac{\nu}{4\nu+3} & \text{else} \end{cases}$$

Then

$$\left(E \left| \mathbb{1}_{(-\infty,K]}(X_T) - \mathbb{1}_{(-\infty,K]}(\hat{X}_N) \right|^2 \right)^{1/2} \leq c_r \cdot \Delta^r$$

for all $0 \leq r < r_0$.

Proof. From Theorem 2.4 in [Avi09] we get

$$\left(E \left(\mathbb{1}_{(-\infty,K]}(X_T) - \mathbb{1}_{(-\infty,K]}(\hat{X}_N) \right)^2 \right)^{1/2} \leq c_p \cdot \left(E \left| X - \hat{X} \right|^p \right)^{\frac{p}{2(p+1)}}$$

for all $p \geq 1$. By Theorem 6.6 this is bounded by

$$\leq c_p \cdot \Delta^{\frac{p}{4(p+1)}}$$

if either $\gamma > 1/2$ or $p < 4\nu/3$. To get the convergence rate $r = p/(4(p+1))$ arbitrarily close to r_0 we just need to choose a sufficiently high p. □

In Theorem 7.2 of [Avi09] it is proven that even in the simple case of the Euler scheme and geometric Brownian motion, an L^2-convergence order[1] of $1/4$ is optimal for the payoff $\mathbb{1}_{(-\infty,K]}$.

[1] Avikainen uses the L^1-norm. Replacing $\|\cdot\|_1$ by $\|\cdot\|_2^2$ in his proof shows the result stated here.

8.1 An Integration by Parts Formula

This section presents a simple integration by parts formula for the Heston model that will be the central tool to compute $E(f(S_T))$ efficiently even for discontinuous f. Note that we smooth the functional only in one direction. For simplicity we choose the first direction, but by reordering the price processes one can smooth in other directions. However, when one does so, one has to change the underlying Brownian motions according to Lemma 5.3. The reason is that we heavily rely on the fact that $D_r^{Z^1} X_t^1 = R_{11}\sqrt{v_r^1}\mathbb{1}_{[0,t]}(r)$ is particularly simple, while $D_r^{Z^1} X_t^i = D_r^{Z^1} v_t^j = 0$ for $i = 2,\dots,d$ and $j = 1,\dots,d$; see Lemma 5.4.

Theorem 8.2 Assume that $f\colon \mathbb{R}_+^d \to \mathbb{R}$ is P_X-a.s. continuous and set $F(x) := F(x_1,\dots,x_d) = \int_0^{x_1} f(\xi, x_2,\dots,x_d)d\xi$. If there exists an $\varepsilon > 0$ such that

$$f(S_T) \in L^1(\Omega) \qquad \text{and} \qquad \frac{F(S_T)}{S_T^1} \in L^{1+\varepsilon}(\Omega)$$

then

$$E(f(S_T)) = E\left(\frac{F(S_T)}{S_T^1}\cdot\left(1 + \frac{1}{R_{11}T}\cdot\int_0^T (v_r^1)^{-1/2} dZ_r^1\right)\right)$$

Proof. We abbreviate D^{Z^1} by D. First we will prove a similar result for the log-price. If G is a function satisfying the assumptions of the chain rule Proposition 3.19 and $g = \partial_1 G$, then

$$\begin{aligned} E(g(X_T)) &= \frac{1}{T}\cdot E\left(\int_0^T g(X_T)\cdot D_r X_t^1 \cdot \frac{1}{D_r X_T^1}\, dr\right) \\ &= \frac{1}{T}\cdot E\left(\int_0^T D_r G(X_T)\cdot \frac{1}{D_r X_T^1}\, dr\right) \\ &= \frac{1}{R_{11}T}\cdot E\left(G(X_T)\cdot\int_0^T (v_r^1)^{-1/2}\, dZ_r^1\right) \end{aligned} \tag{8.1}$$

This formula is at the core of the theorem because it replaces the possibly discontinuous g by an integrated and thus continuous functional G. The first functional to which we will apply it, is $g(x) := f(e^{x_1},\dots,e^{x_d})$ together with $G(x) := \int_0^{x_1} g(\xi, x_2,\dots,x_d)d\xi + F(1, e^{x_2},\dots,e^{x_d})$. Assume f is bounded. Then g and G satisfy the assumptions to apply (8.1). Before we do so, we write G in terms of F:

$$\begin{aligned} G(x) &= \int_0^{x_1} f(e^{\xi}, e^{x_2},\dots,e^{x_d})\cdot e^{\xi}\cdot\frac{1}{e^{\xi}}\, d\xi + F(1, e^{x_2},\dots,e^{x_d}) \\ &= \left[F(e^{\xi}, e^{x_2},\dots,e^{x_d})\cdot\frac{1}{e^{\xi}}\right]_0^{x_1} + \int_0^{x_1} F(e^{\xi}, e^{x_2},\dots,e^{x_d})\cdot\frac{1}{e^{\xi}}d\xi + F(1, e^{x_2},\dots,e^{x_d}) \\ &= \frac{F(e^{x_1},\dots,e^{x_d})}{e^{x_1}} + \int_0^{x_1} F(e^{\xi}, e^{x_2},\dots,e^{x_d})\cdot\frac{1}{e^{\xi}}d\xi \end{aligned} \tag{8.2}$$

Combining this with (8.1) gives

$$\begin{aligned} E(f(S_T)) &= E(g(X_T)) \\ &= \frac{1}{R_{11}T}\cdot E\left(G(X_T)\cdot\int_0^T (v_r^1)^{-1/2}\, dZ_r^1\right) \\ &= \frac{1}{R_{11}T}\cdot E\left(\left(\frac{F(S_T)}{S_T^1} + \int_0^{X_T^1} F(e^{\xi}, S_T^2,\dots,S_T^d)\cdot\frac{1}{e^{\xi}}d\xi\right)\cdot\int_0^T (v_r^1)^{-1/2}\, dZ_r^1\right) \\ &= \frac{1}{R_{11}T}\cdot E\left(\left(\frac{F(S_T)}{S_T^1} + H(X_T)\right)\cdot\int_0^T (v_r^1)^{-1/2}\, dZ_r^1\right) \end{aligned} \tag{8.3}$$

with the functions

$$H(x_1,\dots,x_d) := \int_0^{x_1} h(\xi,x_2,\dots,x_d)d\xi \qquad \text{and} \qquad h(x_1,\dots,x_d) = F(e^{x_1},\dots,e^{x_d})/e^{x_1}$$

Note that h and H again satisfy the assumptions for (8.1) and thus

$$\frac{1}{R_{11}T}\cdot E\left(H(X_T)\cdot\int_0^T (v_r^1)^{-1/2}\,dZ_r^1\right) = E(h(X_T)) = E\left(\frac{F(S_T)}{S_T^1}\right)$$

Applying this in (8.3) gives the result.
It remains to extend the formula to the case where f is not bounded. Choose bounded measurable functions f_n, $n\in\mathbb{N}$, such that $|f_n|\le|f|$ and $f_n\to f$ almost everywhere. For $n\in\mathbb{N}$ we set $F_n(x_1,\dots,x_d) := \int_0^{x_1} f_n(\xi,x_2,\dots,x_d)d\xi$. Dominated convergence shows $F_n\to F$ almost everywhere. By assumption $|F_n(S_T)/S_T^1|\le|F(S_T)/S_T^1|\in L^{1+\varepsilon}(\Omega)$. By Theorems 2.2 and 2.4 the weight term

$$\Pi := 1+\frac{1}{R_{11}T}\cdot\int_0^T (v_t^1)^{-1/2}dZ_t^1$$

is in $L^q(\Omega)$ for all q. Thus Hölder's inequality and dominated convergence prove

$$\frac{F_n(S_T)}{S_T^1}\cdot\Pi\to\frac{F(S_T^1)}{S_T}\cdot\Pi$$

in $L^1(\Omega)$. Because clearly $f_n(S_T)\to f(S_T)$ in $L^1(\Omega)$, this proves the claim. □

If the original functional f is bounded, then the arising functional of the log-price has nice properties:

Proposition 8.3 For $f\colon(\mathbb{R}_{>0})^d\to\mathbb{R}$ define $G\colon\mathbb{R}^d\to\mathbb{R}$ by

$$G(x) := \frac{1}{e^{x_1}}\cdot\int_0^{e^{x_1}} f(\xi,e^{x_2},\dots,e^{x_d})d\xi$$

If f is bounded by $C\ge 0$, then G is also bounded by C and globally $2C$-Lipschitz continuous in $x_1\in\mathbb{R}$:

$$|G(x_1,x_2,\dots,x_d)-G(x_1',x_2,\dots,x_d)|\le 2C|x_1-x_1'|$$

for all $x_1,x_1'\in\mathbb{R}_{>0}$.

Proof. Clearly $|G(x)|\le e^{-x_1}e^{x_1}C = C$. Let $x_1,y_1\in\mathbb{R}$ with $x_1<y_1$. Then for all $x_2,\dots,x_d\in\mathbb{R}$ we have

$$\begin{aligned}
&|G(e^{y_1},e^{x_2},\dots,e^{x_d})-G(e^{x_1},e^{x_2},\dots,e^{x_d})|\\
&\quad= \left|\frac{1}{e^{y_1}}\cdot\int_{e^{x_1}}^{e^{y_1}} f(\xi,e^{x_2},\dots,e^{x_d})d\xi+\left(\frac{1}{e^{y_1}}-\frac{1}{e^{x_1}}\right)\cdot\int_0^{e^{x_1}} f(\xi,e^{x_2},\dots,e^{x_d})d\xi\right|\\
&\quad\le\left|\frac{C}{e^{y_1}}(e^{y_1}-e^{x_1})\right|+\left|\frac{C}{e^{x_1}e^{y_1}}(e^{y_1}-e^{x_1})e^{x_1}\right|\\
&\quad= 2C(1-e^{x_1-y_1})\\
&\quad\le 2C(1-(1+x_1-y_1))\\
&\quad= 2C(y_1-x_1)
\end{aligned}$$

□

8.2 Order of Convergence Using Integration By Parts

To use the integration by parts formula in practice, we need an approximation of the weight term

$$\Pi := 1 + \frac{1}{R_{11}T} \cdot \int_0^T (v_t^1)^{-1/2} dZ_t^1$$

The obvious one is

$$\hat{\Pi} := 1 + \frac{1}{R_{11}T} \cdot \sum_{k=0}^{N-1} (\hat{v}_{k\Delta}^1)^{-1/2} \cdot \Delta_k Z^1$$

It turns out that the integration by parts formula is also valid for functionals of the approximation $\hat{S}_t$. As a consequence, its usage for the quadrature of functionals of the form $E(f(S_T))$ will not introduce an additional bias. In the case $\gamma = 1/2$ however, while the integral within Π has moments of all orders due to Theorem 2.2, this is not true for the sum within $\hat{\Pi}$. This is the reason why the next theorem requires stronger assumptions than Theorem 8.2.

Theorem 8.4 Assume that the log-price X^1 is approximated using the log-Euler scheme. Assume further that $f\colon \mathbb{R}_+^d \to \mathbb{R}$ is P_X-a.s. continuous and bounded. Define $F(x) := F(x_1, \ldots, x_d) = \int_0^{x_1} f(\xi, x_2, \ldots, x_d) d\xi$. Then

$$E(f(\hat{S}_N^1), \ldots, \hat{S}_N^d) = E\left(\frac{F(\hat{S}_N^1, \ldots, \hat{S}_N^d)}{\hat{S}_N^1} \cdot \hat{\Pi} \right)$$

Proof. Like before we will write D for the derivative D^{Z^1}. The log-Euler scheme is given by

$$\hat{X}_{n+1}^1 = \hat{X}_n^1 + \left(\mu - \frac{1}{2}\hat{v}_n^1 \right) \cdot \Delta + \sqrt{\hat{v}_n^1} \cdot \Delta_n B^1$$

because $\Delta_n B^1 = \sum_{i=1}^{2d} R_{1,i} \Delta_n Z^i$ and Z^1 is independent of $\hat{v}$ we can easily calculate the derivative of $\hat{X}_n^1$:

$$D_t \hat{X}_{n+1}^1 = D_t \hat{X}_n^1 + R_{11} \cdot \sqrt{\hat{v}_n^1} \cdot \mathbb{1}_{[n\Delta,(n+1)\Delta)}(t)$$

By recursion this means

$$D_t \hat{X}_{n+1}^1 = R_{11} \cdot \sqrt{\hat{v}_{\eta(t)}^1} \cdot \mathbb{1}_{[0,(n+1)\Delta)}$$

On the other hand, clearly $D\hat{X}_n^i = 0$ for all $i = 2, \ldots, d$. Now we can prove the result analogously to Theorem 8.2, replacing X_t by $\hat{X}_{\eta(t)}$ and v_t by $\hat{v}_{\eta(t)}$. □

Before we can prove the strong error rate of the scheme DISE, we need some preparations in the next lemmas.

Lemma 8.5 Let $\gamma = 1/2$. Assume ε and q satisfy $\varepsilon q < \nu - 1$. Then

$$E\left(\int_0^T \sigma_s^{-2(1+\varepsilon)} \, ds \right)^q < \infty$$

Proof. Choose $\varepsilon' > 0$ such that $\varepsilon q(1+\varepsilon') < \nu - 1$ and set $q' = (1+\varepsilon')/\varepsilon'$. Then

$$\begin{aligned} E\left(\int_0^T \sigma_s^{-2(1+\varepsilon)}\,ds\right)^q &\le E\left(\left(\int_0^T \sigma_s^{-2}\,ds\right)^q \cdot \sup_{s\in[0,T]} \sigma_s^{-2\varepsilon q}\right) \\ &\le \left(E\left(\int_0^T \sigma_s^{-2}\,ds\right)^{qq'}\right)^{1/q'} \cdot \left(E \sup_{s\in[0,T]} \sigma_s^{-2\varepsilon q(1+\varepsilon')}\right)^{1/(1+\varepsilon')} \end{aligned}$$

This is finite due to Theorems 2.2 and 2.1. □

Lemma 8.6 Let $\gamma = 1/2$. Assume that $\nu > 2$ and that ν_t is approximated using the DISE scheme. Let $q \ge 0$. Then there exists $\alpha \in (0,1)$ such that

$$\sup_{k=0,\dots,N} \left(E\hat\sigma_k^{-2q}\right)^{1/q} \le c \cdot \Delta^{-(1-\alpha)}$$

Proof. Choose α such that $4q\alpha < 2\nu/3$. Then

$$\sup_{k=0,\dots,N} \left(E\hat\sigma_k^{-2q}\right)^{1/q} \le \sup_{k=0,\dots,N} \left(E\hat\sigma_k^{-4q\alpha}\right)^{1/2q} \cdot \left(E\hat\sigma_k^{-4q(1-\alpha)}\right)^{1/2q} \le c \cdot \Delta^{-(1-\alpha)}$$

because the first expectation is finite by Theorem 6.2 and the second one can be bounded by Lemma 6.3. □

As a next step we prove an L^2-convergence rate for the weight term in $\hat\Pi$ when using the DISE scheme.

Theorem 8.7 Assume that $\gamma_1 > 1/2$ or $\nu_1 > 3$. Then

$$\left(E(\hat\Pi - \Pi)^2\right)^{1/2} \le c \cdot \sqrt{\Delta}$$

Proof. Because $\hat\Pi$ depends only on the first Heston model, we will omit the index 1.
(i) First consider the case $\gamma = 1/2$ and $\nu > 3$. Then $\hat\Pi = 1 + 1/(R_{11}T) \cdot \int_0^T \hat\sigma_{\eta(s)}^{-1} dZ_s$ and thus

$$\begin{aligned} E|\Pi - \hat\Pi|^2 &\le c \cdot E\left|\int_0^T \frac{1}{\sigma_s} - \frac{1}{\hat\sigma_{\eta(s)}} dZ_s\right|^2 \\ &\le c \cdot E\int_0^T \frac{1}{\sigma_s^2\hat\sigma_{\eta(s)}^2} |\sigma_s - \hat\sigma_{\eta(s)}|^2 ds \\ &\le c \cdot E\int_0^T \frac{1}{\sigma_s^2\hat\sigma_{\eta(s)}^2} |\sigma_s - \sigma_{\eta(s)}|^2 ds + c \cdot E\int_0^T \frac{1}{\sigma_s^2\hat\sigma_{\eta(s)}^2} |\sigma_{\eta(s)} - \hat\sigma_{\eta(s)}|^2 ds \end{aligned} \tag{8.4}$$

We will examine both summands independently and start with

$$\begin{aligned} S_1 &:= E\int_0^T \frac{1}{\sigma_s^2\hat\sigma_{\eta(s)}^2} |\sigma_s - \sigma_{\eta(s)}|^2 ds \\ &\le E\left(\int_0^T \sigma_s^{-2} \cdot |\sigma_s - \sigma_{\eta(s)}|^2\,ds \cdot \sup_{s\in[0,T]} \hat\sigma_{\eta(s)}^{-2}\right) \end{aligned}$$

In the following we will frequently use Hölder's inequality. The main reason for the complexity of the calculations is the unusual fact that σ_s^{-2} does not have all moments, while $\int_0^T \sigma_s^{-2} ds$ does. When using Hölder's inequality we hence must make sure to enlarge not the exponent within the integral but the one between integral and expectation.
Choose $\varepsilon > 0$ such that $1 + \varepsilon < \nu/3$ and set $q = (1 + \varepsilon)/\varepsilon$. Theorem 6.2 guarantees that the necessary moments are bounded with the bound being independent of Δ.

$$\sup_{\Delta>0} E\left(\sup_{s\in[0,T]} \hat{\sigma}_{\eta(s)}^{-2(1+\varepsilon)}\right) < \infty$$

Using Hölder's inequality three times, always with exponents $1 + \varepsilon$ and q, we get

$$\begin{aligned}
S_1 &\le \left(E\left(\int_0^T \sigma_s^{-2} \cdot |\sigma_s - \sigma_{\eta(s)}|^2 \, ds\right)^q\right)^{\frac{1}{q}} \cdot \left(E \sup_{s\in[0,T]} \hat{\sigma}_{\eta(s)}^{-2(1+\varepsilon)}\right)^{\frac{1}{1+\varepsilon}} \\
&\le c \cdot \left(E\left(\int_0^T \sigma_s^{-2} \cdot |\sigma_s - \sigma_{\eta(s)}|^2 \, ds\right)^q\right)^{\frac{1}{q}} \\
&\le c \cdot \left(E\left[\left(\int_0^T \sigma_s^{-2(1+\varepsilon)} ds\right)^{\frac{q}{1+\varepsilon}} \cdot \int_0^T |\sigma_s - \sigma_{\eta(s)}|^{2q} ds\right]\right)^{\frac{1}{q}} \\
&\le c \cdot \left(E\left(\int_0^T \sigma_s^{-2(1+\varepsilon)} ds\right)^q\right)^{\frac{1}{q(1+\varepsilon)}} \cdot \left(E\left(\int_0^T |\sigma_s - \sigma_{\eta(s)}|^{2q} ds\right)^q\right)^{\frac{1}{q^2}}
\end{aligned}$$

Because $\varepsilon q = 1 + \varepsilon < \nu/3 < \nu - 1$ the first factor is finite by Lemma 8.5. Since the CIR process is continuous in the mean (see Theorem 2.3), the last expectation is bounded by $c\Delta$. Thus $S_1 \le c\Delta$.
Now consider the second summand of (8.4):

$$\begin{aligned}
S_2 &:= E\int_0^T \frac{1}{\sigma_s^2 \hat{\sigma}_{\eta(s)}^2} |\sigma_{\eta(s)} - \hat{\sigma}_{\eta(s)}|^2 ds \\
&\le \left(E\left(\int_0^T \sigma_s^{-2} \cdot \hat{\sigma}_{\eta(s)}^{-2} ds\right)^q\right)^{\frac{1}{q}} \cdot \left(E \sup_{k=0,\dots,\lceil T/\Delta\rceil} |\sigma_{k\Delta} - \hat{\sigma}_{k\Delta}|^{2(1+\varepsilon)}\right)^{\frac{1}{1+\varepsilon}}
\end{aligned}$$

The second factor is bounded by $c\Delta^2$ as was shown in Theorem 6.2 (again we need the full strength of the assumption $1 + \varepsilon < \nu/3$). We apply Hölder's inequality twice to split the first factor.

$$\begin{aligned}
&\le c \cdot E\left(\left(\int_0^T \sigma_s^{-2(1+\varepsilon)} ds\right)^{\frac{q}{1+\varepsilon}} \cdot \int_0^T \sigma_{\eta(s)}^{-2q} ds\right)^{\frac{1}{q}} \cdot \Delta^2 \\
&\le c \cdot \left(E\left(\int_0^T \sigma_s^{-2(1+\varepsilon)} ds\right)^q\right)^{\frac{1}{q(1+\varepsilon)}} \cdot \left(E\left(\int_0^T \sigma_{\eta(s)}^{-2q} ds\right)^q\right)^{\frac{1}{q^2}} \cdot \Delta^2
\end{aligned}$$

The first factor is bounded as above, the second factor is bounded by $c\Delta^{-1}$ due to Lemma 6.3. Taken together, we have proven that $S_2 \le c\Delta$ and the proof of the assertion is complete.
(ii) Now assume $\gamma > 1/2$. Set $p = 1/2(1-\gamma)$ so that $\sqrt{v_t} = \sigma_t^p$. Because inverse moments of arbitrarily high powers exist for the CEV process and its approximation (see Theorems

2.4 and 6.2), the proof is much easier in this case.

$$\begin{aligned}
E|\Pi - \hat{\Pi}|^2 &= E\left|\int_0^T \sigma_s^{-p} - \hat{\sigma}_{\eta(s)}^{-p} dZ_s\right|^2 \\
&= c \cdot E\left|\int_0^T (\sigma_s \hat{\sigma}_{\eta(s)})^{-p} \cdot (\hat{\sigma}_{\eta(s)}^p - \sigma_s^p) dZ_s\right|^2 \\
&= c \cdot \int_0^T E\left((\sigma_s \hat{\sigma}_{\eta(s)})^{-2p} \cdot (\hat{\sigma}_{\eta(s)}^p - \sigma_s^p)^2\right) ds \\
&\le c \cdot \sup_{s\in[0,T]} \left[\left(E\,\sigma_s^{-6p}\right)^{1/3} \cdot \left(E\,\hat{\sigma}_{\eta(s)}^{-6p}\right)^{1/3} \cdot \left(E\left(\hat{\sigma}_{\eta(s)}^p - \sigma_s^p\right)^6\right)^{1/3}\right]
\end{aligned}$$

As mentioned above, the first two expectations are finite.

$$\begin{aligned}
&\le c \cdot \left(E\left(\hat{\sigma}_{\eta(s)}^p - \sigma_{\eta(s)}^p\right)^6 + E\left(\sigma_{\eta(s)}^p - \sigma_s^p\right)^6\right)^{1/3} \\
&= c \cdot \left(E\left(\hat{v}_{\eta(s)} - v_{\eta(s)}\right)^6 + E\left(v_{\eta(s)} - v_s\right)^6\right)^{1/3}
\end{aligned}$$

Now we apply Theorems 2.5 and 6.2

$$\begin{aligned}
&\le c \cdot (\Delta^6 + \Delta^3)^{1/3} \\
&\le c \cdot \Delta
\end{aligned}$$

This concludes the proof. □

Under weaker assumptions we can still prove convergence:

Theorem 8.8 Assume that $\gamma_1 > 1/2$ or $\nu_1 > 2$. Then

$$\left(E(\hat{\Pi} - \Pi)^2\right)^{1/2} \to 0 \qquad \text{as } \Delta \to 0$$

Proof. We will again omit the index 1. For $\gamma > 1/2$ the claim follows from the last theorem, so assume $\gamma = 1/2$. Consider again the decomposition $E|\Pi - \hat{\Pi}|^2 \le S_1 + S_2$ that was used in the proof of the previous theorem. Choose $\varepsilon > 0$ such that $1 + \varepsilon < \nu/2$ and set $q = (1+\varepsilon)/\varepsilon$. Then by Hölder's inequality

$$\begin{aligned}
S_1 &= E\int_0^T \frac{1}{\sigma_s^2 \hat{\sigma}_{\eta(s)}^2} |\sigma_s - \sigma_{\eta(s)}|^2 ds \\
&\le \left(E\int_0^T \sigma_s^{-2(1+\varepsilon)} ds\right)^{\frac{1}{1+\varepsilon}} \cdot \left(E\int_0^T \hat{\sigma}_{\eta(s)}^{-4q} ds\right)^{\frac{1}{2q}} \cdot \left(E\int_0^T |\sigma_s - \sigma_{\eta(s)}|^{4q} ds\right)^{\frac{1}{2q}} \\
&\le c \cdot \left(E\int_0^T \sigma_s^{-2(1+\varepsilon)} ds\right)^{\frac{1}{1+\varepsilon}} \cdot \left(\sup_{k=0,\dots,N} E\hat{\sigma}_k^{-4q}\right)^{\frac{1}{2q}} \cdot \left(E\int_0^T |\sigma_s - \sigma_{\eta(s)}|^{4q} ds\right)^{\frac{1}{2q}}
\end{aligned}$$

By our choice of ε the first expectation is finite. In Lemma 8.6 we have shown that the second expectation is bounded by $c\Delta^{-(1-\alpha)}$ for some $\alpha \in (0,1)$. Theorem 2.3 shows that the third factor is bounded by $c\Delta$. In summary, $S_1 \le c \cdot \Delta^\alpha$ which converges to 0.

To bound S_2 we apply Hölder's inequality twice, again using $1+\varepsilon$ and q as exponents:

$$\begin{aligned}
S_2 &= E\int_0^T \frac{1}{\sigma_s^2\hat\sigma^2_{\eta(s)}}\,|\sigma_{\eta(s)}-\hat\sigma_{\eta(s)}|^2 ds \\
&\le E\left(\int_0^T \sigma_s^{-2}\cdot\hat\sigma^{-2}_{\eta(s)}\,ds\cdot \sup_{k=0,\dots,N}|\sigma_{k\Delta}-\hat\sigma_{k\Delta}|^2\right) \\
&\le \left(E\left(\int_0^T \sigma_s^{-2}\hat\sigma^{-2}_{\eta(s)}ds\right)^q\right)^{\frac{1}{q}}\cdot\left(E\sup_{k=0,\dots,N}|\sigma_{k\Delta}-\hat\sigma_{k\Delta}|^{2(1+\varepsilon)}\right)^{\frac{1}{1+\varepsilon}}
\end{aligned}$$

The last factor is bounded by $c\Delta$ due to Theorem 6.2 (note that under the stronger assumptions of the previous theorem we were able to bound it by $c\Delta^2$). Two other applications of Hölder's inequality give

$$\begin{aligned}
&\le c\cdot\left(E\left[\left(\int_0^T\sigma_s^{-2(1+\varepsilon)}ds\right)^{\frac{q}{1+\varepsilon}}\cdot\int_0^T\hat\sigma^{-2q}_{\eta(s)}ds\right]\right)^{\frac{1}{q}}\cdot\Delta \\
&\le c\cdot\left(E\left(\int_0^T\sigma_s^{-2(1+\varepsilon)}ds\right)^q\right)^{\frac{1}{q(1+\varepsilon)}}\cdot\left(E\left(\int_0^T\hat\sigma^{-2q}_{\eta(s)}ds\right)^q\right)^{\frac{1}{q^2}}\cdot\Delta
\end{aligned}$$

Because $\varepsilon q = 1+\varepsilon < \nu/2 < \nu-1$, the first factor is bounded by Lemma 8.5. By Lemma 8.6 there exists $\alpha\in(0,1)$ such that

$$\left(E\left(\int_0^T\hat\sigma^{-2q}_{\eta(s)}ds\right)^q\right)^{\frac{1}{q^2}} \le c\cdot\sup_{k=0,\dots,N}\left(E\hat\sigma_k^{-2q^2}\right)^{\frac{1}{q^2}} \le c\cdot\Delta^{-(1-\alpha)}$$

In summary we have shown that $E|\Pi-\hat\Pi|^2 \le c\cdot\Delta^\alpha$ for some $\alpha>0$. □

After these preparations we can prove that the term on the right-hand side of the integration by parts formula can be approximated with an L^2-convergence rate of $1/2$ which is the same we would get for a Lipschitz continuous functional; see Corollary 6.7. Set

$$\begin{aligned}
P &:= \frac{F(S_T)}{S_T^1}\cdot\Pi = \frac{F(S_T)}{S_T^1}\cdot\left(1+\frac{1}{R_{11}T}\cdot\int_0^T(v_t^1)^{-1/2}dZ_t^1\right) \\
\hat P &:= \frac{F(\hat S_T)}{\hat S_T^1}\cdot\hat\Pi = \frac{F(\hat S_T)}{\hat S_T^1}\cdot\left(1+\frac{1}{R_{11}T}\cdot\sum_{k=0}^{N-1}(\hat v_{k\Delta}^1)^{-1/2}\cdot\Delta_k Z^1\right)
\end{aligned} \tag{8.5}$$

Theorem 8.9 Assume that

- The DISE scheme is used to approximate the (log-)price.
- Either $\gamma_1 > 1/2$ or $\nu_1 > 3$.
- The function f is bounded and globally Lipschitz continuous in all arguments except possibly the first. Denote the Lipschitz constant by L.

Then

$$\|\hat P - P\|_2 \le c\cdot(\sup|f|+L)\cdot\sqrt{\Delta}$$

where the constant is independent of f.

Proof. Let G be the function from Proposition 8.3:

$$G(x) := \frac{F(e^{x_1}, \dots, e^{x_d})}{e^{x_1}} = \frac{1}{e^{x_1}} \cdot \int_0^{e^{x_1}} f(\xi, e^{x_2}, \dots, e^{x_d}) d\xi$$

The proposition shows that G is bounded by $\sup|f|$ and globally Lipschitz continuous in the first argument with Lipschitz constant $2\sup|f|$. Together with the additional assumption of this theorem G must be globally Lipschitz continuous in all arguments with constant $2\sup|f| + L$. Now we can bound $\|P - \hat{P}\|_2$ as follows:

$$\begin{aligned}
\|\hat{P} - P\|_2 &= \|G(\hat{X}_T) \cdot \hat{\Pi} - G(X_T) \cdot \Pi\|_2 \\
&\le \|G(\hat{X}_T) \cdot (\hat{\Pi} - \Pi)\|_2 + \|(G(\hat{X}_T) - G(X_T)) \cdot \Pi\|_2 \\
&\le \sup|f| \cdot \|\hat{\Pi} - \Pi\|_2 + (2\sup|f| + L) \cdot \| \|\hat{X}_T - X_T\|_{\mathbb{R}^d} \cdot \Pi\|_2
\end{aligned}$$

Choose $\varepsilon > 0$ small enough such that $2(1+\varepsilon) < 4\nu/3$ and set $q = (1+\varepsilon)/\varepsilon$. Then use Hölder's inequality.

$$\le \sup|f| \cdot \|\hat{\Pi} - \Pi\|_2 + (2\sup|f| + L) \cdot \|\hat{X}_T - X_T\|_{2(1+\varepsilon)} \cdot \|\Pi\|_q$$

Now the convergence order is provided by Corollary 6.7 and Theorem 8.7. Π has moments of all orders (see Theorem 2.2).

$$\le c \cdot (2\sup|f| + L) \cdot \sqrt{\Delta}$$ □

Again, weaker assumptions suffice to prove convergence:

Theorem 8.10 Assume that

- The DISE scheme is used to approximate the (log-)price.
- Either $\gamma_1 > 1/2$ or $\nu_1 > 2$.
- The function f is bounded and globally Lipschitz continuous in all arguments except possibly the first. Denote the Lipschitz constant by L.

Then

$$\lim_{\Delta \to 0} \|\hat{P} - P\|_2 = 0$$

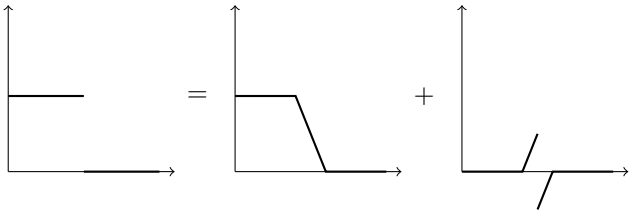
Proof. The proof is exactly the same as that of the previous theorem, but using Theorem 8.8 instead of Theorem 8.7. □

8.3 Payoff Splitting

As will be explained in Chapter 9, the strong rate of convergence of 1/2 established in Theorem 8.9 will lead to a numerical algorithm with costs in $\mathcal{O}(\varepsilon^{-2} \cdot (\log \varepsilon)^2)$, where ε is the root-mean-square error. However, the numerical experiments have shown that without further work the constant involved in the $\mathcal{O}$-notation when using the integration by parts formula to smooth the functional is too large for the algorithm to be of any practical use. The reason here is the variance of the term $\int_0^T \nu_s^{-1/2} dZ_s^1$ within the random variable P from (8.5): Due to Jensen's inequality and $E\nu_t \approx \lambda$, see the formulas in the appendix of [And07], we have

$$V\left(\int_0^T \nu_s^{-1/2} dZ_s^1\right) = E\left(\int_0^T \nu_s^{-1/2} dZ_s^1\right)^2 = \int_0^T E(\nu_s^{-1}) ds \ge \int_0^T (E\nu_s)^{-1} ds \approx T \cdot \lambda^{-1}$$

Figure 8.1: Split discontinuous functional f into $f = f_1 + f_2$ and apply smoothing via Malliavin integration by parts only to the second summand.



Because the parameter λ is typically low (see again [And07]), the variance will be high and this carries over to the approximation $\int \hat{v}_s^{-1/2} dZ_s^1$ and $\hat{P}$.
Fortunately, there is an easy way to weaken the problem. For this we will split the discontinuous functional f into two parts $f = f_1 + f_2$ with f_1 being continuous and f_2 being discontinuous, but with small support. Then we will apply the integration by parts formula only to the second part f_2. Given an antiderivative F_2 of f_2 the final estimator will then be given by

$$\hat{P} := f_1(\hat{S}_T) + \frac{F_2(\hat{S}_T)}{\hat{S}_T^1} \cdot \hat{\Pi} = f_1(\hat{S}_T) + \frac{F_2(\hat{S}_T)}{\hat{S}_T^1} \cdot \left(1 + \frac{1}{R_{11}T} \cdot \sum_{k=0}^{N-1} (\hat{v}_{k\Delta}^1)^{-1/2} \cdot \Delta_k Z^1\right) \tag{8.6}$$

In case of the digital option $f = \mathbb{1}_{[0,K]}$ we will use the following splitting, which depends on a parameter $\delta \in (0,1)$; see Figure 8.1.

$$f_1(x) = \begin{cases} 1 & x \le (1-\delta)K \\ -\frac{1}{2\delta}(x-K) + \frac{1}{2} & x \in [(1-\delta)K, (1+\delta)K] \\ 0 & x \ge (1+\delta)K \end{cases}$$

$$f_2(x) = \begin{cases} 0 & x \le (1-\delta)K \\ -\frac{1}{2\delta}(x-K) - \frac{1}{2} & x \in [(1-\delta)K, K] \\ -\frac{1}{2\delta}(x-K) + \frac{1}{2} & x \in (K, (1+\delta)K] \\ 0 & x \ge (1+\delta)K \end{cases}$$

For typical payoffs the splitting can be done in a way such that f_1 is Lipschitz continuous. In this case the strong convergence rate $1/2$ also holds for the splitted payoff:

Corollary 8.11 Let $f = f_1 + f_2$. Assume that

- The DISE scheme is used to approximate the (log-)price.
- Either $\gamma_1 > 1/2$ or $\nu_1 > 3$.
- The function f_1 is L_1-Lipschitz continuous.
- The function f_2 is bounded and L_2-Lipschitz continuous in all arguments except possibly the first.

Then the estimator in (8.6) satisfies

$$\|\hat{P} - P\|_2 \le c \cdot (\sup|f| + L_1 + L_2) \cdot \sqrt{\Delta}$$

where the constant is independent of f.

Proof. This follows directly from Corollary 6.7 for the Lipschitz continuous summand f_1 and from Theorem 8.9 for the remaining summand. □

Chapter 9

Monte-Carlo Algorithms

In this chapter we will examine algorithms to estimate expectations. We start with the fundamental Monte-Carlo algorithm and proceed to discuss first general multilevel Monte-Carlo, then the adaptive version of the algorithm which we are going to use in the numerical examples of the next chapter.
To simplify notation and because the algorithms are obviously not restricted to the Heston model we will assume that we are approximating the solution S_T of a possibly multidimensional SDE with driving Brownian motion W_t. Also, we define the shorthand notation $p := \mathbb{E}(f(S_T))$ for the quantity that should be estimated.
It should be noted that for one-dimensional models Monte-Carlo algorithms are usually outperformed by PDE-based or FFT-based algorithms. However, the basic Monte-Carlo idea is unsurpassed in its simplicity and trivial to extend to higher dimensions, whereas those other algorithms are mostly confined to a very low number of dimensions. The articles [iHF10] and [CM99] present respectively a PDE-based method and an FFT-based method for the Heston model.

9.1 Standard Monte-Carlo

The Monte-Carlo algorithm is the straightforward algorithm to compute expectations of random variables that can be simulated approximately. It approximates the expectation by the mean

$$\hat{P} := \frac{1}{M} \cdot \sum_{i=1}^{M} f(\hat{S}_N^{\Delta,i})$$

where $\hat{S}_N^{\Delta,i}$, $i = 1, \dots, M$, are independent simulations of $\hat{S}_N$ using stepsize Δ. To find optimal values for N and M, we recall the following small calculation for a random variable X and a number a.

$$\mathbb{E}|X-a|^2 = \mathbb{E}|X-\mathbb{E}X+\mathbb{E}X-a|^2 = \mathbb{E}|X-\mathbb{E}X|^2+|\mathbb{E}X-a|^2+2\mathbb{E}(X-\mathbb{E}X)\cdot(\mathbb{E}X-a) = V(X)+|\mathbb{E}X-a|^2 \tag{9.1}$$

Using this equation the error of a Monte-Carlo algorithm can be divided into two parts: the bias and the variance error. In fact, we have

$$\mathbb{E}|\hat{P}-p|^2 = V(\hat{P}) + |\mathbb{E}\hat{P}-p|^2 = \frac{1}{M}V(f(\hat{S}_T)) + |\mathbb{E}\hat{P}-p|^2$$

Assuming a weak error rate of 1, i.e. $|\mathbb{E}\hat{P}-p| \leq c \cdot \Delta = cT/N$, setting $N := \lceil T/\varepsilon \rceil$ will bound the bias by $c \cdot \varepsilon$, whereas setting $M := \lceil \varepsilon^{-2} \rceil$ will bound the variance by $c \cdot \varepsilon^2$. Together we get that the root-mean-square error is bounded by

$$\|\hat{P}-p\|_2 \leq c \cdot \varepsilon$$

The cost of the algorithm is then $N \cdot M \in \mathcal{O}(\varepsilon^{-3})$.

Algorithm 1: Standard Monte Carlo (assuming a bias of order 1)

Input: error tolerance ε
Output: estimation for $E(f(S_T))$

1 $N := \lceil T/\varepsilon \rceil$ and $\Delta := T/N$
2 $M := \lceil \varepsilon^{-2} \rceil$
3 Simulate M copies $\hat{S}_N^{\Delta,i}$, $i = 1, \dots, M$.
4 **return** $\hat{P} := \frac{1}{M} \cdot \sum_{i=1}^{M} f(\hat{S}_N^{\Delta,i})$

In the Heston model, a rate of 1 for the weak error was proven in Theorem 7.13, albeit under rather strong assumptions on ν. However, as shown in Section 6.6, numerical practice indicates that the rate 1 is valid under much weaker assumptions.

9.2 Multilevel Monte-Carlo (MLMC)

The efficiency of Monte-Carlo can be drastically improved when one uses approximations of different stepsizes and combines them in a clever way. The multilevel Monte-Carlo algorithm was first examined in the context of parametrical integration problems in [Hei01]. In [Gil08] the algorithm was first used in the context of SDEs.
Denote by $\hat{P}_l$, $l \in \mathbb{N}_0$, an approximation to $f(S_T)$ using step size $\Delta_l := T \cdot M^{-l}$, for a fixed refinement parameter M (typically $M \in \{2, 4\}$). The idea of multilevel Monte-Carlo is to split the expectation $E(\hat{P}_L)$ into the telescoping sum

$$E(\hat{P}_L) = E(\hat{P}_0) + \sum_{l=1}^{L} E\big(\hat{P}_l - \hat{P}_{l-1}\big) \tag{9.2}$$

and approximate each expectation on the right-hand side separately and independently from the others. Given estimators $\hat{Y}_0$ for $E(\hat{P}_0)$ and $\hat{Y}_l$ for $E(\hat{P}_l - \hat{P}_{l-1})$ we can then estimate the target expectation $p = E(f(S_T))$ by

$$\hat{Y} := \sum_{l=0}^{L} \hat{Y}_l$$

On the first glance it might seem like a lot of additional work to approximate $L+1$ expectations. However, for small l the approximation $\hat{P}_l$ is cheap to simulate, while for large l the random variables $\hat{P}_l - \hat{P}_{l-1}$ can typically be estimated with a low variance, so that few (expensive) samples are already sufficient to reach high precision.
In our numerical experiments we will simply use standard Monte-Carlo to construct the estimators $\hat{Y}_l$. That is, for $l = 0$ we use

$$\hat{Y}_0 := \frac{1}{N_0} \sum_{i=1}^{N_0} f(\hat{S}_T^{\Delta_0,i}) \tag{9.3}$$

and for $l = 1, \dots, L$ the estimator is given by

$$\hat{Y}_l := \frac{1}{N_l} \sum_{i=1}^{N_l} f(\hat{S}_T^{\Delta_l,i}) - f(\hat{S}_T^{\Delta_{l-1},i}) \tag{9.4}$$

These definitions leave a number of issues open that must be adressed before the algorithm can be implemented in practice: It is still unclear how to choose L and N_l and it is not specified how exactly $f(\hat{S}_T^{\Delta_l,i}) - f(\hat{S}_T^{\Delta_{l-1},i})$ should be simulated. The first issue is solved theoretically in the next section and practically by the adaptive Algorithm 3. The second issue will be adressed in Section 9.2.2.

9.2.1 Costs of Multilevel Monte-Carlo

The generic MLMC algorithm does not require the estimators $\hat{Y}_l$ to be given by (9.3) and (9.4), but can be used with a wide range of estimators, as long as the costs are balanced in a suitable way and weak and strong convergence rates of the error exist. Giles' original article [Gil08] proves a quite general result which is stated in the following.

Theorem 9.1 (Theorem 3.1 in [Gil08]) For $l = 0, \dots, L$ let $\hat{Y}_l = \hat{Y}_l(N_l)$ be an estimator depending on a parameter $N_l \in \mathbb{N}$. Denote the cost of $\hat{Y}_l$ by C_l. Assume that the following conditions are satisfied for some constants $\alpha \geq 1/2$ and $\beta, c_1, c_2, c_3 \geq 0$:

1. $|E\hat{P}_l - p| \leq c_1 \cdot \Delta_l^{\alpha}$,
2. $E(\hat{Y}_0) = E(\hat{P}_0)$ and $E(\hat{Y}_l) = E(\hat{P}_l - \hat{P}_{l-1})$ for $l = 1, \dots, L$,
3. $V(\hat{Y}_l) \leq c_2 \cdot N_l^{-1} \cdot \Delta_l^{\beta}$,
4. $C_l \leq c_3 \cdot N_l \cdot \Delta_l^{-1}$.

Then for every $\varepsilon < e^{-1}$ we can choose L and N_l in a way such that the root-mean-square error of the combined estimator $\hat{Y} = \sum_{l=0}^{L} \hat{Y}_l$ is bounded by ε, i.e.

$$\left(E\,|\hat{Y} - p|^2\right)^{1/2} \leq \varepsilon$$

and the total costs C are bounded by

$$C \leq \begin{cases} c \cdot \varepsilon^{-2}, & \beta > 1 \\ c \cdot \varepsilon^{-2}(\log \varepsilon)^2, & \beta = 1, \\ c \cdot \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1 \end{cases} \tag{9.5}$$

Proof. [Sketch; for the full proof see [Gil08]]
A short calculation reveals that the choice

$$L := \left\lceil \frac{1}{\alpha} \cdot \log_M(\sqrt{2} c_1 T^{\alpha} \varepsilon^{-1}) \right\rceil$$

is large enough to guarantee that the bias is sufficiently small:

$$|E(\hat{Y}) - p| = |E(\hat{P}_L) - p| \leq \frac{\varepsilon}{\sqrt{2}} \tag{9.6}$$

Because the necessary choice of N_l differs depending on β being less, equal or greater than 1, we will now restrict ourselves to the case $\beta = 1$. Set

$$N_l := \lceil 2\varepsilon^{-2}(L+1) c_2 \Delta_l \rceil$$

Then

$$V(\hat{Y}) = \sum_{l=0}^{L} V(\hat{Y}_l) \leq c_2 \cdot \sum_{l=0}^{L} N_l^{-1} \Delta_l \leq \sum_{l=0}^{L} \frac{\varepsilon^2}{2(L+1)} = \frac{\varepsilon^2}{2} \tag{9.7}$$

Now we can split the mean-square error in the usual way (see (9.1)) and combine (9.6) and (9.7) to get the asserted error bound

$$E|\hat{Y} - p|^2 \leq \varepsilon^2$$

It remains to bound the computational cost. To simplify this proof sketch, we will ignore the rounding-up and use that $N_l \approx 2\varepsilon^{-2}(L+1)c_2\Delta_l$. Then

$$C = \sum_{l=0}^{L} C_l \leq c_3 \cdot \sum_{l=0}^{L} N_l \Delta_l^{-1} \lessapprox c_3 2\varepsilon^{-2}(L+1)^2 c_2$$

Under the assumption $\varepsilon < e^{-1}$, hence $|\log \varepsilon| > 1$, it is easy to show that $L \leq c \cdot \log \varepsilon$. This proves $C \leq c \cdot \varepsilon^{-2}(\log \varepsilon)^2$. □

9.2.2 Variance Reduction

To achieve the variance decay required by the third assumption in Theorem 9.1, it will be necessary to simulate the two parts of $\hat{P}_l - \hat{P}_{l-1}$ not independently, but using the same path of the underlying Brownian motion. Because $\hat{P}_{l-1}$ uses a stepsize of $\Delta_{l-1} := T \cdot M^{-l+1} = M \cdot \Delta_l$ we can generate increments of stepsize Δ_l, use them to simulate $\hat{P}_l$, and afterwards sum them up in groups of M increments to generate appropriate increments for the simulation of $\hat{P}_{l-1}$.
Algorithm 2 describes the implementation in detail. We assume that the numerical scheme is given in terms of a function ϕ, i.e. by

$$\hat{S}^{\Delta}_{n+1} = \phi(\hat{S}^{\Delta}_n, \Delta, W_{(n+1)\Delta} - W_{n\Delta})$$

Note that while this variance reduction technique requires a little more implementation effort, the computational effort is actually reduced, because less random numbers need to be generated compared to simulating $\hat{P}_l$ and $\hat{P}_{l-1}$ independently.

Algorithm 2: Simulate $\hat{P}_l - \hat{P}_{l-1}$ using the same path of W.

Input: Level $l \geq 1$
Output: A random variable with expectation $E(\hat{P}_l - \hat{P}_{l-1})$ and low variance.

$N_c := M^{l-1}$ `// Number of steps on coarse level`
$\Delta_c := T/N_c$ `// Coarse step size`
$\Delta_f := \Delta_c/M$ `// Fine step size`
$\hat{S}^{\Delta_c} := 0$
$\hat{S}^{\Delta_f} := 0$
for $n = 1, \ldots, N_c$ **do**
  $\Delta W^c := 0$
  for $i = 1, \ldots, M$ **do**
    Draw ΔW^f from $\mathcal{N}(0, \Delta_f \cdot C)$ `// C is the covariance matrix of` W_1
    $\Delta W^c \leftarrow \Delta W^c + \Delta W^f$ `// sum up M fine increments for one coarse inc.`
    $\hat{S}^{\Delta_f} \leftarrow \phi(\hat{S}^{\Delta_f}, \Delta, \Delta W^f)$
  end
  $\hat{S}^{\Delta_c} \leftarrow \phi(\hat{S}^{\Delta_c}, \Delta, \Delta W^c)$
end
return $f(\hat{S}^{\Delta_f}) - f(\hat{S}^{\Delta_c})$

The next theorem is a version of Theorem 9.1 in the case of our estimators (9.3) and (9.4).

Theorem 9.2 Assume that

1. $\hat{Y}_l$ are given by (9.3) and (9.4),
2. $|Ef(\hat{S}_T^{\Delta}) - p| \leq c \cdot \Delta$,
3. The same Brownian path is used to generate both parts of $f(\hat{S}_T^{\Delta_l,i}) - f(\hat{S}_T^{\Delta_{l-1},i})$ in (9.4).
4. $\|f(\hat{S}_T^{\Delta}) - f(S_T)\|_2 \leq c \cdot \Delta^{\beta/2}$ for some $\beta > 0$.

Then for any $\varepsilon < e^{-1}$ the parameters L and N_l can be chosen in a way such that the resulting estimator $\hat{Y} := \sum_{l=0}^{L} \hat{Y}_l$ achieves a root-mean-square error of

$$\|\hat{Y} - p\|_2 \leq \varepsilon$$

with the costs C being bounded by (9.5).

Proof. We only have to check the assumptions of Theorem 9.1. The second and fourth one follow directly from the construction of the estimators. The required weak convergence rate proves the first assumption (with $\alpha = 1$). Under the assumption that $\hat{P}_l$ and $\hat{P}_{l-1}$ use the same path of the underlying Brownian motion, we can compare both to the exact solution using this path.

$$\hat{P}_l - \hat{P}_{l-1} = \hat{P}_l - f(S_T) - (\hat{P}_{l-1} - f(S_T))$$

Together with the last assumption we can now bound the variance:

$$V(\hat{P}_l - \hat{P}_{l-1}) \leq \left(\|\hat{P}_l - f(S_T)\|_2 + \|\hat{P}_{l-1} - f(S_T)\|_2\right)^2 \leq c \cdot \Delta_l^{\beta}$$

for $l = 1, \ldots, L$. This gives $V(\hat{Y}_l) = V(\hat{P}_l - \hat{P}_{l-1})/N_l \leq c\Delta_l^{\beta}/N_l$ and thus the third assumption of Theorem 9.1 holds. □

In Theorem 8.1 we proved that the third assumption of the previous theorem is satisfied for $f = \mathbb{1}_{[0,K]}$, but only with the low convergence order $\beta = 1/4$. Assuming a weak error rate of 1, this leads us to expect costs of order $\mathcal{O}(\varepsilon^{-5/2})$. On the other hand, for the functional obtained by using the integration by parts rule, Theorem 8.9 proved that we can use the previous theorem with $\beta = 1$ and thus expect costs of order $\mathcal{O}(\varepsilon^{-2} \cdot (\log \varepsilon)^2)$.

9.2.3 Adaptive Multilevel Monte-Carlo

Before the multilevel technique can be applied in practice we still have to find a way of determining the algorithm parameters L and N_l. The proof of Theorem 9.1 was constructive in that it provided exact formulas for these parameters. However, these formulas contained the constants c_1 and c_2 stemming respectively from the weak and strong rate of convergence of the scheme. While it is certainly possible to estimate these parameters in a separate prior computation, the additional computational effort makes this endeavor unpromising. Instead we will use an adaptive algorithm which tries to guess the correct values for L and N_l "on the way". The algorithm was already published in the original multilevel article [Gil08] and is formulated in pseudocode in Algorithm 3.
Instead of fixing the maximum level L in advance, the algorithm starts with $L = 0$ and runs a random amount of rounds. After each round either convergence is detected or L is increased by one and the next round starts. For each level l from 0 to the current maximum level L we store:

- N_l: the number of samples created so far. Samples are either simulations of $\hat{P}_0$ for $l = 0$ or of $\hat{P}_l - \hat{P}_{l-1}$ for $l \geq 1$.
- $\hat{Y}_l$: the mean of the samples,

Algorithm 3: Adaptive Multilevel Monte-Carlo

Input: error tolerance ε, number of initial samples N_{ini}
Output: estimation for $E(f(S_T))$

1 $L := 0$
2 Generate $N_L := N_{ini}$ samples for level L `// start a new round`
3 Set V_L to the sample variance
4 Compute N_l for $l = 0, \dots, L$ using the formula

$$N_l := \max\left\{N_{ini}, \left\lceil 2 \cdot \varepsilon^{-2}\sqrt{V_l \Delta_l}\left(\sum_{l'=0}^{L}\sqrt{V_{l'}/\Delta_{l'}}\right)\right\rceil\right\}$$

5 If this leads to an increase of N_l for some $l = 0, \dots, L$, generate additional samples on level l
6 For $l = 0, \dots, L$ set $\hat{Y}_l$ to the mean of the samples on level l
7 **if** $L \geq 2$ and the convergence condition $\max\left\{M^{-1}|\hat{Y}_{L-1}|, |\hat{Y}_L|\right\} < \frac{1}{\sqrt{2}}(M-1)\varepsilon$ holds
8 **then return** $\sum_{l=0}^{L}\hat{Y}_l$
9 **else** $L := L + 1$ and **goto** step 2

- V_l: the sample variance of the first N_{ini} samples. This will be used as an estimate for $V(\hat{P}_l - \hat{P}_{l-1})$.

In each round the algorithm carries out the following steps.

1. Set $N_L := N_{ini}$ and create N_{ini} samples of $\hat{P}_L - \hat{P}_{L-1}$.
2. Set V_L to the sample variance (with Bessel correction).
3. Use the new variance estimate to update the desired number of iterations per level N_l for each level $l = 0, \dots, L$. The numbers are computed using the formula

$$N_l := \max\left\{N_{ini}, \left\lceil 2 \cdot \varepsilon^{-2}\sqrt{V_l \Delta_l}\left(\sum_{l'=0}^{L}\sqrt{V'_l/\Delta'_l}\right)\right\rceil\right\}$$

   Because the variance estimates V_l never change — except for the newly estimated V_L, of course — the value of N_l will increase by $2\varepsilon^{-2}\sqrt{V_l V_L}$ (ignoring ceiling operation and the lower bound of N_{ini}). The lower bound of N_{ini} makes sure that N_l will never decrease. The formula is chosen so that the estimated variance of the whole estimator is bounded by $\varepsilon^2/2$:

$$\begin{aligned} V(\hat{Y}) = \sum_{l=0}^{L} V(\hat{Y}_l) &= \sum_{l=0}^{L}\frac{V(\hat{P}_l - \hat{P}_{l-1})}{N_l} \\ &\approx \sum_{l=0}^{L}\frac{V_l}{N_l} \leq \sum_{l=0}^{L} V_l \cdot \left(2\varepsilon^{-2}\sqrt{V_l\Delta_l}\sum_{l'=0}^{L}\sqrt{V'_l/\Delta'_l}\right)^{-1} = \frac{\varepsilon^2}{2} \end{aligned} \tag{9.8}$$

4. For each level $l = 0, \dots, L$ generate additional samples of $\hat{P}_l - \hat{P}_{l-1}$ to reach the new N_l and update the mean $\hat{Y}_l$ accordingly.
5. If $L \geq 2$ (i.e. starting in the third round) check for convergence using the condition

$$\max\left\{M^{-1}|\widehat{Y}_{L-1}|, |\widehat{Y}_L|\right\} < \frac{1}{\sqrt{2}}(M-1)\varepsilon$$

This condition is based on a heuristic guess at the remaining bias $|E(\hat{P}_L - P)|$. Assuming a weak rate of 1, the bias on level l is $|E(\hat{P}_l - P)| \approx c_1 \cdot \Delta_l$. Thus

$$|E(\hat{P}_l - \hat{P}_{l-1})| = |E(\hat{P}_l - P - (\hat{P}_{l-1} - P))| \approx c_1(M-1)\Delta_l$$

Now if the condition is fulfilled, the following heuristic reasoning gives hope that the remaining bias is bounded by $\varepsilon/\sqrt{2}$ when estimated based on either of the last two levels.

$$|E(\hat{P}_L - P)| \approx c_1\Delta_L \approx \frac{|E(\hat{P}_L - \hat{P}_{L-1})|}{M-1} \approx \frac{|\hat{Y}_L|}{M-1} < \frac{\varepsilon}{\sqrt{2}}$$
$$|E(\hat{P}_L - P)| \approx \frac{c_1\Delta_{L-1}}{M} \approx \frac{|E(\hat{P}_{L-1} - \hat{P}_{L-2})|}{M(M-1)} \approx \frac{|\hat{Y}_{L-1}|}{M(M-1)} < \frac{\varepsilon}{\sqrt{2}}$$

Together, the preceding estimates, (9.1) and (9.8) provide a heuristic bound to the mean-square error.

$$E|\hat{Y} - p|^2 = V(\hat{Y}) + |E\hat{Y} - p|^2 \lessapprox \varepsilon^2$$

Chapter 10

Numerical Results

In this final chapter we will compare different algorithms to compute the expectation $\mathrm{E}f(S_T)$ for discontinuous f. In particular, we will check whether the integration by parts rule presented in Chapter 8 improves the quadrature of this expectation. Because the theoretical results of that chapter only hold for the drift-implicit square-root Euler scheme (DISE), we will use this scheme throughout this chapter.
The numerical experiment will be the following: We first specify a model, a functional f of the end price(s) and a list of accuracies $\varepsilon_1, \dots, \varepsilon_n$. Then a reference value x^* is computed — by the exact formula of Theorem 2.8 for one-dimensional standard Heston models or by algorithm A3 (see below) on a high input precision $\varepsilon^* = \min\{\varepsilon_1, \dots, \varepsilon_n\}/4$. Finally, for each accuracy ε each algorithm is executed $N \geq 500$ times giving results $\hat{Y}^i$, $i = 1, \dots, N$. After these runs the empirical relative root mean-square error

$$\mathrm{rms} = \frac{1}{x^*} \cdot \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{Y}^i - x^*)^2} \tag{10.1}$$

and the average costs as defined below are computed. The final plot shows the costs on the x-axis and the error on the y-axis, both in $\log_2$-coordinates. To visualize the order of convergence, each graph is accompanied by an affinely linear function that was fitted to the $\log_2$-data using a least-squares fit.

10.1 The Algorithms

We will compare the following three algorithms. Because the theoretical results from chapter 8 only hold for the DISE scheme, we will always use this scheme. Numerical experiments with the DIMIL scheme yield similar results, but with higher errors — as was already indicated in Section 6.6 for model M1.

A1) Standard Monte-Carlo simulating $f(\hat{S}_T)$ directly; see Algorithm 1.

A2) Adaptive multilevel Monte-Carlo (see Algorithm 3) simulating $f(\hat{S}_T)$, i.e. using the discontinuous functional.

A3) Adaptive multilevel Monte-Carlo using Malliavin integration by parts and payoff-splitting as discussed in Chapter 8. To be precise we simulate the estimator from (8.6):

$$\hat{P} := f_1(\hat{S}_T) + \frac{F_2(\hat{S}_T)}{\hat{S}_T^1} \cdot \left(1 + \frac{1}{R_{11} T} \cdot \sum_{k=0}^{N-1} (\hat{v}_{k\Delta}^1)^{-1/2} \cdot \Delta_k Z^1\right)$$

The costs of an algorithm are defined as the total number of approximation steps over all levels. For standard Monte-Carlo (A1) this is simply $d \cdot N \cdot M$, where d is the dimension of the model. For multilevel Monte-Carlo using the estimators (9.3) and (9.4) the costs are defined as

$$\text{cost} = d \cdot \left(N_0 + \sum_{l=1}^{L} N_l \cdot (M^l + M^{l-1}) \right) \tag{10.2}$$

Note that this definition takes into account that we need to generate two paths for each sample of $\hat{P}_l - \hat{P}_{l-1}$.
As discussed in Chapter 9, we expect to find that the costs in terms of the root-mean-square error ε are of order $\text{cost} = \mathcal{O}(\varepsilon^{-3})$ for standard Monte-Carlo (A1), and for multilevel Monte-Carlo either $\text{cost} = \mathcal{O}(\varepsilon^{-5/2})$ when using a discontinuous functional (A2) or $\text{cost} = \mathcal{O}(\varepsilon^{-2}(\log \varepsilon)^2)$ when using a Lipschitz continuous functional (A3), even though the assumptions of Theorem 8.9 are not satisfied for the standard Heston models considered here.
To visualize the convergence order, each of the plots in this chapter also shows affinely linear functions that were fitted to the $\log_2$-data using a least-squares fit. Assuming that $\text{cost} \approx c \cdot \varepsilon^{-\alpha}$ we have in $\log_2$-coordinates

$$\log_2 \text{cost} \approx \log(c) - \alpha \log_2 \varepsilon$$

A least-squares fit will now compute the optimal values for $\log(c)$ and α to fit the linear function to the $(\text{cost}, \varepsilon)$-data points. The slope of the function, i.e. α, gives the measured exponent of the costs for algorithms A1 and A2. To take the additional logarithmic factor of algorithm A3 into account, we additionally fit a logarithmic function to the data of algorithm A3. Assuming $\text{cost} \approx c \cdot \varepsilon^{-\alpha}(\log \varepsilon)^2$ we get in $\log_2$-coordinates

$$\log_2 \text{cost} \approx \log(c) - \alpha \log_2 \varepsilon + 2 \log |\log \varepsilon|$$

Again a least-squares fit will be used to compute $\log(c)$ and α. The measured exponents are listed in Table 10.1. Note that the logarithmic fit is not shown in the plots, because it would be almost identical to the linear fit.

10.2 The Functional

Because the indicator function is the fundamental discontinuous function we will always use the payoff function $f = \mathbb{1}_{[0,K]}$. In multidimensional models the function will be applied to the sum of all prices. To maximize the influence of the discontinuity we set

$$K := E\left(\sum_{i=1}^{d} S_T^i \right) = \sum_{i=1}^{d} S_0^i$$

Because the discontinuity is exactly at the expected price, in financial mathematics this is called an *at-the-money* option. Further numerical tests reveal that the benefit of the Malliavin integration by parts formula decreases when the discontinuity is moved away from the expected price, as one would expect.

10.3 The Models

In our numerical examples we will use two one-dimensional models. In both cases the parameters were taken from [ASK07].

a) Standard Heston model with parameters

$$T = 2,\ \mu = 0,\ \kappa = 5.07,\ \lambda = 0.0457,\ \theta = 0.48,\ \rho = -0.767,\ v_0 = \lambda,\ S_0 = 100$$

These are the same as for model M1 in Section 6.6.

b) Generalized Heston model with parameters $\gamma = 0.6545$ and

$$T = 2,\ \mu = 0,\ \kappa = 4.1031,\ \lambda = 0.0451,\ \theta = 0.8583,\ \rho = -0.760,\ v_0 = \lambda,\ S_0 = 100$$

Figure 10.1 shows that in both cases the price distribution is close to a normal distribution. However, compared to the normal distribution the probability of very high prices is larger, while very low prices are less probable.
In the multidimensional setting we again use two models, covering the cases $\gamma = 1/2$ and $\gamma > 1/2$, respectively.

c) The parameters of the standard model were taken from [DLS11]. To avoid negative correlations close to -1 the parameters ρ_1 and ρ_2 were slightly modified. As in the one-dimensional cases we have $T = 2$ and $\mu_1 = \mu_2 = \mu_3 = 0$. The other parameters are

$$\kappa_1 = 1.0121,\ \lambda_1 = 0.2874,\ \theta_1 = 0.7627,\ \rho_1 = -0.7137,\ v_0^1 = 0.2723,\ S_0^1 = 100,$$
$$\kappa_2 = 0.5217,\ \lambda_2 = 0.2038,\ \theta_2 = 0.4611,\ \rho_2 = -0.8322,\ v_0^2 = 0.2536,\ S_0^2 = 100,$$
$$\kappa_3 = 0.5764,\ \lambda_3 = 0.1211,\ \theta_3 = 0.3736,\ \rho_3 = -0.4835,\ v_0^3 = 0.1539,\ S_0^3 = 100$$

It remains to specify the correlation of the driving Brownian motions which is

$$\Sigma^{(B,W)} = \begin{pmatrix} 1 & 0.0246 & 0.0598 & \rho_1 & 0 & 0 \\ 0.0246 & 1 & 0.6465 & 0 & \rho_2 & 0 \\ 0.0598 & 0.6465 & 1 & 0 & 0 & \rho_3 \\ \rho_1 & 0 & 0 & 1 & 0 & 0 \\ 0 & \rho_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & \rho_3 & 0 & 0 & 1 \end{pmatrix}$$

In particular, the volatility processes are modelled as independent of the other volatility and price processes.

d) Because the literature does not seem to contain parameters for the multidimensional Heston model using CEV processes as volatility, we use the same parameters as for model c) and set additionally $\gamma_1 = 0.63$, $\gamma_2 = 0.68$, $\gamma_3 = 0.71$.

Figure 10.1: Density of the endprice. The left plot uses the standard Heston model a), the right plot the generalized model b). The dashed line is the density of a normal distribution fitted to the price density. Both plots show the sample distribution of 10^6 approximations generated using the DISE scheme and 2^8 steps.

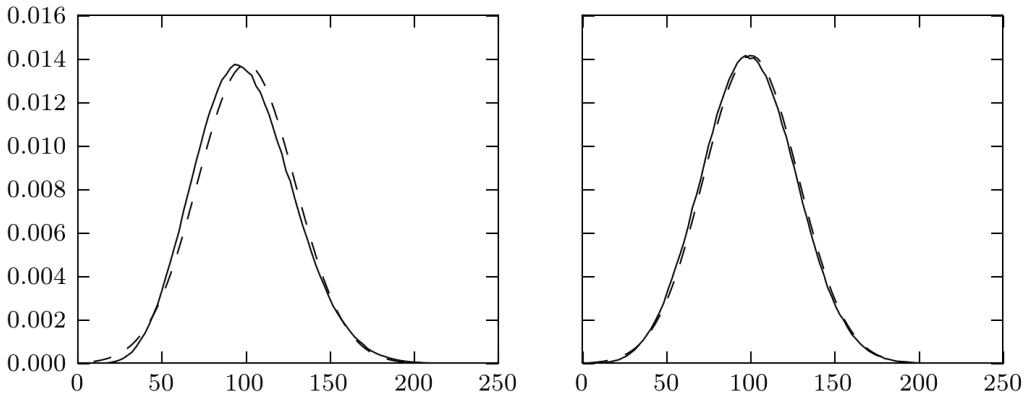
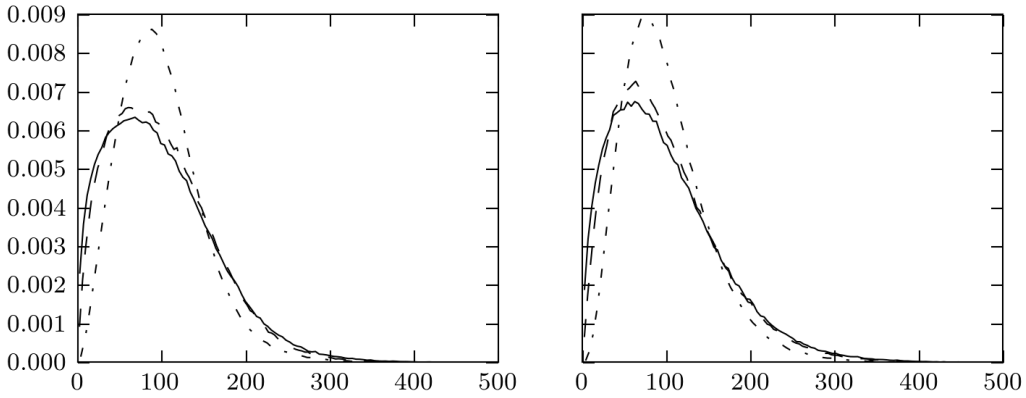


Figure 10.2: Densities of the endprices in multidimensional models. The left plot uses the standard Heston model c), the right plot the generalized model d). The first, second and third price process is drawn solid, dashed or with dashes and dots, respectively. Both plots are based on 500.000 *approximations generated using the DISE scheme and* 2^8 *steps.*



10.4 Numerical Results

Figures 10.3-10.6 plot the measured costs (10.2) versus empirical relative root-mean-square error (10.1) in models a) to d). In the one-dimensional models and the multidimensional generalized model the use of the integration by parts formula leads to an algorithm which is approximately 4 times more efficient than the multilevel Monte-Carlo algorithm which uses the discontinuous functional f directly. In the multidimensional standard model the benefit reduces somewhat to being 2 times more efficient. For the standard Heston model the benefit is comparable to the benefit of MLMC over standard Monte-Carlo, while for the generalized models b) and d) the benefit is significantly larger. However, it turns out to be very important to choose the right splitting parameter δ for the integration by parts formula to be of advantage.

The measurements of the convergence order in Table 10.1 show standard Monte-Carlo to be very close to the theoretical value of 3, while the multilevel Monte-Carlo algorithms have a higher order than theory predicts. For algorithm A3, one needs to take the logarithmic factor of the predicted cost $= \mathcal{O}(\varepsilon^{-2}(\log \varepsilon)^2)$ into account to move the measured exponent close to the theoretical value of 2.

Additionally, we performed a comparison of running times. To this end, all three algorithms were executed 500 times with input error tolerance of $\varepsilon = 2^{-7.5}$ for algorithm A1 and $\varepsilon = 2^{-8}$ for algorithms A2 and A3. The averages of the measured errors, costs and running times are listed in Tables 10.2 to 10.4. Note that the achieved accuracy is approximately $2^{-7.5}$ for all three algorithms and the running times are thus comparable.

Algorithm	Model (a)	Model (b)	Model (c)	Model (d)
A1: Monte Carlo	2.988	2.990	3.010	3.031
A2: Adaptive MLMC	2.634	2.760	2.752	2.991
A3: Adaptive MLMC & Malliavin IBP	2.537	2.616	2.664	2.595
A3 with logarithmic fit	2.144	2.190	2.233	2.140

Table 10.1: Measured convergence order. The method of least-squares was used to fit a function of the form cost $= c \cdot \varepsilon^{-\alpha}$ *to the data, where* ε *is the relative root-mean-square error. In the last line cost* $= c \cdot \varepsilon^{-\alpha}(\log \varepsilon)^2$ *was used instead. The table lists the resulting values of* α*.*

Figure 10.3: Model a) $d = 1, \gamma = 0.5, \delta = 0.3$.
Input accuracies have been $\{2^{-5}, \dots, 2^{-10}\}$ *for algorithm A1 and* $\{2^{-5}, \dots, 2^{-11}\}$ *for algorithms A2 and A3. The reference value was computed using the formula of Theorem 2.8.*

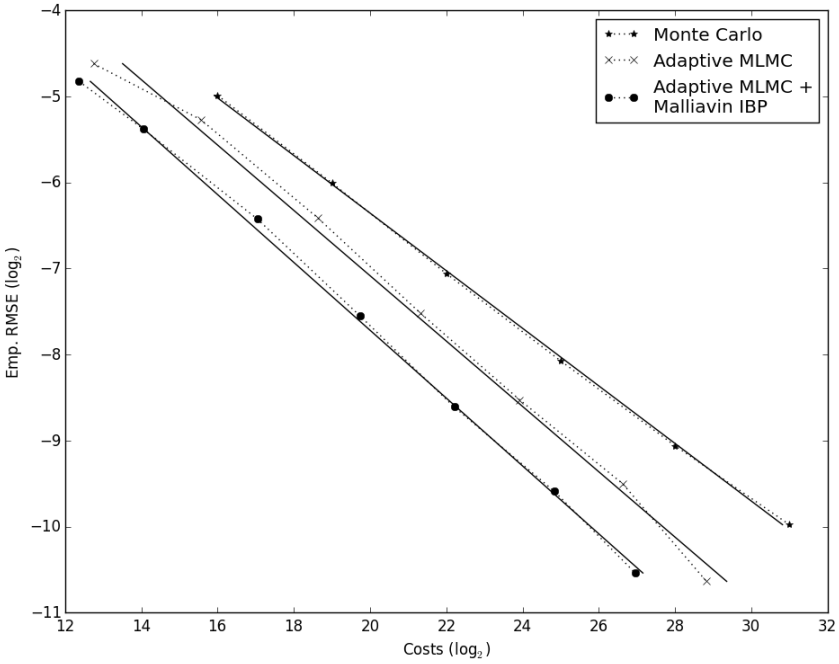


Figure 10.4: Model b) $d = 1, \gamma = 0.6545, \delta = 0.2$.
Input accuracies have been $\{2^{-4}, \dots, 2^{-9}\}$ *for algorithm A1 and* $\{2^{-5}, \dots, 2^{-10}\}$ *for algorithms A2 and A3. The reference value was computed using algorithm A3 and* $\varepsilon = 2^{-12}$.

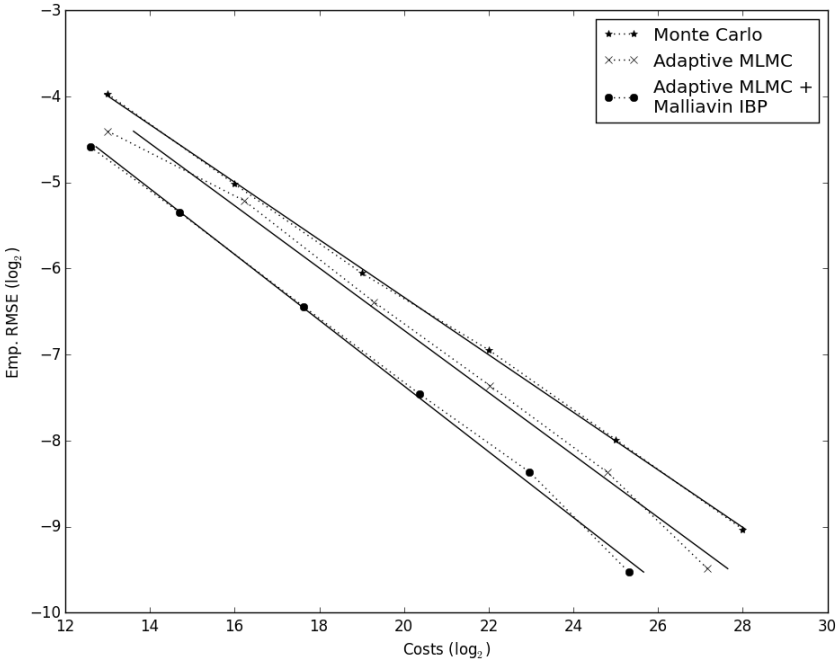


Figure 10.5: Model c) $d = 3, \gamma = 0.5, \delta = 0.1$.
Input accuracies have been $\{2^{-5}, \dots, 2^{-9}\}$ *for algorithm A1 and* $\{2^{-5}, \dots, 2^{-10}\}$ *for algorithms A2 and A3. The reference value was computed using algorithm A3 and* $\varepsilon = 2^{-12}$.

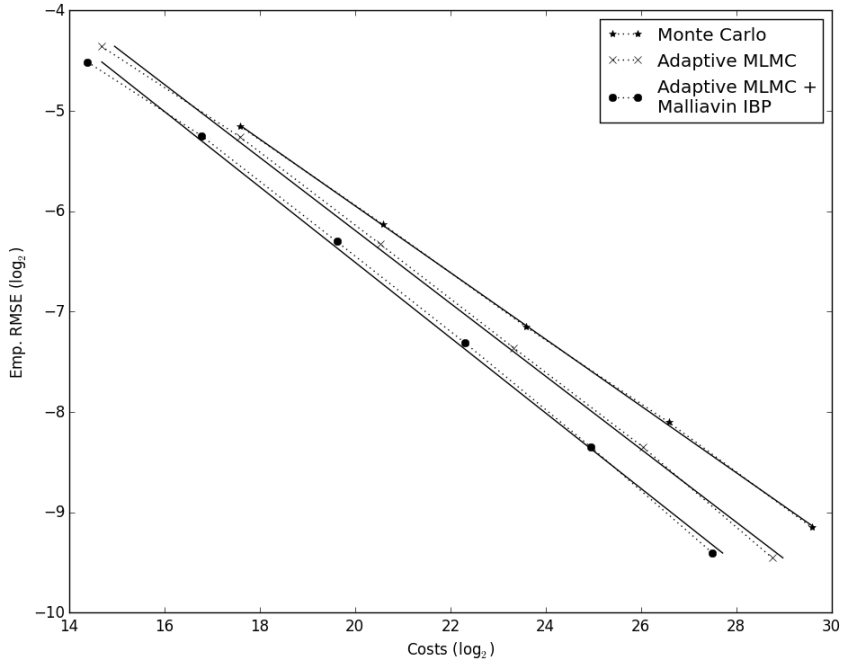
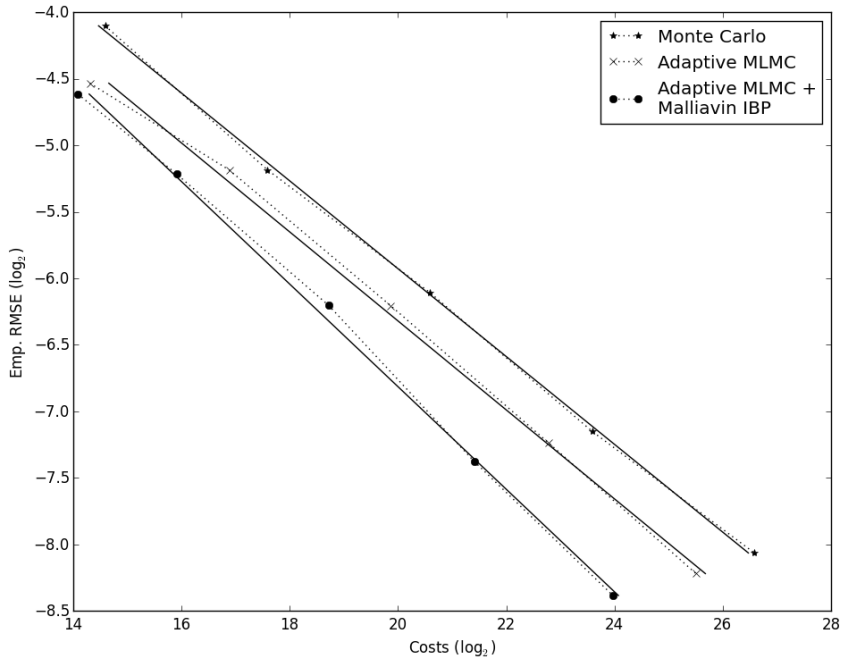


Figure 10.6: Model d) $d = 3, \gamma_1 = 0.63, \gamma_2 = 0.68, \gamma_3 = 0.71, \delta = 0.15$.
Input accuracies have been $\{2^{-4}, \dots, 2^{-8}\}$ *for algorithm A1 and* $\{2^{-5}, \dots, 2^{-9}\}$ *for algorithms A2 and A3. The reference value was computed using algorithm A3 and* $\varepsilon = 2^{-11}$.



Algorithm	Model (a)	Model (b)	Model (c)	Model (d)
A1: Monte Carlo	0.005 (0.003)	0.005 (0.004)	0.003 (0.003)	0.004 (0.003)
A2: Adaptive MLMC	0.005 (0.003)	0.005 (0.003)	0.005 (0.003)	0.006 (0.005)
A3: Adaptive MLMC & Malliavin IBP	0.004 (0.003)	0.005 (0.003)	0.005 (0.004)	0.006 (0.004)

Table 10.2: Empirical relative root-mean-square errors (with standard deviations); see (10.1)*. Input accuracy was $\varepsilon = 2^{-7.5}$ for algorithm A1 and $\varepsilon = 2^{-8}$ for algorithms A2 and A3.*

Algorithm	Model (a)	Model (b)	Model (c)	Model (d)
A1: Monte Carlo	11.9	11.9	35.7	35.7
A2: Adaptive MLMC	2.8 (1.3)	4.3 (1.6)	9.7 (2.9)	6.9 (2.6)
A3: Adaptive MLMC & Malliavin IBP	0.9 (0.4)	1.4 (0.4)	5.3 (1.5)	2.7 (0.9)

Table 10.3: Measured costs (number of discretization steps) in million steps (with standard deviations); see (10.2)*. Input accuracy was $\varepsilon = 2^{-7.5}$ for algorithm A1 and $\varepsilon = 2^{-8}$ for algorithms A2 and A3.*

Algorithm	Model (a)	Model (b)	Model (c)	Model (d)
A1: Monte Carlo	0.92 (0.01)	7.90 (0.04)	5.24 (0.05)	28.92 (0.16)
A2: Adaptive MLMC	0.31 (0.16)	3.15 (1.17)	1.23 (0.44)	5.31 (2.34)
A3: Adaptive MLMC & Malliavin IBP	0.10 (0.04)	0.97 (0.28)	0.64 (0.17)	2.00 (0.62)

Table 10.4: Running times in seconds (with standard deviations). Input accuracy was $\varepsilon = 2^{-7.5}$ for algorithm A1 and $\varepsilon = 2^{-8}$ for algorithms A2 and A3. Higher running times in b) and d) are due to the bisection method necessary in each step of the implicit scheme.

10.5 Conclusion

In this thesis we have seen several applications of the Malliavin integration by parts rule in the Heston model. Most notably, the rule can be used to replace discontinuous functionals of the Heston price by continuous ones. The exact formula is

$$E(f(S_T)) = E\left(\frac{F(S_T)}{S_T^1} \cdot \left(1 + \frac{1}{R_{11}T} \cdot \int_0^T (v_s^1)^{-1/2} dZ_s^1\right)\right)$$

with F being an antiderivative of f. An analogous formula holds for the price approximated with the DISE scheme. When using multilevel Monte-Carlo, this leads to a significantly faster algorithm, as we have shown both in theory and in practice.
Other applications of the integration by parts rule have been a proof that the price in the multidimensional Heston model has a density, and the derivation of the weak error rate of 1 for discontinuous payoff functionals in Chapter 7. All these applications show that the rule is a versatile tool to remove undesired derivatives from functionals of the Heston price.
In Chapter 7 one of the major problems when dealing with the Heston model became visible: It is often difficult to obtain results without making strong assumptions on $\nu = 2\kappa\lambda/\theta^2$, although numerical experiments indicate that these restrictions are rarely necessary in practice. In the case of the weak error, the critical requirement stems from the necessary inverse moments of the drift-implicit Milstein scheme. In a future article we will prove the same result, but with moderate smoothness assumptions on the payoff f and essentially no assumption on ν. And again, the Malliavin integration by parts rule will play a role.

Appendix A

The following theorem — known as Yamada–Watanabe–Comparison Lemma — allows to compare two diffusions with the same diffusion coefficient but different drifts. In the proof of Lemma 4.2 it was used with $h(x) = x^\gamma$, $\gamma \in [1/2, 1)$. Because some work was necessary to check the five conditions, we state the theorem including the conditions here.

Theorem A.1 (Prop. 5.2.18 in [KS10]) For $j \in \{1, 2\}$ let X^j be a continuous adapted process such that

$$X_t^j = X_0^j + \int_0^t b_j(s, X_s^j)ds + \int_0^t \sigma(s, X_s^j)dW_s, \qquad t \in \mathbb{R}_{\geq 0}$$

Assume that the following conditions hold:

1. the coefficients $\sigma(t, x)$ and $b_j(t, x)$ are continuous functions on $\mathbb{R}_{\geq 0} \times \mathbb{R}$,
2. $\sigma(t, x)$ satisfies $|\sigma(t, x) - \sigma(t, y)| \leq h(|x - y|)$ for every $t \geq 0$ and all $x, y \in \mathbb{R}$ and a strictly increasing function $h\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $h(0) = 0$ and

   $$\int_0^\varepsilon h^{-2}(u)du = \infty$$

   for all $\varepsilon > 0$,
3. $X_0^1 \leq X_0^2$ almost surely,
4. $b_1(t, x) \leq b_2(t, x)$ for all $t \geq 0$, $x \in \mathbb{R}$,
5. either b_1 or b_2 is Lipschitz continuous, i.e. for $j = 1$ or $j = 2$ there exists a $K > 0$ such that $|b_j(t, x) - b_j(t, y)| \leq K \cdot |x - y|$ for all $x, y \in \mathbb{R}$.

Then $P(X_t^1 \leq X_t^2 \ \forall\ t \geq 0) = 1$.

Bibliography

[AE08] E. Alòs and C.-O. Ewald. Malliavin differentiability of the Heston volatility and applications to option pricing. *Adv. Appl. Probab.*, 40(1):144–162, 2008.

[Alf05] A. Alfonsi. On the discretization schemes for the CIR (and Bessel squared) processes. *Monte Carlo Methods Appl.*, 11(4):355–384, 2005.

[Alt11] M. Altmayer. *Malliavin Calculus and Approximations*. Unpublished diploma thesis, Technical University of Kaiserslautern, 2011.

[AN15] M. Altmayer and A. Neuenkirch. Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model Using Malliavin Integration by Parts. *SIAM J. Finan. Math.*, 6(1):22–52, 2015.

[And07] L.B.G. Andersen. Efficient Simulation of the Heston Stochastic Volatility Model. 2007.

[AP06] L.B.G. Andersen and V.V. Piterbarg. Moment explosions in stochastic volatility models. *Fin.Stoch.*, 11:29–50, 2006.

[AS64] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. U.S. Departement of Commerce, Washington, 1964.

[ASK07] Y. Aït-Sahalia and R. Kimmel. Maximum likelihood estimation of stochastic volatility models. *J. Financial Economics*, 83:413–452, 2007.

[Avi09] R. Avikainen. On irregular functionals of SDEs and the Euler scheme. *Fin.Stoch.*, 13:381–401, 2009.

[Bau74] H. Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter, 2nd edition, 1974.

[BBD07] A. Berkaoui, M. Bossy, and A. Diop. Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence. *ESAIM Probab. Stat.*, 12:1–11, 2007.

[BK06] M. Broadie and Ö. Kaya. Exact simulation of stochastic volatility and other affine jump diffusion models. *Operations Research*, 54(2), 2006.

[BRFCU10] S. Baño-Rollin, A. Ferreiro-Castilla, and F. Utzet. On the density of log-spot in the Heston volatility model. *Stochastic Processes and their Applications*, 120(10):2037–2063, 2010.

[Bru57] F. Bruno. Note sur une nouvelle formule de calcul differentiel. *The Quaterly Journal of Pure and Applied Mathematics*, (1):359–360, 1857.

[BS73] F. Black and M.S. Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[CIR85] J.C. Cox, J.E. Ingersoll, and S.A. Ross. A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2), 1985.

[CM99] P. Carr and D.B. Madan. Option valuation using the fast Fourier transform. *J. Comput. Finance*, 2:61–73, 1999.

[DD98] G. Deelstra and F. Delbaen. Convergence of Discretized Stochastic (Interest Rate) Processes with Stochastic Drift Term. *Appl. Stochastic Models Data Anal.*, 14(1):77–84, 1998.

[DLS11] G. Dimitroff, S. Lorenz, and A. Szimayer. A parsimonious multi-asset Heston model: Calibration and derivative pricing. *Int. J. Theor. Appl. Finance*, 14(08):1299–1333, 2011.

[DNOP09] G. Di Nunno, B. Øksendal, and F. Proske. *Malliavin Calculus for Lévy Processes with Applications to Finance*. Universitext. Springer, 2009.

[DNS12] S. Dereich, A. Neuenkirch, and L. Szpruch. An Euler-type method for the strong approximation of the Cox–Ingersoll–Ross process. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 468:1105–1115, 2012.

[Eva10] L.C. Evans. *Partial Differential Equations*. Amer. Math. Soc., Providence, Rhode Island, 2nd edition, 2010.

[Fel51] W. Feller. Two Singular Diffusion Processes. *Annals of Mathematics*, 54(1):173–182, 1951.

[FKR10] P.K. Friz and M. Keller-Ressel. Moment explosions in stochastic volatility models. In *Encyclopedia of Quantitative Finance*. John Wiley & Sons, Ltd, 2010.

[FLL+99] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.L. Lions, and N. Touzi. Applications of malliavin calculus to monte carlo methods in finance. *Finance and Stochastics*, 3:391–412, 1999.

[Fri64] A. Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, New Jersey, 1964.

[Gil08] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[GNR14] M. Giles, T. Nagapetyan, and K. Ritter. Multi-Level Monte Carlo approximation of distribution functions and densities. *DFG SPP 1324 Preprint*, 2014.

[GP51] J. Gil-Pelaez. Note on the inversion theorem. *Biometrika*, 38:481–482, 1951.

[Gy98] I. Gyöngy. A Note on Euler's Approximations. *Potential Analysis*, 8(3):205–216, 1998.

[Hei01] S. Heinrich. Multilevel Monte Carlo methods. In *Large-scale scientific computing, Margenov, S. (ed.) et al., Lect. Notes Comput. Sci. 2179*, pages 58–67, 2001.

[Hes93] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financial Studies*, 6(2):327–343, 1993.

[Heu06] Harro Heuser. *Gewöhnliche Differentialgleichungen: Einführung in Lehre und Gebrauch*. Lehrbuch : Mathematik. Teubner, 2006.

[HK08] T.R. Hurd and A. Kuznetsov. Explicit formulas for Laplace transforms of stochastic integrals. *Markov Process. Related Fields*, 14(2):277–290, 2008.

[HM05] D.J. Higham and X. Mao. Convergence of Monte Carlo simulations involving the mean-reverting square root process. *Journal of Computational Finance*, 8(3):35–61, 2005.

[HW87] J. Hull and A. White. The Pricing of Options on Assets with Stochastic Volatilities. *The Journal of Finance*, 42:281–300, 1987.

[Iac08] S.M. Iacus. *Simulation and Inference for Stochastic Differential Equations*. Springer, 2008.

[iHF10] K.J. in't Hout and S. Foulon. ADI finite difference schemes for option pricing in the Heston model with correlation. *Int. J. Numer. Anal. Model*, 7:303–320, 2010.

[IW89] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*. North-Holland, 1989.

[Kar81] Karlin, S. and Taylor, H.M. *A Second Course in Stochastic Processes*. Academic Press, 1981.

[Keb05] A. Kebaier. Statistical Romberg Extrapolation: A New Variance Reduction Mtehod and Applications to Option Pricing. *Annals of Applied Probability*, 15(4):2681–2705, 2005.

[KJ05] C. Kahl and P. Jäckel. Not-so-complex logarithms in the Heston model. *Wilmott*, pages 94–103, September 2005.

[KJ06] C. Kahl and P. Jäckel. Fast strong approximation Monte-Carlo schemes for stochastic volatility models. *Journal of Quantitative Finance*, 6(6):513–536, 2006.

[KK99] R. Korn and E. Korn. *Optionsbewertung und Portfolio-Optimierung*. Vieweg+Teubner, 2nd edition, 1999.

[KP92] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Applications of mathematics : stochastic modelling and applied probability. Springer, 1992.

[KS10] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, N.Y., 2nd edition, 2010.

[LKvD10] R. Lord, R. Koekkoek, and D. van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *Journal of Quantitative Finance*, 10(2):177–194, 2010.

[LM07] P.-L. Lions and M. Musiela. Correlations and Bounds for Stochastic Volatility Models. *Annales de l'Institut Henri Poincaré (C) Non Linear Analysis*, 24(1):1–16, 2007.

[Mal76] P. Malliavin. Stochastic calculus of variations and hypoelliptic operators. *Proc. Int. Symp. Stochastic Diff. Eq. (Kyoto, 1976)*, pages 195–263, 1976.

[NS14] A. Neuenkirch and L. Szpruch. First order strong approximations of scalar SDEs defined in a domain. *Numerische Mathematik*, 128(1):103–136, 2014.

[Nua06] D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, 2nd edition, 2006.

[Øk97] B. Øksendal. An introduction to Malliavin calculus with applications to economics, 1997. Available online at `http://bora.nhh.no/bitstream/2330/841/1/oksendal%20bernt%20wp0396.pdf`; visited on July 14th 2010.

[Øk07] B. Øksendal. *Stochastic Differential Equations*. Springer, 6th edition, 2007.

[Sco87] L.O. Scott. Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application. *Journal of Financial and Quantitative Analysis*, 22:419–438, 12 1987.

[SS91] E.M. Stein and J.C. Stein. Stock Price Distributions with Stochastic Volatility: An Analytic Approach. *The Review of Financial Studies*, 4(4):727–752, 1991.

[TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509, 1990.

[Wer07] D. Werner. *Funktionalanalysis*. Springer, Berlin, 6th edition, 2007.