Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model using Malliavin Integration by Parts

M. Altmayer, A. Neuenkirch

Preprint 144
Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg
Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model using Malliavin Integration by Parts

M. Altmayer, A. Neuenkirch

Preprint 144
The consecutive numbering of the publications is determined by their chronological order.

The aim of this preprint series is to make new research rapidly available for scientific discussion. Therefore, the responsibility for the contents is solely due to the authors. The publications will be distributed by the authors.
Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model using Malliavin Integration by Parts

Martin Altmayer∗, Andreas Neuenkirch†

Abstract. In this manuscript, we establish an integration by parts formula for the quadrature of discontinuous payoffs in a multidimensional Heston model. For its derivation we use Malliavin calculus techniques and work under mild integrability conditions on the payoff and under the assumption of a strictly positive volatility. Since the integration by parts procedure smooths the original functional, our formula in combination with a payoff-splitting allows to construct efficient multilevel Monte Carlo estimators, which is illustrated by several numerical examples.

Key words. Generalized Heston model, Malliavin calculus, Multilevel Monte Carlo, Drift-implicit Euler scheme

AMS subject classifications. 60H07; 60H35; 65C05; 91G60

1. Introduction. The Heston Model is a popular stochastic volatility model in mathematical finance. It was introduced in [18] and further developed since then, see e.g. the recent work [16] on the Heston model with stochastic interest rates or [15] on time-dependent coefficients. In its classical form the Heston model uses a Cox-Ingersoll-Ross process (CIR) as volatility and is given by the SDEs

\[ \begin{align*}
    dS_t &= bS_t dt + \sqrt{v_t} S_t (\rho dW^1_t + \sqrt{1-\rho^2} dW^2_t), & t \in [0,T] \\
    dv_t &= \kappa (\lambda - v_t) dt + \theta \sqrt{v_t} dW^1_t, & t \in [0,T]
\end{align*} \]

with \( S_0, v_0, \kappa, \lambda, \theta > 0, b \in \mathbb{R}, \rho \in [-1,1] \) and independent Brownian motions \( W^1, W^2 \). In this article we will also consider a generalized Heston model (introduced in [5]) that uses a mean-reverting constant elasticity of variance process (CEV) as volatility, i.e.

\[ dv_t = \kappa (\lambda - v_t) dt + \theta v_t^\gamma dW^1_t \]

with \( \gamma \in (1/2,1) \).

An efficient method to compute expectations of (smooth) functionals of SDEs is the multilevel Monte Carlo method, see [17] and [13]. Combining approximations using different step-sizes in a way that reduces the overall variance this method usually is significantly more efficient than standard Monte Carlo. However, the method requires a good \( L^2 \)-convergence rate for the approximations which is often not easy to achieve for discontinuous functionals, see [6, 14].

In this article we are interested to compute

\[ Ef(S_T) \]
for general, e.g. discontinuous, payoff functionals \( f : [0, \infty) \to \mathbb{R} \) by means of multilevel Monte Carlo methods. For this, we will prove that under certain assumptions such an expected payoff can be written as

\[
Ef(S_T) = E \left( \frac{F(S_T)}{S_T} \cdot \Pi \right)
\]

where \( F : [0, \infty) \to \mathbb{R} \) is an antiderivative of \( f \) and

\[
\Pi = 1 + \frac{1}{\sqrt{1 - \rho^2 T}} \int_0^T \frac{1}{\sqrt{v_r}} \mathrm{d}W_r^2
\]

Thus even for discontinuous \( f \) the expectation on right hand side of (1.1) contains a continuous functional of the price as well as a weight term independent of the functional. The proof of this representation formula relies on a Malliavin integration by parts technique. Moreover, we will show that this approach can be easily extended to multidimensional Heston models.

Based on the representation formula (1.1), which smoothes the original payoff functional, we will then construct a multilevel estimator for \( p = Ef(S_T) \) that uses a log-Euler method for the Heston price, an Euler method for the Malliavin weight, and a drift-implicit Lamperti-Euler method for the volatility (see e.g. [2, 9]). The latter method preserves the strict positivity for the volatility, which is crucial for the discretization of the Malliavin weight (1.2). Combined with a payoff-splitting to reduce the variance of the Malliavin weight, this estimator outperforms the direct multilevel Monte Carlo estimator for \( Ef(S_T) \) in our numerical experiments. We strongly suppose that this a general feature of the presented Malliavin integration by parts multilevel estimator and a complete error analysis will be carried out in a forthcoming project to justify this conjecture.

The remainder of this manuscript is structured as follows: In the following three sections we derive the required Malliavin regularity of the CIR- and CEV-processes and also of the Heston price. Note that we could treat the CIR- and CEV-process simultaneously (using the notation \( \gamma \in [1/2, 1) \)), but in our opinion this would be less accessible to the reader. The proofs for the CEV-process follow the same ideas as for the CIR-process, so in the analysis of the CEV-process (which is a bit more technical) we only outline the differences. Formula (1.1) is established in Section 5, while Section 6 studies the multidimensional case. Finally, in Section 7 we present the used multilevel Monte Carlo method, while Section 8 contains our numerical results. The appendix lists some required results from Malliavin calculus and other stochastic analysis tools.

2. The CIR-process. The first volatility process considered in this paper is the well known Cox-Ingersoll-Ross (CIR) process defined by the SDE

\[
dv_t = \kappa (\lambda - v_t) dt + \theta \sqrt{v_t} dW_t, \quad v_0 > 0
\]

for \( \kappa, \lambda, \theta > 0 \). A theorem of Yamada and Watanabe (see [22, Prop. 5.2.13]) implies that this SDE possesses a unique strong solution. We will always work under the so called Feller condition for the CIR-process:

Assumption 2.1.

\[
2\kappa \lambda > \theta^2
\]
It turns out that this assumption is necessary for our mathematical analysis and often met in stock markets, see e.g. [10, 27, 1]. This condition in particular guarantees that

$$P(v_t > 0 \ \forall t \in [0, T]) = 1$$

for all $T > 0$, see e.g. [21, Lemma 2.2], and the existence of inverse moments up to a certain order:

**Lemma 2.2.** Assume that $2\kappa\lambda > p\theta^2$ for some $p \in \mathbb{R}$. Then

$$\sup_{t \in [0, T]} E(v_t^{-p}) < \infty$$

**Proof.** This follows from [19, Thm. 3.1]. See also [9].

Moreover, all moments of the maximum of the CIR-process exist:

**Lemma 2.3.** We have $E \sup_{t \in [0, T]} v_t^p < \infty$ for all $p \geq 1$.

**Proof.** See Lemma 3.2 in [9].

For a major part of this article we will consider the square root of the volatility, which we denote $\sigma_t := \sqrt{v_t}$, instead of the volatility itself. The Itô formula shows that $\sigma_t$ follows the SDE

$$d\sigma_t = \left(\frac{\kappa \lambda}{2} - \frac{\theta^2}{8} \frac{1}{\sigma_t} - \frac{\kappa}{2} \sigma_t\right) dt + \frac{\theta}{2} dW_t$$

(2.2)

The constant diffusion term is no surprise because the square root is up to the factor $2/\theta$ the Lamperti transformation of the CIR-process. In the following we set $C_\sigma := \kappa \lambda/2 - \theta^2/8$, which is positive by (2.1).

**2.1. Approximation of the square root CIR-process.** To prove that $\sigma_t$ and $v_t$ are Malliavin differentiable, we will follow an article of Alós and Ewald [4] and approximate $\sigma_t$ by a process whose SDE has Lipschitz coefficients. This allows us to use standard results from Malliavin calculus. Because we later want to show the differentiability of $1/\sigma_t$ and $S_t$ we cannot use the results of [4] directly and have to choose a different function $f_\varepsilon$ than Alós and Ewald.

For each $\varepsilon > 0$ let $f_\varepsilon$ be a function that satisfies:

(i) $f_\varepsilon \in C^\infty(\mathbb{R}; \mathbb{R}_{>0})$ with bounded derivative,

(ii) $f_\varepsilon(x) = \frac{1}{x}$ for all $x \geq \varepsilon$,

(iii) $f_\varepsilon$ is monotonically decreasing,

(iv) $f_\varepsilon(x) \leq \frac{1}{x}$ for all $x \in \mathbb{R}_{>0}$,

(v) $|f_\varepsilon'(x)| \leq \frac{1}{x} = |\left(\frac{1}{2}\right)'|$ for all $x \in \mathbb{R}_{>0}$.

Now define a process $\sigma_{\varepsilon}^t$ by the SDE

$$d\sigma_{\varepsilon}^t = \left(C_\sigma f_\varepsilon'(\sigma_{\varepsilon}^t) - \frac{\kappa}{2} \sigma_{\varepsilon}^t\right) dt + \frac{\theta}{2} dW_t, \quad \sigma_{\varepsilon}^0 = \sqrt{v_0}$$

(2.3)

Because the coefficients satisfy the usual Lipschitz and linear growth conditions this SDE has a unique strong solution.
Proposition 2.4. For almost all $\omega \in \Omega$ we have $\sigma^\varepsilon_t(\omega) = \sigma_t(\omega)$ for all $t \in [0,T]$ and all $0 < \varepsilon < \varepsilon_0(\omega) := \inf_{t \in [0,T]} \sigma_t(\omega)$. In particular

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

holds almost surely for $\varepsilon \to 0$.

Proof. Comparing the SDEs in integral form yields for all $\varepsilon > 0$ and almost all $\omega \in \Omega$ that

$$|\sigma^\varepsilon_t(\omega) - \sigma_t(\omega)| \leq \int_0^t C_\sigma \left| f_\varepsilon(\sigma^\varepsilon_\tau(\omega)) - \frac{1}{\sigma_\tau(\omega)} \right| \, d\tau + \int_0^t \frac{\kappa}{2} |\sigma^\varepsilon_\tau(\omega) - \sigma_\tau(\omega)| \, d\tau, \quad t \in [0,T]$$

Fix an $\omega$ such that the path $\sigma(.)(\omega)$ is positive. Then for $0 < \varepsilon < \varepsilon_0(\omega)$ the functions $f_\varepsilon$ and $g : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, $g(x) = 1/x$ coincide on the path $\sigma(.)(\omega)$, i.e. $f_\varepsilon(\sigma_t(\omega)) = g(\sigma_t(\omega))$ for all $t \in [0,T]$. We have

$$|\sigma^\varepsilon_t(\omega) - \sigma_t(\omega)| \leq \int_0^t \left( C_\sigma L_\varepsilon + \frac{\kappa}{2} \right) |\sigma^\varepsilon_\tau(\omega) - \sigma_\tau(\omega)| \, d\tau, \quad t \in [0,T],$$

where $L_\varepsilon$ is a Lipschitz constant for $f_\varepsilon$. By Gronwall’s lemma it follows that $\sigma_t(\omega) = \sigma^\varepsilon_t(\omega)$ for all $t \in [0,T]$. \hfill \blackslug

Note that the last proposition does not imply that $\sigma^\varepsilon$ has almost surely positive paths. In fact, when $\sigma$ falls below $\varepsilon$, the approximation $\sigma^\varepsilon$ might fall below 0, too.

Lemma 2.5. Let $u_t$ be an Ornstein-Uhlenbeck process given by the SDE

$$du_t = -\frac{\kappa}{2} u_t dt + \frac{\theta}{2} dW_t, \quad u_0 = \sigma_0$$

(2.4)

Then $P(u_t \leq \sigma^\varepsilon_t \leq \sigma_t \ \forall t \in [0,T]) = 1$.

Proof. (This follows the proof of [4, Prop. 2.1], but with $f_\varepsilon$ instead of $\Lambda_\varepsilon$.)

(1) We first show $u_t(\omega) \leq \sigma^\varepsilon_t(\omega)$ uniformly in $t$ and for almost all $\omega$. The diffusion coefficients of (2.3) and (2.4) are the same while the drift coefficient of $u_t$ is smaller than the one of $\sigma^\varepsilon_t$. By the Yamada-Watanabe comparison lemma (see Theorem 9.1 in the Appendix) we have $P(u_t \leq \sigma^\varepsilon_t \ \forall t \in [0,T]) = 1$.

(2) The comparison lemma cannot be used to prove the second inequality since the drift term in the SDE for $\sigma_t$ is not continuous on $\mathbb{R}$. But because $\sigma_t > 0$ it suffices to show the inequality $v^\varepsilon_t := (\sigma^\varepsilon_t)^2 \leq \sigma_t^2 = v_t$. The SDE for $v^\varepsilon_t$ is

$$dv^\varepsilon_t = \left( 2C_\sigma \sqrt{v^\varepsilon_t} f_\varepsilon(\sqrt{v^\varepsilon_t}) - \kappa v^\varepsilon_t + \frac{\theta^2}{4} \right) dt + \theta \sqrt{v^\varepsilon_t} dW_t, \quad v^\varepsilon_0 = v_0$$

(2.5)

Now the drift and diffusion terms are continuous on $\mathbb{R}$ (set $\sqrt{x} = 0$ for $x \leq 0$ to define the drift and diffusion coefficients on $\mathbb{R}_{\leq 0}$) and we can apply the comparison lemma after we have verified its remaining conditions. Condition (ii) of the comparison lemma can be fulfilled.
choosing $h(x) = \theta \sqrt{x}$ because of $\sqrt{|x - y|} \geq \sqrt{|x|} - \sqrt{|y|}$ for all $x, y \geq 0$. Since we assumed $f_\varepsilon(x) \leq 1/x$ for all $x > 0$ and thus $f_\varepsilon(\sqrt{x})/\sqrt{x} \leq 1$, condition (iv) follows from

$$
\kappa \lambda \geq \frac{\theta^2}{4} \Rightarrow \left( \kappa \lambda - \frac{\theta^2}{4} \right) \left( \sqrt{x} \cdot f_\varepsilon(\sqrt{x}) \right) \leq \kappa \lambda - \frac{\theta^2}{4} \\
\Rightarrow C_\sigma \left( \sqrt{x} \cdot f_\varepsilon(\sqrt{x}) \right) - \kappa x + \frac{\theta^2}{4} \leq \kappa (\lambda - x)
$$

For the final condition (v) it suffices if only one of the drift coefficients is Lipschitz-continuous, which $\mathbb{R} \ni x \mapsto \kappa (\lambda - x) \in \mathbb{R}$ clearly is. ■

Proposition 2.6. For each $p \geq 1$ we have $\sup_{t \in [0,T]} |\sigma_t^\varepsilon - \sigma_t| \to 0$ in $L^p$.

Proof. From the previous lemma we obtain

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

We have $\sup_{t \in [0,T]} \sigma_t \in L^p$ by Lemma 2.3 and it is a standard application of Doob’s martingale inequality that an Ornstein-Uhlenbeck process $u_t$ satisfies $\sup_{t \in [0,T]} |u_t| \in L^p$. Thus the claim follows from Proposition 2.4 using dominated convergence. ■

2.2. Malliavin differentiability of the CIR-process. Because the volatility depends on only one Brownian motion, we will use one-dimensional Malliavin calculus in this section. The underlying Hilbert space is thus given by $H = L^2([0,T])$.

Proposition 2.7. We have $\sigma_t^\varepsilon \in \mathbb{D}^{1,\infty}$ and

$$
D_r \sigma_t^\varepsilon = \frac{\theta}{2} \cdot \exp \left( \int_r^t \left( -\frac{\kappa}{2} + \frac{C_\sigma}{\sigma^2} f_\varepsilon'(\sigma_s^\varepsilon) \right) ds \right) \cdot 1_{[0,t]}(r) \leq \frac{\theta}{2}
$$

(2.6)

Proof. The claim $\sigma_t^\varepsilon \in \mathbb{D}^{1,\infty}$ and the form of the derivative follow from Theorem 9.4. Note that we can only apply this result because the coefficients of the SDE for $\sigma^\varepsilon$ are differentiable with bounded derivatives on the whole of $\mathbb{R}$. The $\theta/2$-bound is clear because $f_\varepsilon'(x) \leq 0$ (this is not true for the corresponding function $\Lambda_\varepsilon$ in [4]). ■

Proposition 2.8. We have $\sigma_t \in \mathbb{D}^{1,\infty}$ and the derivative is given by

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

(2.7)

Additionally the following uniform convergence holds true almost surely and in $L^p$ for all $p \geq 1$:

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

Proof. Set

$$
v_{r,t} = \frac{\theta}{2} \cdot \exp \left( \int_r^t \left( -\frac{\kappa}{2} - \frac{C_\sigma}{\sigma^2} \right) ds \right) \cdot 1_{[0,t]}(r)
$$
From Proposition 2.4 we know that for almost all \( \omega \in \Omega \) we have \( \varepsilon_0(\omega) := \inf_{s \in [0,T]} \sigma_s(\omega) = 0 \) and \( \sigma_s^2(\omega) = \sigma_s(\omega) \) for all \( 0 < \varepsilon < \varepsilon_0(\omega) \) and \( s \in [0,T] \). Hence comparing the integrals in (2.6) and (2.7) and using the definition of \( f_\varepsilon \) we have

\[
D_r \sigma_t^\varepsilon(\omega) = v_{r,t}(\omega)
\]

for \( r, t \in [0,T] \) and all \( 0 < \varepsilon < \varepsilon_0(\omega) \). Since \( \varepsilon \) does not depend on \( r, t \in [0,T] \) this shows

\[
\sup_{r,t \in [0,T]} |D_r \sigma_t^\varepsilon - v_{r,t}| \to 0 \quad \text{a.s.}
\]

as \( \varepsilon \to 0 \) and the \( L^p(\Omega) \)-convergence follows by dominated convergence. Hence \( D_r \sigma_t^\varepsilon \to v_{r,t} \) in \( L^p(\Omega;H) \) and this implies \( D_r \sigma_t = v_{r,t} \) due to the closedness of the Malliavin derivative. \( \blacksquare \)

Theorem 2.9. The Heston volatility \( \sigma_t \) is in \( \mathbb{D}^{1,\infty} \) and has the derivative

\[
D_r \sigma_t = 2\sigma_t \cdot D_r(\sigma_t) = \theta \cdot \exp \left( \int_r^t \left( -\frac{\kappa}{2} \cdot \frac{C_2}{\sigma_s} \right) \, ds \right) \cdot \sqrt{\sigma_t} \cdot 1_{[0,t]}(r)
\]

(2.8)

Proof. This follows from the Malliavin chain rule Theorem 9.2. \( \blacksquare \)

Our quadrature rule will contain the inverse of the square root volatility. As a preparation we prove now its differentiability.

Proposition 2.10. The inverse square root volatility \( \sigma_t^{-1} \) is in \( \mathbb{D}^{1,2} \) with derivative \( D_r \sigma_t^{-1} = -\sigma_t^{-2} \cdot D_r \sigma_t \).

Proof. Fix \( t \in [0,T] \). By the chain rule Theorem 9.2 we have \( f_\varepsilon(\sigma_t) \in \mathbb{D}^{1,2} \) with derivative \( f'_\varepsilon(\sigma_t) \cdot D_r \sigma_t \). This converges almost surely to \( -1/\sigma_t^2 \cdot D_r \sigma_t \) and its absolute value is bounded by \( \frac{\eta}{2} \cdot 1/\sigma_t^2 \) (see (2.7)) which is in \( L^p(\Omega) \) due to Lemma 2.2. This implies \( 1/\sigma_t \in \mathbb{D}^{1,2} \) and the given form of the derivative. \( \blacksquare \)

3. The CEV-process. In this section we will prove the same results as in the previous one for a new volatility process which differs from the CIR-process only in the exponent of the diffusion term. For \( \gamma \in (1/2, 1) \) and \( \kappa, \lambda, \theta > 0 \) define \( \nu_t(\gamma) \) by the SDE

\[
d\nu_t(\gamma) = \kappa(\lambda - \nu_t(\gamma))dt + \theta(\nu_t(\gamma))^{\gamma}dW_t
\]

(3.1)

Again [22, Prop. 5.2.13] guarantees the existence of a strong solution. In the literature this process is denoted as mean-reverting constant elasticity of variance process (CEV), see e.g. [21, 24]. In the following, we will usually omit the index \( \gamma \) on each process.

In contrast to the CIR-process we do not need further restrictions on the parameters to ensure that \( \nu_t \) remains positive, i.e. we have

\[
P(\nu_t > 0 \ \forall t \in [0,T]) = 1
\]

for all \( T > 0 \), see [5, Prop. 2.1]. Also its moments of any order exist:

Lemma 3.1. Let \( p \geq 0 \). Then

\[
E \sup_{t \in [0,T]} \nu^p_t < \infty
\]
and
\[ \sup_{t \in [0,T]} E(v_t^{-p}) < \infty \]

**Proof.** See Lemma 2.1 in [7]. ■

As for the CIR-process we will use a scaled Lamperti’s transformation to replace the SDE by one with a constant diffusion coefficient. In this case the transformation takes the form
\[ \sigma_t := v_t^{1-\gamma} \] and leads to the SDE
\[
\begin{align*}
\sigma_t &= (1-\gamma) \left( \kappa \lambda \sigma_t^{\gamma} - \kappa \sigma_t - \frac{\gamma \theta^2}{2} \sigma_t \right) dt + \theta (1-\gamma) dW_t \\
&= (1-\gamma) \left( \frac{\kappa \lambda \sigma_t^{\gamma} - \kappa \sigma_t - \frac{\gamma \theta^2}{2} \sigma_t}{\frac{1}{1-\gamma}} \right) dt + \theta (1-\gamma) dW_t \\
&= (1-\gamma) \left( \frac{\kappa \lambda x^{\gamma} - \kappa x - \frac{\gamma \theta^2}{2} x}{\frac{1}{1-\gamma}} \right) dt + \theta (1-\gamma) dW_t
\end{align*}
\]
(3.2)

Now set
\[ f(x) := \kappa \lambda x^{\gamma} - \kappa x - \frac{\gamma \theta^2}{2} x \]
so that
\[ \sigma_t = (1-\gamma) (f(\sigma_t) - \kappa \sigma_t) dt + \theta (1-\gamma) dW_t \]

Note that
\[ \sup_{x \in \mathbb{R}_+} f(x) = \infty \quad \text{and} \quad \sup_{x \to \infty} f(x) = 0, \]
so there exists a \( C_f \geq 0 \)
\[ f(x) \geq -C_f x \quad \forall x \in \mathbb{R}_+ \]

Finally, since
\[ f'(x) = \frac{\kappa \lambda \gamma x^{\gamma-1} - \frac{1}{1-\gamma} + \frac{\gamma \theta^2}{2} x^{-1}}{x} \]
we have
\[ \sup_{x \in \mathbb{R}_+} f'(x) < \infty \]

### 3.1. Approximation of the CEV-volatility

Again we will modify (3.2), replace \( f \) by a \( C^\infty \)-function and approximate \( \sigma \) by the solution of this modified SDE. For each \( \varepsilon > 0 \) choose a function \( f_\varepsilon \) with
\begin{enumerate}
\item \( f_\varepsilon \in C^\infty(\mathbb{R}; \mathbb{R}) \) with bounded derivative,
\item \( f_\varepsilon = f \) on \( [\varepsilon, \infty) \),
\item \( f_\varepsilon \) is monotonically decreasing on \( (-\infty, 0] \),
\item \( f_\varepsilon(x) \leq f(x) \) for \( x \in \mathbb{R}_+ \),
\item \( f'_\varepsilon(x) \leq f'(x) \) for all \( x \in \mathbb{R}_+ \),
\item \( f_\varepsilon(x) \geq -C_f x \) for all \( x \in \mathbb{R}_+ \).
\end{enumerate}

Now define \( \sigma_\varepsilon_t \) by the modified SDE
\[
\begin{align*}
\sigma_\varepsilon_t &= (1-\gamma)(f_\varepsilon(\sigma_\varepsilon_t) - \kappa \sigma_\varepsilon_t) dt + \theta (1-\gamma) dW_t, \\
\sigma_\varepsilon_0 &= \sigma_0
\end{align*}
\]
(3.3)
Proposition 3.2. For almost all $\omega \in \Omega$ we have $\sigma_t^\varepsilon(\omega) = \sigma_t(\omega)$ for all $t \in [0, T]$ and $0 < \varepsilon < \varepsilon_0(\omega) := \inf_{t \in [0, T]} \sigma_t(\omega)$. In particular we have

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

almost surely for $\varepsilon \to 0$.

Proof. This can be shown completely analogously to Proposition 2.4. ■

Lemma 3.3. Let $u_t$ be the solution of the SDE

$$du_t = -(1 - \gamma)(C_f + \kappa)u_t dt + \theta(1 - \gamma)dW_t, \quad u_0 = \sigma_0 \quad \text{(3.4)}$$

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

Proof. (1) We first show $u_t \leq \sigma_t^\varepsilon$. The diffusion coefficients of (3.3) and (3.4) are the same while the drift coefficient of $u_t$ is smaller than that of $\sigma_t^\varepsilon$. So Theorem 9.1 gives $P(u_t \leq \sigma_t^\varepsilon \ \forall t \in [0, T]) = 1$.

(2) The comparison lemma cannot be used to prove $\sigma_t^\varepsilon \leq \sigma_t$ directly because the drift term in the SDE for $\sigma_t$ is not continuous on $\mathbb{R}$. But because $\sigma_t > 0$ it suffices to show the corresponding inequality after the back transformation $v_t := (\sigma_t^\varepsilon)^{1/(1-\gamma)} \leq \sigma_t^{1/(1-\gamma)} = v_t$. The SDE for $v_t$ is

$$dv_t = \left[\left((v_t^\varepsilon)^{1/(1-\gamma)} - \kappa v_t^\varepsilon + \frac{\gamma}{2} (v_t^\varepsilon)^{2\gamma-1}\right)dt + \theta(v_t^\varepsilon)^\gamma dW_t\right] v_t^\varepsilon \to v_0 \quad \text{(3.5)}$$

Now the drift and diffusion terms are continuous on $\mathbb{R}$, if we set $x^\alpha = 0$ for $\alpha \in (0, 1)$ and $x < 0$, and it remains to verify the conditions (ii)-(v) of the comparison lemma. Condition (ii) can be fulfilled choosing $h(x) = \theta \cdot x^\gamma$ since this function is concave and satisfies $\int_0^\infty x^{-2\gamma}dx = \infty$.

For condition (iv) note that the drift coefficient of $v_t$ is

$$x \mapsto x^\gamma f(x^{1-\gamma}) - \kappa x + \frac{\theta^2 \gamma}{2} x^{2\gamma-1}$$

for $x > 0$, while for $x < 0$ it is

$$x \mapsto -\kappa x$$

Due to assumption (iv) for $f_\varepsilon$ we have

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

for $x > 0$, while for $x < 0$ we clearly have $-\kappa x \leq \kappa(\lambda - x)$. For the final condition (v) it suffices that the drift coefficient of $v_t$ is Lipschitz-continuous. ■

The following Proposition can be shown again identically as its analogon for the square root CIR-process.

Proposition 3.4. For each $p \geq 1$ we have $\sup_{t \in [0, T]} |\sigma_t^\varepsilon - \sigma_t| \to 0$ in $L^p$ as $\varepsilon \to 0$. 

3.2. Differentiability of the CEV-volatility. As in Proposition 2.7 we use Theorem 9.4 to obtain \( \sigma_t^\varepsilon \in \mathbb{D}^{1,\infty} \) and to compute the derivative as

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

Note that this derivative is not necessarily bounded by \( \theta(1 - \gamma) \) as it was with the CIR-process. However, it is bounded by \( \theta(1 - \gamma) \cdot \exp \left( T(1 - \gamma) \cdot \sup_{x \in \mathbb{R}_{>0}} f'(x) \right) \) because \( f'_\varepsilon(x) \leq f'(x) \) for \( x > 0 \) and \( f'_\varepsilon(x) \leq 0 \) for \( x \leq 0 \). Using this bound and the previous approximation results one can prove the following proposition exactly as Proposition 2.8.

Proposition 3.5. We have \( \sigma_t \in \mathbb{D}^{1,\infty} \) and the derivative is given by

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

(3.6)

Additionally the following uniform convergence holds true almost surely and in \( L^p \) for all \( p \geq 1 \):

\[
\sup_{r,t \in [0,T]} |D_r \sigma_t^\varepsilon - D_r \sigma_t| \to 0
\]
as \( \varepsilon \to 0 \).

The chain rule Theorem 9.2 shows:

Theorem 3.6. The volatility process \( v_t \) is in \( \mathbb{D}^{1,\infty} \) and has the derivative

\[
D_r v_t = \frac{1}{1 - \gamma} (\sigma_t)^{\frac{\gamma}{1 - \gamma}} \cdot D_r \sigma_t
\]

(3.7)

Later on we need:

Proposition 3.7. The inverse square root volatility \( 1/\sqrt{v_t} \) is in \( \mathbb{D}^{1,2} \) with derivative

\[
D_r \left( \frac{1}{\sqrt{v_t}} \right) = -\frac{1}{2(1 - \gamma)} v_t^{-\frac{3}{2} + \gamma} \cdot D_r \sigma_t
\]

Proof. By the chain rule Theorem 9.2 we have \( \sqrt{f_\varepsilon'(v_t)} \in \mathbb{D}^{1,2} \) with

\[
D_r(\sqrt{f_\varepsilon'(v_t)}) = \frac{1}{2} \cdot \frac{f'_\varepsilon(v_t)}{\sqrt{f_\varepsilon'(v_t)}} \cdot D_r v_t = \frac{1}{2(1 - \gamma)} \cdot \frac{f'_\varepsilon(v_t)}{\sqrt{f_\varepsilon'(v_t)}} \cdot (\sigma_t)^{\frac{\gamma}{1 - \gamma}} \cdot D_r \sigma_t
\]

where \( f_\varepsilon \) is the approximation of \( \mathbb{R}_{>0} \ni x \mapsto 1/x \in \mathbb{R}_{>0} \) from Section 2. By dominated convergence \( \sqrt{f_\varepsilon'(v_t)} \to 1/\sqrt{v_t} \) in \( L^p(\Omega) \) and

\[
D_r(\sqrt{f_\varepsilon'(v_t)}) \to -\frac{1}{2(1 - \gamma)} v_t^{-3/2} \cdot (\sigma_t)^{\frac{\gamma}{1 - \gamma}} \cdot D_r \sigma_t
\]
in \( L^p(\Omega; H) \). Because the Malliavin derivative is a closed operator, the last term must be the derivative \( D_r(1/\sqrt{v_t}) \).
4. The Price Process. This section deals with the price process $S_t$ given by
\[ dS_t = bS_t dt + \sqrt{\sigma_t} S_t dB_t \]  
(4.1)

Here $v_t$ is either the CIR- or the CEV-process and $B_t = \rho \cdot W_t^1 + \sqrt{1 - \rho^2} \cdot W_t^2$ is a Brownian motion composed of two independent Brownian motions, with $W^1$ being the Brownian motion driving $v_t$. To calculate the derivative of the price we will concentrate on the log-price $X_t := \ln S_t$. Itô’s formula shows that $X_t$ can be given explicitly in terms of the volatility:
\[ dX_t = \left( b - \frac{1}{2} v_t \right) dt + \sqrt{\sigma_t} dB_t \]  
(4.2)

We will again use an approximation by an SDE with Lipschitz coefficients. For $\varepsilon > 0$ let $\psi_\varepsilon$ be a function such that

(i) $\psi_\varepsilon : \mathbb{R} \to \mathbb{R}$ is bounded and continuously differentiable,
(ii) $|\psi_\varepsilon(x)| \leq |x|$ for all $x \in \mathbb{R}$,
(iii) $\psi_\varepsilon(x) = x$ on $[0, 1/\varepsilon]$,
(iv) $|\psi_\varepsilon'(x)| \leq 1$ for all $x \in \mathbb{R}$.

Denote by $\sigma^\varepsilon$ the approximation to the transformed volatility $\sigma_t = v_t^{1-\gamma}$ of Section 2 (with $\gamma = 1/2$) or Section 3. To obtain the derivative we define $X^\varepsilon_t$ by
\[ X^\varepsilon_t = X_0 + \int_0^t \left( b - \frac{1}{2} \psi_\varepsilon^{1/\gamma}(\sigma^\varepsilon_s) \right) ds + \int_0^t \psi_\varepsilon^{1-1/\gamma}(\sigma^\varepsilon_s) dB_s \]

Lemma 4.1. For each $p, q \geq 1$ we have
\[ E \sup_{t \in [0,T]} |\sigma^\varepsilon_t - \psi_\varepsilon(\sigma^\varepsilon_t)|^p \to 0 \]
as well as
\[ E \sup_{t \in [0,T]} |\sigma^\varepsilon_t - \psi_\varepsilon(\sigma^\varepsilon_t) \cdot \psi_\varepsilon'(\sigma^\varepsilon_t)|^p \to 0 \]

Proof. By Propositions 2.4 and 3.2 for almost all $\omega$ there exists a $\varepsilon_0(\omega) > 0$ such that $\varepsilon < \sigma_t(\omega) = \sigma^\varepsilon_t(\omega) < 1/\varepsilon$ for all $t \in [0,T]$ and $0 < \varepsilon < \varepsilon_0(\omega)$. For such $\omega$ and $\varepsilon$ we have
\[ \sup_{t \in [0,T]} |\sigma^\varepsilon_t(\omega) - \psi_\varepsilon(\sigma^\varepsilon_t(\omega))| = \sup_{t \in [0,T]} |\sigma^\varepsilon_t(\omega) - \psi_\varepsilon(\sigma^\varepsilon_t(\omega)) \cdot \psi_\varepsilon'(\sigma^\varepsilon_t(\omega))| = 0 \]

Also, both suprema are bounded by $2 \sup_{t \in [0,T]} |\sigma_t|^q$. Since this is $L^p$-integrable, the assertion follows by dominated convergence. \[ \blacksquare \]

Proposition 4.2. We have $\sup_{t \in [0,T]} |X^\varepsilon_t - X_t| \to 0$ in $L^p$ for all $p \geq 1$.

Proof. From the SDEs we obtain the following inequality for some constant $c > 0:
\[ E \sup_{t \in [0,T]} |X^\varepsilon_t - X_t|^p \leq c \cdot E \sup_{t \in [0,T]} \left| \int_0^t \left( \sigma^{1/\gamma}_s - \psi_\varepsilon^{1/\gamma}(\sigma^\varepsilon_s) \right) ds \right|^p \]
+ $c \cdot E \sup_{t \in [0,T]} \left| \int_0^t \left( \sigma^{1-1/\gamma}_s - \psi_\varepsilon^{1-1/\gamma}(\sigma^\varepsilon_s) \right) dB_s \right|^p$
The log-price converges to zero by the previous lemma. For the second item we use the Burkholder-Davis-Gundy inequality as well as \(|x - y|^2 \leq |x^2 - y^2|\) for \(x, y \geq 0\): There is a constant \(c\) such that

\[
E \sup_{t \in [0, T]} \left| \int_0^t \left( \frac{\rho}{\sigma_s^{1-\gamma}} - \frac{\psi^2}{\sigma_s^{1-\gamma}} (\sigma_s^2) \right) dB_s \right|^p \leq c \cdot E \left( \int_0^T \left| \frac{\rho}{\sigma_s^{1-\gamma}} - \frac{\psi^2}{\sigma_s^{1-\gamma}} (\sigma_s^2) \right| ds \right)^{p/2}
\]

which converges to zero again by the previous lemma.

Theorem 4.3. The log-price \(X_t\) is in \(D^{1,p}\) for all \(p \geq 1\) and its derivative is given by

\[
D^1_t X_t = \rho \sqrt{\nu_t} - \frac{1}{2(1-\gamma)} \cdot \int_0^t \frac{\psi^2}{\sigma_s^{1-\gamma}} D^1 \sigma_s ds + \frac{1}{2(1-\gamma)} \cdot \int_0^t \frac{\rho}{\sigma_s^{1-\gamma}} D^1 \sigma_s dB_s
\]

\[
D^2_t X_t = \sqrt{1 - \rho^2 \sigma_t^{1-\gamma}} (\sigma_t^2)
\]

for \(r \leq t\) and \(D^1_t X_t = D^2_t X_t = 0\) else. Moreover for each \(r \in [0,T]\) and \(i \in \{1, 2\}\) the uniform convergence

\[
\sup_{t \in [0,T]} |D^i_t X_t - D^i_r X_t| \to 0
\]

holds in \(L^p(\Omega)\).

Proof. Because the system of SDEs for \(X_t^\gamma\) and \(\sigma_t^\gamma\) has globally Lipschitz coefficients, we can use Theorem 9.4 to calculate the derivatives: For \(r \leq t\) we have

\[
D^1_t X_t^\gamma = \rho \psi^2 \frac{1}{\sigma_s^{1-\gamma}} (\sigma_s^2) - \frac{1}{2(1-\gamma)} \cdot \int_r^t \psi^2 \frac{1}{\sigma_s^{1-\gamma}} (\sigma_s^2) \cdot \frac{1}{\sigma_s^{1-\gamma}} (\sigma_s^2) D^1 \sigma_s ds
\]

\[
+ \frac{1}{2(1-\gamma)} \cdot \int_r^t \frac{1}{\sigma_s^{1-\gamma}} (\sigma_s^2) \cdot \frac{1}{\sigma_s^{1-\gamma}} (\sigma_s^2) D^1 \sigma_s dB_s
\]

\[
D^2_t X_t^\gamma = \sqrt{1 - \rho^2 \psi^2 \frac{1}{\sigma_s^{1-\gamma}} (\sigma_s^2)}
\]

The second derivative and the first term of the first one converge to \(\rho \sqrt{\nu_t}\) and \(\sqrt{1 - \rho^2 \sqrt{\nu_t}}\), respectively, using Lemma 4.1. Now consider the second term of the first derivative:

\[
\sup_{t \in [0,T]} \left| \int_r^t \left( \psi^2 \frac{1}{\sigma_s^{1-\gamma}} (\sigma_s^2) \cdot \frac{1}{\sigma_s^{1-\gamma}} (\sigma_s^2) D^1 \sigma_s - \sigma_s^{1-\gamma} D^1 \sigma_s \right) ds \right|
\]

\[
\leq T \cdot \sup_{s \in [0,T]} \left| \psi^2 \frac{1}{\sigma_s^{1-\gamma}} (\sigma_s^2) - \sigma_s^{1-\gamma} \right| \cdot \sup_{s \in [0,T]} |D^1 \sigma_s|
\]

\[
+ T \cdot \sup_{s \in [0,T]} \left| \sigma_s^{1-\gamma} \right| \cdot \sup_{s \in [0,T]} |D^1 \sigma_s - D^1 \sigma_s|
\]

This converges to 0 in \(L^p(\Omega)\) using Lemma 4.1, Lemma 2.2/3.1 and Proposition 2.8/3.5. Convergence of the third term of the first derivative can be shown analogously, using additionally the Burkholder-Davis-Gundy inequality.

It is well known that in the generalized Heston model moment explosions may appear, see e.g. [5]. For the existence of the moments for all \(T > 0\), one has the following result (see [12]):
Theorem 4.4. Let $u \in [1, \infty)$ and define

$$T^*(u) = \inf\{t \geq 0 : ES^u_t = \infty\}$$

(1) For the Heston model with the CIR-process ($\gamma = \frac{1}{2}$) as volatility process we have

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

(2) For the Heston model with the CEV-process ($\gamma \in (\frac{1}{2}, 1)$) as volatility process we have

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

Based on the following result one can check the Malliavin smoothness of $S_t$:

Theorem 4.5. Assume $S_t \in L^{p+\varepsilon}(\Omega)$ for some $p \geq 1$ and $\varepsilon > 0$. Then the Heston price $S_t$ is in $D^{1,p}$ and its derivative is $DS_t = S_t \cdot DX_t$.

Proof. Because $X_t \in D^{1,\infty}$ we have $S_t \cdot DX_t \in L^p(\Omega; H)$. Now the claim follows directly from the chain rule Theorem 9.2. ■

5. Quadrature. Now we can turn to the derivation of our quadrature formula. We will make the following assumption on the payoff function $f$:

Assumption 5.1. Assume that $f: \mathbb{R}_{\geq 0} \to \mathbb{R}$ is a measurable function bounded by a polynomial $p$ and that $F: \mathbb{R}_{\geq 0} \to \mathbb{R}$ is an antiderivative of $f$ given by $F(x) = \int_0^x f(z)dz$. Furthermore assume that there is an $\varepsilon > 0$ such that $S_T \in L^{\max\{2, \deg p\}+\varepsilon}$.

The following Lemma is a final technical preparation.

Lemma 5.2. For all $p \geq 0$ we have

$$E\left(\int_0^T \frac{1}{\sqrt{v_s}} ds\right)^p < \infty$$

Proof. If $v$ is the CIR-process this is [9, Lemma 3.1]. In the CEV-case the claim follows from Lemma 3.1. ■

Theorem 5.3. If Assumption 5.1 holds true and if in the case $\gamma = 1/2$ Assumption 2.1 is satisfied, then we have

$$E(f(S_T)) = E\left(\frac{F(S_T)}{S_T} \cdot \left(1 + \frac{1}{\sqrt{1-\rho^2 T}} \cdot \int_0^T \frac{1}{\sqrt{v_s}} dW^2_s\right)\right)$$

Proof. (1) First assume that $f$ is bounded and that there exists an $\varepsilon > 0$ such that $f(x) = 0$ for all $x < 3\varepsilon$. Moreover, let $\psi = \psi_\varepsilon$ be a function such that

(i) $\psi: \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is continuously differentiable,

(ii) $\psi(x) = 1$ for $x > 2\varepsilon$,
Now we can use the integration by parts rule of Malliavin calculus and Theorems 4.3 and 4.5 to obtain

\[
E(f(S_T)) = \frac{1}{T} \cdot E \left( \int_0^T f(S_T) \cdot D_r^2 S_T \cdot \frac{1}{D_r^2 S_T} \psi(S_T) dr \right)
\]

Under our assumptions Theorem 9.2 together with Theorem 9.3 prove that \(F(S_T)\) belongs to \(\mathbb{D}^{1,2}\) with derivative \(DF(S_T) = f(S_T) \cdot DS_T\). Thus

\[
E(f(S_T)) = \frac{1}{T} \cdot E \left( \int_0^T f(S_T) D_r^2 S_T \cdot \frac{1}{D_r^2 S_T} \psi(S_T) dr \right)
\]

where the Skorohod integral integrates over \(r\) and with respect to \(W^2\). In order to calculate this integral, we use Proposition 1.3.3 from [26], whose prerequisites we will check now (see Lemma 9.5 in the appendix of this manuscript):

(i) Due to our assumptions we have \(S_T \in \mathbb{D}^{1,2}\). Since \(\mathbb{R}_{\geq 0} \ni x \mapsto \frac{\psi(x)}{x^2} \in \mathbb{R}\) is bounded with bounded first derivative, the chain rule implies that \(\frac{\psi(S_T)}{S_T^2} \in \mathbb{D}^{1,2}\).

(ii) Proposition 2.10 or 3.7 and the discussion of \(L^{1,2}\) in [26, p. 42] imply that \(1/\sqrt{v_r} \in \text{dom} \delta\).

(iii) \(\psi(S_T)/(S_T \sqrt{v_r}) \in L^2(\Omega; H)\): Here we have

\[
E \left( \int_0^T \left( \frac{\psi(S_T)}{S_T \sqrt{v_r}} \right)^2 dr \right) \leq \epsilon^{-2} \cdot E \left( \int_0^T \frac{1}{v_r} dr \right) < \infty
\]

by Lemma 5.2.

(iv) \(\psi(S_T)/S_T \cdot \delta(1/\sqrt{v}) \in L^2(\Omega)\) follows similarly using the Itô-isometry because the Skorohod integral equals \(\int_0^T 1/\sqrt{v_r} dW_r^2\).

Since moreover

\[
\left< D^2 \left( \frac{\psi(S_T)}{S_T} , \frac{1}{\sqrt{v_r}} \right) \right>_H = \left< \left( \frac{\psi'(S_T)}{S_T} - \frac{\psi(S_T)}{S_T^2} \right) D^2 S_T \cdot \frac{1}{\sqrt{v_r}} \right>_H
\]

\[
= \sqrt{1 - \rho^2} \left< \psi'(S_T) \sqrt{v} \cdot \frac{1}{\sqrt{v_r}} \right> - \sqrt{1 - \rho^2} \left< \psi(S_T) \sqrt{v} \cdot \frac{1}{\sqrt{v_r}} \right>
\]

\[
= \sqrt{1 - \rho^2} \cdot \left( \psi'(S_T) - \psi(S_T) \right).
\]
which is square-integrable by our assumptions, we have
\[
\delta \left( \frac{\psi(S_T)}{S_T} \cdot \frac{1}{\sqrt{v_r}} \right) = \frac{\psi(S_T)}{S_T} \cdot \int_0^T \frac{1}{\sqrt{v_r}} \, dW_r^2 - \sqrt{1 - \rho^2 T} \cdot \left( \psi'(S_T) - \frac{\psi(S_T)}{S_T} \right)
\]
Hence we have shown that
\[
E(f(S_T)) = E \left( F(S_T) \cdot \frac{1}{\sqrt{1 - \rho^2 T}} \cdot \frac{\psi(S_T)}{S_T} \cdot \int_0^T \frac{1}{\sqrt{v_r}} \, dW_r^2 \right) - E \left( F(S_T) \cdot \left( \psi'(S_T) - \frac{\psi(S_T)}{S_T} \right) \right)
\]
But since \( F(x) = 0 \) for \( x \leq 3 \varepsilon \) and \( \psi(x) = 1 \) for \( x > 2 \varepsilon \) we have
\[
F(S_T) \frac{\psi(S_T)}{S_T} = F(S_T) = F(S_T) \cdot \left( \frac{\psi(S_T)}{S_T} - \psi'(S_T) \right)
\]
and the assertion follows.

(2) Now let \( f \) be bounded. Then there exists functions \( f_n \) such that \( |f_n(x)| \leq |f(x)| \) for \( x \in \mathbb{R}_{\geq 0} \), \( f_n(x) = 0 \) for \( x < 1/n \) and \( f_n \to f \) Lebesgue-almost everywhere. From (1) we have
\[
E(f_n(S_T)) = E \left( \frac{F_n(S_T)}{S_T} \cdot \Pi \right)
\]
where \( F_n(x) := \int_0^x f_n(z) \, dz \) and the weight
\[
\Pi := 1 + \left( \sqrt{1 - \rho^2 T} \right)^{-1} \int_0^T \sqrt{v_r}^{-1} \, dW_r^2
\]
is in \( L^q(\Omega) \) for all \( q \geq 0 \) by the Burkholder-Davis-Gundy inequality and Lemma 5.2. Since \( f_n \to f \) almost everywhere and \( |f_n| \leq |f| \), we have \( F_n \to F \) almost everywhere due to dominated convergence. Moreover, \( \sup_{x \in \mathbb{R}_{\geq 0}} |F_n(x)/x| \) is uniformly bounded in \( n \in \mathbb{N} \).

Using that \( S_T \) is absolute continuous with respect to the Lebesgue measure by Theorem 9.3 we have by dominated convergence
\[
E(f(S_T)) = \lim_{n \to \infty} E(f_n(S_T)) = \lim_{n \to \infty} E \left( \frac{F_n(S_T)}{S_T} \cdot \Pi \right) = E \left( \frac{F(S_T)}{S_T} \cdot \Pi \right)
\]
(3) Now consider the general case. Choose bounded functions \( f_n \) such that \( |f_n| \leq |f| \) and \( f_n \to f \) almost everywhere. Define \( F_n(x) := \int_0^x f_n(z) \, dz \) and \( P(x) := \int_0^x p(z) \, dz \). Then \( F_n \to F \) almost everywhere due to dominated convergence. By our assumptions we have \( |f_n(S_T)| \leq |f(S_T)| \leq |p(T)| \in L^1(\Omega) \) and, because \( P(x)/x \) is a polynomial of degree \( \deg p \),
\[
\left| \frac{F_n(S_T)}{S_T} \right| \leq \left| \frac{F(S_T)}{S_T} \right| \leq \left| \frac{P(S_T)}{S_T} \right| \in L^{1+\varepsilon/\deg p}(\Omega)
\]
Since the weight \( \Pi \) is in \( L^q(\Omega) \) for all \( q \geq 0 \) we obtain
\[
\frac{F(S_T)}{S_T} \cdot \Pi \in L^1(\Omega)
\]
Again the assertion follows by dominated convergence.
6. Multidimensional Heston Models. In this section we will extend our quadrature formula of Theorem 5.3 to functionals in the multidimensional Heston model, i.e. to \((S^i, v^i)\), \(i = 1, \ldots, d\), given by the SDEs

\[
\begin{aligned}
\left(\frac{dS^i_t}{dv^i_t}\right) &= \left(\begin{array}{cc}
 b_i & S^i_t \\
 0 & \kappa_i(\lambda_i - v^i_t)
\end{array}\right) dt + \left(\begin{array}{c}
 \sqrt{v^i_t} S^i_t \\
 0
\end{array}\right) \theta_i(v^i_t)^\gamma dt \\
& \quad + \left(\begin{array}{c}
 \sqrt{v^i_t} S^i_t \\
 0
\end{array}\right) \gamma dB^i_t dW^i_t
\end{aligned}
\]  

(6.1)

with \(b_i \in \mathbb{R}, \kappa_i, \lambda_i, \theta_i > 0\) and \(\gamma_i \in [1/2, 1)\) for \(i = 1, \ldots, d\). Here \(B^i, W^i, i = 1, \ldots, d\), are (possibly) correlated Brownian motions.

It remains to specify the covariance matrix of \((B, W)\), i.e.

\[
\Sigma^{(B,W)} = \begin{pmatrix}
(EB^i_1 B^j_1)_{i,j=1,\ldots,d} & (EB^i_1 W^j_1)_{i,j=1,\ldots,d} \\
(EW^i_1 B^j_1)_{i,j=1,\ldots,d} & (EW^i_1 W^j_1)_{i,j=1,\ldots,d}
\end{pmatrix}
\]

Assumption 6.1.

(a) Assume \(2\kappa_i \lambda_i > 1\) or \(\gamma_i > 1/2\) for all \(i = 1, \ldots, d\).

(b) Assume that \(\Sigma^{(B,W)}\) is positive definite.

Under assumption (b), we can find an upper \(2d \times 2d\)-triangular matrix \(R\) with positive values on the diagonal such that

\[
Z := R^{-1} \begin{pmatrix} B \\ W \end{pmatrix}
\]

is a standard \(2d\)-dimensional Brownian motion, i.e. the components of \(Z\) are independent Brownian motions. For this define the linear transformation

\[
g: \mathbb{R}^{2d \times 2d} \rightarrow \mathbb{R}^{2d \times 2d}, \quad g(A)_{ij} := A_{2d+1-i, 2d+1-j}, \quad i, j = 1, \ldots, 2d
\]

Then \(g(\Sigma) := g(\Sigma^{(B,W)})\) is the covariance matrix of \((W_n, \ldots, W_1, B_n, \ldots, B_1)\), and let \(LL^T = g(\Sigma)\) be its Cholesky decomposition. Then \(R = g(L)\) is an upper triangular matrix and fulfills \(RR^T = \Sigma^{(B,W)}\).

Lemma 6.2. Let \(Z\) be defined as above and \(S^i_t \in L^{2+\varepsilon}, i = 1, \ldots, d\). The Brownian motion \(Z^1\) is independent of \((S^i_t)_{t \in [0,T]}\) if and only if \(j \neq 1\). Using the partial Malliavin derivative \(D^{Z^1}\) defined in the appendix we have in particular

\[
D^{Z^1}_{r} X^i_t = \begin{cases}
R_{11} \cdot \sqrt{v^i_t} \cdot 1_{[0,t]}(r) & j = 1 \\
0 & j \neq 1
\end{cases}
\]

(6.2)

and

\[
D^{Z^1}_{r} S^i_t = \begin{cases}
S^i_t \cdot R_{11} \cdot \sqrt{v^i_t} \cdot 1_{[0,t]}(r) & j = 1 \\
0 & j \neq 1
\end{cases}
\]

(6.3)

Proof. By construction we have that \(Z^1\) and \(B^j, W^i, j = 2, \ldots, d, i = 1, \ldots, d\), are independent, which implies the first claim.
These independence relations allow us to easily calculate Malliavin derivatives with respect to $Z^1$ (see the appendix for this type of derivative).

$$D^2_r B^1_t = \sum_{k=1}^{2d} R_{1,k} \cdot D^2_r Z^k_t = R_{11} \cdot 1_{[0,t]}(r)$$

Similarly we obtain

$$D^2(B^j_t) = 0 \text{ for } j \neq 1$$

$$D^2(W^i_t) = 0 \text{ for } i = 1, \ldots, d$$

Now we can easily calculate the derivatives of the price processes. ■

An application of Theorem 4.5 and Theorem 9.3 shows that for fixed $i = 1, \ldots, d$ the Heston price $S^i_t$ is absolutely continuous with respect to the Lebesgue measure if $S^i_t \in L^{2+\varepsilon}(\Omega)$.

Now we can easiliy calculate the derivatives of the price processes. ■

An application of Theorem 4.5 and Theorem 9.3 shows that for fixed $i = 1, \ldots, d$ the Heston price $S^i_t$ is absolutely continuous with respect to the Lebesgue measure if $S^i_t \in L^{2+\varepsilon}(\Omega)$.

For our multidimensional quadrature formula we also need the existence of a joint density.

Theorem 6.3. Let $t \in [0,T]$ and $S^i_t \in L^{2+\varepsilon}(\Omega)$, $i = 1, \ldots, d$. Then the law of the random vector $S_t$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$.

Proof. It suffices to prove that the law of $X_t$ is absolutely continuous. This assertion follows from [26, Lemma 2.1.1], if we can show that

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

for all $\varphi \in C^\infty_0(\mathbb{R}^d; \mathbb{R})$ and each $i = 1, \ldots, d$ for some constant $c > 0$.

Case 1: $i = 1$. Applying the standard chain rule ($\varphi$ is differentiable) and writing $D^1$ for $D^2$ and $\delta^1$ for $D^2$, we have

$$|E(\partial_i \varphi(X_t))| = \frac{1}{t} \cdot \left| E \left( \int_0^t \partial_i \varphi(X_t) \cdot D^1_r X^1_t \cdot \frac{1}{D^1_r X^1_t} dr \right) \right|$$

By (6.2) we have $(D^1_r X^1_t)^{-1} = (R_{11} \sigma^1_r)^{-1}$. As seen in the proof of Theorem 5.3, this is in the domain of $\delta^1$ and thus we can apply the integration by parts rule.

$$= \frac{1}{t} \cdot \left| E \left( \varphi(X_t) \cdot \delta^1 \left( \frac{1}{D^1 X^1_t} \right) \right) \right|$$

$$\leq \frac{1}{t} \cdot \|\varphi\|_{\infty} \cdot E \left| \delta^1 \left( \frac{1}{D^1 X^1_t} \right) \right|$$

Case 2: $i \neq 1$. This can be shown analogously to the first case by using a reordering of the Brownian motions so that $B^1$ is the first one.

Now we will extend the quadrature formula to the multidimensional setting. For notational convenience we restrict ourselves to the smoothing of the first component.
Assumption 6.4. Assume that \( f: (\mathbb{R}_{\geq 0})^n \to \mathbb{R} \) is a measurable function, which is almost everywhere continuous and is bounded by a polynomial \( p \). Furthermore let \( F: (\mathbb{R}_{\geq 0})^n \to \mathbb{R} \) be given by \( F(x) = \int_0^{x_1} f(\xi, x_2, \ldots, x_n) d\xi \) for all \( x \in (\mathbb{R}_{\geq 0})^n \). Finally assume that there is an \( \varepsilon > 0 \) such that \( S_i^T \in L^{\max(2, \deg p) + \varepsilon} \) for all \( i = 1, \ldots, d \).

Theorem 6.5. If Assumptions 6.1 and 6.4 hold true, we have

\[
E(f(S_T)) = E\left( \frac{F(S_T)}{S_T^1} \cdot \left( 1 + \frac{1}{R_{11}^T} \cdot \int_0^T \frac{1}{\psi(S_T^1)} \ dZ^1_r \right) \right)
\]

Proof. First assume again that \( f \) is bounded and that there exists an \( \varepsilon > 0 \) such that \( f(x) = 0 \) for all \( |x| < 3\varepsilon \). Moreover, let \( \psi = \psi_\varepsilon \) be again the smooth localizing function with \( \psi(x) = 0 \) for \( x < \varepsilon \) and \( \psi(x) = 0 \) for \( x > 2\varepsilon \) from the proof of Theorem 5.3. By Lemma 6.2 the first Brownian motion \( Z_1 \) is independent of all price processes but the first. To simplify notation we will write \( D \) for the partial derivative \( D_{Z_1} \) and \( \delta \) for \( \delta_{Z_1} \). The previous theorem allows us to apply the chain rule Proposition 9.6 to \( F(S_T) \). Using the integration by parts rule of Malliavin calculus and formula (6.3) we obtain

\[
E(f(S_T)) = \frac{1}{T} \cdot E\left( \int_0^T f(S_T) \cdot D_r S^1_T \cdot \frac{1}{D_r S^1_T} \ dr \right)
= \frac{1}{T} \cdot E\left( \int_0^T D_r F(S_T) \cdot \psi(S_T^1) \cdot \frac{1}{D_r S^1_T} \ dr \right)
= \frac{1}{T} \cdot E\left( F(S_T) \cdot \delta\left( \psi(S_T^1) \cdot \frac{1}{S^1_T} \right) \right)
= \frac{1}{T} \cdot E\left( F(S_T) \cdot \frac{1}{R_{11}^T} \cdot \delta\left( \psi(S_T^1) \cdot \frac{1}{S^1_T} \cdot \frac{1}{\sqrt{v^*_1}} \right) \right)
\]

where the Skorohod integral integrates over \( r \). To compute the Skorohod integral we can use the same proof as in Theorem 5.3 but using \( D^2 \) instead of \( D^2 \) and

\[
D\left( \frac{\psi(S_T^1)}{S^1_T} \right) = \left( \frac{\psi'(S_T^1)}{S_T^1} - \frac{\psi(S_T^1)}{(S_T^1)^2} \right) \cdot DS_T^1
\]

The extension to general \( f \) is exactly the same as in Theorem 5.3. □

In the case of a so called basket option

\[
f(S_T) = g\left( \sum_{i=1}^d a_i S^i_T \right)
\]

with \( g: \mathbb{R}_{\geq 0} \to \mathbb{R} \) and \( a_i \geq 0, i = 1, \ldots, d \), the function \( F: (\mathbb{R}_{\geq 0})^d \to \mathbb{R} \) is given by

\[
F(x) = \frac{1}{a_1} \left( G\left( \sum_{i=1}^d a_i x_i \right) - G\left( \sum_{i=2}^d a_i x_i \right) \right)
\]
with an arbitrary antiderivative $G$ of $g$. For options of the form
\[ f(S_T) = g(S_T^1)h(S_T^2, \ldots, S_T^d) \]
with $h: (\mathbb{R}_{\geq 0})^{d-1} \to \mathbb{R}$ we get
\[ F(x) = G(x_1)h(x^2, \ldots, x^d) \]
for $G(x) = \int_0^x g(\xi)\,d\xi$. Turning to log-coordinates again, i.e. $X_T^i = \log(S_T^i)$, we have the following smoothness result:

**Proposition 6.6.** If $f: (\mathbb{R}_{\geq 0})^d \to \mathbb{R}$ is bounded by $C \geq 0$, then
\[ \tilde{G}: \mathbb{R} \to \mathbb{R} \quad \tilde{G}(x) := \frac{1}{e^{x_1}} \cdot \int_0^{e^{x_1}} f(\xi, x_2, \ldots, x_d)\,d\xi \]
is bounded by $C$ and globally Lipschitz-continuous.

**Proof.** Clearly $\tilde{G}(x) \leq e^{-x_1}Ce^{x_1} = C$. For $x, y \in (\mathbb{R}_{\geq 0})^d$ with $x_1 < y_1$ we have
\[
|\tilde{G}(e^{y_1}) - \tilde{G}(e^{x_1})| = \left| \frac{1}{e^{y_1}} \cdot \int_{e^{x_1}}^{e^{y_1}} f(\xi, x_2, \ldots, x_d)\,d\xi + \left( \frac{1}{e^{y_1}} - \frac{1}{e^{x_1}} \right) \cdot \int_0^{e^{x_1}} f(\xi, x_2, \ldots, x_d)\,d\xi \right|
\leq \frac{C}{e^{y_1}}(e^{y_1} - e^{x_1}) + \frac{C}{e^{x_1}e^{y_1}}(e^{y_1} - e^{x_1})e^{x_1}
= 2C(1 - e^{x_1 - y_1})
\leq 2C(1 - (1 + x_1 - y_1))
= 2C(y_1 - x_1)
\]

### 7. Multilevel Monte Carlo


Assume we want to compute
\[ p = Ef(S_T), \]
where $f: (\mathbb{R}_{\geq 0})^d \to \mathbb{R}$ and $S_T$ is given by the $d$-dimensional generalized Heston model (6.1). While in the one-dimensional case PDE methods, see e.g. [20], or FFT methods [8] can be applied, one has to rely on Monte Carlo methods in higher-dimensional settings. The standard Monte Carlo approach uses a simulatable approximation $S_T^{(h)}$ of $S_T$ based on a discretization of the driving Brownian motions with stepsize $h$ and an average of $N$ independent copies of $S_T^{(h)}$ to estimate $p$, i.e.
\[ p \approx \frac{1}{N} \sum_{i=1}^{N} f(S_T^{(h),(i)}) \]

The multilevel Monte Carlo method instead relies on approximations $S_T^{(h_l)}$ to $S_T$ with different stepsizes $h_l = M^{-l}T$, $l = 0, \ldots, L$, with $M \in \mathbb{N}, M \geq 2$. Its idea is to use the
telescoping sum
\[
Ef(S_T^{(hL)}) = Ef(S_T^{(h_0)}) + \sum_{l=1}^L Ef\left(f(S_T^{(h_l)}) - f(S_T^{(h_{l-1})})\right)
\]

to estimate the expectations on the right hand side using independent standard Monte Carlo estimators
\[
Ef(S_T^{(h_0)}) \approx \hat{P}_0 = \frac{1}{N_0} \sum_{i=1}^{N_0} f(S_T^{(h_0),(i)}),
\]
\[
Ef(S_T^{(h_l)}) \approx \hat{P}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} \left(f(S_T^{(h_l),(i)}) - f(S_T^{(h_{l-1}),(i)})\right), \quad l = 1, \ldots, L,
\]

and to balance the variance and the computational cost of the summands, i.e. the number of random variables, arithmetic operations and function evaluations, in an optimal way. Here the \(S_T^{(h_0),(i)}\) are independent copies of \(S_T^{(h_0)}\) and the \((S_T^{(h_l),(i)}, S_T^{(h_{l-1}),(i)})\) are independent copies of \((S_T^{(h_l)}, S_T^{(h_{l-1})})\), which use the same sample paths of the Brownian motion. The multilevel Monte Carlo estimator of the quantity \(p\) is then given by
\[
\hat{P} = \sum_{l=0}^L \hat{P}_l,
\]
where the estimators \(\hat{P}_l, l = 0, \ldots, L\), of the different levels have to be independent.

Based on estimates for the weak error
\[
|Ef(S_T^{(h)}) - p| \leq c_\alpha \cdot h^\alpha
\]
and the strong error
\[
E|f(S_T^{(h)}) - f(S_T)|^2 \leq c_\beta \cdot h^{2\beta}
\]
with \(c_\alpha, c_\beta, \alpha, \beta > 0\), the number of levels \(L\) and the number of repetitions \(N_l, l = 1, \ldots, L\), can be chosen such that the mean square error satisfies
\[
E|p - \hat{P}|^2 \leq \varepsilon^2
\]
with
\[
\text{computational cost} \leq c_{\alpha,\beta} \begin{cases} 
\varepsilon^{-2} & \text{if } \beta > 1 \\
(\log(\varepsilon))^2 \varepsilon^{-2} & \text{if } \beta = 1 \\
\varepsilon^{-2-\frac{1-\beta}{\alpha}} & \text{if } \beta < 1
\end{cases}
\]
for a given accuracy \(\varepsilon > 0\). Here the constant \(c_{\alpha,\beta}\) depends only on \(\alpha, \beta, c_\alpha, c_\beta\). In particular, the multilevel Monte Carlo estimator outperforms the standard estimator for moderate and high accuracies \(\varepsilon > 0\).

The above multilevel estimator has two drawbacks if applied to the generalized Heston model with general payoffs. It relies on good strong convergence properties, i.e. \(\beta \geq 1\),
and also on the knowledge of $c_\alpha, c_\beta, \alpha$ and $\beta$. However for discontinuous functionals $f$ the strong convergence rates of approximations schemes deteriorate: In [6, 14] it is shown that a discontinuity leads to a halving of the strong convergence rate of the approximation scheme. Also, sharp weak convergence rates for the generalized Heston model have not been established so far up to the best of our knowledge.

Our integration by parts result, i.e. Theorem 6.5, circumvents one of these problems. Here we have shown that

$$E(f(S_T)) = E\left(\frac{F(S_T)}{S_T^2} \cdot \left(1 + \frac{1}{R_{11}T} \cdot \int_0^T \frac{1}{\sqrt{v_1^s}} dZ_1^r\right)\right) \quad (7.1)$$

and for several types of functionals $f$, the arising functional

$$(\mathbb{R}^d \ni x \mapsto \frac{F(x)}{x_1} \in \mathbb{R})$$

is locally or globally Lipschitz, see Section 6. Clearly, in the globally Lipschitz case the strong convergence rate of the approximation scheme for $S_T$ is retained when applied to $F(S_T)/S_T^2$.

Concerning the problem of knowing $c_\alpha, c_\beta, \alpha$ and $\beta$, M. Giles presented in [13] a heuristic algorithm, which chooses the $N_l$’s adaptively and can be used without knowledge of $c_\alpha, c_\beta, \alpha$ and $\beta$, see Algorithm 1.

**Algorithm 1: Adaptive Multilevel Monte Carlo**

We will use this algorithm with $M = 2$ and $N_{ini} = 500$.

Before we can finally apply this adaptive multilevel estimator together with a suitable approximation scheme for

$$E\left(\frac{F(S_T)}{S_T^2} \cdot \left(1 + \frac{1}{R_{11}T} \cdot \int_0^T \frac{1}{\sqrt{v_1^s}} dZ_1^r\right)\right)$$
we need to take care of the fact that the variance of the weight
\[
\Pi = 1 + \frac{1}{R_{11}} \cdot \int_0^T \frac{1}{\sqrt{v_s}} \, dZ_s^1
\]
is typically very high, because it contains the inverse of the typically low volatility. When applied directly, our multilevel estimator would benefit from the smoothing of the functional only for very small error tolerances \( \varepsilon \). To solve this problem we split the payoff in a Lipschitz continuous part \( g \) and a discontinuous part \( h \) with small support and apply the quadrature formula only to the latter part. To be precise, given \( f = g + h \) and \( H \) with \( H(x) = \int_0^x h(\xi, x_2, \ldots, x_d) \, d\xi \), where \( f, g, h, H : \mathbb{R}_{\geq 0}^d \to \mathbb{R} \), we apply the multilevel estimator to the functional
\[
g(S_T) + \frac{H(S_T)}{S_T^\beta} \cdot \Pi
\]

8. Numerical Results. It remains to specify the numerical methods for the approximation of the Heston prices and the weight \( \Pi \). As pointed out the diffusion coefficients of the volatility processes do not satisfy the usual global Lipschitz assumption, so the standard theory for the approximation of SDEs (see e.g. [23]) cannot be applied. However, recently ([2, 9, 3, 25]) it has been shown that the discretization of the Lamperti-transformed volatility process with a backward Euler-scheme leads to a positivity preserving scheme that also attains the standard convergence rates from the global Lipschitz case.

So we consider the transformed volatility processes \( \sigma_i^t = (v_i^t)^{1-\gamma_i} \) and price processes \( X_i^t = \log(S_i^t) \), i.e.
\[
(\frac{dX_i^t}{d\sigma_i^t}) = \begin{pmatrix}
   b_i - \frac{1}{2}(\sigma_i^t)^2 \\
   (1 - \gamma_i)\left(f_i(\sigma_i^t) - \kappa_i\sigma_i^t\right)
\end{pmatrix} dt + \begin{pmatrix}
   \sigma_i^t & 0 \\
   0 & \theta_i(1 - \gamma_i)
\end{pmatrix}
\begin{pmatrix}
   dB_i^t \\
   dW_i^t
\end{pmatrix}
\]

with
\[
f_i(x) := \kappa_i \lambda_i x^{-\frac{1}{\gamma_i}} - \gamma \theta_i^2 x^{-1}
\]

Recall that the Brownian motions \( B^i, W^i \) are given as \( \begin{pmatrix} B \ W \end{pmatrix} = R \cdot Z \) where \( Z \) is a 2d-dimensional Brownian motion and \( R \in \mathbb{R}^{2n \times 2n} \) is an invertible upper triangular matrix. Using these notations we finally set
\[
P = g(e^{X_1^T}, \ldots, e^{X_d^T}) + \frac{H(e^{X_1^T}, \ldots, e^{X_d^T})}{e^{X_1^T}} \cdot \Pi
\]

with
\[
\Pi = 1 + \frac{1}{R_{11}} \cdot \int_0^T \frac{1}{\sigma_s^1} \, dZ_s^1
\]

As already mentioned, we discretize the transformed volatility processes with the backward Euler scheme with step size \( \Delta > 0 \), i.e.
\[
\hat{\sigma}_{k+1} = \hat{\sigma}_k + (1 - \gamma_i)\left(f_i(\hat{\sigma}_{k+1}) - \kappa_i\hat{\sigma}_{k+1}\right) \Delta \\
+ (1 - \gamma_i)\theta_i(W_{(k+1)\Delta} - W_{k\Delta}), \quad k = 0, 1 \ldots
\]
where $\hat{\sigma}_0 = \sigma_0^i$. Under Assumption 6.1(a) this scheme is positivity-preserving, i.e.

$$P \left( \hat{\sigma}_k > 0, \ k = 0, 1, \ldots \right) = 1 \quad (8.3)$$

Moreover, in the case $\gamma = 1/2$ the above implicit equation has the explicit solution

$$\hat{\sigma}_{k+1}^i = \frac{\hat{\sigma}_k^i + \theta \Delta (W^i_{(k+1)\Delta} - W^i_{k\Delta})}{2 + \kappa_i \Delta} + \sqrt{\frac{(\hat{\sigma}_k^i + \theta \Delta (W^i_{(k+1)\Delta} - W^i_{k\Delta}))^2}{(2 + \kappa_i \Delta)^2} + \frac{(\kappa_i \lambda - \theta^2 \Delta)}{2 + \kappa_i \Delta}},$$

while in the general case it can be solved using standard zero-finding methods. We discretize then the log-price $X_T$ and the weight $\Pi$ using the Euler scheme, which yields

$$\hat{X}_N^i = \log(S_0^i) + bT - \frac{1}{2} \sum_{k=0}^{N-1} (\hat{\sigma}_k^i)^2 \Delta + \sum_{k=0}^{N-1} \hat{\sigma}_k^i (B_{(k+1)\Delta} - B_{k\Delta})$$

and

$$\hat{\Pi} = 1 + \frac{1}{R_{11} T} \sum_{k=0}^{N-1} \frac{1}{\hat{\sigma}_k^i \Delta} (Z_{(k+1)\Delta}^i - Z_{k\Delta}^i)$$

where $N \in \mathbb{N}$ with $N \Delta = T$. Note that our approximation of $\Pi$ is well defined due to (8.3) and that the results of [9, 3] imply the strong error estimate

$$\left( E \left| X_T^i - \hat{X}_N^i \right|^p \right)^{1/p} \leq K_{p,i} \cdot \sqrt{\Delta} \quad \text{for} \quad \begin{cases} 1 \leq p < \frac{2 \kappa_i \lambda}{\theta^2} & \text{if } \gamma = \frac{1}{2} \\ 1 \leq p < \infty & \text{else} \end{cases}$$

with $K_{p,i} > 0$. Furthermore set $\hat{S}_T^i := \exp(\hat{X}_T^i)$, $\hat{Y}_T = \sum_{i=1}^d \hat{S}_T^i$, and $Y_T = \sum_{i=1}^d S_T^i$.

As prototype examples for options with discontinuous payoffs we will consider digital options in the one-dimension (extended) Heston model as well as digital basket options in a three-dimensional model. More precisely, we will consider the functional $1_{[0,d,S_0]}$ and approximate $p = E(1_{[0,d,S_0]}(Y_T))$ in the following models:

(a) One-dimensional Heston model. Here the parameters are taken from [1]:

$$T = 2, \ b = 0, \ \kappa = 5.07, \ \lambda = 0.0457, \ \theta = 0.48, \ \rho = -0.767, \ \nu_0 = \lambda, \ S_0 = 100$$

Note that this choice of parameters ensures that $S_T \in L^{2+\varepsilon}$ for some $\varepsilon > 0$ due to Theorem 4.4.

(b) One-dimensional extended Heston model. We use $T = 2, \ b = 0$, and again parameters from [1]:

$$\kappa = 4.1031, \ \lambda = 0.0451, \ \theta = 0.8583, \ \gamma = 0.6545, \ \rho = -0.760, \ \nu_0 = \lambda, \ S_0 = 100$$

The correlation is again sufficiently negative to ensure that $S_T \in L^{2+\varepsilon}$. 
(c) Three-dimensional Heston model. Again we set $T = 2$, $b_1 = b_2 = b_3 = 0$. Here the parameters are taken from [10]. However, we modified $\rho_1, \rho_2$ to avoid negative correlations close to $-1$.

- $\kappa_1 = 1.0121$, $\lambda_1 = 0.2874$, $\theta_1 = 0.7627$, $\rho_1 = -0.7137$, $v_0^1 = 0.2723$, $S_0^1 = 100$,
- $\kappa_2 = 0.5217$, $\lambda_2 = 0.2038$, $\theta_2 = 0.4611$, $\rho_2 = -0.8322$, $v_0^2 = 0.2536$, $S_0^2 = 100$,
- $\kappa_3 = 0.5764$, $\lambda_3 = 0.1211$, $\theta_3 = 0.3736$, $\rho_3 = -0.4835$, $v_0^3 = 0.1539$, $S_0^3 = 100$

The different volatility processes are independent while the different price processes are correlated according to the following correlation matrix

$$
\Sigma^S = \begin{pmatrix}
1 & 0.0246 & 0.0598 \\
0.0246 & 1 & 0.6465 \\
0.0598 & 0.6465 & 1
\end{pmatrix}
$$

Again all prices belong to $L^{2+\varepsilon}$.

(d) Three-dimensional extended Heston model. Because we did not find parameters for this model in the literature, we took the same as in (c), but with $\gamma_1 = 0.63$, $\gamma_2 = 0.68$, $\gamma_3 = 0.71$.

Figures 1 show a comparison of the following three algorithms at different error tolerances $\varepsilon$ in these models.

1. Standard Monte Carlo (using $\lceil \varepsilon^{-2} \rceil$ iterations and a stepsize of $\lceil T/\varepsilon \rceil$)
2. Adaptive Multilevel Monte Carlo using the discontinuous payoff directly, i.e. simulating $\mathbb{1}_{[0,d_S]}(\hat{Y}_T)$
3. Adaptive Multilevel Monte Carlo using the quadrature formula and payoff splitting. For a $\delta \in (0,1)$ we use the splitting

$$
g(x) = \begin{cases} 
1 & x < (1 - \delta)K \\
-\frac{1}{2\delta}(x - K) + 0.5 & x \in [(1 - \delta)K, (1 + \delta)K] \\
0 & x > (1 + \delta)K
\end{cases}
$$

and $h = \mathbb{1}_{[0,K]} - g$, $H(x) = \int_0^x h(z)dz$. As discussed before the algorithm then simulates

$$
g(\hat{Y}_T) + \frac{H(\hat{Y}_T)}{\hat{S}_T^1} \cdot \hat{\Pi}
$$

For each algorithm a series of error tolerances was chosen and the algorithm was executed 500 times for each tolerance. The figures plot the the mean cost against the relative mean-square error, both on logarithmic scale (base 2). Since we the exact value of $p$ is unknown, the reference value $p_{ref}$ has been computed using algorithm (iii) with a very small error tolerance. For each algorithm the figures contain a least squares line that has been fitted to the data points to estimate the convergence rates.

The (random) cost of the multilevel algorithms was measured by the total number of discretization steps, i.e.

$$
cost := d \cdot \left( N_0 + \sum_{i=1}^L N_i \cdot (M^l + M^{l-1}) \right)
$$

(8.4)
where \( d \) is the dimension of the model. Note that we account for the additional cost of computing two approximations for each level \( l \geq 1 \). The mean cost is then given by the average over 500 repetitions of the algorithm. Analogously, the cost of the standard Monte Carlo is also measured by the total number of discretization steps, i.e. \( \lceil \varepsilon^{-2} \rceil \cdot \lceil T/\varepsilon \rceil \).

Finally, the empirical relative root mean-square error is given by

\[
\text{rmsq error} = \sqrt{\frac{1}{500} \sum_{j=1}^{500} \frac{|\hat{P}(j) - p_{\text{ref}}|^2}{p_{\text{ref}}^2}}
\]

Figure 8.1 shows that algorithm (iii) is between 2 and 4 times faster than algorithm (ii), i.e. achieves the same accuracy with between one half and one fourth of the computational cost, and both are much faster than the standard Monte Carlo algorithm (i). However, further numerical tests have shown that it is important to choose a good splitting parameter: For the figures the parameter \( \delta \) which governs the payoff-splitting was set to values between 0.1 and 0.4. Higher values of \( \delta \) lead in fact to an algorithm worse than algorithm (ii).

Table 8.1 contains the estimated convergence rates. As already mentioned, a detailed error analysis will be carried out in a forthcoming project.

\[
\begin{array}{|l|c|c|c|c|}
\hline
\text{Algorithm} & \text{Model (a)} & \text{Model (b)} & \text{Model (c)} & \text{Model (d)} \\
\hline
(i) Monte Carlo & 0.333 & 0.327 & 0.335 & 0.334 \\
(ii) Adaptive MLMC & 0.379 & 0.370 & 0.353 & 0.355 \\
(iii) Ad. MLMC+Malliavin Quadrature & 0.392 & 0.385 & 0.381 & 0.388 \\
\hline
\end{array}
\]

Table 8.1: Estimated convergence rates

Table 8.2 contains the measured running times for a single run (averaged over 100 runs) using error tolerance \( \varepsilon = 2^{-8} \) for algorithms (ii) and (iii) and \( \varepsilon = 2^{-7.5} \) for algorithm (i). Using these tolerances all three algorithms given an empirical relative mean-square error of approximately \( 2^{-7.5} \) and the running times are thus comparable.

\[
\begin{array}{|l|c|c|c|c|}
\hline
\text{Algorithm} & \text{Model (a)} & \text{Model (b)} & \text{Model (c)} & \text{Model (d)} \\
\hline
(i) Monte Carlo & 0.91 & 5.27 & 7.84 & 28.85 \\
(ii) Adaptive MLMC & 0.32 & 1.20 & 1.47 & 5.68 \\
(iii) Ad. MLMC+Malliavin Quadrature & 0.11 & 0.62 & 0.96 & 2.03 \\
\hline
\end{array}
\]

Table 8.2: Measured running times in seconds
9. Appendix.

9.1. Comparison Theorem. The following theorem from [22] allows to compare two diffusions with the same diffusion coefficient but different drifts. In Lemma 2.5 and 3.4 we applied it with $h(x) = x^\gamma$ with $\gamma \in [1/2, 1)$.

Theorem 9.1 (Prop. 5.2.18 in [22]). For $j \in \{1, 2\}$ let $X_j$ be a continuous adapted process such that

$$X_t^j = X_0^j + \int_0^t b_j(s, X_s^j)ds + \int_0^t \sigma(s, X_s^j)dW_s, \quad t \in \mathbb{R}_{\geq 0}$$

Assume that the following conditions hold:

(i) the coefficients $\sigma(t, x)$ and $b_j(t, x)$ are continuous functions on $\mathbb{R}_{\geq 0} \times \mathbb{R}$,

(ii) $\sigma(t, x)$ 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 \( h: \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \) with \( h(0) = 0 \) and
\[
\int_0^\varepsilon h^{-2}(u)du = \infty
\]
for all \( \varepsilon > 0 \).

(iii) \( X^1_t \leq X^0_0 \) almost surely,
(iv) \( b_1(t,x) \leq b_2(t,x) \) for all \( t \geq 0, x \in \mathbb{R} \),
(v) 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 \( P(X^1_t \leq X^0_0 \forall t \geq 0) = 1 \).

9.2. Malliavin Calculus. The Malliavin calculus extends stochastic analysis by adding a derivative operator, which can be interpreted to measure the influence of the underlying Brownian motion at a specific time on the differentiated random variable.

Let \( W = (W^1, \ldots, W^d) \) be a \( d \)-dimensional Brownian motion on a probability space \((\Omega, \mathcal{F}, P)\) and fix an endtime \( T > 0 \). Denote by \( C^\infty_{pol}(\mathbb{R}^n; \mathbb{R})\) the functions \( f: \mathbb{R}^n \to \mathbb{R} \) which are infinitely often differentiable such that \( f \) and all of its partial derivatives of any order are bounded by polynomials. We call a random variable \( X \) smooth, if it can be written as
\[
X = f \left( \int_0^T h_1(s)dW_s, \ldots, \int_0^T h_n(s)dW_s \right)
\]
with \( f \in C^\infty_{pol}(\mathbb{R}^n; \mathbb{R}) \) and \( h_i \in H := L^2([0,T];\mathbb{R}^d) \) for \( i = 1, \ldots, n \). The set of smooth random variables is denoted by \( \mathcal{S} \).

For this class of random variables the Malliavin derivative is defined by the \( H \)-valued random variable
\[
DX = \sum_{i=1}^n \partial_i f \left( \int_0^T h_1(s)dW_s, \ldots, \int_0^T h_n(s)dW_s \right) \cdot h_i
\]
In particular \( DW^i_t = 1_{[0,t]}(\cdot) \cdot e_i \) where \( e_i \) is the \( i \)-th unit vector in \( \mathbb{R}^d \). It can be shown that this definition is independent of the representation of \( X \) in terms of the used function \( f \) and the Wiener integrals \( \int_0^T h(s)dW_s \). Furthermore, the above operator is closable from \( \mathcal{S} \subset L^p(\Omega, \mathcal{F}, P) \) to \( L^p(\Omega, \mathcal{F}, P; H) \) for every \( p \geq 1 \) and the final Malliavin derivative is defined as the closure of the above operator. The class of differentiable random variables \( \mathbb{D}^{1,p} \) is then given by
\[
\mathbb{D}^{1,p} = \left\{ X \in L^p(\Omega) : \exists (X_n)_{n \in \mathbb{N}} \subset \mathcal{S}, Y \in L^p(\Omega; H) \text{ s.t. } X_n \to X \text{ in } L^p(\Omega), DX_n \to Y \text{ in } L^p(\Omega; H) \right\}
\]
and the derivative \( DX \in L^p(\Omega; H) \) is defined as the limit \( Y = \lim_{n \to \infty} DX_n \) for any approximating sequence \( X_n \) (the closability of the operator precisely guarantees that this is well-defined). The derivative is a \( d \)-dimensional random function and we write \( D^i_p X \) for the \( i \)-th component evaluated at \( r \in [0,T] \). Note that while the set of differentiable random variables depends on the chosen \( L^p \)-norm, the actual derivative does not depend on it. In any
case only random variables which are measurable with respect to the underlying Brownian motions can be treated in Malliavin calculus. We denote by $\mathbb{D}^{1,\infty} := \cap_{p \geq 1} \mathbb{D}^{1,p}$ the set of random variables differentiable w.r.t. all $p \geq 1$.

A main tool for the derivation of the quadrature rule is the chain rule. Because we apply it to the antiderivative of a discontinuous function, it is important that the chain rule does not require the function to be differentiable.

Theorem 9.2. Let $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$ be a locally Lipschitz function. Assume $X = (X_1, \ldots, X_m)$ is a random vector with components in $\mathbb{D}^{1,p}$. If

- $\varphi(X) \in L^p(\Omega)$,
- $\partial_i \varphi(X) \cdot DX_i \in L^p(\Omega; H)$ for each $i = 1, \ldots, m$,
- and the set of $x \in \mathbb{R}^m$ where $\varphi$ is not partially continuously differentiable is a $P_X$-null set,

then the chain rule holds: $\varphi(X) \in \mathbb{D}^{1,p}$ and its derivative is given by

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

Here we define $\partial_i \varphi := 0$ on the set of points, where $\partial_i \varphi$ does not exist.

Note that by the Theorem of Rademacher [11, §5.8.3, Theorem 6] this set is a Lebesgue-null set and hence the third assumption is always fulfilled, if the law of $X$ is absolute continuous with respect to the Lebesgue measure.

Proof. The derivative is a closed operator, so in order to prove the chain rule for a function $\varphi$, it is sufficient to find functions $\varphi_n$ such that $\varphi_n(X) \rightarrow \varphi(X)$ in $L^p(\Omega)$, the chain rule holds for all $\varphi_n$ and $\partial_i \varphi_n(X) \cdot DX_i \rightarrow \partial_i \varphi(X) \cdot DX_i$ in $L^p(\Omega; H)$ for all $i = 1, \ldots, m$.

In the first step assume that $\varphi$ is globally Lipschitz and bounded. For each $n \in \mathbb{N}$ choose a function $\psi_n : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$ such that $\psi_n \in C^\infty$, supp $\psi_n \subset [-1/n, 1/n]^m$ and $\int \psi_n dx = 1$ (see for example [28, pp. 82-83]). Define $\varphi_n := \varphi \ast \psi_n$. It is easily shown that $\varphi_n(X) \rightarrow \varphi(X)$ in $L^p(\Omega)$ and that $\partial_i \varphi_n$ is bounded by the Lipschitz-constant of $\varphi$. By Proposition 1.2.3 in [26] the chain rule holds for functions which are continuously differentiable with bounded partial derivatives and thus for all $\varphi_n$. Furthermore

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

For all $\omega \in \Omega$ such that $\partial_i \varphi$ is continuous at $X(\omega)$ this converges to zero. By dominated convergence we have $\partial_i \varphi_n(X) \cdot DX_i \rightarrow \partial_i \varphi(X) \cdot DX_i$ in $L^p(\Omega; H)$ and the chain rule is shown.

In the next step assume that $\varphi$ is only locally Lipschitz, but bounded. Choose an increasing sequence of natural numbers $n_k$ such that $P(\|X\|_{\infty} = n_k) = 0$ for all $k \in \mathbb{N}$. Define $\varphi_k(x) := \varphi(-n_k \vee x \wedge n_k)$. Note that by construction the set of $x \in \mathbb{R}^m$ where $\varphi_k$ is not differentiable, is still a $P_X$-null set. By the first part of the proof the chain rule holds for $\varphi_k$. Because $\varphi$ is bounded we have $\varphi_k(X) \rightarrow \varphi(X)$ in $L^p(\Omega)$ and by dominated convergence $\partial_i \varphi_k(X) \cdot DX_i \rightarrow \partial_i \varphi(X) \cdot DX_i$ in $L^p(\Omega; H)$. This proves the chain rule for $\varphi$.

To extend the chain rule finally to unbounded functions choose an increasing sequence of natural numbers $n_k$ such that $P(|\varphi(X)| = n_k) = 0$ and use the approximation $\varphi_k(x) := -n \vee \varphi(x) \wedge n$ in a similar way as above.
To check the third condition of the previous theorem, the following result is particularly useful (Theorem 2.1.3 in [26]):

Theorem 9.3. Let \( X \in D^{1,p} \) for some \( p \geq 1 \) and assume \( \|DX\|_H > 0 \) almost surely. Then \( X \) is absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R} \).

Finally, we need a theorem to compute derivatives of stochastic processes given by SDEs.

Theorem 9.4. Let \( W_t, t \in [0,T], \) be an \( m \)-dimensional Brownian motion and let \( X_t, t \in [0,T], \) be given by the \( d \)-dimensional SDE

\[
dX_t = b(X_t)dt + \sigma(X_t)dW_t
\]

Assume that both \( b: \mathbb{R}^d \to \mathbb{R}^d \) and \( \sigma: \mathbb{R}^d \to \mathbb{R}^{d \times m} \) are continuously differentiable with bounded derivative. Then \( X_i(t) \in D^{1,\infty} \) for all \( t \in [0,T], i = 1, \ldots, d \). The derivative \( D^r_iX^i(t) \) satisfies almost everywhere the equation

\[
D^r_iX^i(t) = \sigma_{ij}(X(r)) + \sum_{k=1}^d \sum_{l=1}^m \int_r^t \partial_k \sigma_{il}(s)D^r_kX^k(s)ds + \sum_{k=1}^d \int_r^t \partial_k b_i(X(s))D^r_kX^k(s)ds
\]

for \( r \leq t \) and the equation \( D^r_iX^i(t) = 0 \) for \( r > t \).

Proof. This is a simplified version of Theorem 2.2.1 in [26] together with the remark following that theorem. \( \square \)

The second important operator in Malliavin calculus is the divergence operator which in the case of an underlying Brownian motion is called Skorohod integral. It is denoted by \( \delta \) and defined as adjoint operator of the derivative on \( D^{1,2}, \) i.e. the domain is given by

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

and on this set the Skorohod integral is defined (by Riesz’ representation theorem) as the unique \( \delta(u) \in L^2(\Omega) \) such that

\[
\langle DX, u \rangle_{L^2(\Omega; H)} = \langle X, \delta(u) \rangle_{L^2(\Omega)} \quad \forall X \in D^{1,2}
\]

This is also known as the integration by parts rule of Malliavin calculus and can be written as

\[
E \left( \int_0^T \langle D_r(X), u(r) \rangle_{\mathbb{R}^d} dr \right) = E(X \cdot \delta(u)) \quad \forall X \in D^{1,2}, u \in \dom \delta
\]

Further analysis shows that for an adapted process \( u \) the Skorohod integral \( \delta(u) \) coincides with \( \int_0^T u(s)dW_s \) and hence it can be viewed as an extension of the Itô-integral.

To compute Skorohod integrals of non-adapted integrands, the following lemma is useful:

Lemma 9.5 (Prop. 1.3.3 in [26]). Let \( F \in L^2(\Omega) \) and \( u \in L^2(\Omega; H) \) such that the following conditions hold:

(i) \( F \in D^{1,2}, \)

(ii) \( u \in \dom \delta, \)

...
(iii) $F u \in L^2(\Omega; H)$,
(iv) $F \delta(u) - \langle DF, u \rangle_H \in L^2(\Omega)$.

Then $F u \in \text{dom} \delta$ and $\delta(F u) = F \delta(u) - \langle DF, u \rangle_H$.

To derive the quadrature formula for a large class of multidimensional functionals in Theorem 6.5, we will need a derivative operator that differentiates only with respect to one specific Brownian motion. Let $a \in \mathbb{R}^d$ with $\|a\|_2 = 1$. Then $B_t := \langle W_t, a \rangle_{\mathbb{R}^d}$ defines a Brownian motion. For $X \in \mathbb{D}^{1,p}$ we define the derivative with respect to $B$ as the $L^2([0, T])$-valued random variable given by

$$D^B_r X = \langle D_r X, a \rangle_{\mathbb{R}^d}, \quad r \in [0, T]$$

On random variables which are measurable w.r.t. to $B$ this derivative coincides with the usual derivative in the one-dimensional Malliavin calculus w.r.t. $B$. If $X \in L^p(\Omega)$ is independent of $B$, then $D^B = 0$.

The adjoint of $D^B$ is defined for all $u \in L^p(\Omega; L^2([0, T]))$ such that $u \cdot a \in \text{dom} \delta$. For such $u$ it is given by $\delta^B(u) = \delta(u \cdot a)$. The chain rule Theorem 9.2 and Lemma 9.5 are easily extended to $D^B$ and $\delta^B$.

The chain rule can even be extended to functions which are not locally Lipschitz continuous in some directions, as long as the random variables in those directions are independent of $B$.

Proposition 9.6. Let $B$ be a Brownian motion on $(\Omega, \mathcal{F}, P)$ and $X = (X_1, \ldots, X_m)$ a random vector with components in $\mathbb{D}^{1,p}$ and set $A := \{i = 1, \ldots, m : D^B X_i \neq 0\}$. Assume $\varphi: \mathbb{R}^d \to \mathbb{R}$ is continuous except on a $P_X$-zero set and assume that $\varphi$ is locally Lipschitz continuous in all directions $\epsilon_i$, $i \in A$: For all $i \in A$, $x \in \mathbb{R}^d$ there exists a constant $L_i(x)$ and $\epsilon_i(x) > 0$ such that $|\varphi(x + h \epsilon_i) - \varphi(x)| \leq L_i \cdot h$ for all $0 \leq h \leq \epsilon_i(x)$. Then $\varphi(X) \in \mathbb{D}^{1,p}$ and its derivative is given by

$$D^B \varphi(X) = \sum_{i \in A} \partial_i \varphi(X) \cdot D^B X_i$$

Proof. First assume that $\varphi$ is bounded and globally Lipschitz continuous in all directions $\epsilon_i$, $i \in A$. We will reuse the mollifier functions from the proof of Theorem 9.2. Because the chain rule holds for $\varphi \ast \psi_n$ we only have to prove $\varphi \ast \psi_n(X) \to \varphi(X)$ in $L^p(\Omega)$ and $\partial_i \varphi \ast \psi_n(X) \cdot D^B X_i \to \partial_i \varphi(X) \cdot D^B X_i$ in $L^p(\Omega; L^2([0, T]))$ for $i \in A$. As in the proof of Theorem 9.2 we have

$$|\varphi \ast \psi_n(X) - \varphi(X)|^p \to 0,$$

because $\varphi$ is continuous $P_X$-almost everywhere. Moreover, this expression is bounded by $2^p \sup_{x \in \mathbb{R}^d} |\varphi(x)|^p$, so we have also convergence to 0 in $L^p(\Omega)$. Furthermore

$$|\partial_i(\varphi \ast \psi_n(X) - \partial_i \varphi(X)) \cdot D^B X_i|^p \leq \int_{\mathbb{R}^d} |\partial_i \varphi(y) - \partial_i \varphi(X)|^p \cdot |\psi_n(X - y)|^p \cdot |D^B X_i|^p \, dy$$

This converges to 0 for all $\omega$ such that $\partial_i \varphi$ is continuous at $X(\omega)$. By dominated convergence we also have convergence in $L^p(\Omega; H)$ and the assertion is proven.

The extension to unbounded and locally Lipschitz functions is the same as in Theorem 9.2. ■
REFERENCES

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports


