SIAM J. FINANCIAL MATH.
Vol. 6, pp. 22–52

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

Martin Altmayer† and Andreas Neuenkirch†

Abstract. In this article, 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. The integration by parts procedure smoothes the original functional, and thus our formula in combination with a payoff splitting allows us to construct efficient multilevel Monte Carlo estimators. This is confirmed by our numerical analysis and 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

DOI. 10.1137/130933629

1. Introduction.

The Heston model is a popular stochastic volatility model in mathematical finance. It was introduced in [20] and further developed since then; see, e.g., the recent work [18] on the Heston model with stochastic interest rates or [17] on time-dependent coefficients. In its classical form the Heston model uses a Cox–Ingersoll–Ross (CIR) process as volatility and is given by the SDEs

$$\begin{aligned} \mathrm{d}S_t &= bS_t\mathrm{d}t + \sqrt{v_t}S_t(\rho\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2), && t \in [0,T],\\ \mathrm{d}v_t &= \kappa(\lambda - v_t)\mathrm{d}t + \theta\sqrt{v_t}\mathrm{d}W_t^1, && t \in [0,T], \end{aligned}$$

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 (CEV) process as volatility, i.e.,

$$\mathrm{d}v_t = \kappa(\lambda - v_t)\mathrm{d}t + \theta v_t^{\gamma}\mathrm{d}W_t^1$$

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

An efficient method to compute expectations of (smooth) functionals of SDEs is the multilevel Monte Carlo method; see [19] and [14]. Combining approximations using different stepsizes 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 β for the approximations, preferably $\beta \geq 1/2$, which is often not easy to achieve for discontinuous functionals; see [6, 15].

*Received by the editors August 19, 2013; accepted for publication (in revised form) September 15, 2014; published electronically January 29, 2015.
http://www.siam.org/journals/sifin/6/93362.html

†Institut für Mathematik, Universität Mannheim, A5,6, D-68131 Mannheim, Germany (altmayer@uni-mannheim.de, neuenkirch@kiwi.math.uni-mannheim.de).

In this article we are interested in computing

$$Ef(S_T)$$

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

$$Ef(S_T) = E\left(\frac{F(S_T)}{S_T} \cdot \Pi\right), \tag{1.1}$$

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

$$\Pi = 1 + \frac{1}{T\sqrt{1-\rho^2}} \cdot \int_0^T \frac{1}{\sqrt{v_r}}\, \mathrm{d}W_r^2. \tag{1.2}$$

Thus even for discontinuous f the expectation on the 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 $p = Ef(S_T)$. The estimator 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, 10]). The latter method preserves the strict positivity for the volatility, which is crucial for the discretization of the Malliavin weight Π, and the arising approximation of the Malliavin smoothed payoff, i.e., of $(F(S_T)/S_T)\cdot\Pi$, has strong order $\beta = 1/2$. Hence we are able to achieve an L^2-error of order ε with computational cost of order $\log(\varepsilon)^2\varepsilon^{-2}$ for our multilevel estimator.

This numerical analysis is also supported by several numerical experiments. In particular, 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.

Note that recently a different smoothing approach for the functional f has been studied in [16].

The remainder of this manuscript is structured as follows. In the following two sections we derive the required Malliavin regularity of the CIR and CEV processes and also of the Heston price. Formula (1.1) is established in section 4, while section 5 studies the multidimensional case. Finally, in section 6 we present the used multilevel Monte Carlo method, while sections 7 and 8 contain our numerical analysis and numerical tests, respectively. The appendix gives a short introduction to Malliavin calculus and lists some required other results.

2. The CIR and CEV processes.

The volatility process is given by the SDE

$$\mathrm{d}v_t = \kappa(\lambda - v_t)\mathrm{d}t + \theta v_t^{\gamma}\mathrm{d}W_t, \qquad v_0 > 0, \tag{2.1}$$

for $\kappa, \lambda, \theta > 0$ and $\gamma \in [1/2, 1)$. For $\gamma = 1/2$ this is the well-known CIR process, while the general case is known as the mean-reverting CEV process; see, e.g., [23, 29]. A theorem of

Yamada and Watanabe (see [25, Prop. 5.2.13]) implies that this SDE possesses a unique strong solution.

We will always make the following assumption on the parameters.

Assumption 2.1. Assume that either $\gamma > 1/2$ or $2\kappa\lambda > \theta^2$.

It turns out that this assumption is necessary for the mathematical analysis of the CIR process and is often met in stock markets; see, e.g., [11, 32, 1]. This condition in particular guarantees that

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

for all $T > 0$; see, e.g., [23, Lem. 2.2] or [5, Prop. 2.1].

Lemma 2.2.

(i) *For all $p \geq 0$ we have*

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

(ii) *For $\gamma = 1/2$ we have that*

$$\sup_{t\in[0,T]} E(v_t^{-p}) < \infty \qquad \forall \quad 0 \leq p < 2\kappa\lambda/\theta^2$$

$$E \sup_{t\in[0,T]} v_t^{-p} < \infty \qquad \forall \quad 0 \leq p < \frac{2\kappa\lambda}{\theta^2} - 1$$

and

$$E\left(\int_0^T \frac{1}{v_t}\mathrm{d}t\right)^p < \infty \qquad \forall \quad p \geq 0.$$

(iii) *For $\gamma > 1/2$ it holds that*

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

Proof. For $\gamma = 1/2$ see, e.g., section 3.1 in [10] or [21, Thm. 3.1] and Lemma 2.13 in [30]. For $\gamma > 1/2$ see, e.g., Lemma 2.1 in [8]. ∎

For a major part of this article we will consider the Lamperti transformation of the volatility, which is defined by $\sigma_t := v_t^{1-\gamma}$, instead of the volatility itself. The Itō formula shows that σ_t follows the SDE

$$\mathrm{d}\sigma_t = (1-\gamma)\left(\kappa\lambda\sigma_t^{-\frac{\gamma}{1-\gamma}} - \kappa\sigma_t - \frac{\gamma\theta^2}{2}\sigma_t^{-1}\right)\mathrm{d}t + \theta(1-\gamma)\mathrm{d}W_t. \tag{2.2}$$

Now set

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

so that

$$\mathrm{d}\sigma_t = (1-\gamma)\left(f(\sigma_t) - \kappa\sigma_t\right)\mathrm{d}t + \theta(1-\gamma)\mathrm{d}W_t.$$

Note that

$$\lim_{x\searrow 0} f(x) = \infty \quad \text{and} \quad \lim_{x\to\infty} f(x) = 0,$$

so there exists a $C_f \geq 0$ with

$$f(x) \geq -C_f x \qquad \forall x \in \mathbb{R}_{>0}.$$

Finally, since

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

we have

$$\sup_{x\in\mathbb{R}_{>0}} f'(x) < \infty.$$

2.1. Approximation of the transformed process. To prove that σ_t and v_t are Malliavin differentiable, we will follow an article of Alòs and Ewald [4] and approximate σ_t by a process whose SDE has Lipschitz coefficients. This allows us to use standard results from Malliavin calculus. Because we allow $\gamma > 1/2$ and later also 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_ε than Alòs and Ewald.

For each $\varepsilon > 0$ choose a function f_ε with

(i) $f_\varepsilon \in C^\infty(\mathbb{R};\mathbb{R})$ with bounded derivative,
(ii) $f_\varepsilon = f$ on $[\varepsilon,\infty)$,
(iii) f_ε is monotonically decreasing on $(-\infty,0]$,
(iv) $f_\varepsilon(x) \le f(x)$ for $x \in \mathbb{R}_{>0}$,
(v) $f_\varepsilon'(x) \le \sup_{y\in\mathbb{R}_{>0}} f'(y)$ for all $x \in \mathbb{R}$,
(vi) $f_\varepsilon(x) \ge -C_f x$ for all $x \in \mathbb{R}_{>0}$.

Now define σ_t^ε by the modified SDE

$$\mathrm{d}\sigma_t^\varepsilon = (1-\gamma)(f_\varepsilon(\sigma_t^\varepsilon) - \kappa\sigma_t^\varepsilon)\mathrm{d}t + \theta(1-\gamma)\mathrm{d}W_t, \qquad \sigma_0^\varepsilon = \sigma_0. \tag{2.3}$$

Because the coefficients satisfy the usual Lipschitz and linear growth conditions this SDE has a unique strong solution.

Proposition 2.3. *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. Comparing the SDEs in integral form yields for all $\varepsilon > 0$ and almost all $\omega \in \Omega$ that

$$|\sigma_t^\varepsilon(\omega) - \sigma_t(\omega)| \le (1-\gamma)\int_0^t |f_\varepsilon(\sigma_\tau^\varepsilon(\omega)) - f(\sigma_\tau(\omega))|\mathrm{d}\tau + (1-\gamma)\kappa\int_0^t |\sigma_\tau^\varepsilon(\omega) - \sigma_\tau(\omega)|\mathrm{d}\tau.$$

Fix an ω such that the path $\sigma_\cdot(\omega)$ is positive. Then for $0 < \varepsilon < \varepsilon_0(\omega)$ the functions f_ε and f coincide on the path $\sigma_\cdot(\omega)$, i.e., $f_\varepsilon(\sigma_t(\omega)) = f(\sigma_t(\omega))$ for all $t \in [0,T]$. Thus we have

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

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 $t \in [0,T]$. ∎

Note that the last proposition does not imply that σ^ε has almost surely positive paths. In fact, when σ falls below ε, the approximation σ^ε might fall below 0.

Lemma 2.4. *Let u_t be the solution of the SDE*

$$\mathrm{d}u_t = -(1-\gamma)(C_f+\kappa)u_t\mathrm{d}t + \theta(1-\gamma)\mathrm{d}W_t, \qquad u_0 = \sigma_0. \tag{2.4}$$

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

Proof. The proof uses a comparison lemma which allows us to compare solutions of SDEs with the same diffusion coefficient but different drift coefficients; see Proposition 5.2.18 in [25].

(i) Because the drift coefficient of (2.3) is smaller than that of (2.4) the comparison lemma gives $P(u_t \le \sigma_t^\varepsilon \ \ \forall\, t \in [0,T]) = 1$.

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

The SDE of v_t^ε has drift coefficient

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

for $x \ge 0$ and 0 for $x \le 0$. This is smaller than $\kappa(\lambda - x)$ because $f_\varepsilon(x) \le f(x)$. The diffusion coefficient is the same as in the SDE for v_t and satisfies condition (ii) in Proposition 5.2.18 in [25] choosing $h(x) = \theta \cdot x^\gamma$. Thus we can apply the comparison lemma to prove $P(v_t^\varepsilon \le v_t \ \ \forall\, t \in [0,T]) = 1$ and the claim follows. ∎

Proposition 2.5. *For each $p \ge 1$ we have $\sup_{t\in[0,T]} |\sigma_t^\varepsilon - \sigma_t| \to 0$ in L^p as $\varepsilon \to 0$.*

Proof. From the previous lemma we obtain

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

We have $\sup_{t\in[0,T]} \sigma_t \in L^p$ by Lemma 2.2 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 2.3 using dominated convergence. ∎

2.2. Differentiability of the volatility. 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 2.6. *We have $\sigma_t^\varepsilon \in \mathbb{D}^{1,\infty}$ and*

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

Moreover this derivative is bounded by

$$\theta(1-\gamma)\cdot\exp\left(T(1-\gamma)\cdot\sup_{y\in\mathbb{R}_{>0}} f'(y)\right).$$

Proof. The claim $\sigma_t^\varepsilon \in \mathbb{D}^{1,\infty}$ and the form of the derivative follow from Theorem 2.2.1 in [31]. Note that we can only apply this result because the coefficients of the SDE for σ^ε

are differentiable with bounded derivatives on the whole of $\mathbb{R}$. The bound holds because $f'_\varepsilon(x) \le \sup_{y\in\mathbb{R}_{>0}} f'(y)$. ∎

Proposition 2.7. *We have $\sigma_t \in \mathbb{D}^{1,\infty}$ and the derivative is given by*

$$D_r\sigma_t = \theta(1-\gamma)\cdot\exp\left(\int_r^t (1-\gamma)(f'(\sigma_s)-\kappa)\mathrm{d}s\right)\cdot 1_{[0,t]}(r). \tag{2.5}$$

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

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

as $\varepsilon \to 0$.

Proof. Set

$$v_{r,t} := \theta(1-\gamma)\cdot\exp\left(\int_r^t (1-\gamma)(f'(\sigma_s)-\kappa)\mathrm{d}s\right)\cdot 1_{[0,t]}(r).$$

From Proposition 2.3 we know that for almost all $\omega\in\Omega$ we have $\varepsilon_0(\omega) := \inf_{s\in[0,T]}\sigma_s(\omega) > 0$ and $\sigma_s^\varepsilon(\omega) = \sigma_s(\omega)$ for all $0 < \varepsilon < \varepsilon_0(\omega)$ and $s\in[0,T]$. Hence $D_r\sigma_t^\varepsilon(\omega) = v_{r,t}(\omega)$ for all $r,t\in[0,T]$ and $0<\varepsilon<\varepsilon_0(\omega)$. This implies almost sure convergence and the claim follows using the bound from Proposition 2.6 and the fact that the Malliavin derivative is a closed operator. ∎

The chain rule Theorem 9.1 shows the following.

Theorem 2.8. *The volatility process v_t is in $\mathbb{D}^{1,\infty}$ and has the derivative*

$$\begin{aligned} D_r v_t &= \frac{1}{1-\gamma}(\sigma_t)^{\frac{\gamma}{1-\gamma}}\cdot D_r\sigma_t \\ &= \theta v_t^\gamma\cdot\exp\left(\int_r^t\left(-\frac{\kappa\lambda\gamma}{v_s}+\frac{\theta^2}{2}\frac{\gamma(1-\gamma)}{v_s^{2(1-\gamma)}}-\kappa(1-\gamma)\right)\mathrm{d}s\right)\cdot 1_{[0,t]}(r). \end{aligned} \tag{2.6}$$

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

Proposition 2.9. *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. For each $\varepsilon>0$ choose a function $g_\varepsilon\in C^1(\mathbb{R};\mathbb{R}_{>0})$ satisfying $g_\varepsilon(x) = x^{-1/2}$ for all $x>\varepsilon$, $g_\varepsilon(x)\le x^{-1/2}$ for all $x>0$, $|g'_\varepsilon(x)|\le\frac{1}{2}x^{-3/2}$ for all $x>0$. The chain rule Theorem 9.1 can be applied to $g_\varepsilon(v_t)$ and dominated convergence using Lemma 2.2 yields $g_\varepsilon(v_t)\to v_t^{-1/2}$ and $D_r g_\varepsilon(v_t) = g'_\varepsilon(v_t)D_r v_t \to -\frac{1}{2}v_t^{-3/2}D_r v_t$ as $\varepsilon\to 0$. The claim follows because the Malliavin derivative is a closed operator. ∎

3. The price process. This section deals with the price process S_t given by

$$\mathrm{d}S_t = bS_t\mathrm{d}t + \sqrt{v_t}S_t\mathrm{d}B_t. \tag{3.1}$$

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 $X_t := \log S_t$. Itō's formula shows that X_t can be given explicitly in terms of the volatility:

$$\mathrm{d}X_t = \left(b - \frac{1}{2}v_t\right)\mathrm{d}t + \sqrt{v_t}\mathrm{d}B_t. \tag{3.2}$$

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

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

Denote by σ^ε the approximation to the transformed volatility $\sigma_t = v_t^{1-\gamma}$ of section 2. To obtain the derivative we define X_t^ε by

$$X_t^\varepsilon = X_0 + \int_0^t \left(b - \frac{1}{2}\psi_\varepsilon^{\frac{1}{1-\gamma}}(\sigma_s^\varepsilon)\right)\mathrm{d}s + \int_0^t \psi_\varepsilon^{\frac{1}{2(1-\gamma)}}(\sigma_s^\varepsilon)\mathrm{d}B_s.$$

Lemma 3.1. *For each $p, q \ge 1$ we have*

$$E \sup_{t\in[0,T]} |\sigma_t^q - \psi_\varepsilon^q(\sigma_t^\varepsilon)|^p \to 0$$

as well as

$$E \sup_{t\in[0,T]} \left|\sigma_t^q - \psi_\varepsilon^q(\sigma_t^\varepsilon)\cdot\psi_\varepsilon'(\sigma_t^\varepsilon)\right|^p \to 0.$$

Proof. By Proposition 2.3 for almost all ω there exists an $\varepsilon_0(\omega) > 0$ such that $\varepsilon < \sigma_t(\omega) = \sigma_t^\varepsilon(\omega) < 1/\varepsilon$ for all $t \in [0,T]$ and $0 < \varepsilon < \varepsilon_0(\omega)$. For such ω and ε we have

$$\sup_{t\in[0,T]} |\sigma_t^q(\omega) - \psi_\varepsilon^q(\sigma_t^\varepsilon(\omega))| = \sup_{t\in[0,T]} \left|\sigma_t^q(\omega) - \psi_\varepsilon^q(\sigma_t^\varepsilon(\omega))\cdot\psi_\varepsilon'(\sigma_t^\varepsilon(\omega))\right| = 0.$$

For all ω and ε 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. ∎

Proposition 3.2. *We have $\sup_{t\in[0,T]}|X_t^\varepsilon - X_t| \to 0$ in L^p for all $p \ge 1$.*

Proof. From the SDEs we obtain the following inequality for some constant $c > 0$:

$$\begin{aligned} E \sup_{t\in[0,T]} |X_t^\varepsilon - X_t|^p \le{}& c \cdot E \sup_{t\in[0,T]} \left|\int_0^t \left(\sigma_s^{\frac{1}{1-\gamma}} - \psi_\varepsilon^{\frac{1}{1-\gamma}}(\sigma_s^\varepsilon)\right)\mathrm{d}s\right|^p \\ &+ c \cdot E \sup_{t\in[0,T]} \left|\int_0^t \left(\sigma_s^{\frac{1}{2(1-\gamma)}} - \psi_\varepsilon^{\frac{1}{2(1-\gamma)}}(\sigma_s^\varepsilon)\right)\mathrm{d}B_s\right|^p. \end{aligned}$$

The first term converges to zero by the previous lemma. For the second term we use the Burkholder–Davis–Gundy inequality as well as $|x-y|^2 \le |x^2-y^2|$ for $x, y \ge 0$: There is a constant c such that

$$E \sup_{t\in[0,T]} \left| \int_0^t \left(\sigma_s^{\frac{1}{2(1-\gamma)}} - \psi_\varepsilon^{\frac{1}{2(1-\gamma)}}(\sigma_s^\varepsilon) \right) \mathrm{d}B_s \right|^p \le c \cdot E \left(\int_0^T \left| \sigma_s^{\frac{1}{1-\gamma}} - \psi_\varepsilon^{\frac{1}{1-\gamma}}(\sigma_s^\varepsilon) \right| \mathrm{d}s \right)^{p/2}$$

which converges to zero again by the previous lemma. ∎

Theorem 3.3. *The log-price X_t is in $\mathbb{D}^{1,p}$ for all $p \ge 1$ and its derivative is given by*

$$D_r^1 X_t = \rho\sqrt{v_r} - \frac{1}{2(1-\gamma)} \cdot \int_0^t \sigma_s^{\frac{\gamma}{1-\gamma}} D_r\sigma_s \mathrm{d}s + \frac{1}{2(1-\gamma)} \cdot \int_0^t \sigma_s^{\frac{1}{2(1-\gamma)}-1} D_r\sigma_s \mathrm{d}B_s,$$
$$D_r^2 X_t = \sqrt{1-\rho^2}\sqrt{v_r}$$

for $r \le t$ and $D_r^1 X_t = D_r^2 X_t = 0$ else. Moreover for each $r \in [0,T]$ and $i \in \{1,2\}$ the uniform convergence

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

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

Proof. Because the system of SDEs for X_t^ε and σ_t^ε has globally Lipschitz coefficients, we can use Theorem 2.2.1 in [31] to calculate the derivatives: For $r \le t$ we have

$$\begin{aligned} D_r^1 X_t^\varepsilon &= \rho\psi_\varepsilon^{\frac{1}{2(1-\gamma)}}(\sigma_r^\varepsilon) - \frac{1}{2(1-\gamma)} \cdot \int_r^t \psi_\varepsilon^{\frac{\gamma}{1-\gamma}}(\sigma_s^\varepsilon)\psi_\varepsilon'(\sigma_s^\varepsilon) D_r^1\sigma_s^\varepsilon \mathrm{d}s \\ &\quad + \frac{1}{2(1-\gamma)} \cdot \int_r^t \psi_\varepsilon^{\frac{1}{2(1-\gamma)}-1}(\sigma_s^\varepsilon)\psi_\varepsilon'(\sigma_s^\varepsilon) D_r^1\sigma_s^\varepsilon \mathrm{d}B_s, \\ D_r^2 X_t^\varepsilon &= \sqrt{1-\rho^2}\psi_\varepsilon^{\frac{1}{2(1-\gamma)}}(\sigma_r^\varepsilon). \end{aligned}$$

The first term of the first deriviative and the second one converge to $\rho\sqrt{v_r}$ and $\sqrt{1-\rho^2}\sqrt{v_r}$, respectively, using Lemma 3.1. Now consider the second term of the first derivative:

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

This converges to 0 in $L^p(\Omega)$ using Lemma 3.1, Lemma 2.2, and Proposition 2.7. 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 [13, 28]).

Theorem 3.4. *Let $u \in (1,\infty)$ and define*

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

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

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

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

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

The behavior for the Heston model with the CEV process for the borderline case $\rho = -\sqrt{(u-1)/u}$ is unknown, to the best of our knowledge. Based on the following result one can check the Malliavin smoothness of S_t.

Theorem 3.5. *Assume $S_t \in L^{p+\varepsilon}(\Omega)$ for some $p \geq 1$ and $\varepsilon > 0$. Then the Heston price S_t is in $\mathbb{D}^{1,p}$ and its derivative is $DS_t = S_t \cdot DX_t$.*

Proof. Because $X_t \in \mathbb{D}^{1,\infty}$ we have $S_t \cdot DX_t \in L^p(\Omega; H)$ by Hölder's inequality. Now the claim follows directly from the chain rule Theorem 9.1. ∎

4. Quadrature. Now we can turn to the derivation of our quadrature formula. First we prove a similar formula for the log-price.

Proposition 4.1. *If $g\colon \mathbb{R} \to \mathbb{R}$ is bounded and G is an antiderivative of g, then*

$$E(g(X_T)) = \frac{1}{T\sqrt{1-\rho^2}} \cdot E\left(G(X_T) \cdot \int_0^T \frac{1}{\sqrt{v_t}}\, \mathrm{d}W_t^2\right).$$

Proof. Theorem 2.1.3 in [31] shows that X_T has a density which is continuous w.r.t. the Lebesgue measure. Thus we can apply the chain rule given in Theorem 9.1 and get $G(X_T) \in \mathbb{D}^{1,2}$ with $DG(X_T) = g(X_T) \cdot DX_T$:

$$\begin{aligned} E(g(X_T)) &= \frac{1}{T} \cdot E\left(\int_0^T g(X_T) \cdot D_t^2 X_T \cdot \frac{1}{D_t^2 X_T}\, \mathrm{d}t\right) \\ &= \frac{1}{T\sqrt{1-\rho^2}} \cdot E\left(\int_0^T D_t^2 G(X_T) \cdot \frac{1}{\sqrt{v_t}}\, \mathrm{d}t\right). \end{aligned}$$

By Lemma 2.2 the process $t \mapsto 1/\sqrt{v_t}$ is square-integrable. Thus we can apply the Malliavin integration by parts formula with a simple Itō integral instead of a Skorohod integral:

$$E(g(X_T)) = \frac{1}{T\sqrt{1-\rho^2}} \cdot E\left(G(X_T) \cdot \int_0^T \frac{1}{\sqrt{v_t}}\, \mathrm{d}W_t^2\right). \qquad \blacksquare$$

Theorem 4.2. *Let $f\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ be measurable and let $F(x) := \int_0^x f(y)\mathrm{d}y$ be an antiderivative of it. Assume that f is bounded or there is an $\epsilon > 0$ such that*

$$f(S_T) \in L^{1+\epsilon}(\Omega) \qquad \text{and} \qquad \frac{F(S_T)}{S_T} \in L^{1+\epsilon}(\Omega)$$

(a sufficient condition for this is $|f| \leq |p|$ and $S_T \in L^{\deg p+\epsilon}(\Omega)$ for some polynomial p). Then

$$E(f(S_T)) = E\left(\frac{F(S_T)}{S_T} \cdot \left(1 + \frac{1}{T\sqrt{1-\rho^2}} \cdot \int_0^T \frac{1}{\sqrt{v_t}}\mathrm{d}W_t^2\right)\right).$$

Proof. First assume that f is bounded. Define $g(x) = f(e^x)$ and $G(x) = \int_0^x g(y)\mathrm{d}y + F(1)$. By the previous proposition we have

$$E(f(S_T)) = E(g(X_T)) = \frac{1}{T\sqrt{1-\rho^2}} \cdot E\left(G(X_T) \cdot \int_0^T \frac{1}{\sqrt{v_t}}\mathrm{d}W_t^2\right).$$

Moreover we can write G using ordinary partial integration as

$$G(x) = \int_0^x f(e^y)e^y \cdot \frac{1}{e^y}\mathrm{d}y + F(1) = \frac{F(e^x)}{e^x} + \int_0^x \frac{F(e^y)}{e^y}\mathrm{d}y.$$

Now the quadrature formula follows from applying the previous proposition backward to get

$$E\left(\int_0^{X_T} \frac{F(e^y)}{e^y}\mathrm{d}y \cdot \frac{1}{T\sqrt{1-\rho^2}} \int_0^T \frac{1}{\sqrt{v_t}}\mathrm{d}W_t^2\right) = E\left(\frac{F(e^{X_T})}{e^{X_T}}\right).$$

Next consider the general case. Choose bounded functions f_n such that $|f_n| \leq |f|$ and $f_n \to f$ almost everywhere. Define $F_n(x) := \int_0^x f_n(z)\mathrm{d}z$. Then $F_n \to F$ almost everywhere due to dominated convergence. By our assumptions we have $|F_n(S_T)/S_T| \leq |F(S_T)/S_T| \in L^{1+\epsilon}(\Omega)$. Since the weight Π is in $L^q(\Omega)$ for all $q \geq 0$ we obtain

$$\frac{F(S_T)}{S_T} \cdot \Pi \in L^1(\Omega).$$

The theorem follows using dominated convergence. ∎

5. Multidimensional Heston models. In this section we will extend our quadrature formula of Theorem 4.2 to functionals in the multidimensional Heston model, i.e., to (S^i, v^i), $i = 1, \ldots, d$, given by the SDEs

$$\begin{pmatrix}\mathrm{d}S_t^i \\ \mathrm{d}v_t^i\end{pmatrix} = \begin{pmatrix}b_i S_t^i \\ \kappa_i(\lambda_i - v_t^i)\end{pmatrix}\mathrm{d}t + \begin{pmatrix}\sqrt{v_t^i}S_t^i & 0 \\ 0 & \theta_i(v_t^i)^{\gamma_i}\end{pmatrix}\begin{pmatrix}\mathrm{d}B_t^i \\ \mathrm{d}W_t^i\end{pmatrix} \tag{5.1}$$

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.,

$$\Sigma^{(B,W)} = \begin{pmatrix}(EB_1^iB_1^j)_{i,j=1,\ldots,d} & (EB_1^iW_1^j)_{i,j=1,\ldots,d} \\ (EW_1^iB_1^j)_{i,j=1,\ldots,d} & (EW_1^iW_1^j)_{i,j=1,\ldots,d}\end{pmatrix}.$$

We will work under the following assumption.

Assumption 5.1.

(a) For each $i = 1, \dots, d$ we have either $\frac{2\kappa_i\lambda_i}{\theta_i^2} > 1$ or $\gamma_i > 1/2$.

(b) $\Sigma^{(B,W)}$ 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

$$g \colon \mathbb{R}^{2d,2d} \to \mathbb{R}^{2d,2d}, \qquad g(A)_{ij} := A_{2d+1-i,\,2d+1-j}, \qquad i, j = 1, \dots, 2d.$$

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

Lemma 5.2. *Let Z be defined as above and $S_t^i \in L^{2+\varepsilon}$, $i = 1, \dots, d$. The Brownian motion Z^1 is independent of $(S_t^j)_{t\in[0,T]}$ if and only if $j \neq 1$. Using the partial Malliavin derivative D^{Z^1} defined in the appendix we have in particular*

$$D_r^{Z^1} X_t^j = \begin{cases} R_{11} \cdot \sqrt{v_r^1} \cdot 1_{[0,t]}(r), & j = 1, \\ 0, & j \neq 1, \end{cases} \tag{5.2}$$

and

$$D_r^{Z^1} S_t^j = \begin{cases} S_t^1 \cdot R_{11} \cdot \sqrt{v_r^1} \cdot 1_{[0,t]}(r), & j = 1, \\ 0, & j \neq 1. \end{cases} \tag{5.3}$$

Proof. By construction we have that Z^1 and $B^j, W^i, j = 2, \dots, d, i = 1, \dots, d$, are independent, which implies the first claim.

These independence relations allow us to easily calculate Malliavin derivatives w.r.t. Z^1 (see the appendix for this type of derivative):

$$D_r^{Z^1} B_t^1 = \sum_{k=1}^{2d} R_{1,k} \cdot D_r^{Z^1} Z_t^k = R_{11} \cdot 1_{[0,t]}(r).$$

Similarly we obtain

$$D^{Z^1}(B_t^j) = 0 \quad \text{for } j \neq 1,$$
$$D^{Z^1}(W_t^i) = 0 \quad \text{for } i = 1, \dots, d.$$

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

As mentioned before Theorem 2.1.3 in [31] proves that for fixed $i = 1, \dots, d$ the Heston log-price S_t^i is absolutely continuous w.r.t. the Lebesgue measure. For our multidimensional quadrature formula we also need the existence of a joint density.

Theorem 5.3. *For all $t \in [0,T]$ the law of the random vector X_t is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^d$.*

Proof. The assertion follows from [31, Lem. 2.1.1] if we can show 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 $i = 1,\dots,d$ for some constant $c > 0$.

Case 1: $i = 1$. Applying the standard chain rule (φ is differentiable) and writing D^1 for D^{Z^1}, we have

$$\begin{aligned}|E(\partial_1 \varphi(X_t))| &= \frac{1}{t} \cdot \left| E\left(\int_0^t \partial_1 \varphi(X_t) \cdot D_r^1 X_t^1 \cdot \frac{1}{D_r^1 X_t^1} \mathrm{d}r \right) \right| \\ &= \frac{1}{t} \cdot \left| E\left(\int_0^t D_r^1 \varphi(X_t) \cdot \frac{1}{D_r^1 X_t^1} \, \mathrm{d}r \right) \right|.\end{aligned}$$

By (5.2) we have $(D_\cdot^1 X_t^1)^{-1} = (R_{11}\sqrt{v_\cdot^1})^{-1}$. Thus the integration by parts rule gives

$$\begin{aligned}&= \frac{1}{R_{11} t} \cdot \left| E\left(\varphi(X_t) \cdot \int_0^t \frac{1}{\sqrt{v_r^1}} \, dZ_r^1 \right) \right| \\ &\le \frac{1}{R_{11} t} \cdot \|\varphi\|_\infty \cdot E\left| \int_0^t \frac{1}{\sqrt{v_r^1}} \, dZ_r^1 \right|.\end{aligned}$$

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^i 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.

Theorem 5.4. *Let $f\colon (\mathbb{R}_{\ge 0})^d \to \mathbb{R}$ be almost everywhere continuous and set*

$$F(x) = F(x_1,\dots,x_d) = \int_0^{x_1} f(\xi, x_2, \dots, x_d) \mathrm{d}\xi.$$

Assume that either f is bounded or that there exists an $\varepsilon > 0$ such that

$$f(S_T) \in L^{1+\epsilon}(\Omega) \qquad \text{and} \qquad \frac{F(S_T)}{S_T^1} \in L^{1+\epsilon}(\Omega).$$

Then

$$E(f(S_T)) = E\left(\frac{F(S_T)}{S_T^1} \cdot \left(1 + \frac{1}{R_{11} T} \cdot \int_0^T \frac{1}{\sqrt{v_r^1}} \, \mathrm{d}Z_r^1 \right) \right).$$

Proof. To simplify notation we will write D for the partial derivative D^{Z^1}. The previous theorem allows us to apply the chain rule Proposition 9.2 to $G(X_T)$ for any almost everywhere differentiable function G. Thus for each bounded function $g\colon \mathbb{R}^d \to \mathbb{R}$ and $G(x) = \int_0^T g(\xi, x_2, \dots, x_d)\mathrm{d}\xi$ we have

$$\begin{aligned}E(g(X_T)) &= \frac{1}{T} \cdot E\left(\int_0^T g(X_T) \cdot D_r X_T^1 \cdot \frac{1}{D_r X_T^1} \mathrm{d}r \right) \\ &= \frac{1}{T} \cdot E\left(\int_0^T D_r(G(X_T)) \cdot \frac{1}{D_r X_T^1} \mathrm{d}r \right).\end{aligned}$$

Applying the integration by parts rule gives

$$= \frac{1}{R_{11}T} \cdot E\left(G(S_T) \cdot \int_0^T \frac{1}{\sqrt{v_r^1}} dZ_r^1 \right).$$

With this analogue of Proposition 4.1 the theorem can be proved exactly like in the one-dimensional case; see Theorem 4.2. ∎

5.1. Lipschitz continuity after Malliavin integration by parts. Using again the log-price $X_T^i = \log(S_T^i)$, we have the following smoothness result for the functional $G(S_T) = F(S_T)/S_T^1$ on the right-hand side of the quadrature formula.

Proposition 5.5. *If $f\colon (\mathbb{R}_{\geq 0})^d \to \mathbb{R}$ is bounded by $C \geq 0$, then*

$$G\colon \mathbb{R}^d \to \mathbb{R}, \qquad G(x) := \frac{1}{e^{x_1}} \cdot \int_0^{e^{x_1}} f(\xi, e^{x_2}, \dots, e^{x_d}) \mathrm{d}\xi$$

is bounded by C and globally Lipschitz continuous in $x_1 \in \mathbb{R}$, uniformly in $x_2, \dots, x_d \in \mathbb{R}$.

Proof. Clearly $G(x) \leq e^{-x_1} C e^{x_1} = C$. For $x_1, y_1 \in \mathbb{R}$ with $x_1 < y_1$ we have

$$\begin{aligned}
&|G(e^{y_1}, e^{x_2}, \dots, e^{x_d}) - G(e^{x_1}, e^{x_2}, \dots, e^{x_d})| \\
&\quad = \left| \frac{1}{e^{y_1}} \cdot \int_{e^{x_1}}^{e^{y_1}} f(\xi, e^{x_2}, \dots, e^{x_d}) \mathrm{d}\xi + \left(\frac{1}{e^{y_1}} - \frac{1}{e^{x_1}} \right) \cdot \int_0^{e^{x_1}} f(\xi, e^{x_2}, \dots, e^{x_d}) \mathrm{d}\xi \right| \\
&\quad \leq \left| \frac{C}{e^{y_1}} (e^{y_1} - e^{x_1}) \right| + \left| \frac{C}{e^{x_1} e^{y_1}} (e^{y_1} - e^{x_1}) e^{x_1} \right| \\
&\quad = 2C(1 - e^{x_1 - y_1}) \\
&\quad \leq 2C(1 - (1 + x_1 - y_1)) \\
&\quad = 2C(y_1 - x_1)
\end{aligned}$$

for all $x_2, \dots, x_d \in \mathbb{R}$. ∎

In particular, for $d = 1$ and bounded payoffs, G is globally Lipschitz continuous. For so-called basket options

$$f\left(\sum_{i=1}^d a_i e^{X_T^i} \right)$$

with $f\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ and $a_i \geq 0$, $i = 1, \dots, d$, the function $G\colon \mathbb{R}^d \to \mathbb{R}$ also can be given as

$$G(x) = \frac{1}{e^{x_1}} \int_0^{\sum_{i=1}^d a_i e^{x_i}} f(\xi) \mathrm{d}\xi, \qquad x \in \mathbb{R}^d. \tag{5.4}$$

By straightforward calculations we have the following.

Proposition 5.6. *Let $G : \mathbb{R}^d \to \mathbb{R}$ be given by (5.4). If f has compact support contained in $(0, \infty)$, then G is bounded and globally Lipschitz continuous.*

Since we will later use the technique of payoff splitting, which leads to discontinuous payoffs with small support, the above proposition will be useful.

6. Multilevel Monte Carlo. Multilevel Monte Carlo is a powerful method for variance reduction introduced by Heinrich [19] for parametric integration and by Giles [14] for the quadrature of SDEs. The latter generalizes the two-level method developed in [26].

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 (5.1). While in the one-dimensional case PDE methods (see, e.g., [22]) or FFT methods [9] 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.,

$$p \approx \frac{1}{N} \sum_{i=1}^{N} f\big(S_T^{(h),(i)}\big).$$

The multilevel Monte Carlo method instead relies on approximations $S_T^{(h_l)}$ to S_T with different stepsizes $h_l = M^{-l}T$, $l = 0, \dots, L$, with $M \in \mathbb{N}, M \geq 2$. Its idea is to use the telescoping sum

$$Ef\big(S_T^{(h_L)}\big) = Ef\big(S_T^{(h_0)}\big) + \sum_{l=1}^{L} E\left(f\big(S_T^{(h_l)}\big) - f\big(S_T^{(h_{l-1})}\big)\right)$$

to estimate the expectations on the right-hand side using independent standard Monte Carlo estimators

$$Ef\big(S_T^{(h_0)}\big) \approx \widehat{P}_0 = \frac{1}{N_0} \sum_{i=1}^{N_0} f\big(S_T^{(h_0),(i)}\big),$$

$$Ef\big(S_T^{(h_l)}\big) - Ef\big(S_T^{(h_{l-1})}\big) \approx \widehat{P}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} \left(f\big(S_T^{(h_l),(i)}\big) - f\big(S_T^{(h_{l-1}),(i)}\big)\right), \qquad l = 1, \dots, L,$$

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

$$\widehat{P} = \sum_{l=0}^{L} \widehat{P}_l,$$

where the estimators $\widehat{P}_l$, $l = 0, \dots, L$, of the different levels have to be independent.

Based on estimates for the weak error

$$\big|Ef(S_T^{(h)}) - p\big| \leq c_\alpha \cdot h^\alpha$$

and the strong error

$$E\big|f(S_T^{(h)}) - f(S_T)\big|^2 \leq c_\beta \cdot h^{2\beta}$$

with $c_\alpha, c_\beta, \alpha \geq 1/2, \beta > 0$, the number of levels L and the number of repetitions N_l, $l = 1, \dots, L$, can be chosen such that the mean-square error satisfies

$$E|p - \widehat{P}|^2 \leq \varepsilon^2$$

with

$$\text{computational cost} \leq c_{\alpha,\beta} \begin{cases} \varepsilon^{-2} & \text{if} \quad \beta > 1/2, \\ (\log(\varepsilon))^2 \varepsilon^{-2} & \text{if} \quad \beta = 1/2, \\ \varepsilon^{-2-\frac{1-2\beta}{\alpha}} & \text{if} \quad \beta < 1/2 \end{cases} \tag{6.1}$$

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., $\beta \geq 1/2$ preferably, and also on the knowledge of $c_\alpha, c_\beta, \alpha$, and β. However, for discontinuous functionals f the strong convergence rates of approximation schemes deteriorate: In [6, 15] it is shown that a discontinuity leads (at least) 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, to the best of our knowledge.

Our integration by parts result, i.e., Theorem 5.4, circumvents one of these problems. Here we have shown that

$$E(f(S_T)) = E\left(\frac{F(S_T)}{S_T^1} \cdot \left(1 + \frac{1}{R_{11}T} \cdot \int_0^T \frac{1}{\sqrt{v_s^1}} \mathrm{d}Z_r^1\right)\right) \tag{6.2}$$

and for several types of functionals f (e.g., $d = 1$, f bounded or basket option, f compact support in $(0, \infty)$), the arising functional

$$G\colon \mathbb{R}^d \to \mathbb{R}, \quad G(x) = \frac{F(e^{x_1}, \dots, e^{x_d})}{e^{x_1}}$$

is globally Lipschitz continuous; see subsection 5.1. Clearly, in this case the strong convergence rate of the approximation scheme for X_T is retained when applied to $G(S_T)$.

Concerning the problem of knowing $c_\alpha, c_\beta, \alpha$, and β, Giles presented in [14] a heuristic algorithm, which chooses the N_l's adaptively and can be used without knowledge of $c_\alpha, c_\beta, \alpha$, and β; see Algorithm 1.

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

$$E\left(\frac{F(S_T)}{S_T^1} \cdot \left(1 + \frac{1}{R_{11}T} \cdot \int_0^T \frac{1}{\sqrt{v_s^1}} \, \mathrm{d}Z_r^1\right)\right)$$

ALGORITHM 1. ADAPTIVE MULTILEVEL MONTE CARLO.

Input: error tolerance ε, initial samples N_{ini}
Output: estimation for $E(f(S_T))$
$L := 0$
Generate $N_L = N_{ini}$ samples for level L
Estimate the variance V_L from these samples
Compute N_l for $l = 0, \dots, L$ using the formula $N_l := \left\lceil 2 \cdot \varepsilon^{-2} \sqrt{V_l h_l} \left(\sum_{l=0}^{L} \sqrt{V_l / h_l} \right) \right\rceil$
If an N_l surpasses the number of samples generated on level l so far, generate additional samples
For $l = 0, \dots, L$ set $\widehat{P}_l$ to the mean of the samples on level l
if $L \geq 2$ and the convergence condition $\max\left\{ M^{-1} |\widehat{P}_{L-1}|, |\widehat{P}_L| \right\} < \frac{1}{\sqrt{2}} (M-1)\varepsilon$ holds
then return $\sum_{l=0}^{L} \widehat{P}$

else $L := L + 1$ and **goto** step 2

we need to take care of the fact that the variance of the weight

$$\Pi = 1 + \frac{1}{R_{11} T} \cdot \int_0^T \frac{1}{\sqrt{v_s^1}} \, \mathrm{d}Z_r^1$$

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 ε. Besides reordering the processes such that R_{11} and $\kappa_1 \lambda_1 / \theta_1^2$ are as large as possible, this problem can be addressed in the following way: We split the payoff in a Lipschitz continuous part f_1 and a discontinuous part f_2 with small support and apply the quadrature formula only to the latter part. To be precise, given $f = f_1 + f_2$ and F_2 with $\frac{\partial}{\partial x_1} F_2 = f_2$, where $f, f_1, f_2, F_2 \colon (\mathbb{R}_{\geq 0})^d \to \mathbb{R}$, we apply the multilevel estimator to the functional

$$f_1(S_T) + \frac{F_2(S_T)}{S_T^1} \cdot \Pi. \tag{6.3}$$

7. Discretization and numerical analysis. It remains to specify the numerical methods for the approximation of the Heston prices and the weight Π. 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., [27]) cannot be applied. However, recently [2, 10, 3, 30] it has been shown that the discretization of the Lamperti-transformed volatility process with a drift-implicit (also called 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 $\sigma_t^i = (v_t^i)^{1-\gamma_i}$ and price processes $X_t^i = \log(S_t^i)$, i.e.,

$$\begin{pmatrix} \mathrm{d}X_t^i \\ \mathrm{d}\sigma_t^i \end{pmatrix} = \begin{pmatrix} b_i - \frac{1}{2}(\sigma_t^i)^{\frac{1}{1-\gamma_i}} \\ (1-\gamma_i)\left(f_i(\sigma_t^i) - \kappa_i \sigma_t^i\right) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} (\sigma_t^i)^{\frac{1}{2(1-\gamma_i)}} & 0 \\ 0 & \theta_i(1-\gamma_i) \end{pmatrix} \begin{pmatrix} \mathrm{d}B_t^i \\ \mathrm{d}W_t^i \end{pmatrix}$$

with

$$f_i(x) := \kappa_i \lambda_i x^{-\frac{\gamma_i}{1-\gamma_i}} - \frac{\gamma_i \theta_i^2}{2} x^{-1}.$$

Recall that the Brownian motions B^i, W^i are given as $\binom{B}{W} = R \cdot Z$, where Z is a $2d$-dimensional Brownian motion and $R \in \mathbb{R}^{2n \times 2n}$ is an invertible upper triangular matrix. Using this notation we finally set

$$P = f_1(e^{X_T^1}, \dots, e^{X_T^d}) + \frac{F_2(e^{X_T^1}, \dots, e^{X_T^d})}{e^{X_T^1}} \cdot \Pi \tag{7.1}$$

with

$$\Pi = 1 + \frac{1}{R_{11} T} \cdot \int_0^T \frac{1}{\sqrt{v_s^1}} \mathrm{d}Z_s^1. \tag{7.2}$$

As mentioned, we discretize the transformed volatility processes with the backward Euler scheme with stepsize $\Delta > 0$, i.e.,

$$\hat{\sigma}_{k+1}^i = \hat{\sigma}_k^i + (1-\gamma_i)\left(f_i(\hat{\sigma}_{k+1}^i) - \kappa_i \hat{\sigma}_{k+1}^i\right)\Delta + (1-\gamma_i)\theta_i \Delta_k W^i, \qquad k = 0, 1 \dots,$$

where $\hat{\sigma}_0^i = \sigma_0^i$ and $\Delta_k W^i = W_{(k+1)\Delta}^i - W_{k\Delta}^i$. Under Assumption 5.1(a) this scheme preserves positivity, i.e.,

$$P\left(\hat{\sigma}_k^i > 0,\ k = 0, 1, \dots\right) = 1. \tag{7.3}$$

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

$$\hat{\sigma}_{k+1}^i = \frac{\hat{\sigma}_k^i + \frac{\theta_i}{2}\Delta_k W^i}{2 + \kappa_i \Delta} + \sqrt{\frac{\left(\hat{\sigma}_k^i + \frac{\theta_i}{2}\Delta_k W^i\right)^2}{(2+\kappa_i\Delta)^2} + \frac{(\kappa_i \lambda_i - \frac{\theta_i^2}{4})\Delta}{2 + \kappa_i \Delta}}, \tag{7.4}$$

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 Π using the Euler scheme, which yields

$$\hat{X}_N^i = \log(S_0^i) + b_i T - \frac{1}{2}\sum_{k=0}^{N-1} (\hat{\sigma}_k^i)^{\frac{1}{1-\gamma_i}} \Delta + \sum_{k=0}^{N-1} (\hat{\sigma}_k^i)^{\frac{1}{2(1-\gamma_i)}} \Delta_k B^i$$

and

$$\hat{\Pi} = 1 + \frac{1}{R_{11} T} \sum_{k=0}^{N-1} (\hat{\sigma}_{k\Delta}^i)^{-\frac{1}{2(1-\gamma_i)}} \Delta_k Z^1,$$

where $N \in \mathbb{N}$ with $N\Delta = T$. Note that our approximation of Π is well defined due to (7.3). Finally, set

$$\hat{P} = f_1(e^{\hat{X}_N^1}, \dots, e^{\hat{X}_N^d}) + \frac{F_2(e^{\hat{X}_N^1}, \dots, e^{\hat{X}_N^d})}{e^{\hat{X}_N^1}} \cdot \hat{\Pi}. \tag{7.5}$$

We start with a result on the strong approximation of the price process itself.

Proposition 7.1. *Let $p \geq 2$. We have*

$$\left(E\left|X_T^i - \hat{X}_N^i\right|^p\right)^{1/p} \leq C_{p,i} \cdot \sqrt{\Delta} \qquad \text{for} \qquad \begin{cases} p < \frac{8\kappa_i\lambda_i}{3\theta_i^2} & \text{if} \quad \gamma_i = \frac{1}{2},\ \frac{\kappa_i\lambda_i}{\theta_i^2} > 1, \\ p < \infty & \text{else,} \end{cases}$$

with $C_{p,i} > 0$, $i = 1, \dots, d$.

Proof. For notational simplicity we drop the index i in the proof and we will denote constants, whose value is not important, by c, regardless of their value. Moreover, let $\eta(s) := \lfloor \frac{s}{\Delta} \rfloor \cdot \Delta$. First note that

$$|x^\alpha - y^\alpha| \leq \alpha \left(x^{\alpha-1} + y^{\alpha-1}\right) |x - y| \tag{7.6}$$

for all $\alpha \geq 1$ and $x, y \geq 0$. Together with the smoothness in the pth mean of the CIR/CEV process and Hölder's inequality this gives

$$E\left|\sigma_t^{\frac{1}{(1-\gamma)}} - \sigma_s^{\frac{1}{(1-\gamma)}}\right|^p + E\left|\sigma_t^{\frac{1}{2(1-\gamma)}} - \sigma_s^{\frac{1}{2(1-\gamma)}}\right|^p \leq c \cdot \left(E(|\sigma_s - \sigma_t|^{2p})\right)^{\frac{1}{2}} \leq c \cdot \Delta^{p/2}$$

for $|t-s| \leq \Delta$. Analogously (7.6) and (9.2) from Theorem 9.3 together with Hölder's inequality imply

$$E\left|\sigma_{\eta(s)}^{\frac{1}{1-\gamma}} - \hat{\sigma}_{\eta(s)}^{\frac{1}{1-\gamma}}\right|^p + E\left|\sigma_{\eta(s)}^{\frac{1}{2(1-\gamma)}} - \hat{\sigma}_{\eta(s)}^{\frac{1}{2(1-\gamma)}}\right|^p \leq c \cdot \Delta^{p/2}.$$

Thus, using Jensen's inequality for the Riemann integral and the Burkholder–Davis–Gundy inequality for the Itō integral leads to

$$\begin{aligned} E|X_T - \hat{X}_N|^p &\leq c \cdot E \int_0^T \left|\sigma_s^{\frac{1}{1-\gamma}} - \sigma_{\eta(s)}^{\frac{1}{1-\gamma}}\right|^p + \left|\sigma_{\eta(s)}^{\frac{1}{1-\gamma}} - \hat{\sigma}_{\eta(s)}^{\frac{1}{1-\gamma}}\right|^p \mathrm{d}s \\ &\quad + c \cdot E \int_0^T \left|\sigma_s^{\frac{1}{2(1-\gamma)}} - \sigma_{\eta(s)}^{\frac{1}{2(1-\gamma)}}\right|^p + \left|\sigma_{\eta(s)}^{\frac{1}{2(1-\gamma)}} - \hat{\sigma}_{\eta(s)}^{\frac{1}{2(1-\gamma)}}\right|^p \mathrm{d}s \\ &\leq c \cdot \Delta^{p/2}, \end{aligned}$$

which finishes the proof. ∎

As pointed out in the introduction, the convergence rate deteriorates for discontinuous functionals. For example, we have the following result.

Corollary 7.2. *Let $K > 0$ and $i \in \{1, \dots, d\}$. (i) For $\gamma_i > 1/2$ there exists $C_{i,\lambda} > 0$ such that*

$$\left(E\left|1_{(-\infty,K]}(X_T^i) - 1_{(-\infty,K]}(\hat{X}_N^i)\right|^2\right)^{1/2} \leq C_i \cdot \Delta^\lambda \qquad \forall \quad \lambda < 1/4.$$

(ii) *For* $\gamma_i = 1/2$, $\frac{\kappa_i\lambda_i}{\theta_i^2} > 1$ *there exists* $C_{i,\lambda} > 0$ *such that*

$$\left(E\left|1_{(-\infty,K]}(X_T^i) - 1_{(-\infty,K]}(\hat{X}_N^i)\right|^2\right)^{1/2} \leq C_{i,\lambda} \cdot \Delta^\lambda \qquad \forall \quad \lambda < \frac{1}{4}\frac{\kappa_i\lambda_i/\theta_i^2}{\kappa_i\lambda_i/\theta_i^2 + 3/8}.$$

This result follows from the above proposition and an application of Theorem 2.4 in [6]. Moreover, similar to Theorem 7.4 in [6] it can be shown that the convergence order 1/4 is best possible in the above result.

We will now turn to the numerical analysis of (7.5), i.e., of

$$\hat{P} = f_1(e^{\hat{X}_N^1}, \dots, e^{\hat{X}_N^d}) + \frac{F_2(e^{\hat{X}_N^1}, \dots, e^{\hat{X}_N^d})}{e^{\hat{X}_N^1}} \cdot \hat{\Pi}.$$

For this, we will need the following assumptions.

Assumption 7.3.

(a) The functions $\tilde{f}_1, \tilde{f}_2, \tilde{G}_2 : \mathbb{R}^d \to \mathbb{R}$ given by

$$\tilde{f}_1(x_1, \dots, x_d) = f_1(e^{x_1}, \dots, e^{x_d}), \qquad \tilde{f}_2(x_1, \dots, x_d) = f_2(e^{x_1}, \dots, e^{x_d}),$$
$$\tilde{G}_2(x_1, \dots, x_d) = \frac{F_2(e^{x_1}, \dots, e^{x_d})}{e^{x_1}}$$

are bounded.

(b) The functions $\tilde{f}_1, \tilde{G}_2 : \mathbb{R}^d \to \mathbb{R}$ are globally Lipschitz continuous.

(c) For all $i = 2, \dots, d$ we have $\gamma_i > 1/2$ or $\frac{\kappa_i\lambda_i}{\theta_i^2} > 1$. For $i = 1$ we have $\gamma_1 > 1/2$ or $\frac{\kappa_1\lambda_1}{\theta_1^2} > \frac{3}{2}$.

The last assumption is stronger than Assumption 5.1(a) in the derivation of the quadrature formula and arises in the numerical analysis of the CIR process; see Theorem 9.3 in the appendix. The first assumption is, e.g., implied by f_1 and f_2 being bounded (see subsection 5.1), while the second depends on the chosen payoff splitting. We require the boundedness Assumption 7.3(a) here mainly to avoid another technical difficulty: since moment explosions might appear in the Heston model itself, they might also arise in its approximation.

The next result shows that the biases of the direct approximation and of (7.5) coincide, i.e., the integration by parts procedure does not change the bias of the approximation.

Proposition 7.4. *Let Assumptions* 5.1(a) *and* 7.3(a) *hold. Then we have*

$$Ef_2(e^{\hat{X}_N^1}, \dots, e^{\hat{X}_n^d}) = E\left(\frac{F_2(e^{\hat{X}_N^1}, \dots, e^{\hat{X}_N^d})}{e^{\hat{X}_N^1}} \cdot \hat{\Pi}\right).$$

Proof. Recall that

$$\hat{X}_{n+1}^1 = \hat{X}_n^1 + \left(b_1 - \frac{1}{2}(\hat{\sigma}_n^1)^{\frac{1}{1-\gamma_1}}\right)\Delta + (\hat{\sigma}_n^1)^{\frac{1}{2(1-\gamma_1)}}\Delta_n B^1$$

for $n = 0, 1, 2, \dots$. Because $\hat{\sigma}_1$ and Z^1 are independent by Assumption 5.1(b), we can easily calculate the derivative of $\hat{X}_n^i$:

$$D_t^{Z^1}\hat{X}_{n+1}^1 = D_t^{Z^1}\hat{X}_n^1 + R_{11}(\hat{\sigma}_n^1)^{\frac{1}{2(1-\gamma_1)}}1_{[n\Delta,(n+1)\Delta]}(t).$$

By recursion we have

$$D_t^{Z_1}\hat{X}_n^1 = R_{11}(\hat{\sigma}^1_{\lfloor t/\Delta\rfloor})^{\frac{1}{2(1-\gamma_1)}}.$$

In particular, it follows that $D_t^{Z_1}\hat{X}_n^1 \in \mathbb{D}^{1,p}$ for all $p \geq 1$ and $t \in [0,T]$. Using this result we can prove the quadrature formula for the approximation proceeding analogously to section 5, replacing X_t by $\hat{X}_{\lfloor t/\Delta\rfloor}$ and σ_t by $\hat{\sigma}_{\lfloor t/\Delta\rfloor}$. ■

The weak convergence rate α in this setting is still unknown. Numerical experiments indicate $\alpha = 1$. Note that to establish such a result would require the generalization of the seminal results of [7] to the nonhypoelliptic case.

The last piece of our error analysis is the strong approximation of the Malliavin weight.

Proposition 7.5. *Let Assumption 7.3(c) hold. Then, there exists a $C > 0$ such that*

$$E|\hat{\Pi} - \Pi|^2 \leq C \cdot \Delta.$$

Proof. We will again use c to denote constants, regardless of their value, and drop the indices.

(i) $\gamma = 1/2$. We have

$$\begin{aligned} E|\Pi - \hat{\Pi}|^2 &\leq c\,E\left|\int_0^T \left(\frac{1}{\sigma_s} - \frac{1}{\hat{\sigma}_{\eta(s)}}\right) \mathrm{d}Z_s\right|^2 \\ &\leq c\,E\int_0^T \frac{1}{\sigma_s^2\hat{\sigma}^2_{\eta(s)}}|\sigma_s - \sigma_{\eta(s)}|^2 ds + c\,E\int_0^T \frac{1}{\sigma_s^2\hat{\sigma}^2_{\eta(s)}}|\sigma_\eta(s) - \hat{\sigma}_{\eta(s)}|^2 ds \\ &=: S_1 + S_2. \end{aligned}$$

Choose $\varepsilon_1 > 0$ such that $1 + \varepsilon_1 < \frac{2\kappa\lambda}{3\theta^2}$ and set $q_1 = \frac{1+\varepsilon_1}{\varepsilon_1}$, then choose $\varepsilon_2 > 0$ such that $q_1\varepsilon_2 < \kappa\lambda/\theta^2 - 1/2$ and set $q_2 = \frac{1+\varepsilon_2}{\varepsilon_2}$. Note that $\frac{1}{1+\varepsilon_i} + \frac{1}{q_i} = 1$ for $i \in \{1,2\}$. Our choice of ε_1 and Theorem 9.3(i) give that

$$E\sup_{s\in[0,T]} \hat{\sigma}^{-2(1+\varepsilon_1)}_{\eta(s)} < \infty.$$

Thus we obtain by Hölder's inequality that

$$\begin{aligned} S_1 &\leq E\left(\int_0^T \sigma_s^{-2}\cdot|\sigma_s - \sigma_{\eta(s)}|^2 \mathrm{d}s \cdot \sup_{s\in[0,T]} \hat{\sigma}^{-2}_{\eta(s)}\right) \\ &\leq \left(E\left(\int_0^T \sigma_s^{-2}\cdot|\sigma_s - \sigma_{\eta(s)}|^2 \mathrm{d}s\right)^{q_1}\right)^{\frac{1}{q_1}} \cdot \left(E\sup_{s\in[0,T]} \hat{\sigma}^{-2(1+\varepsilon_1)}_{\eta(s)}\right)^{\frac{1}{1+\varepsilon_1}} \\ &\leq c\left(E\left(\int_0^T \sigma_s^{-2}\cdot|\sigma_s - \sigma_{\eta(s)}|^2 \mathrm{d}s\right)^{q_1}\right)^{\frac{1}{q_1}}. \end{aligned}$$

Applying Hölder's inequality twice and using Jensen's inequality then gives

$$\begin{aligned} S_1 &\leq c \cdot \left(E\left[\left(\int_0^T \sigma_s^{-2(1+\varepsilon_2)} \mathrm{d}s \right)^{\frac{q_1}{1+\varepsilon_2}} \cdot \left(\int_0^T |\sigma_s - \sigma_{\eta(s)}|^{2q_2} \mathrm{d}s \right)^{\frac{q_1}{q_2}} \right] \right)^{\frac{1}{q_1}} \\ &\leq c \cdot \left(E\left(\int_0^T \sigma_s^{-2(1+\varepsilon_2)} \mathrm{d}s \right)^{q_1} \right)^{\frac{1}{q_1(1+\varepsilon_2)}} \cdot \left(E\left(\int_0^T |\sigma_s - \sigma_{\eta(s)}|^{2q_2} \mathrm{d}s \right)^{q_1} \right)^{\frac{1}{q_1 q_2}} \\ &\leq c \cdot \left(E\left(\int_0^T \sigma_s^{-2(1+\varepsilon_2)} \mathrm{d}s \right)^{q_1} \right)^{\frac{1}{q_1(1+\varepsilon_2)}} \cdot \left(E \int_0^T |\sigma_s - \sigma_{\eta(s)}|^{2q_1 q_2} \mathrm{d}s \right)^{\frac{1}{q_1 q_2}}. \end{aligned}$$

By the mean-square smoothness of the CIR process, the last expectation is bounded by $c\Delta$. It remains to show that the first factor is finite. In fact, we have

$$\begin{aligned} E\left(\int_0^T \sigma_s^{-2(1+\varepsilon_2)} \mathrm{d}s \right)^{q_1} &\leq E\left(\left(\int_0^T \sigma_s^{-2} \mathrm{d}s \right)^{q_1} \cdot \sup_{s\in[0,T]} \sigma_s^{-2\varepsilon_2 q_1} \right) \\ &\leq \left(E\left(\int_0^T \sigma_s^{-2} \mathrm{d}s \right)^{2q_1} \right)^{\frac{1}{2}} \cdot \left(E \sup_{s\in[0,T]} \sigma_s^{-4q_1\varepsilon_2} \right)^{\frac{1}{2}}. \end{aligned}$$

Both terms are finite due to Lemma 2.2(ii) and $2q_1\varepsilon_2 < 2\kappa\lambda/\theta^2 - 1$.

Now consider the second summand S_2. Choose $q_3, \varepsilon_3 > 0$ such that $\frac{1}{q_3} + \frac{1}{1+\varepsilon_3} = 1$ and $1 + \varepsilon_3 < \frac{2}{3}\frac{\kappa\lambda}{\theta^2}$ as well as $(1+\varepsilon_3)^2 < 2\kappa\lambda/\theta^2 - 1$. Applying Hölder's inequality yields

$$S_2 \leq c \cdot \left(E\left(\int_0^T \sigma_s^{-2} \hat{\sigma}_{\eta(s)}^{-2} \mathrm{d}s \right)^{q_3} \right)^{1/q_3} \cdot \left(E \sup_{k=0,\dots,T/\Delta} |\sigma_{k\Delta} - \hat{\sigma}_{k\Delta}|^{2(1+\varepsilon_3)} \right)^{\frac{1}{1+\varepsilon_3}}.$$

The last factor is bounded by $c \cdot \Delta^2$ due to (9.1) in Theorem 9.3 and thus we have

$$S_2 \leq c \cdot \left(E\left(\int_0^T \sigma_s^{-2} \hat{\sigma}_{\eta(s)}^{-2} \mathrm{d}s \right)^{q_3} \right)^{1/q_3} \cdot \Delta^2.$$

We apply Hölder's inequality twice and use Jensen's inequality to get

$$\begin{aligned} E\left(\int_0^T \sigma_s^{-2} \hat{\sigma}_{\eta(s)}^{-2} \mathrm{d}s \right)^{q_3} &\leq E\left(\left(\int_0^T \sigma_s^{-2(1+\varepsilon_3)} \mathrm{d}s \right)^{\frac{q_3}{1+\varepsilon_3}} \cdot \int_0^T \hat{\sigma}_{\eta(s)}^{-2q_3} \mathrm{d}s \right) \\ &\leq E\left(\left(\int_0^T \sigma_s^{-2(1+\varepsilon_3)} \mathrm{d}s \right)^{q_3} \right)^{\frac{1}{1+\varepsilon_3}} \cdot \left(E \int_0^T \hat{\sigma}_s^{-2q_3^2} \mathrm{d}s \right)^{\frac{1}{q_3}}. \end{aligned}$$

Here the last term is bounded by $c \cdot \Delta^{-q_3}$ due to Lemma 9.4. Applying Hölder's inequality again yields

$$E\left(\int_0^T \sigma_s^{-2(1+\varepsilon_3)}\mathrm{d}s\right)^{q_3} \le E\left(\left(\int_0^T \sigma_s^{-2}\mathrm{d}s\right)^{q_3} \cdot \sup_{s\in[0,T]} \sigma_s^{-2\varepsilon_3 q_3}\right)$$
$$\le \left(E\left(\int_0^T \sigma_s^{-2}\mathrm{d}s\right)^{q_3^2}\right)^{\frac{1}{q_3}} \cdot \left(E \sup_{s\in[0,T]} \sigma_s^{-2\varepsilon_3 q_3(1+\varepsilon_3)}\right)^{\frac{1}{1+\varepsilon_3}}.$$

Since $\varepsilon_3 q_3(1+\varepsilon_3) = (1+\varepsilon_3)^2 < 2\kappa\lambda/\theta^2 - 1$, both terms are finite due to Lemma 2.2(ii).

In summary we have shown that $S_2 \le c \cdot \Delta$, which concludes the proof of this assertion.

(ii) $\gamma > 1/2$. Here we have

$$E\left|\int_0^T \left(\sigma_s^{-\frac{1}{2(1-\gamma)}} - \hat\sigma_{\eta(s)}^{-\frac{1}{2(1-\gamma)}}\right)\mathrm{d}Z_s\right|^2$$
$$\le cE\left|\int_0^T \left(\sigma_s^{-\frac{1}{2(1-\gamma)}} - \sigma_{\eta(s)}^{-\frac{1}{2(1-\gamma)}}\right)\mathrm{d}Z_s\right|^2 + cE\left|\int_0^T \left(\sigma_{\eta(s)}^{-\frac{1}{2(1-\gamma)}} - \hat\sigma_{\eta(s)}^{-\frac{1}{2(1-\gamma)}}\right)\mathrm{d}Z_s\right|^2.$$

Since all inverse moments of σ and its Lamperti–backward Euler approximation exist, the result can be established now by straightforward estimates. ∎

Combining Propositions 7.1, 7.4, and 7.5 we have shown the following.

Theorem 7.6. *Let Assumption 7.3 hold. Then we have for $\hat P$ given by (7.5) that*

$$E\hat P = Ef(e^{\hat X_N^1}, \dots, e^{\hat X_N^d})$$

and

$$E|\hat P - P|^2 \le C \cdot \Delta$$

for some $C > 0$.

This allows in particular to apply multilevel Monte Carlo and the complexity result (6.1) with $\beta = 1/2$. If Assumption 7.3(c) is replaced by the following weaker assumption, then we still have L^2-convergence of $\hat P$.

Assumption 7.7. For all $i = 1, \dots, d$ we have $\gamma_i > 1/2$ or $\frac{\kappa_i\lambda_i}{\theta_i^2} > 1$.

Proposition 7.8. *Let Assumptions* 7.3(a) and (b) *and Assumption 7.7 hold. Then we have for $\hat P$ given by (7.5) that*

$$E\hat P = Ef(e^{\hat X_N^1}, \dots, e^{\hat X_N^d})$$

and

$$\lim_{\Delta\to 0} E|\hat P - P|^2 = 0.$$

Proof. The weaker assumption 7.7 has no influence on Propositions 7.1 and 7.4. We can show that $\lim_{\Delta\to 0} E|\hat\Pi - \Pi|^2 = 0$ along the lines of the proof of Proposition 7.5 by choosing $1 + \varepsilon_1 < \frac{4\kappa\lambda}{3\theta^2}$ and using (9.2) to estimate S_2. ∎

8. Numerical results. In the case of the Heston model we additionally use the Kahl–Jäckel scheme [24] for a comparison with the scheme introduced above. In the Kahl–Jäckel scheme the variance process is discretized with a drift-implicit Milstein scheme, i.e.,

$$\hat{v}^i_{k+1} = \frac{\hat{v}^i_k + \kappa_i\lambda_i\Delta + \theta_i\sqrt{\hat{v}^i_k}\Delta_k W^i + \frac{\theta_i^2}{4}((\Delta_k W^i)^2 - \Delta)}{1+\kappa_i\Delta},$$

while the log-price is approximated using the following formula:

$$\begin{aligned}\hat{X}^i_{k+1} = \hat{X}^i_k - \frac{\Delta}{4}(\hat{v}^i_{k+1} + \hat{v}^i_k) + \rho_i\sqrt{\hat{v}^i_k}\Delta_k W^i \\ + \frac{1}{2}\left(\sqrt{\hat{v}^i_{k+1}} + \sqrt{\hat{v}^i_k}\right)(\Delta_k B^i - \rho_i\Delta_k W^i) + \frac{1}{4}\theta_i\rho_i((\Delta_k W^i)^2 - \Delta).\end{aligned}$$

Note that the main aim of this section is to test the numerical benefit of using the quadrature formula and not the comparison of the many different simulation schemes, which are available for the generalized Heston model.

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

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

$$T = 2,\ b = 0,\ \kappa = 5.07,\ \lambda = 0.0457,\ \theta = 0.48,\ \rho = -0.767,\ v_0 = \lambda,\ S_0 = 100.$$

Note that we have $\frac{\kappa\lambda}{\theta^2} > 1$.

(a′) Same as above, but using the Kahl–Jäckel scheme instead.

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

$$\kappa = 4.1031,\ \lambda = 0.0451,\ \theta = 0.8583,\ \gamma = 0.6545,\ \rho = -0.760,\ v_0 = \lambda,\ S_0 = 100.$$

(c) Three-dimensional Heston model. Again we set $T = 2$, $b_1 = b_2 = b_3 = 0$. Here the parameters are taken from [11]. However, we modified ρ_1, ρ_2 to avoid negative correlations close to -1.

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

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 $L^{2+\varepsilon}$. Furthermore, we have $\frac{\kappa_i\lambda_i}{\theta_i^2} > 1$, $i = 1, 2, 3$.

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

In our numerical tests we perform a comparison of the following three algorithms at different error tolerances ε in these models:

A1. Standard Monte Carlo (using $\lceil \varepsilon^{-2} \rceil$ iterations and $\lceil T/\varepsilon \rceil$ steps).

A2. Adaptive multilevel Monte Carlo using the discontinuous payoff directly, i.e., simulating $1_{[0,d\cdot S_0]}(\sum_{i=1}^d e^{\hat{X}_N^i})$.

A3. Adaptive multilevel Monte Carlo using the quadrature formula and payoff splitting. For a $\delta \in (0,1)$ we use the splitting

$$f_1(x) = \begin{cases} 1, & x < (1-\delta)K, \\ -\frac{1}{2\delta K}(x-K) + 0.5, & x \in [(1-\delta)K, (1+\delta)K], \\ 0, & x > (1+\delta)K, \end{cases}$$

and $f_2 = 1_{[0,K]} - f_1$, $F_2(x) = \int_0^x f_2(z)\mathrm{d}z$. As discussed before the algorithm then simulates

$$f_1\left(\sum_{i=1}^d e^{\hat{X}_N^i}\right) + \frac{F_2\left(\sum_{i=1}^d e^{\hat{X}_N^i}\right)}{e^{\hat{X}_N^1}} \cdot \hat{\Pi}.$$

Note that Assumptions 7.3(a) and (b) are satisfied here.

For each algorithm a series of error tolerances was chosen and the algorithm was executed 500 times for each tolerance. Figures 1 and 2 plot the mean cost against the relative mean-square error, both on logarithmic scale (base 2). Since the exact value of p is unknown, the reference value p_{ref} has been computed using algorithm A3 with an error tolerance of 25% of the smallest tolerance in the plot. For each algorithm the figures contain a least squares line that has been fitted to the data points to estimate the convergence rates.

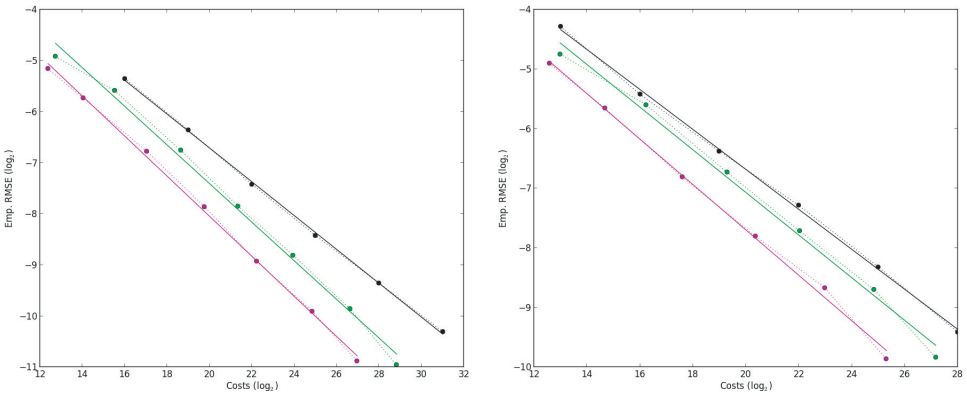
The (random) cost of the multilevel algorithms was measured by the total number of discretization steps, i.e.,

$$\text{cost} := d \cdot \left(N_0 + \sum_{l=1}^{L} N_l \cdot (M^l + M^{l-1}) \right), \tag{8.1}$$

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

$$\text{rmsq error} = \sqrt{\frac{1}{500} \sum_{j=1}^{500} \frac{|\widehat{P}^{(j)} - p_{ref}|^2}{p_{ref}^2}}.$$

Figure 1 that algorithm A3 (pink markers) is between two and four times faster than algorithm A2 (green markers), 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 A1 (black markers). However, further numerical tests have shown that it



(a) $d = 1$, $\gamma = 0.5, \delta = 0.3$ (b) $d = 3$, $\gamma = 0.5, \delta = 0.1$

Figure 1. *Numerical results and fitted convergence rates.*

Table 1

Measured convergence rates (with least squares residual). Note that the residuals are computed from the logarithms of the errors (or—in the last line—the costs); see Figure 1. MLMC = multilevel Monte Carlo.

Algorithm	Model (a)	Model (b)	Model (c)	Model (d)	Model (a′)
A1: Monte Carlo	0.33 (0.01)	0.34 (0.02)	0.33 (0.01)	0.33 (0.01)	0.34 (0.00)
A2: Adaptive MLMC	0.38 (0.16)	0.36 (0.12)	0.36 (0.03)	0.34 (0.04)	0.38 (0.03)
A3: Adaptive MLMC & Malliavin quadrature	0.39 (0.05)	0.38 (0.05)	0.38 (0.04)	0.40 (0.03)	0.40 (0.03)
A3 with logarithmic fit	0.46 (0.12)	0.45 (0.18)	0.44 (0.09)	0.47 (0.09)	0.48 (0.04)

is important to choose a good splitting parameter: for the figures, the parameter δ which governs the payoff splitting was set to values between 0.1 and 0.4. Higher values of δ lead in fact to an algorithm worse than algorithm A2.

Note that here we have only considered at-the-money options. The smoothing can also be applied to out-of-the-money options; however, the numerical benefit is smaller in this case, since most paths will end up far away from the strike.

Table 1 contains the estimated convergence rates, which are in good accordance with the theoretical rates given by (6.1). Assuming $\alpha = 1$ for A1 and $\alpha = 1$, $\beta = 1/4$ for A2 we obtain a convergence order of 1/3 for standard Monte Carlo, respectively, 2/5 for direct multilevel Monte Carlo. For algorithm A3 the convergence order 1/2 indicated by Theorem 7.6 is visible when the logarithmic term is taken into account. In fact, when fitting a function $\varepsilon \mapsto c \cdot \varepsilon^r \cdot (\log \varepsilon)^2$ to the data points of algorithm A3, the rate r takes values between 0.44 and 0.48 for the different models; see the last row in Table 1.

Tables 2–4 contains the measured errors, running times, and costs for a single run (averaged over 500 runs) using error tolerance $\varepsilon = 2^{-8}$ for algorithms A2 and A3 and $\varepsilon = 2^{-7.5}$ for A1. Using these tolerances all three algorithms give an empirical relative mean-square error of approximately $2^{-7.5}$ (see Table 2) and the running times are thus comparable. Note that for the generalized Heston model, solving the implicit equation creates a computational overhead which is visible in the running times.

Table 2
Relative errors (with standard deviations).

Algorithm	Model (a)	Model (b)	Model (c)	Model (d)	Model (a′)
A1	0.005 (0.003)	0.005 (0.004)	0.003 (0.003)	0.004 (0.003)	0.005 (0.003)
A2	0.005 (0.003)	0.005 (0.003)	0.005 (0.003)	0.006 (0.005)	0.005 (0.004)
A3	0.004 (0.003)	0.005 (0.003)	0.005 (0.004)	0.006 (0.004)	0.006 (0.004)

Table 3
Running times in seconds (with standard deviations).

Algorithm	Model (a)	Model (b)	Model (c)	Model (d)	Model (a′)
A1	0.92 (0.01)	7.90 (0.04)	5.24 (0.05)	28.92 (0.16)	1.03 (0.00)
A2	0.31 (0.16)	3.15 (1.17)	1.23 (0.44)	5.31 (2.34)	0.65 (0.20)
A3	0.10 (0.04)	0.97 (0.28)	0.64 (0.17)	2.00 (0.62)	0.17 (0.04)

Table 4
Measured costs (number of discretization steps) in million steps (with standard deviation).

Algorithm	Model (a)	Model (b)	Model (c)	Model (d)	Model (a′)
A1: Monte Carlo	11.9	11.9	35.7	35.7	11.9
A2: Adaptive MLMC	2.8 (1.3)	4.3 (1.6)	9.7 (2.9)	6.9 (2.6)	5.9 (2.0)
A3: Adaptive MLMC & Malliavin Quadrature	0.9 (0.4)	1.4 (0.4)	5.3 (1.5)	2.7 (0.9)	1.5 (0.4)

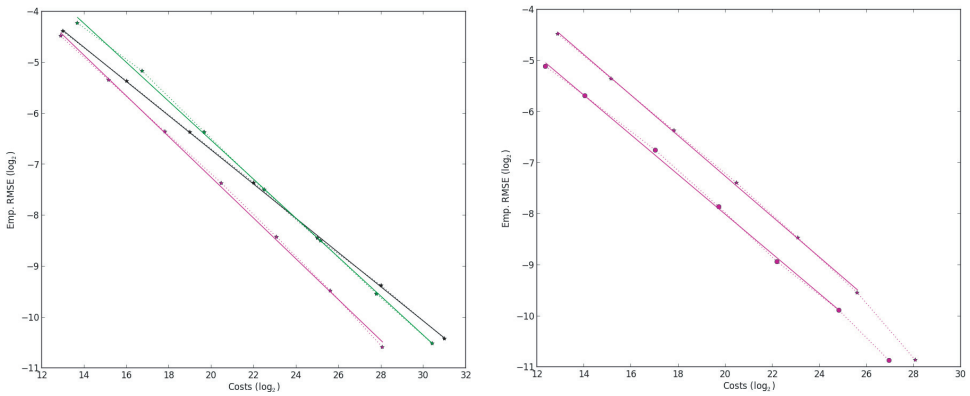


Figure 2. *Left: Comparison of algorithms in model* (a′). *Right: Comparison of algorithm* A3 *using Kahl–Jäckel scheme (stars) versus log-Euler/Lamperti–backward Euler scheme (dots).*

Figure 2 shows that the Kahl–Jäckel method gives similar results for the performance of the different algorithms, the only difference being a lower initial variance for standard Monte Carlo.

9. Appendix.

9.1. 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 $W = (W^1, \dots, W^d)$ be a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$ and fix an endtime $T > 0$. Denote by $C^\infty_{\mathrm{pol}}(\mathbb{R}^n; \mathbb{R})$ the functions $f\colon \mathbb{R}^n \to \mathbb{R}$ which are infinitely often differentiable such that f and all its partial derivatives of any order are bounded by polynomials. We call a random variable X *smooth* if it can be written as

$$X = f\left(\int_0^T h_1(s)\mathrm{d}W_s, \dots, \int_0^T h_n(s)\mathrm{d}W_s\right)$$

with $f \in C^\infty_{\mathrm{pol}}(\mathbb{R}^n; \mathbb{R})$ and $h_i \in H := L^2([0,T]; \mathbb{R}^d)$ for $i = 1, \dots, n$. 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

$$DX = \sum_{i=1}^n \partial_i f\left(\int_0^T h_1(s)\mathrm{d}W_s, \dots, \int_0^T h_n(s)\mathrm{d}W_s\right) \cdot h_i.$$

In particular $DW_t^i = 1_{[0,t]}(\cdot) \cdot e_i$, where e_i is the ith 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

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

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 $D_r^i X$ for the ith 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 w.r.t. 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 9.1. *Let $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ be a locally Lipschitz function. Assume $X = (X_1, \dots, X_m)$ 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 $i = 1, \dots, m$,*
- *and the set of $x \in \mathbb{R}^m$ where φ 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 [12, sect. 5.8.3, Thm. 6] this set is a Lebesgue-null set and hence the third assumption is always fulfilled, if the law of X is absolute continuous w.r.t. the Lebesgue measure.

Proof. By Proposition 1.2.3 in [31] the chain rule holds for functions which are continuously differentiable with bounded partial derivatives. The claim follows by approximating φ with such functions. ∎

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

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

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

$$\langle DX, u\rangle_{L^2(\Omega;H)} = \langle X, \delta(u)\rangle_{L^2(\Omega)} \qquad \forall\, X \in \mathbb{D}^{1,2}.$$

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

$$E\left(\int_0^T \langle D_r(X), u(r)\rangle_{\mathbb{R}^d}\,\mathrm{d}r\right) = E(X\cdot\delta(u)) \qquad \forall\, X \in \mathbb{D}^{1,2},\ \ u \in \operatorname{dom}\delta.$$

Further analysis shows that for an adapted process u the Skorohod integral $\delta(u)$ coincides with $\int_0^T u(s)\mathrm{d}W_s$ and hence it can be viewed as an extension of the Itō integral.

To derive the quadrature formula for a large class of multidimensional functionals in Theorem 5.4, we will need a derivative operator that differentiates only w.r.t. 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 w.r.t. B as the $L^2([0,T])$-valued random variable given by

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

On random variables which are measurable w.r.t. 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 $u\cdot a \in \operatorname{dom}\delta$. For such u it is given by $\delta^B(u) = \delta(u\cdot a)$. The chain rule Theorem 9.1 and the integration by parts rule are easily extended to D^B and δ^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 9.2. *Let B be a Brownian motion on $(\Omega,\mathcal{F},P)$ and $X = (X_1,\dots,X_m)$ a random vector with components in $\mathbb{D}^{1,p}$ and set $A := \{i = 1,\dots,m : D^B X_i \neq 0\}$. Assume $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ is continuous except on a P_X-zero set and assume that φ is globally Lipschitz continuous in all directions e_i, $i \in A$, i.e., for all $i \in A$ there exists a constant L_i such that $|\varphi(x+he_i) - \varphi(x)| \le L_i\cdot h$ for all $h \in \mathbb{R}$. 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. This can be shown using the previous chain rule and a standard approximation procedure. ∎

9.2. Properties of the Lamperti–backward Euler scheme for CIR/CEV. Here we collect some properties of the backward Euler approximation

$$\begin{aligned}\hat{\sigma}^i_{k+1} &= \hat{\sigma}^i_k + (1-\gamma_i)\left(f_i(\hat{\sigma}^i_{k+1}) - \kappa_i\hat{\sigma}^i_{k+1}\right)\Delta + (1-\gamma_i)\theta_i\Delta_k W^i, \qquad k = 0, 1\ldots,\\ \hat{\sigma}^i_0 &= \sigma^i_0\end{aligned}$$

of the transformed CIR/CEV process. The following error bounds have been derived in [3, 30].

Theorem 9.3.

(i) *Let* $\gamma_i = 1/2$ *and* $\frac{\kappa_i\lambda_i}{\theta_i^2} > 1$. *Then we have*

$$E\sup_{k=0,\ldots,\lceil T/\Delta\rceil}|\hat{\sigma}_{k\Delta}|^p < \infty \qquad \forall \quad p > -\frac{4\kappa\lambda}{3\theta^2}.$$

Moreover, it holds that

$$E\sup_{k=0,\ldots,\lceil T/\Delta\rceil}|\sigma_{k\Delta} - \hat{\sigma}_{k\Delta}|^p \le c_p\cdot\Delta^p \tag{9.1}$$

for some constant $c_p > 0$ *and all* $1 \le p < \frac{4\kappa\lambda}{3\theta^2}$ *and*

$$E\sup_{k=0,\ldots,\lceil T/\Delta\rceil}|\sigma_{k\Delta} - \hat{\sigma}_{k\Delta}|^p \le c_p\cdot\Delta^{p/2} \tag{9.2}$$

for some constant $c_p > 0$ *and all* $1 \le p < \frac{8\kappa\lambda}{3\theta^2}$.

(ii) *Let* $\gamma_i > 1/2$. *Then it holds that*

$$E\sup_{k=0,\ldots,\lceil T/\Delta\rceil}|\hat{\sigma}_{k\Delta}|^p < \infty \qquad \forall \quad p \in \mathbb{R}$$

and we have

$$E\sup_{k=0,\ldots,\lceil T/\Delta\rceil}|\sigma_{k\Delta} - \hat{\sigma}_{k\Delta}|^p \le c_p\cdot\Delta^p \tag{9.3}$$

for some constant $c_p > 0$ *and all* $p \ge 1$.

In the case of the backward Euler approximation of the transformed CIR process the following lower bound will be helpful.

Lemma 9.4. *Let* $\gamma_i = 1/2$. *For all* $p \ge 1$ *there exists* $c_p > 0$ *such that*

$$E|\hat{\sigma}^i_{(k+1)\Delta}|^{-p} \le c_p\cdot\Delta^{-p/2}$$

for all $k = 0, 1, \ldots$.

Proof. Drop the index i and set

$$u_k = \hat{\sigma}_{k\Delta} + \frac{\theta}{2}\Delta W^1_k, \qquad c_\Delta = \left(\kappa\lambda - \frac{\theta^2}{4}\right)(2+\kappa\Delta).$$

Then we can rewrite (7.4) as

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

If $u_k \geq 0$, then obviously

$$\hat{\sigma}_{(k+1)\Delta} \geq \frac{\sqrt{c_\Delta \Delta}}{2 + \kappa\Delta}.$$

Moreover, if $u_k < 0$, then the mean value theorem implies that

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

So we have

$$\frac{1}{\hat{\sigma}_{(k+1)\Delta}} \leq \frac{2(2 + \kappa\Delta)\sqrt{u_k^2 + c_\Delta \Delta}}{c_\Delta \Delta}.$$

Since $\hat{\sigma}_{k\Delta} > 0$, the condition $u_k < 0$ gives that $|u_k| \leq \frac{\theta}{2}|\Delta_k W|$. Hence it follows that

$$\frac{1}{\hat{\sigma}_{(k+1)\Delta}} \leq \frac{\theta(2 + \kappa\Delta)\sqrt{|\Delta_k W|^2 + c_\Delta \Delta}}{c_\Delta \Delta},$$

which yields the assertion. ■

Acknowledgment. The authors would like thank the referees for their valuable comments and suggestions.

REFERENCES

[1] Y. Aït-Sahalia and R. Kimmel, *Maximum likelihood estimation of stochastic volatility models*, J. Financial Econom., 83 (2007), pp. 413–452.
[2] A. Alfonsi, *On the discretization schemes for the CIR (and Bessel squared) processes*, Monte Carlo Methods Appl., 11 (2005), pp. 355–384.
[3] A. Alfonsi, *Strong order one convergence of a drift implicit Euler scheme: Application to the CIR process*, Statist. Probab. Lett., 83 (2013), pp. 602–607.
[4] E. Alòs and C.-O. Ewald, *Malliavin differentiability of the Heston volatility and applications to option pricing*, Adv. Appl. Probab., 40 (2008), pp. 144–162.
[5] L.B.G. Andersen and V.V. Piterbarg, *Moment explosions in stochastic volatility models*, Finance Stoch., 11 (2006), pp. 29–50.
[6] R. Avikainen, *On irregular functionals of SDEs and the Euler scheme*, Finance Stoch., 13 (2009), pp. 381–401.
[7] V. Bally and D. Talay, *The law of the Euler scheme for stochastic differential equations.* I: *Convergence rate of the distribution function*, Probab. Theory Related Fields, 104 (1996), pp. 43–60.
[8] A. Berkaoui, M. Bossy, and A. Diop, *Euler scheme for SDEs with non-Lipschitz diffusion coefficient: Strong convergence*, ESAIM Probab. Stat., 12 (2007), pp. 1–11.
[9] P. Carr and D.B. Madan, *Option valuation using the fast Fourier transform*, J. Comput. Finance, 2 (1999), pp. 61–73.
[10] 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 (2012), pp. 1105–1115.

[11] G. Dimitroff, S. Lorenz, and A. Szimayer, *A parsimonious multi-asset Heston model: Calibration and derivative pricing*, Int. J. Theor. Appl. Finance, 14 (2011), pp. 1299–1333.
[12] L.C. Evans, *Partial Differential Equations*, 2nd ed., AMS, Providence, RI, 2010.
[13] P.K. Friz and M. Keller-Ressel, *Moment explosions in stochastic volatility models*, in Encyclopedia of Quantitative Finance, John Wiley, Chichester, UK, 2010, pp. 1247–1253.
[14] M.B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.
[15] M.B. Giles, D.J. Higham, and X. Mao, *Analysing multilevel Monte Carlo for options with non-globally Lipschitz payoff*, Finance Stoch., 13 (2009), pp. 403–413.
[16] M. Giles, T. Nagapetyan, and K. Ritter, *Multi-level Monte Carlo Approximation of Distribution Functions and Densities*, preprint 157, DFG-Schwerpunktprogramm 1324, 2014.
[17] E. Gobet, E. Benhamou, and M. Miri, *Time dependent Heston model*, SIAM J. Financial Math., 1 (2010), pp. 289–325.
[18] L.A. Grzelak and C.W. Oosterlee, *On the Heston model with stochastic interest rates*, SIAM J. Financial Math., 2 (2011), pp. 255–286.
[19] S. Heinrich, *Multilevel Monte Carlo methods*, in Large-Scale Scientific Computing, S. Margenov, et al., eds., Lecture Notes in Comput. Sci. 2179, Springer, New York, 2001, pp. 58–67.
[20] S.L. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, Rev. Financial Studies, 6 (1993), pp. 327–343.
[21] T.R. Hurd and A. Kuznetsov, *Explicit formulas for Laplace transforms of stochastic integrals*, Markov Process. Related Fields, 14 (2008), pp. 277–290.
[22] 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 (2010), pp. 303–320.
[23] C. Kahl, M. Günther, and T. Rossberg, *Structure preserving stochastic integration schemes in interest rate derivative modeling*, Appl. Numer. Math., 58 (2008), pp. 284–295.
[24] C. Kahl and P. Jäckel, *Fast strong approximation Monte Carlo schemes for stochastic volatility models*, Quant. Finance, 6 (2006), pp. 513–536.
[25] I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer, New York, 2010.
[26] A. Kebaier, *Statistical Romberg extrapolation: A new variance reduction method and applications to option pricing*, Ann. Appl. Prob., 15 (2005), pp. 2681–2705.
[27] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, 1992.
[28] P.-L. Lions and M. Musiela, *Correlations and bounds for stochastic volatility models*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 24 (2007), pp. 1–16.
[29] R. Lord, R. Koekkoek, and D. Van Dijk, *A comparison of biased simulation schemes for stochastic volatility models*, Quant. Finance, 10 (2010), pp. 177–194.
[30] A. Neuenkirch and L. Szpruch, *First order strong approximations of scalar SDEs with values in a domain*, Numer. Math., 128 (2014), pp. 103–136.
[31] D. Nualart, *The Malliavin Calculus and Related Topics*, 2nd ed., Springer, Berlin, 2006.
[32] W. Schoutens, E. Simons, and J. Tistaert, *A perfect calibration! Now what?*, Wilmott Magazine, March, 2004.