

Numerical Smoothing with Hierarchical Adaptive Sparse Grids and Quasi-Monte Carlo Methods for Efficient Option Pricing

Christian Bayer¹, Chiheb Ben Hammouda*² and Raúl Tempone^{3,4}

¹Weierstrass Institute for Applied Analysis and Stochastics (WIAS), Berlin, Germany.

²Chair of Mathematics for Uncertainty Quantification, RWTH Aachen University, Aachen, Germany.

³King Abdullah University of Science and Technology (KAUST), Computer, Electrical and Mathematical Sciences & Engineering Division (CEMSE), Thuwal, Saudi Arabia.

⁴Alexander von Humboldt Professor in Mathematics for Uncertainty Quantification, RWTH Aachen University, Aachen, Germany.

Abstract

When approximating the expectations of a functional of a solution to a stochastic differential equation, the numerical performance of deterministic quadrature methods, such as sparse grid quadrature and quasi-Monte Carlo (QMC) methods, may critically depend on the regularity of the integrand. To overcome this issue and improve the regularity structure of the problem, we consider cases in which analytic smoothing (bias-free mollification) cannot be performed and introduce a novel numerical smoothing approach by combining a root-finding method with a one-dimensional numerical integration with respect to a single well-chosen variable. We prove that, under appropriate conditions, the resulting function of the remaining variables is highly smooth, potentially affording the improved efficiency of adaptive sparse grid quadrature (ASGQ) and QMC methods, particularly when combined with hierarchical transformations (*i.e.*, the Brownian bridge and Richardson extrapolation on the weak error). This approach facilitates the effective treatment of high dimensionality. Our study is motivated by option pricing problems, focusing on dynamics where the discretization of the asset price is necessary. Based on our analysis and numerical experiments, we demonstrate the advantages of combining numerical smoothing with the ASGQ and QMC methods over these methods without smoothing and the Monte Carlo approach. Finally, our approach is generic and can be applied to solve a broad class of problems, particularly approximating distribution functions, computing financial Greeks, and estimating risk quantities.

Keywords Adaptive sparse grid quadrature, quasi-Monte Carlo, numerical smoothing, Brownian bridge, Richardson extrapolation, option pricing, Monte Carlo, distribution functions, Greeks, risk estimation

2010 Mathematics Subject Classification 65C05, 65D30, 65D32, 65Y20, 91G20, 91G60.

*benhammouda@uq.rwth-aachen.de

1 Introduction

In several applications, such as pricing digital and barrier options, computing financial Greeks, and estimating risk quantities and distribution functions, one is interested in efficiently computing the expectation of a functional g of a solution X to a stochastic differential equation (SDE):

$$(1.1) \quad \mathbb{E}[g(X)].$$

Approximating (1.1) is usually challenging due to the combination of two complications:

1. An input space can have high dimensionality for many reasons, including (i) the time discretization of an SDE that describes the dynamics or (ii) having numerous underlying assets.
2. The payoff function, g , exhibits low regularity with respect to (w.r.t.) the input parameters.

The first class of methods for approximating (1.1) relies on Monte Carlo (MC) techniques. Although the convergence rate of the standard MC method is insensitive to the input space dimensionality and the regularity of the observable g , the convergence may be very slow. Moreover, it may not exploit the available regularity structure that could help achieve better convergence rates, except for multilevel MC methods [19, 7], where Lipschitzity is necessary to obtain optimal convergence rates. Another class of methods relies on deterministic quadrature techniques (e.g., sparse grid quadrature [18, 4, 11], adaptive sparse grid quadrature (ASGQ) [8, 6, 9, 5], and quasi-MC (QMC) [30, 8, 6]). In this work, we introduce a numerical smoothing technique to improve the performance of deterministic quadrature approaches by improving the regularity structure of the problem.

The high dimensionality of the input space and existence of discontinuities¹ in the integrand considerably degrade the performance of deterministic quadrature methods. Some studies [21, 22, 23, 8, 34] have addressed cases involving integrands with discontinuities; however, the emphasis was on the QMC method. In particular, [21, 22, 23] focused on the theoretical aspects of employing the QMC method in such a setting. An adaptive version of the QMC method combined with geometric random splitting was employed for pricing multidimensional vanilla options for the Black-Scholes model [14]. Moreover, the low regularity of the integrand was addressed in previous studies by (i) performing bias-free mollification using the conditional expectation over a subset of integration variables [8, 34, 6], or (ii) mapping the problem to the frequency space [5], implying a better regularity structure compared to the physical space, when applicable.²

This work considers cases where bias-free mollification cannot be performed. We introduce a novel numerical smoothing technique based on (i) identifying discontinuity locations in a lower-dimensional space using hierarchical path generation and a linear transformation of the coarsest factors, (ii) solving the discontinuities using root-finding algorithms, (iii) employing suitable transformations of the integration domain, and (iv) a numerical preintegration step w.r.t. the dimension containing discontinuities. We prove that, under appropriate conditions, the resulting function of the remaining variables is highly smooth, potentially affording improved efficiency of the ASGQ and QMC methods, particularly when combined with hierarchical transformations to treat the high dimensionality effectively [8]. Given that ASGQ and QMC methods benefit from anisotropy, the first technique involves employing a hierarchical path generation method based on the Brownian bridge construction to reduce the effective dimension. The second technique involves employing

¹We consider discontinuities either in the gradients (kinks) or in the function (jumps).

²The Fourier transform of the density function is available and inexpensive to compute.

the Richardson extrapolation to reduce the bias (weak error), subsequently reducing the number of time steps required at the coarsest level to achieve a certain error tolerance and decreasing the total number of dimensions required for the integration problem. Our analysis and numerical experiments demonstrate the advantage of our approach, substantially outperforming the ASGQ and QMC methods without smoothing and the MC approach, for high-dimensional examples and dynamics where discretization is needed, such as the Heston model.

The outline of this study is as follows: Section 2 explains the technique of numerical smoothing, the selection of the optimal smoothing direction, and the different building blocks that constitute our hierarchical quadrature methods. Section 3.1 presents the smoothness analysis of the resulting integrand after numerical smoothing. Next, Section 3.2 discusses the error and work for the ASGQ method with numerical smoothing. Finally, Section 4 reports the results of the numerical experiments conducted using the ASGQ, QMC, and MC methods. These results verify the considerable computational gains achieved using the ASGQ and QMC methods (both combined with numerical smoothing) over the MC method and the standard (without smoothing) ASGQ and QMC methods.

2 Problem Setting and Approach Formulation

To demonstrate the application of our approach, we work mainly with two possible structures of the observable g :

$$(2.1) \quad \text{(i) } g(\mathbf{x}) = \max(\phi(\mathbf{x}), 0); \text{ (ii) } g(\mathbf{x}) = \mathbf{1}_{(\phi(\mathbf{x}) \geq 0)}, \mathbf{x} \in \mathbb{R}^d,$$

where the function $\phi : \mathbb{R}^d \mapsto \mathbb{R}$ is assumed to be smooth.

We introduce the notation \mathbf{x}_{-j} to denote a vector with length $d - 1$ representing all variables other than x_j in \mathbf{x} . Abusing the notation, we define $\phi(\mathbf{x}) = \phi(x_j, \mathbf{x}_{-j})$, and for ease of presentation, we assume that, for fixed \mathbf{x}_{-j} , the function $\phi(x_j, \mathbf{x}_{-j})$ either has a simple root or is positive for all $x_j \in \mathbb{R}$. This is guaranteed by the monotonicity condition (2.2) and infinite growth condition (2.3), which are assumed for some $j \in \{1, \dots, d\}$.

$$(2.2) \quad \frac{\partial \phi}{\partial x_j}(\mathbf{x}) > 0, \forall \mathbf{x} \in \mathbb{R}^d \text{ (Monotonicity condition)}^3$$

$$(2.3) \quad \lim_{x_j \rightarrow +\infty} \phi(\mathbf{x}) = \lim_{x_j \rightarrow +\infty} \phi(x_j, \mathbf{x}_{-j}) = +\infty, \forall \mathbf{x}_{-j} \in \mathbb{R}^{d-1} \text{ or } \frac{\partial^2 \phi}{\partial x_j^2}(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathbb{R}^d \text{ (Growth condition)}.$$

Our approach can be easily extended to the case of finitely many roots without accumulation. We explain this extension in Remark 2.5.

2.1 Continuous-time formulation and optimal smoothing direction

In this section, we characterize the optimal smoothing direction using the continuous-time formulation. The purpose of this work is to approximate $\mathbb{E}[g(\mathbf{X}_T)]$ at final time T , where g is a low-regular

³We present the monotonicity condition for an increasing function without loss of generality. However, the assumption still holds for a decreasing function, which may be the case when considering a spread option.

payoff function and $\mathbf{X} := (X^{(1)}, \dots, X^{(d)})$ is described using the following SDE:⁴

$$(2.4) \quad dX_t^{(i)} = a_i(\mathbf{X}_t)dt + \sum_{j=1}^d b_{ij}(\mathbf{X}_t)dW_t^{(j)}.$$

First, we hierarchically represent $\mathbf{W} := (W^{(1)}, \dots, W^{(d)})$ as follows:

$$(2.5) \quad W^{(j)}(t) = \frac{t}{T}W^{(j)}(T) + B^{(j)}(t) = \frac{t}{\sqrt{T}}Z_j + B^{(j)}(t), \quad 1 \leq j \leq d,$$

where $\{Z_j\}_{j=1}^d$ are independent and identically distributed (i.i.d.) standard Gaussian random variables (rdvs), and $\{B^{(j)}\}_{j=1}^d$ are independent Brownian bridges.

We can hierarchically represent $\mathbf{Z} := (Z_1, \dots, Z_d)$ as

$$\mathbf{Z} = \underbrace{P_0\mathbf{Z}}_{\text{One dimensional projection}} + \underbrace{P_\perp\mathbf{Z}}_{\text{Projection on the complementary}},$$

where $P_0\mathbf{Z} := (\mathbf{Z}, \mathbf{v})\mathbf{v}$, with $\|\mathbf{v}\| = 1$, and $Z_v := (\mathbf{Z}, \mathbf{v})$ is a standard Gaussian rdv.⁵

Furthermore, defining $\mathbf{w} := \mathbf{Z} - Z_v\mathbf{v}$ yields

$$(2.6) \quad Z_j = Z_v v_j + (P_\perp\mathbf{Z})_j = Z_v v_j + w_j, \quad 1 \leq j \leq d.$$

Using (2.5) and (2.6) in (2.4) implies that

$$(2.7) \quad dX_t^{(i)} = \left(a_i(\mathbf{X}_t) + \sum_{j=1}^d b_{ij}(\mathbf{X}_t) \frac{Z_v v_j}{\sqrt{T}} \right) dt + \left(\sum_{j=1}^d b_{ij}(\mathbf{X}_t) \frac{w_j}{\sqrt{T}} \right) dt + \sum_{j=1}^d b_{ij}(\mathbf{X}_t) dB_t^{(j)}.$$

If we define $H_{\mathbf{v}}(Z_v, \mathbf{w}) := g(\mathbf{X}(T))$, then (2.6) and (2.7) can be used to yield

$$(2.8) \quad \mathbb{E}[g(\mathbf{X}(T))] = \mathbb{E}[\mathbb{E}[H_{\mathbf{v}}(Z_v, \mathbf{w}) \mid \mathbf{w}]],$$

$$(2.9) \quad \text{Var}[g(\mathbf{X}(T))] = \mathbb{E}[\text{Var}[H_{\mathbf{v}}(Z_v, \mathbf{w}) \mid \mathbf{w}]] + \text{Var}[\mathbb{E}[H_{\mathbf{v}}(Z_v, \mathbf{w}) \mid \mathbf{w}]].$$

Using (2.8) and (2.9), the optimal smoothing direction is characterized as the one that maximizes the smoothing effect at T , that is, \mathbf{v} solves the following equivalent optimization problem:

$$(2.10) \quad \min_{\substack{\mathbf{v} \in \mathbb{R}^d \\ \|\mathbf{v}\|=1}} \text{Var}[\mathbb{E}[H_{\mathbf{v}}(Z_v, \mathbf{w}) \mid \mathbf{w}]] \iff \max_{\substack{\mathbf{v} \in \mathbb{R}^d \\ \|\mathbf{v}\|=1}} \mathbb{E}[\text{Var}[H_{\mathbf{v}}(Z_v, \mathbf{w}) \mid \mathbf{w}]].$$

The left-hand side of (2.10) corresponds to reducing the variance of the original estimator by conditioning w.r.t. a specific subset of rdvs, where the best conditioning direction (i.e., leading to the least variance) depends on the choice of \mathbf{v} . Moreover, because $\mathbb{E}[g(\mathbf{X}(T))]$ is constant, the right-hand side of (2.10) can be motivated as follows: $H_{\mathbf{v}}(Z_v, \mathbf{w})$ can be understood as convoluting g with a Gaussian density whose scale parameter depends on the choice of \mathbf{v} . A larger scale parameter for the corresponding Gaussian density results in better regularity for the resulting function.

Solving (2.10) is difficult, and \mathbf{v} is dependent on the problem. In this work, we aim to heuristically determine \mathbf{v} by considering the structure of the problem. In the following section, we provide more insight on selecting \mathbf{v} and performing numerical smoothing in the time-stepping setting.

⁴We assume that $\{W^{(j)}\}_{j=1}^d$ are uncorrelated and the correlation terms are included in the diffusion terms b_{ij} .

⁵The notation (\cdot, \cdot) denotes the scalar product operator.

2.2 Motivation and idea of numerical smoothing

We consider \mathbf{X} the solution of the SDE (2.4). To illustrate our numerical smoothing idea, we consider, for ease of presentation, the discretized d -dimensional geometric Brownian motion (GBM) model given by⁶

$$(2.11) \quad dX_t^{(j)} = \sigma^{(j)} X_t^{(j)} dW_t^{(j)}, \quad 1 \leq j \leq d,$$

where $\{W^{(1)}, \dots, W^{(d)}\}$ are correlated Brownian motions with correlations ρ_{ij} , and $\{\sigma^{(j)}\}_{j=1}^d$ denote the volatilities of the different assets.

We denote by $(Z_1^{(j)}, \dots, Z_N^{(j)})$ the N standard Gaussian independent rdvs that will be used to construct the approximate path of the j -th asset $\bar{X}^{(j)}$, where N represents the number of time steps ($\Delta t = \frac{T}{N}$). Moreover, we denote by $\psi^{(j)} : (Z_1^{(j)}, \dots, Z_N^{(j)}) \mapsto (B_1^{(j)}, \dots, B_N^{(j)})$ the mapping of the Brownian bridge construction, and by $\Phi : (\Delta t, \mathbf{B}) \mapsto \bar{\mathbf{X}}^{\Delta t}(T)$ the mapping of the time-stepping scheme, where $\mathbf{B} := (B_1^{(1)}, \dots, B_N^{(1)}, \dots, B_1^{(d)}, \dots, B_N^{(d)})$ is the noncorrelated Brownian bridge⁷ and $\bar{\mathbf{X}}^{\Delta t}(T) := (\bar{X}_T^{(1)}, \dots, \bar{X}_T^{(d)})$. Then, the option price can be expressed as

$$(2.12) \quad \begin{aligned} \mathbb{E}[g(\mathbf{X}(T))] &\approx \mathbb{E}\left[g\left(\bar{X}_T^{(1)}, \dots, \bar{X}_T^{(d)}\right)\right] = \mathbb{E}\left[g\left(\bar{\mathbf{X}}^{\Delta t}(T)\right)\right] \\ &= \mathbb{E}\left[g \circ \Phi\left(B_1^{(1)}, \dots, B_N^{(1)}, \dots, B_1^{(d)}, \dots, B_N^{(d)}\right)\right] \\ &= \mathbb{E}\left[g \circ \Phi\left(\psi^{(1)}(Z_1^{(1)}, \dots, Z_N^{(1)}), \dots, \psi^{(d)}(Z_1^{(d)}, \dots, Z_N^{(d)})\right)\right] \\ &= \int_{\mathbb{R}^{d \times N}} G(z_1^{(1)}, \dots, z_N^{(1)}, \dots, z_1^{(d)}, \dots, z_N^{(d)}) \rho_{d \times N}(\mathbf{z}) dz_1^{(1)} \dots dz_N^{(1)} \dots dz_1^{(d)} \dots dz_N^{(d)}, \end{aligned}$$

where⁸ $G := g \circ \Phi \circ (\psi^{(1)}, \dots, \psi^{(d)})$ and $\rho_{d \times N}$ represents the $d \times N$ multivariate Gaussian density.

Moreover, the numerical approximation of $X^{(j)}(T)$, using the Euler–Maruyama scheme, satisfies

$$(2.13) \quad \bar{X}^{(j)}(T) = X_0^{(j)} \prod_{n=0}^{N-1} \underbrace{\left[1 + \frac{\sigma^{(j)}}{\sqrt{T}} Z_1^{(j)} \Delta t + \sigma^{(j)} \Delta B_n^{(j)}\right]}_{:= f_n^{(j)}(Z_1^{(j)}; \mathbf{Z}_{-1}^{(j)})}, \quad 1 \leq j \leq d,$$

where $\Delta B_n^{(j)} := B_{n+1}^{(j)} - B_n^{(j)}$.

Remark 2.1. Equation (2.13) holds even for stochastic volatility models, where $\sigma^{(j)}$ is a nonconstant and changes at each time step.

⁶For ease of presentation, we set the drift term in (2.11) to 0.

⁷Without loss of generality, the correlated Brownian bridge can be obtained via simple matrix multiplication.

⁸The formulation of our method is generic; for instance the mapping ψ^j may be based on Haar basis functions as in (3.1) instead of the Brownian bridges. Moreover, a different scheme for the mapping Φ may be considered instead of the Euler–Maruyama scheme used in this work.

2.2.1 Step 1 of numerical smoothing: Root finding for the discontinuity location

In this step, the discontinuity location is determined by solving the corresponding root-finding problem in one dimension after adopting suboptimal linear mapping for the coarsest factors of the Brownian increments $\mathbf{Z}_1 := (Z_1^{(1)}, \dots, Z_1^{(d)})$:

$$(2.14) \quad \mathbf{Y} = \mathcal{A}\mathbf{Z}_1,$$

where \mathcal{A} is a $d \times d$ matrix representing a linear mapping. To connect with Section 2.1, the smoothing direction \mathbf{v} is expressed using the first row of \mathcal{A} , which is generally orthogonal, selected from a family of rotations. For instance, if we consider an arithmetic basket call option, a sufficiently suitable selection of \mathcal{A} is a rotation matrix, with the first row (corresponding to the smoothing direction \mathbf{v} introduced in Section 2.1) leading to $Y_1 = \sum_{i=1}^d Z_1^{(i)}$ up to rescaling without any constraint for the remaining rows. In practice, we construct \mathcal{A} by fixing the first row to⁹ $\frac{1}{\sqrt{d}}\mathbf{1}_{1 \times d}$, and the remaining rows are obtained using the Gram-Schmidt procedure.

From (2.13), using (2.14), we obtain

$$\overline{X}^{(j)}(T) = X_0^{(j)} \prod_{n=0}^{N-1} f_n^{(j)} \left((\mathcal{A}^{-1}\mathbf{Y})_j; \mathbf{Z}_{-1}^{(j)} \right) = X_0^{(j)} \prod_{n=0}^{N-1} F_n^{(j)}(Y_1; \mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(j)}), \quad 1 \leq j \leq d,$$

where, by defining $\mathcal{A}^{\text{inv}} := \mathcal{A}^{-1}$, we have

$$F_n^{(j)}(Y_1; \mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(j)}) = \left[1 + \frac{\sigma^{(j)}\Delta t}{\sqrt{T}} A_{j1}^{\text{inv}} Y_1 + \frac{\sigma^{(j)}}{\sqrt{T}} \left(\sum_{i=2}^d A_{ji}^{\text{inv}} Y_i \right) \Delta t + \sigma^{(j)} \Delta B_n^{(j)} \right].$$

Considering that the irregularity is located at $\phi(\overline{\mathbf{X}}^{\Delta t}(T)) = 0$ (see (2.1))¹⁰ then to determine the discontinuity location $y_1^* := y_1^*(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)})$, we must find, for fixed $\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}$, the roots of $P(y_1^*)$:

$$(2.15) \quad \phi(\overline{\mathbf{X}}^{\Delta t}(T)) = \phi \left(X_0^{(1)} \prod_{n=0}^{N-1} F_n^{(1)}(y_1^*; \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}), \dots, X_0^{(d)} \prod_{n=0}^{N-1} F_n^{(d)}(y_1^*; \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(d)}) \right) := P(y_1^*) = 0.$$

We use the Newton iteration method to determine the approximated discontinuity location, $\overline{y}_1^* := \overline{y}_1^*(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)})$.

Remark 2.2. We chose \mathbf{Z}_1 for the numerical smoothing direction in (2.14) for two reasons: (i) in this work, we consider European options whose payoff functionals depend only on the assets prices at the final time T ; and (ii) the Brownian bridge construction creates a hierarchy of importance for the rdvs such that \mathbf{Z}_1 tends to be the random factor most contributing to the information in $\mathbf{X}(T)$. It may be more appropriate to consider a linear combination of $Z_1^{(j)}, \dots, Z_N^{(j)}, 1 \leq j \leq d$ for the smoothing direction when considering payoff functionals that depend on the whole path of the asset price, such as Asian options. Investigating this possibility is left for future work. Finally, the selection of \mathcal{A} creates a new hierarchy of smoothness that depends more on the payoff structure.

⁹Note that $\mathbf{1}_{1 \times d}$ denotes the row vector with dimension d , where all its coordinates are 1.

¹⁰The locations may differ depending on the considered payoff function; for instance, many payoffs in quantitative finance have kinks at the strike price.

Remark 2.3. We recall that \mathcal{A} is generally selected from a family of rotations, depending on the problem and payoff structure. The investigation of optimal choices of \mathcal{A} for various settings is left for future work, where we intend to perform a sensitivity analysis regarding possible choices.

2.2.2 Step 2 of numerical smoothing: Numerical preintegration

In this stage, we perform the numerical preintegrating step w.r.t. the direction considered for finding the root to determine y_1^* . Using Fubini's theorem and (2.12), we obtain

$$(2.16) \quad \begin{aligned} \mathbb{E}[g(\mathbf{X}(T))] &\approx \mathbb{E}\left[g\left(\overline{X}_T^{(1)}, \dots, \overline{X}_T^{(d)}\right)\right] := \mathbb{E}\left[I\left(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)}\right)\right] \\ &\approx \mathbb{E}\left[\overline{I}\left(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)}\right)\right], \end{aligned}$$

where

$$(2.17) \quad \begin{aligned} I\left(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right) &= \int_{\mathbb{R}} G\left(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right) \rho_1(y_1) dy_1 \\ &= \int_{-\infty}^{y_1^*} G\left(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right) \rho_1(y_1) dy_1 + \int_{y_1^*}^{+\infty} G\left(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right) \rho_1(y_1) dy_1, \end{aligned}$$

and \overline{I} is the approximation of I obtained using Newton iteration and the two-sided Laguerre quadrature rule, expressed as

$$(2.18) \quad \overline{I}\left(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right) := \sum_{k=0}^{M_{\text{Lag}}} \eta_k G\left(\zeta_k(\overline{y}_1^*), \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}\right),$$

where \overline{y}_1^* denotes the approximated discontinuity location and M_{Lag} represents the number of Laguerre quadrature points $\zeta_k \in \mathbb{R}$ with $\zeta_0 = \overline{y}_1^*$ and corresponding weights η_k .¹¹

The numerical smoothing treatment enables us to obtain a highly smooth integrand \overline{I} (see Section 3.1 for the smoothness analysis).

Remark 2.4 (Extending the numerical smoothing idea to other payoffs and dynamics). Although we consider the case of the multivariate GBM model to illustrate our numerical smoothing approach, we believe that this concept is generic and can be extended straightforwardly to several types of payoff functions and dynamics (see Section 4 for different tested examples).

Remark 2.5 (Extending the numerical smoothing approach to the case of multiple roots). The aforementioned preintegration step can be generalized when finitely many discontinuities exist without accumulation, occurring either because of the payoff structure or the use of the Richardson extrapolation. If we have R multiple roots, $\{y_i^*\}_{i=1}^R$ with the following order $y_1^* < y_2^* < \dots < y_R^*$,

¹¹Of course, the points ζ_k must be selected in a systematic manner depending on \overline{y}_1^* .

the smoothed integrand in (2.16) is expressed as follows:

$$\begin{aligned}
I(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) &= \int_{\mathbb{R}} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 \\
&= \int_{-\infty}^{y_1^*} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 \\
&\quad + \sum_{i=1}^{R-1} \int_{y_i^*}^{y_{i+1}^*} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 \\
(2.19) \quad &\quad + \int_{y_R^*}^{+\infty} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1,
\end{aligned}$$

and its approximation \bar{I} is given by

$$\begin{aligned}
\bar{I}(\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) &:= \sum_{k=0}^{M_{\text{Lag},1}} \eta_k^{\text{Lag}} G(\zeta_{k,1}^{\text{Lag}}(\bar{y}_1^*), \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \\
&\quad + \sum_{i=1}^{R-1} \left(\sum_{k=0}^{M_{\text{Leg},i}} \eta_k^{\text{Leg}} G(\zeta_{k,i}^{\text{Leg}}(\bar{y}_i^*, \bar{y}_{i+1}^*), \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \right) \\
&\quad + \sum_{k=0}^{M_{\text{Lag},R}} \eta_k^{\text{Lag}} G(\zeta_{k,R}^{\text{Lag}}(\bar{y}_R^*), \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}),
\end{aligned}$$

where $\{\bar{y}_i^*\}_{i=1}^R$ are the approximated discontinuity locations, $M_{\text{Lag},1}$ and $M_{\text{Lag},R}$ are the number of Laguerre quadrature points $\zeta_{\dots}^{\text{Lag}} \in \mathbb{R}$ with corresponding weights $\eta_{\dots}^{\text{Lag}}$, and $\{M_{\text{Leg},i}\}_{i=1}^{R-1}$ are the numbers of Legendre quadrature points $\zeta_{\dots}^{\text{Leg}}$ with corresponding weights $\eta_{\dots}^{\text{Leg}}$ ¹². Moreover, \bar{I} can be approximated further depending on the decay of $G \times \rho_1$ in the semi-infinite domains in (2.19) and how close the roots are. This approximation enables dealing with a countable number of discontinuities by keeping them toward infinity and then truncating the domain.

Remark 2.6. Our approach can be extended to solve a broad class of problems, particularly for estimating risk quantities and computing Greeks (sensitivities) for discontinuous financial payoffs (e.g., for the low-factor London interbank offer rate (LIBOR) market [17, 12]). The numerical smoothing idea can be used to design novel efficient methods based on the pathwise approach relying on the differentiability of the payoff. We intend to explore these directions in the future, where we should also address additional challenges related to pathwise simulation.

Example 2.7 (Simple illustration: Single digital call under the GBM model). We let $dX = \sigma X dW$, with $\sigma > 0$, and W is a Brownian motion. We consider g given by (ii) in (2.1), where $\phi(x) = x - K$, and K is the strike price. From previous section, the discontinuity is located at y_1^* , which is an invertible function satisfying $X(T; y_1^*(x), \mathbf{z}_{-1}) = x$. In this particular case, y_1^* is deterministic (it does not depend on the Brownian bridge increments) and is given by

$$y_1^* = (\log(K/X_0) + T\sigma^2/2) \frac{1}{\sqrt{T}\sigma}.$$

¹²The points $\zeta_{\dots}^{\text{Lag}}$ and $\zeta_{\dots}^{\text{Leg}}$ must be selected systematically depending on $\{\bar{y}_i^*\}_{i=1}^R$.

Then,

$$(2.20) \quad I(\mathbf{z}_{-1}) = \int_{\mathbb{R}} \mathbf{1}_{\{X(T; y, \mathbf{z}_{-1}) > K\}} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) dy = \mathbb{P}(Y > y^*(K)),$$

where $Y \sim \mathcal{N}(0, 1)$. Generally, y_1^* is not deterministic; for instance when considering the Heston model (see (4.1)) because the volatility is stochastic.

2.3 Hierarchical quadrature methods combined with numerical smoothing

After performing the numerical smoothing step, we end up with an integration problem (2.16) of a highly regular integrand \bar{I} in a $(dN - 1)$ -dimensional space (see Section 3.1 for the regularity analysis). The second stage of our approach involves approximating (2.16) efficiently. Thus, we employ the ASGQ and QMC methods (see Section 2.3.1 for a brief description of the methods in our context; refer to [6, 5] for more details). In general, given (2.15), we must to compute this root for any quadrature or QMC point. However, there are cases in which the root is deterministic (see Remark 2.7), where it is only computed once for all quadrature or QMC points.

The dimension of the integration problem (2.16) may become very large because of (i) numerous time steps N in the discretization scheme or (ii) a large number of assets, d . To overcome the high-dimensionality issue, we apply an idea similar to that introduced in [6] and combine the ASGQ and QMC methods with two hierarchical transformations. We first employ a hierarchical path generation based on the Brownian bridge construction to reduce the effective dimension and then use the Richardson extrapolation to reduce the bias and, consequently, the dimension of the integration problem. More details on these two hierarchical representations are available in [6].

2.3.1 Brief description of ASGQ and QMC methods

We denote by $\beta := (\beta_n)_{n=1}^{dN-1} \in \mathbb{N}^{dN-1}$ a multi-index and by $\bar{I}_\beta := \sum_{j=1}^{\#\mathcal{T}^{m(\beta)}} \bar{\omega}_j \bar{I}(\hat{\mathbf{y}}_j)$ the Cartesian quadrature estimator¹³ of $E[\bar{I}]$ in the tensor grid $\mathcal{T}^{m(\beta)} = \prod_{n=1}^{dN-1} \mathcal{H}^{m(\beta_n)}$, with $\mathcal{H}^{m(\beta_n)} := \{y_{\beta_n}^1, \dots, y_{\beta_n}^{m(\beta_n)}\} \subset \mathbb{R}$, $\hat{\mathbf{y}}_j \in \mathcal{T}^{m(\beta)}$ are the quadrature points and $\bar{\omega}_j$ denotes the products of the weights of the univariate rules. Then, using a construction similar to that described in [6, 5], the ASGQ estimator for approximating (2.16) using a set of multi-indices $\mathcal{I}_{\text{ASGQ}} \subset \mathbb{N}^{dN-1}$ is

$$(2.21) \quad Q^{\text{ASGQ}} := Q^{\mathcal{I}_{\text{ASGQ}}}[\bar{I}] = \sum_{\beta \in \mathcal{I}_{\text{ASGQ}}} \Delta[\bar{I}_\beta],$$

with

$$\Delta[\bar{I}_\beta] = \left(\prod_{i=1}^{dN-1} \Delta_i \right) \bar{I}_\beta, \quad \text{and} \quad \Delta_i \bar{I}_\beta := \begin{cases} \bar{I}_\beta - \bar{I}_{\beta'}, & \text{with } \beta' = \beta - \mathbf{e}_i, \text{ if } \beta_i > 0, \\ \bar{I}_\beta, & \text{otherwise,} \end{cases}$$

where \mathbf{e}_i denotes the i th $(dN - 1)$ -dimensional unit vector.

¹³The cardinality of $\mathcal{T}^{m(\beta)}$ is $\#\mathcal{T}^{m(\beta)} = \prod_{n=1}^{dN-1} m(\beta_n)$ with $m(\beta_n)$ quadrature points along the n th dimension, and $m : \mathbb{N} \rightarrow \mathbb{N}$ is a strictly increasing function with $m(0) = 0$ and $m(1) = 1$.

The construction of $\mathcal{I}_{\text{ASGQ}}$ is done a posteriori and adaptively by profit thresholding, such that $\mathcal{I}_{\text{ASGQ}} = \{\boldsymbol{\beta} \in \mathbb{N}_+^{dN-1} : P_{\boldsymbol{\beta}} \geq \bar{T}\}$, where $P_{\boldsymbol{\beta}} = \frac{|\Delta E_{\boldsymbol{\beta}}|}{\Delta \mathcal{W}_{\boldsymbol{\beta}}}$ is the profit of a hierarchical surplus, and

$$(2.22) \quad \begin{aligned} \Delta E_{\boldsymbol{\beta}} &= \left| Q^{\mathcal{I}_{\text{ASGQ}} \cup \{\boldsymbol{\beta}\}} - Q^{\mathcal{I}_{\text{ASGQ}}} \right| \quad (\text{the error contribution}) \\ \Delta \mathcal{W}_{\boldsymbol{\beta}} &= \text{Work} \left[Q^{\mathcal{I}_{\text{ASGQ}} \cup \{\boldsymbol{\beta}\}} \right] - \text{Work} \left[Q^{\mathcal{I}_{\text{ASGQ}}} \right] \quad (\text{the work contribution}). \end{aligned}$$

We also use the randomized QMC (rQMC) method based on lattice rules [32, 31], as described in Section 4.2 in [6]. The rQMC estimator is defined as follows:

$$(2.23) \quad Q^{\text{rQMC}} = \frac{1}{q} \sum_{i=0}^{q-1} \left(\frac{1}{n} \sum_{k=0}^{n-1} (\bar{T} \circ F^{-1}) \left(\frac{k\mathbf{u} + \Delta^{(i)} \bmod n}{n} \right) \right),$$

where $\{\Delta^{(i)}\}_{i=0}^{q-1}$ are q independent random shifts from the uniform distribution of $[0, 1]^{dN-1}$, $\mathbf{u} = (u_1, \dots, u_{dN-1})$ is the fixed lattice generating vector, and $F^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function. The total number of rQMC samples is $M^{\text{rQMC}} = q \times n$.

3 Smoothness Analysis and Error Discussion

3.1 Smoothness analysis

To achieve the optimal performance of the ASGQ and QMC methods, the integrand should be highly smooth. In this section, we analyze the smoothness of the integrand of interest after employing our numerical smoothing approach. First, we introduce the notation and then state the smoothness theorem, Theorem 3.4.

For simplicity, we assume that we work on a fixed time interval $[0, T]$, with $T = 1$. Using the Haar mother wavelet

$$\psi(t) := \begin{cases} 1, & 0 \leq t < \frac{1}{2}, \\ -1, & \frac{1}{2} \leq t < 1, \\ 0, & \text{else,} \end{cases}$$

we construct the Haar basis functions of $L^2([0, 1])$ by setting

$$\psi_{-1}(t) := \mathbf{1}_{[0,1]}(t); \quad \psi_{n,k}(t) := 2^{n/2} \psi(2^n t - k), \quad n \in \mathbb{N}_0, \quad k = 0, \dots, 2^n - 1.$$

The support of $\psi_{n,k}$ is $[2^{-n}k, 2^{-n}(k+1)]$. Moreover, we define a grid $\mathcal{D}^n := \{t_\ell^n \mid \ell = 0, \dots, 2^{n+1}\}$ by $t_\ell^n := \frac{\ell}{2^{n+1}}T$. The Haar basis functions up to level n are piecewise constants with points of discontinuity given by \mathcal{D}^n . Next, we define the antiderivatives of the Haar basis functions:

$$\Psi_{-1}(t) := \int_0^t \psi_{-1}(s) ds; \quad \Psi_{n,k}(t) := \int_0^t \psi_{n,k}(s) ds.$$

For an i.i.d. set of standard normal rdvs (*coefficients*) $Z_{-1}, Z_{n,k}, n \in \mathbb{N}_0, k = 0, \dots, 2^n - 1$, we define the standard Brownian motion

$$W_t := Z_{-1} \Psi_{-1}(t) + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} Z_{n,k} \Psi_{n,k}(t),$$

and the truncated version

$$(3.1) \quad W_t^N := Z_{-1} \Psi_{-1}(t) + \sum_{n=0}^N \sum_{k=0}^{2^n-1}, Z_{n,k} \Psi_{n,k}(t).$$

where W^N already coincides with W along the grid \mathcal{D}^N . We define the corresponding increments for any function or process F as follows:

$$\Delta_\ell^N F := F(t_{\ell+1}^N) - F(t_\ell^N).$$

For simplicity, we consider a one-dimensional SDE for the process X as follows:

$$(3.2) \quad dX_t = b(X_t) dW_t, \quad X_0 = x \in \mathbb{R}.$$

We assume that b and its derivatives for all orders are bounded. Recall that we want to compute, for $g : \mathbb{R} \rightarrow \mathbb{R}$, which is not necessarily smooth, $E[g(X_T)]$. Furthermore, we define the solution of the Euler–Maruyama scheme along \mathcal{D}^N by $X_0^N := X_0 = x$; for convenience, we also define $X_T^N := X_{2^N}^N$.

$$(3.3) \quad X_{\ell+1}^N := X_\ell^N + b(X_\ell^N) \Delta_\ell^N W, \quad \ell = 0, \dots, 2^N - 1.$$

The rdv X_ℓ^N is a deterministic function of the rdvs Z_{-1} and $\mathbf{Z}^N := (Z_{n,k})_{n=0, \dots, N, k=0, \dots, 2^n-1}$. Using this notation, we write

$$(3.4) \quad X_\ell^N = X_\ell^N(Z_{-1}, \mathbf{Z}^N),$$

for the appropriate (now deterministic) map $X_\ell^N : \mathbb{R} \times \mathbb{R}^{2^{N+1}-1} \rightarrow \mathbb{R}$. We write $y := z_{-1}$ and \mathbf{z}^N for the (deterministic) arguments of the function X_ℓ^N .¹⁴

We define the deterministic function $H^N : \mathbb{R}^{2^{N+1}-1} \rightarrow \mathbb{R}$, expressed as follows:

$$(3.5) \quad H^N(\mathbf{z}^N) := E[g(X_T^N(Z_{-1}, \mathbf{z}^N))].$$

Before stating the main theorem that H^N satisfies, we motivate its proof and underlying assumptions. We consider a mollified version g_δ of g and the corresponding function H_δ^N (defined by replacing g with g_δ in (3.5)). Tacitly, assuming that we can interchange the integration and differentiation (refer to Lemma A.1 for justification), we achieve

$$\frac{\partial H_\delta^N(\mathbf{z}^N)}{\partial z_{n,k}} = E \left[g'_\delta(X_T^N(Z_{-1}, \mathbf{z}^N)) \frac{\partial X_T^N(Z_{-1}, \mathbf{z}^N)}{\partial z_{n,k}} \right].$$

Multiplying and dividing by $\frac{\partial X_T^N(Z_{-1}, \mathbf{z}^N)}{\partial y}$ and replacing the expectation by an integral w.r.t. the standard normal density, we obtain

$$(3.6) \quad \frac{\partial H_\delta^N(\mathbf{z}^N)}{\partial z_{n,k}} = \int_{\mathbb{R}} \frac{\partial g_\delta(X_T^N(y, \mathbf{z}^N))}{\partial y} \left(\frac{\partial X_T^N}{\partial y}(y, \mathbf{z}^N) \right)^{-1} \frac{\partial X_T^N}{\partial z_{n,k}}(y, \mathbf{z}^N) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

¹⁴We offer a note of caution regarding the convergence as $N \rightarrow \infty$. Although the sequence of random processes X^N converges to the solution of (3.2) (under the usual assumptions on b), this is not true in any sense for deterministic functions.

If integration by parts is possible, we can discard the mollified version and obtain the smoothness of H^N because

$$\frac{\partial H^N(\mathbf{z}^N)}{\partial z_{n,k}} = - \int_{\mathbb{R}} g(X_T^N(y, \mathbf{z}^N)) \frac{\partial}{\partial y} \left[\left(\frac{\partial X_T^N}{\partial y}(y, \mathbf{z}^N) \right)^{-1} \frac{\partial X_T^N}{\partial z_{n,k}}(y, \mathbf{z}^N) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \right] dy.$$

However, there are situations in which there may be a potential problem looming in the inverse of the derivative w.r.t. y ¹⁵. This observation motivates the introduction of Assumptions 3.2 and 3.3.

Notation 3.1. For sequences of rdvs F_N , we write that $F_N = \mathcal{O}(1)$ if there exists a rrv C with finite moments of all orders, such that for all N , we have $|F_N| \leq C$ a.s.

Assumption 3.2. There are positive rdvs C_p with finite moments of all orders¹⁶ such that

$$\forall N \in \mathbb{N}, \forall \ell_1, \dots, \ell_p \in \{0, \dots, 2^N - 1\} : \left| \frac{\partial^p X_T^N}{\partial X_{\ell_1}^N \dots \partial X_{\ell_p}^N} \right| \leq C_p \text{ a.s.}$$

In terms of notation 3.1, this means that $\frac{\partial^p X_T^N}{\partial X_{\ell_1}^N \dots \partial X_{\ell_p}^N} = \mathcal{O}(1)$.

Assumption 3.2 is natural because it is fulfilled if the diffusion coefficient $b(\cdot)$ is smooth. This situation is valid for many option pricing models. Besides Assumption 3.2, we make another assumption, Assumption 3.3, which might be challenging to verify in practice for some models. In Appendix B, we explain cases with sufficient conditions where this assumption is valid.

Assumption 3.3. For any $p \in \mathbb{N}$ we obtain

$$\left(\frac{\partial X_T^N}{\partial y}(Z_{-1}, \mathbf{Z}^N) \right)^{-p} = \mathcal{O}(1).$$

We are now in a position to state Theorem 3.4 for H^N . We refer to Appendix A for its proof.

Theorem 3.4. Assume that X_T^N , defined by (3.3) and (3.4), satisfies Assumptions 3.2 and 3.3. Then, for any $p \in \mathbb{N}$ and indices n_1, \dots, n_p and k_1, \dots, k_p (satisfying $0 \leq k_j < 2^{n_j}$), the function H^N defined in (3.5) satisfies the following (with constants independent of n_j, k_j)

$$\frac{\partial^p H^N}{\partial z_{n_1, k_1} \dots \partial z_{n_p, k_p}}(\mathbf{z}^N) = \mathcal{O}\left(2^{-\sum_{j=1}^p n_j/2}\right).$$

In particular, H^N is of class C^∞ .

Remark 3.5 (Regarding the analyticity of H^N). We expect that H^N is analytic; however, the formal proof is subtle. In particular, our proof in Appendix A relies on successively applying the technique of dividing by $\frac{\partial X_T^N}{\partial y}$ and then integrating by parts. Thus, the constant in $\mathcal{O}\left(2^{-\sum_{j=1}^p n_j/2}\right)$ depends on p and increases in p . In other words, Theorem 3.4 should be interpreted as an assertion of the anisotropy in the variables $z_{n,k}$ rather than a statement on the behavior of higher derivatives

¹⁵As an example, let us assume that $X_T^N(y, \mathbf{z}^N) = \cos(y) + z_{n,k}$. Then, (3.6) is generally not integrable.

¹⁶It is probably difficult to argue that a deterministic constant C may exist in Assumption 3.2.

of H^N . Our proof reveals that the number of summands increases as $p!$. Therefore, the statement of the theorem does not already imply analyticity. This problem is an artifact of our construction, and there is no reason to assume such behavior in general. Finally, we expect the analyticity region to shrink as $N \rightarrow \infty$, which motivates the use of the Richardson extrapolation to keep N as small as possible while achieving the desired accuracy.

Remark 3.6. The analysis of the smoothness direction and sufficient conditions for Theorem 3.4 to be valid at high dimensions is an open problem and is beyond the study scope.

3.2 Error and work discussion for ASGQ combined with numerical smoothing

In this section, we analyze the errors in the proposed approach when using the ASGQ method combined with numerical smoothing. The error analysis of the QMC method combined with numerical smoothing is almost similar, as explained in Remark 3.7. Following the notation in Sections 2.2 and 2.3.1, we obtain the following error decomposition for the ASGQ estimator, Q^{ASGQ} :

$$\begin{aligned}
\text{E} [g(X(T)) - Q^{\text{ASGQ}}] &= \underbrace{\text{E} [g(X(T))] - \text{E} [g(\bar{\mathbf{X}}^{\Delta t}(T))]}_{\text{Error I: bias or weak error}} \\
&+ \underbrace{\text{E} [I(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)})] - \text{E} [\bar{I}(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)})]}_{\text{Error II: numerical smoothing error}} \\
(3.7) \quad &+ \underbrace{\text{E} [\bar{I}(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)})] - Q^{\text{ASGQ}}}_{\text{Error III: ASGQ error}}.
\end{aligned}$$

Because we use the Euler–Maruyama scheme to simulate asset dynamics, we achieve

$$(3.8) \quad \text{Error I} = \mathcal{O}(\Delta t).$$

We denote by $\text{TOL}_{\text{Newton}}$ the tolerance of the Newton method to approximate the discontinuity location by finding the roots of $P(y_1^*)$ defined in (2.15). Thus, $|P(\bar{y}_1^*)| \leq \text{TOL}_{\text{Newton}}$, and using the Taylor expansion, $(y_1^* - \bar{y}_1^*) = \mathcal{O}(\text{TOL}_{\text{Newton}})$. Consequently, Error II in (3.7) is expressed as

$$\begin{aligned}
\text{Error II} &:= \text{E} [I(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)})] - \text{E} [\bar{I}(\mathbf{Y}_{-1}, \mathbf{Z}_{-1}^{(1)}, \dots, \mathbf{Z}_{-1}^{(d)})] \\
&= E \left[\int_{-\infty}^{y_1^*} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 + \int_{y_1^*}^{+\infty} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 \right. \\
&\quad - \left(\int_{-\infty}^{\bar{y}_1^*} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 + \int_{\bar{y}_1^*}^{+\infty} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 \right) \\
&\quad + \left(\int_{-\infty}^{\bar{y}_1^*} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 + \int_{\bar{y}_1^*}^{+\infty} G(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \rho_1(y_1) dy_1 \right) \\
&\quad \left. - \sum_{k=0}^{M_{\text{Lag}}} \eta_k G(\zeta_k(\bar{y}_1^*), \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}) \right] \\
(3.9) \quad &= \mathcal{O}(\text{TOL}_{\text{Newton}}) + \mathcal{O}(M_{\text{Lag}}^{-s/2}),
\end{aligned}$$

where $s > 0$ is related to the degree of regularity of the integrand, G , w.r.t. y_1 .¹⁷

The first error contribution in (3.9) originates from the gap created by integrating G over domains separated by the approximated discontinuity location \bar{y}_1^* instead of y_1^* , which is the exact location. We consider without loss of generality that $\bar{y}_1^* < y_1^*$, then we obtain¹⁸

$$\begin{aligned} & E \left[\int_{-\infty}^{y_1^*} G \left(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)} \right) \rho_1(y_1) dy_1 - \int_{-\infty}^{\bar{y}_1^*} G \left(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)} \right) \rho_1(y_1) dy_1 \right] \\ &= E \left[\int_{\bar{y}_1^*}^{y_1^*} G \left(y_1, \mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)} \right) \rho_1(y_1) dy_1 \right] \\ &= E \left[\int_{\bar{y}_1^*}^{y_1^*} \left(G(\bar{y}_1^*; \cdot) + G'(\bar{y}_1^*; \cdot) (y_1 - \bar{y}_1^*) + \mathcal{O}((y_1 - \bar{y}_1^*)^2) \right) \rho_1(y_1) dy_1 \right] \\ &= E \left[G(\bar{y}_1^*; \cdot) (F(y_1^*) - F(\bar{y}_1^*)) + \mathcal{O}(\text{TOL}_{\text{Newton}}^2) \right] = \mathcal{O}(\text{TOL}_{\text{Newton}}), \end{aligned}$$

where $F(\cdot)$ is the standard normal cumulative distribution function. For the error bound above to hold, we assume that G and all its derivatives at \bar{y}_1^* , appearing in the constant in $\mathcal{O}(\text{TOL}_{\text{Newton}})$, are integrable.

The second error contribution in (3.9) originates from the one-dimensional preintegration step using the Laguerre quadrature, as explained in Section 2.2.2. Considering that G is a smooth function in parts of the integration domain separated by the discontinuity, we achieve a spectral convergence of the quadrature [29], justifying the term $M_{\text{Lag}}^{-s/2}$. For the error bound in (3.9) to hold, we assume that the constant in $\mathcal{O}(M_{\text{Lag}}^{-s/2})$, which depends on $\mathbf{y}_{-1}, \mathbf{z}_{-1}^{(1)}, \dots, \mathbf{z}_{-1}^{(d)}$, is integrable.

Finally, considering M_{ASGQ} quadrature points used in the ASGQ method, we achieve

$$(3.10) \quad \text{Error III} = \mathcal{O}\left(M_{\text{ASGQ}}^{-p/2}\right),$$

where the bound in (3.10) is justified by the analysis in [13, 16] and $p := p(N, d) > 0$ is related to the degree of regularity of \bar{I} , as defined in (2.16) and (2.17), in the $(dN - 1)$ -dimensional space.¹⁹ In this case, our smoothness analysis (Section 3.1) implies that $p \gg 1$, under the assumption that \bar{I} converges to I for large values of M_{Lag} and small $\text{TOL}_{\text{Newton}}$. Nevertheless, the optimal performance for ASGQ can deteriorate (i) if p and s are not sufficiently large, or (ii) due to the adverse effect of the high dimension that may severely affect the rates. Finally, although we work in the preasymptotic regime (small number of time steps, N), the regularity parameter p may deteriorate when increasing the dimension of the integration problem by increasing N , justifying the use of Richardson extrapolation.

Considering (3.7), (3.8), (3.9) and (3.10), the total error estimate of our approach is

$$(3.11) \quad \mathcal{E}_{\text{total, ASGQ}} := E[g(X(T))] - Q^{\text{ASGQ}} = \mathcal{O}(\Delta t) + \mathcal{O}\left(M_{\text{ASGQ}}^{-p/2}\right) + \mathcal{O}\left(M_{\text{Lag}}^{-s/2}\right) + \mathcal{O}(\text{TOL}_{\text{Newton}}).$$

¹⁷For the parts of the domain separated by the discontinuity location, the derivatives of G w.r.t. y_1 are bounded up to order s .

¹⁸A similar argument holds for the part of G located on the right of the discontinuity.

¹⁹We refer to [13, 16] for a clear characterization of p .

To achieve optimal performance, we need to optimize the parameters in (3.11) to satisfy a certain error tolerance, TOL, with the least amount of work, which can be achieved by solving (3.12):

$$(3.12) \quad \begin{cases} \min_{(M_{\text{ASGQ}}, M_{\text{Lag}}, \text{TOL}_{\text{Newton}})} \text{Work}_{\text{ASGQ}} \propto \Delta t^{-1} \times M_{\text{ASGQ}} \times M_{\text{Lag}} \\ \text{s.t. } \mathcal{E}_{\text{total,ASGQ}} = \text{TOL}. \end{cases}$$

We do not solve (3.12) in our experiments in Section 4 (we select the parameters heuristically to achieve a suboptimal performance). However, in Appendix C we reveal that, for a given error tolerance TOL, under certain conditions of the regularity parameters s and p ($p, s \gg 1$), a lower bound on the computational work of the ASGQ method is of order $\text{Work}_{\text{ASGQ}} = \mathcal{O}(\text{TOL}^{-1})$. This complexity is significantly better than $\mathcal{O}(\text{TOL}^{-3})$ achieved by the MC method.

Remark 3.7 (On the error of the QMC method combined with numerical smoothing). Let Q^{rQMC} denote the rQMC estimator to approximate $\mathbb{E}[\bar{I}]$ in (2.16) with M_{rQMC} samples. Then, we achieve an error decomposition similar to that in (3.7), with Error III being the rQMC statistical error in this case [30], and expressed as follows:

$$\text{Error III (rQMC error)} = \mathcal{O}\left(M_{\text{rQMC}}^{-\frac{1}{2}-\delta} (\log M_{\text{rQMC}})^{d \times N-1}\right),$$

where $0 \leq \delta \leq \frac{1}{2}$ is related to the degree of regularity of \bar{I} , defined in (2.18).

Moreover, the analysis of QMC with randomly shifted lattice rules in [33, 15] indicates that convergence rates close to the optimal rates $\mathcal{O}\left(M_{\text{rQMC}}^{-1}\right)$ can be observed if our integrand, $\bar{I}(\cdot)$, belongs to the $(dN - 1)$ -dimensional weighted Sobolev space of functions with square-integrable mixed first derivatives, $\mathcal{W}_{dN-1, \gamma}$,²⁰ where $\gamma := \{\gamma_{\alpha} > 0 : \alpha \subseteq \{1, 2, \dots, dN - 1\}\}$ is a given collection of weights.

Remark 3.8. Although we did not use the Richardson extrapolation in the previous analysis, this hierarchical representation improves the complexity rate of the ASGQ method (as observed in our numerical experiments in Section 4).

Remark 3.9. As an alternative method to approximate (2.12), we can use the multilevel MC (MLMC) method, which also benefits from the numerical smoothing (see [7]) in terms of complexity and robustness, where we recover complexities obtained for Lipschitz functionals, $\mathcal{O}(\text{TOL}^{-2}(\log(\text{TOL}))^2)$ when using the Euler–Maruyama scheme with numerical smoothing. The comparison of MLMC and deterministic quadrature methods, such as ASGQ, is not straightforward and is problem-dependent because there is a compromise between the regularity class of the integrand and the anisotropy w.r.t. the different dimensions. We intend to conduct this systematic comparison in future work. We also plan to explore the idea of numerical smoothing with multilevel QMC [20], where we can profit from the good features of QMC and MLMC in this setting.

²⁰ $\mathcal{W}_{dN-1, \gamma}$ is equipped with the (unanchored) norm $\|f\|_{\mathcal{W}_{dN-1, \gamma}}^2 = \sum_{\alpha \subseteq \{1:dN-1\}} \frac{1}{\gamma_{\alpha}} \int_{[0,1]^{|\alpha|}} \left(\int_{[0,1]^{dN-1-|\alpha|}} \frac{\partial^{|\alpha|}}{\partial \mathbf{y}_{-\alpha}} f(\mathbf{y}) d\mathbf{y}_{-\alpha} \right)^2 d\mathbf{y}_{\alpha}$, where $\mathbf{y} := (y_j)_{j \in \alpha}$ and $\mathbf{y}_{-\alpha} := (y_j)_{j \in \{1:dN-1\} \setminus \alpha}$.

4 Numerical Experiments

We conduct experiments using three examples of payoffs: a single-asset digital option, a single-asset call option, and a four-asset arithmetic basket call option.²¹ These examples are tested under two dynamics for the asset price: the discretized GBM model (a didactic example) and the Heston model, which is a relevant application of our approach (discretization is required). Table 4.1 lists the specifications of each example. Further details of the models and discretization schemes are described in Section 4.1. Sections 4.2 and 4.3 demonstrate the advantage of combining numerical smoothing with the ASGQ and rQMC methods over the ASGQ and rQMC without smoothing. In Section 4.4, we study the effect of the numerical smoothing parameters on the numerical smoothing error and consequently on the quadrature error of the ASGQ method. Finally, Section 4.5 compares the MC and ASGQ methods in terms of errors and computational times. Our ASGQ implementation was based on <https://sites.google.com/view/sparse-grids-kit>.

Example	Parameters	Reference solution
Single-asset digital option under GBM	$\sigma = 0.4, r = 0, T = 1, S_0 = K = 100$	0.42074
Single-asset digital option under Heston	$v_0 = 0.04, \mu = 0, \rho = -0.9, \kappa = 1, \xi = 0.1,$ $\theta = 0.0025, S_0 = K = 100$	0.5146 ($2.0e-05$)
Single-asset call option under GBM	$\sigma = 0.4, r = 0, T = 1, S_0 = K = 100$	15.8519
Single-asset call option under Heston	$v_0 = 0.04, \mu = 0, \rho = -0.9, \kappa = 1, \xi = 0.1,$ $\theta = 0.0025, S_0 = K = 100$	6.33254
4-asset basket call option under GBM	$\sigma_{1,2,3,4} = 0.4, \rho = 0.3, r = 0, T = 1,$ $S_0^{1,2,3,4} = K = 100, c_{1,2,3,4} = 1/4$	11.04 ($1.0e-03$)

Table 4.1: Model and option parameters of the tested examples with their reference solutions. The reference solution for the call option under the Heston model is computed using Premia software using the method in [25]. The numbers between parentheses correspond to the statistical error estimates when the reference solution is estimated using the MC estimator.

4.1 Experiments setting

Regarding the numerical experiments under the GBM model, the assets dynamics follow (2.11) and are simulated using the Euler–Maruyama scheme. Moreover, we test options under the Heston model [25, 10, 27, 2], providing the following dynamics:

$$(4.1) \quad \begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dW_t^S = \mu S_t dt + \rho \sqrt{v_t} S_t dW_t^v + \sqrt{1 - \rho^2} \sqrt{v_t} S_t dW_t \\ dv_t &= \kappa(\theta - v_t) dt + \xi \sqrt{v_t} dW_t^v, \end{aligned}$$

where S_t denotes the asset price, v_t represents the instantaneous variance, (W_t^S, W_t^v) are the correlated Wiener processes with correlation ρ , μ represents the asset’s rate of return, θ is the mean variance, κ is the rate at which v_t reverts to θ , and ξ denotes the volatility of the volatility.

Many simulation schemes of (4.1) have been proposed in the literature. These methods primarily differ in how they simulate the volatility process to ensure positivity. Appendix D provides an overview of the most popular methods in this context.

²¹The payoff g is expressed by $g(\mathbf{x}) = \max\left(\sum_{j=1}^d c_j x^{(j)} - K, 0\right)$, where $\{c_j\}_{j=1}^d$ denote the weights of the basket.

The ASGQ and rQMC methods are extremely sensitive to the smoothness of the integrand. In particular, we numerically found (Appendix D.4) that using a nonsmooth transformation to ensure the positivity of the volatility process deteriorates the performance of the ASGQ method. To overcome this undesirable feature, we propose using an alternative scheme, namely, the Heston Ornstein–Uhlenbeck (OU)-based scheme, in which the volatility is simulated as the sum of the OU or Bessel processes (Appendix D.3). In the literature [2, 28, 1], the focus has been on designing schemes that ensure the positivity of the volatility process and exhibit a good weak error behavior. In our setting, an optimal scheme is determined based on two criteria: (i) the behavior of the rates of mixed differences, which is an important feature for ensuring the optimal performance of the ASGQ method (see Appendix D.4.2 for more details), and (ii) the weak error behavior to apply the Richardson extrapolation when necessary. Comparing the different schemes (see Appendices D.4.1 and D.4.2) suggests that the Heston OU-based scheme yields the best results based on our criteria. Therefore, we used this scheme with the ASGQ and rQMC methods in our numerical experiments. For the MC method, we used the full truncation scheme (explained in Appendix D.1).

Remark 4.1. In this work, our primary focus is the numerical smoothing idea with the implied additional regularity for the quantity of interest and its benefits on the performance of ASGQ and QMC. As a byproduct, for the examples under the Heston model, we numerically found that using a nonsmooth transformation to ensure the positivity of the volatility process deteriorates the performance of ASGQ even after applying the numerical smoothing because it affects the path regularity of the process. To overcome this undesirable feature for the parameters settings that we consider ($4\kappa\theta/\xi^2$ is an integer), we suggest using the Heston OU-based scheme as an alternative. We expect our observations to still be valid for cases with tiny perturbations of the Heston model parameters ($4\kappa\theta/\xi^2$ is very close to an integer). An extensive analysis of the proposed Heston OU-based scheme is left for future work, where we plan to conduct a systematic investigation of its performance compared to the popular existing schemes, in the same spirit as [28], and examine various challenging settings of model parameters (e.g., when $\xi^2 \gg 4\kappa\theta$ and $4\kappa\theta/\xi^2$ is not an integer). The above observations suggest that, besides smoothing out the observable, the regularity of the discretization scheme is also essential. Numerical smoothing works perfectly if the scheme has sufficient path regularity. Otherwise, besides numerical smoothing, a smooth discretization scheme must preserve the process path regularity to ensure the optimal performance of ASGQ and QMC.

4.2 Comparison of the ASGQ method with and without numerical smoothing

This section illustrates the advantage of combining numerical smoothing with the ASGQ method. Figures 4.1 and 4.2 show comparisons of the relative quadrature error convergence for the examples under the Heston model in Table 4.1, with and without the Richardson extrapolation.²² Numerical smoothing significantly improves the quadrature error convergence for all cases, which agrees with Theorem 3.4. For instance, for the call option under the Heston model (left plot in Figure 4.2), the ASGQ method without smoothing cannot achieve a relative quadrature error below 10%, even in the case of more than 10^3 quadrature points. Alternatively, the ASGQ method with numerical smoothing achieves a relative quadrature error below 1% with the same number of quadrature points. The gains are more evident in the digital option case than in the call option case (Figure 4.1). Further, using the Richardson extrapolation, the ASGQ method with numerical smoothing

²²The dimension of the integration problem is N for the GBM examples and $2N$ for the Heston examples.

yields a smaller quadrature error. For all cases of options and models, with or without Richardson extrapolation, we observe that numerical smoothing also reduces the constant in the quadrature error besides improving the convergence rate. This observation can be explained by the analysis in [13], indicating that the constant in the error estimate depends on the weighted sum of the mixed derivatives of the integrand. From this perspective, the numerical smoothing enables a faster decay of mixed derivatives than the case without smoothing (see Proposition 3.4 in [13]).

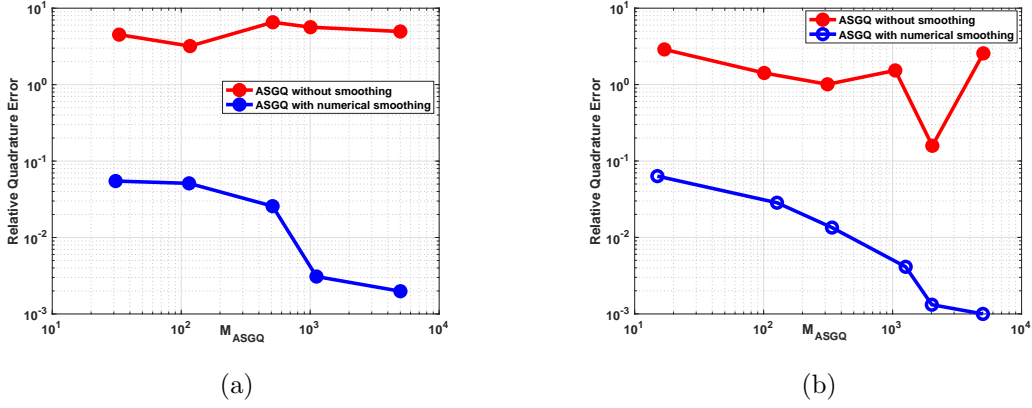


Figure 4.1: Digital option under the Heston model: Comparison of the relative quadrature error convergence for the ASGQ method with and without numerical smoothing. (a) Without the Richardson extrapolation ($N = 8$), and (b) with the Richardson extrapolation ($N_{\text{fine level}} = 8$).

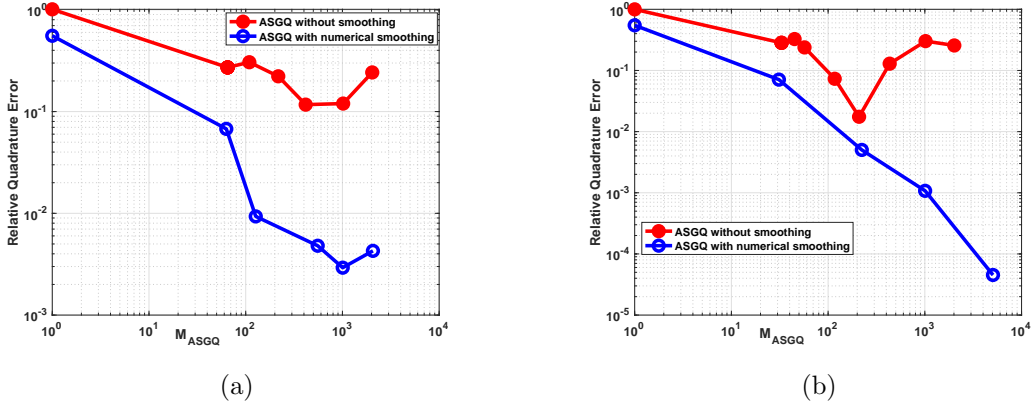


Figure 4.2: Call option under the Heston model: Comparison of the relative quadrature error convergence for the ASGQ method with and without numerical smoothing. (a) Without the Richardson extrapolation ($N = 16$), and (b) with the Richardson extrapolation ($N_{\text{fine level}} = 8$).

4.3 Comparison of the rQMC method with and without numerical smoothing

In this section, we demonstrate the advantage of combining numerical smoothing with the rQMC method. Figures 4.3 and 4.4 display comparisons of the statistical error convergence for the ex-

amples in Table 4.1 and for some number of time steps N . Because regularity was regained using numerical smoothing, an improvement in the statistical error convergence of the rQMC method occurs, which agrees with Remark 3.7 (see left plots in Figures 4.3 and 4.4). For the Heston and GBM models, the convergence rate for QMC was improved more significantly for the digital option than the call option payoff.

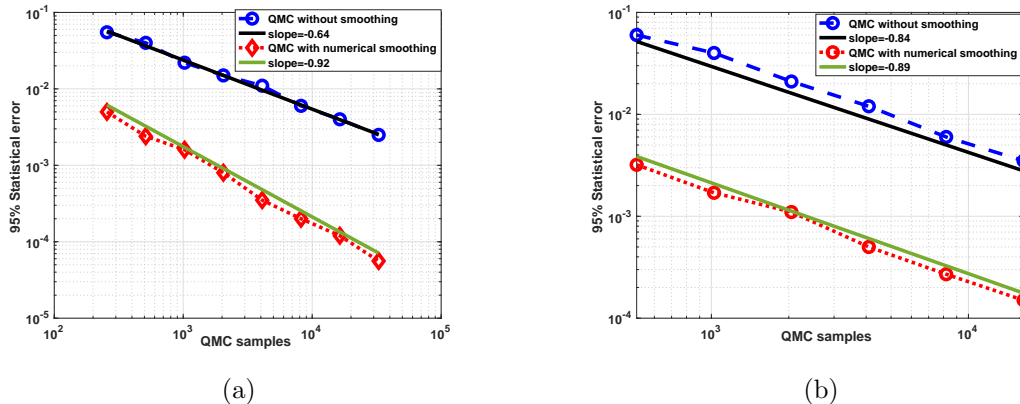


Figure 4.3: Comparison of the 95% statistical error convergence for rQMC with and without numerical smoothing with $N = 8$. (a) Digital option under GBM, (b) call option under GBM.

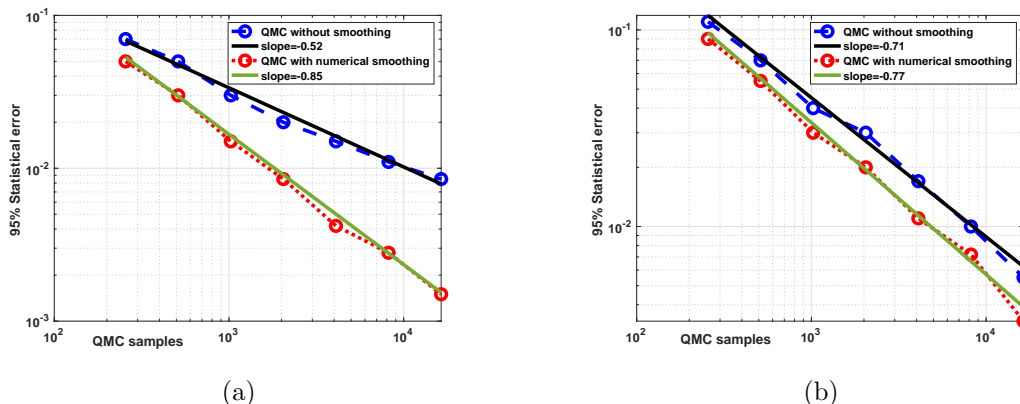


Figure 4.4: Comparison of the 95% statistical error convergence for rQMC with and without numerical smoothing with $N = 4$. (a) Digital option under Heston, (b) call option under Heston.

4.4 Study of the numerical smoothing parameters

We study the effect of the smoothing parameters on the relative numerical smoothing error for sufficiently large ASGQ points $M_{ASGQ} = 10^3$. These parameters are (i) the number of Laguerre points in the preintegration step, M_{Lag} , and (ii) the Newton tolerance in the root-finding step, TOL_{Newton} . Figures 4.5 and 4.6 present the digital and call option results under the GBM model in Table 4.1, for the case without the Richardson extrapolation and when $N = 4$. These plots

show that a faster convergence of the root-finding and quadrature errors can be achieved for the call option compared to the digital option. Moreover, we observe that the numerical smoothing procedure is cheap because few Laguerre quadrature points and large values of Newton tolerance are required to achieve a certain accuracy. Similar observations have been obtained for other examples.

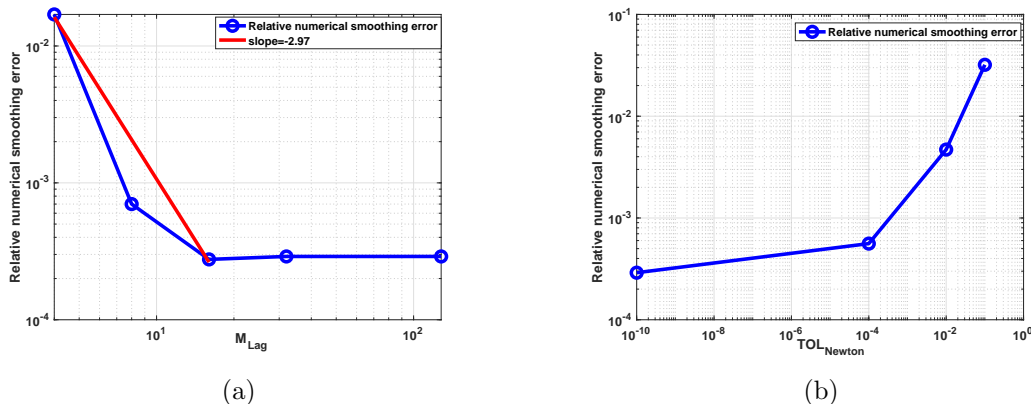


Figure 4.5: Digital option under GBM with $N = 4$: The relative numerical smoothing error for a fixed number of ASGQ points $M_{ASGQ} = 10^3$ plotted against (a) different values of M_{Lag} with a fixed Newton tolerance $TOL_{Newton} = 10^{-10}$, (b) different values of TOL_{Newton} with a fixed number of Laguerre quadrature points $M_{Lag} = 128$.

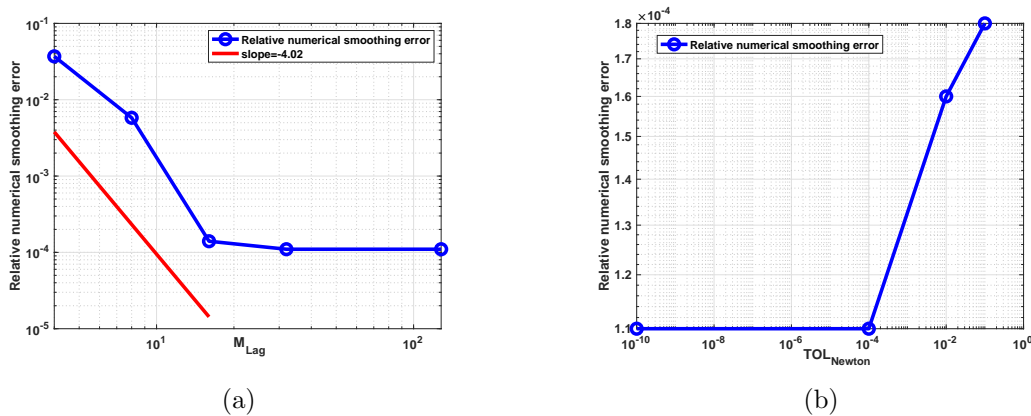


Figure 4.6: Call option under GBM with $N = 4$: The relative numerical smoothing error for a fixed number of ASGQ points $M_{ASGQ} = 10^3$ plotted against (a) different values of M_{Lag} with a fixed Newton tolerance $TOL_{Newton} = 10^{-10}$, (b) different values of TOL_{Newton} with a fixed number of Laguerre quadrature points $M_{Lag} = 128$.

4.5 ASGQ method with numerical smoothing versus the MC method

For a sufficiently fixed small error tolerance in the price estimates, we compare the computational time needed for the MC method and the ASGQ method with numerical smoothing to meet the desired error tolerance. The reported errors are relative errors normalized using the reference solutions. Furthermore, we conduct our numerical experiments for two scenarios: without the Richardson extrapolation, and with level-1 Richardson extrapolation. The actual work (runtime) is obtained using a 3,2 GHz 8-Core Intel Xeon W architecture.

The numerical findings are summarized in Table 4.2. The reported results highlight the computational gains achieved using the ASGQ method with numerical smoothing compared to the MC method to meet a relative error below 1%. These results correspond to the best configuration with the Richardson extrapolation for each method. More details for each case are provided in Figures 4.7, 4.8 and 4.9, comparing the numerical complexity of each method under the two Richardson extrapolation scenarios. These figures illustrate that, to achieve a relative error of less than 1%, the optimal configuration is the level-1 Richardson extrapolation for both the MC and ASGQ methods, except for the four-asset basket call option under the GBM model.

Example	Total relative error	CPU time (ASGQ/MC) in %
Single-asset digital option (GBM)	0.4%	0.2%
Single-asset call option (GBM)	0.5%	0.3%
Single-asset digital option (Heston)	0.4%	3.2%
Single-asset call option (Heston)	0.5%	0.4%
4-asset basket call option (GBM)	0.8%	7.4%

Table 4.2: Summary of the relative errors and computational gains achieved using ASGQ with numerical smoothing compared to the MC method, to realize a certain error tolerance. The CPU time ratios are computed for the best configuration with Richardson extrapolation for each method.

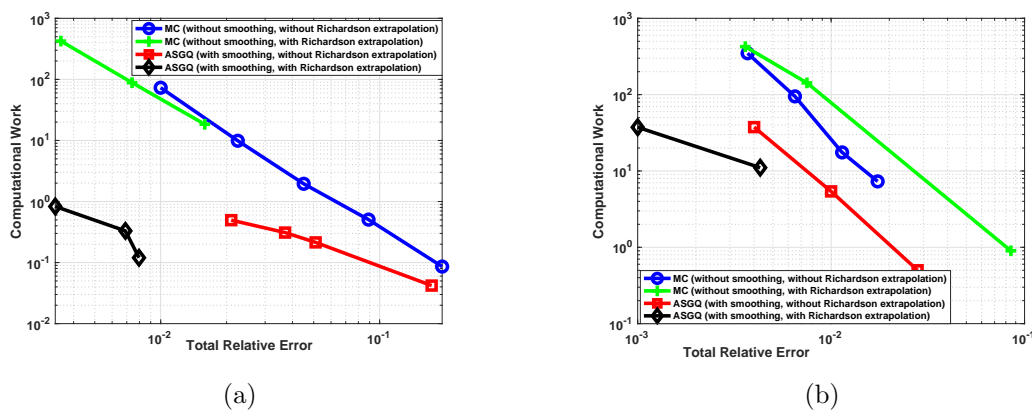


Figure 4.7: Computational work comparison for the different methods with the various configurations in terms of the Richardson extrapolation level. To achieve a relative error below 1%, ASGQ combined with numerical smoothing and level-1 Richardson extrapolation significantly outperforms the other methods. (a) Digital option under GBM, (b) digital option under Heston.

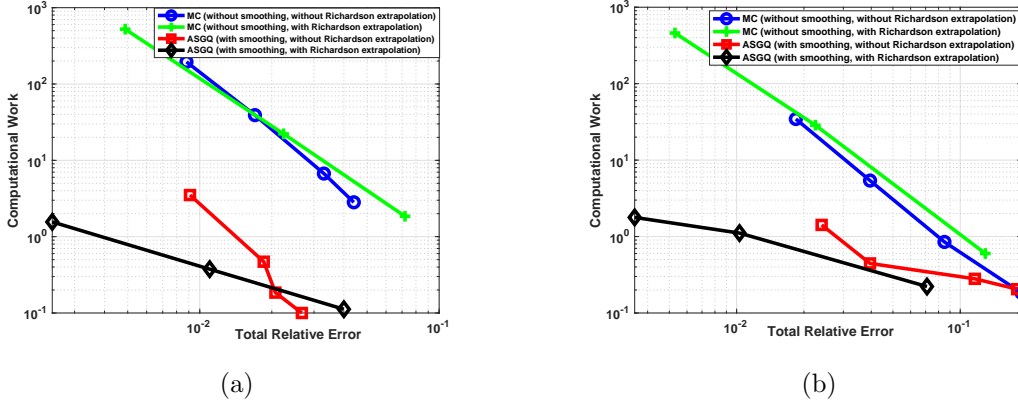


Figure 4.8: Computational work comparison for the different methods with the various configurations in terms of the Richardson extrapolation level. To achieve a relative error below 1%, ASGQ combined with numerical smoothing and level-1 Richardson extrapolation significantly outperforms the other methods. (a) Call option under GBM, (b) call option under Heston.

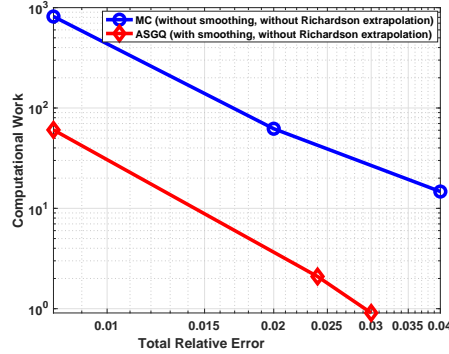


Figure 4.9: Four-asset basket call option under GBM: Computational work comparison for the different methods. To achieve a relative error below 1%, ASGQ combined with numerical smoothing significantly outperforms the MC method.

Remark 4.2 (Regarding rQMC with numerical smoothing). We also combined numerical smoothing with the rQMC method and observed an improvement in the performance compared to the case without smoothing (Section 4.3). Moreover, the rQMC method with numerical smoothing consistently outperforms the MC method to achieve a relative error below 1%. However, we consistently observe that the ASGQ method outperforms the rQMC method in all our numerical examples, when both are combined with numerical smoothing. In particular, as an illustration, Figure 4.10 shows the comparison for the example of the digital option under the GBM model.

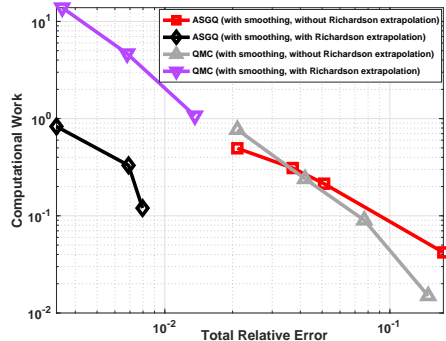


Figure 4.10: Digital option under GBM: Computational work comparison of rQMC and ASGQ, both combined with numerical smoothing, with different Richardson extrapolation level configurations. To achieve a relative error below 1%, ASGQ combined with numerical smoothing and level-1 Richardson extrapolation significantly outperforms the other methods.

Acknowledgments C. Bayer gratefully acknowledges support from the German Research Foundation (DFG) via the Cluster of Excellence MATH+ (project AA4-2) and the individual grant BA5484/1. This publication is based on work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-2019-CRG8-4033 and the Alexander von Humboldt Foundation. The authors are incredibly grateful to the anonymous referees for their valuable comments and suggestions that greatly contributed to shaping the final version of the paper.

References Cited

- [1] Aurélien Alfonsi. High order discretization schemes for the CIR process: application to affine term structure and Heston models. *Mathematics of Computation*, 79(269):209–237, 2010.
- [2] Leif Andersen. Efficient simulation of the Heston stochastic volatility model. *Available at SSRN 946405*, 2007.
- [3] Leif Andersen and Rupert Brotherton-Ratcliffe. Extended libor market models with stochastic volatility. *Available at SSRN 294853*, 2001.
- [4] Volker Barthelmann, Erich Novak, and Klaus Ritter. High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mathematics*, 12(4):273–288, 2000.
- [5] Christian Bayer, Chiheb Ben Hammouda, Antonis Papapantoleon, Michael Samet, and Raúl Tempone. Optimal damping with hierarchical adaptive quadrature for efficient Fourier pricing of multi-asset options in Lévy models. *arXiv preprint arXiv:2203.08196*, 2022.
- [6] Christian Bayer, Chiheb Ben Hammouda, and Raúl Tempone. Hierarchical adaptive sparse grids and quasi-Monte Carlo for option pricing under the rough Bergomi model. *Quantitative Finance*, 20(9):1457–1473, 2020.

- [7] Christian Bayer, Chiheb Ben Hammouda, and Raúl Tempone. Multilevel Monte Carlo combined with numerical smoothing for robust and efficient option pricing and density estimation. *arXiv preprint arXiv:2003.05708*, 2020.
- [8] Christian Bayer, Markus Siebenmorgen, and Raúl Tempone. Smoothing the payoff for efficient computation of basket option pricing. *Quantitative Finance*, 18(3):491–505, 2018.
- [9] Chiheb Ben Hammouda. *Hierarchical Approximation Methods for Option Pricing and Stochastic Reaction Networks*. PhD thesis, 2020.
- [10] Mark Broadie and Özgür Kaya. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations research*, 54(2):217–231, 2006.
- [11] Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta numerica*, 13:147–269, 2004.
- [12] Jiun Hong Chan and Mark Joshi. Fast Monte Carlo Greeks for financial products with discontinuous pay-offs. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 23(3):459–495, 2013.
- [13] Peng Chen. Sparse quadrature for high-dimensional integration with Gaussian measure. *ESAIM: Mathematical Modelling and Numerical Analysis*, 52(2):631–657, 2018.
- [14] Christophe De Luigi, Jérôme Lelong, and Sylvain Maire. Robust adaptive numerical integration of irregular functions with applications to basket and other multi-dimensional exotic options. *Applied Numerical Mathematics*, 100:14–30, 2016.
- [15] Josef Dick, Frances Y Kuo, and Ian H Sloan. High-dimensional integration: the quasi-Monte Carlo way. *Acta Numerica*, 22:133–288, 2013.
- [16] Oliver G Ernst, Bjorn Sprungk, and Lorenzo Tamellini. Convergence of sparse collocation for functions of countably many Gaussian random variables (with application to elliptic pdes). *SIAM Journal on Numerical Analysis*, 56(2):877–905, 2018.
- [17] Christian P Fries and Mark S Joshi. Conditional analytic Monte-Carlo pricing scheme of auto-callable products. *Available at SSRN 1125725*, 2008.
- [18] Thomas Gerstner and Michael Griebel. Numerical integration using sparse grids. *Numerical algorithms*, 18(3):209–232, 1998.
- [19] Michael B Giles, Tigran Nagapetyan, and Klaus Ritter. Multilevel Monte Carlo approximation of distribution functions and densities. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):267–295, 2015.
- [20] Michael B Giles and Benjamin J Waterhouse. Multilevel quasi-Monte Carlo path simulation. *Advanced Financial Modelling, Radon Series on Computational and Applied Mathematics*, 8:165–181, 2009.
- [21] Michael Griebel, Frances Kuo, and Ian Sloan. The smoothing effect of integration in \mathbb{R}^d and the ANOVA decomposition. *Mathematics of Computation*, 82(281):383–400, 2013.

- [22] Michael Griebel, Frances Kuo, and Ian Sloan. Note on “the smoothing effect of integration in \mathbb{R}^d and the ANOVA decomposition”. *Mathematics of Computation*, 86(306):1847–1854, 2017.
- [23] Andreas Griewank, Frances Y Kuo, Hernan Leövey, and Ian H Sloan. High dimensional integration of kinks and jumps-smoothing by preintegration. *Journal of Computational and Applied Mathematics*, 344:259–274, 2018.
- [24] Abdul-Lateef Haji-Ali, Fabio Nobile, Lorenzo Tamellini, and Raúl Tempone. Multi-index stochastic collocation for random PDEs. *Computer Methods in Applied Mechanics and Engineering*, 306:95–122, 2016.
- [25] Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.
- [26] Monique Jeanblanc, Marc Yor, and Marc Chesney. *Mathematical methods for financial markets*. Springer Science & Business Media, 2009.
- [27] Christian Kahl and Peter Jäckel. Fast strong approximation Monte Carlo schemes for stochastic volatility models. *Quantitative Finance*, 6(6):513–536, 2006.
- [28] Roger Lord, Remmert Koekoek, and Dick Van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2):177–194, 2010.
- [29] G Mastroianni and G Monegato. Error estimates for Gauss-Laguerre and Gauss-Hermite quadrature formulas. In *Approximation and Computation: A Festschrift in Honor of Walter Gautschi*, pages 421–434. Springer, 1994.
- [30] Harald Niederreiter. *Random number generation and quasi-Monte Carlo methods*, volume 63. Siam, 1992.
- [31] Dirk Nuyens. The construction of good lattice rules and polynomial lattice rules. In *Uniform distribution and quasi-Monte Carlo methods*, pages 223–256. De Gruyter, 2014.
- [32] Ian H. Sloan. Lattice methods for multiple integration. *Journal of Computational and Applied Mathematics*, 12-13:131–143, 1985.
- [33] Ian H Sloan and Henryk Woźniakowski. When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *Journal of Complexity*, 14(1):1–33, 1998.
- [34] Ye Xiao and Xiaoqun Wang. Conditional quasi-Monte Carlo methods and dimension reduction for option pricing and hedging with discontinuous functions. *Journal of Computational and Applied Mathematics*, 343:289–308, 2018.

A Details of the Proof of Theorem 3.4 in Section 3.1

In this section, we state and prove the theoretical results for the proof of Theorem 3.4 in Section 3.1. We use the same notation as in Section 3.1. In particular, we recall that X_T^N denotes the numerical solution of the SDE (3.2) using the Euler–Maruyama scheme along the grid \mathcal{D}^N .

Lemma A.1. *If Assumption 3.2 holds, we have the following:*

$$\frac{\partial X_T^N}{\partial z_{n,k}}(Z_{-1}, \mathbf{Z}^N) = 2^{-n/2+1} \mathcal{O}(1)$$

in the sense that the $\mathcal{O}(1)$ term does not depend on n or k .

Proof. First, Assumption 3.2 implies that $\frac{\partial X_T^N}{\partial \Delta_\ell^N W} = \mathcal{O}(1)$. Indeed,

$$\frac{\partial X_T^N}{\partial \Delta_\ell^N W} = \frac{\partial X_T^N}{\partial X_{\ell+1}^N} \frac{\partial X_{\ell+1}^N}{\partial \Delta_\ell^N W} = \mathcal{O}(1) b(X_\ell^N) = \mathcal{O}(1).$$

Next, we must identify the increments Δ_ℓ^N that depend on $Z_{n,k}$. This is the case if and only if (iff) the support of $\psi_{n,k}$ has a nonempty intersection with $]t_\ell^N, t_{\ell+1}^N[$. Explicitly, this means that

$$\ell 2^{-(N-n+1)} - 1 < k < (\ell + 1) 2^{-(N-n+1)}.$$

If we fix N , k , and n , the derivative of $\Delta_\ell^N W$ w.r.t. $Z_{n,k}$ does not vanish iff $2^{N-n+1}k \leq \ell < 2^{N-n+1}(k+1)$ because

$$(A.1) \quad \left| \frac{\partial \Delta_\ell^N W}{\partial Z_{n,k}} \right| = |\Delta_\ell^N \Psi_{n,k}| \leq 2^{-(N-n/2)}.$$

Thus, we obtain the following:

$$(A.2) \quad \frac{\partial X_T^N}{\partial z_{n,k}}(Z_{-1}, \mathbf{Z}^N) = \sum_{\ell=2^{N-n+1}k}^{2^{N-n+1}(k+1)-1} \frac{\partial X_T^N}{\partial \Delta_\ell^N W} \frac{\partial \Delta_\ell^N W}{\partial Z_{n,k}} = 2^{N-n+1} 2^{-(N-n/2)} \mathcal{O}(1) = 2^{-n/2+1} \mathcal{O}(1). \quad \square$$

Lemma A.2. *If Assumption 3.2 holds, then similar to Lemma A.1, we have*

$$\frac{\partial^2 X_T^N}{\partial y \partial z_{n,k}}(Z_{-1}, \mathbf{Z}^N) = 2^{-n/2+1} \mathcal{O}(1).$$

Proof. $\Delta_\ell^N W$ is a linear function in Z_{-1} and \mathbf{Z}^N , implying that all mixed derivatives $\frac{\partial^2 \Delta_\ell^N W}{\partial Z_{n,k} \partial Z_{-1}}$ vanish. From equation (A.2) we hence obtain

$$\frac{\partial^2 X_T^N}{\partial z_{n,k} \partial y}(Z_{-1}, \mathbf{Z}^N) = \sum_{\ell=2^{N-n+1}k}^{2^{N-n+1}(k+1)-1} \frac{\partial^2 X_T^N}{\partial \Delta_\ell^N W \partial Z_{-1}} \frac{\partial \Delta_\ell^N W}{\partial Z_{n,k}}.$$

Further,

$$\frac{\partial^2 X_T^N}{\partial \Delta_\ell^N W \partial Z_{-1}} = \sum_{j=0}^{2^{N+1}-1} \frac{\partial^2 X_T^N}{\partial \Delta_\ell^N W \partial \Delta_j^N W} \frac{\partial \Delta_j^N W}{\partial Z_{-1}}.$$

Note that

$$(A.3) \quad \frac{\partial^2 X_T^N}{\partial \Delta_\ell^N W \partial \Delta_j^N W} = \frac{\partial^2 X_T^N}{\partial X_{\ell+1}^N \partial X_{j+1}^N} b(X_\ell^N) b(X_j^N) + \mathbb{1}_{j < \ell} \frac{\partial X_T^N}{\partial X_\ell^N} b'(X_\ell^N) \frac{\partial X_\ell^N}{\partial X_{j+1}^N} b(X_j^N) = \mathcal{O}(1)$$

using Assumption 3.2. We also have $\frac{\partial \Delta_j^N W}{\partial Z_{-1}} = \mathcal{O}(2^{-N})$, implying the statement of the lemma. \square

Remark A.3. Lemmas A.1 and A.2 also hold (mutatis mutandis) for $z_{n,k} = y$ (with $n = 0$).

Proposition A.4. *if Assumptions 3.2 and 3.3 hold, we have $\frac{\partial H^N(\mathbf{z}^N)}{\partial z_{n,k}} = \mathcal{O}(2^{-n/2})$ such that the constant in front of $2^{-n/2}$ does not depend on n or k .*

Proof. if Assumptions 3.2 and 3.3 hold, we obtain the following:

$$\begin{aligned} \frac{\partial H^N(\mathbf{z}^N)}{\partial z_{n,k}} &= - \int_{\mathbb{R}} g(X_T^N(y, \mathbf{z}^N)) \frac{\partial}{\partial y} \left[\left(\frac{\partial X_T^N}{\partial y}(y, \mathbf{z}^N) \right)^{-1} \frac{\partial X_T^N}{\partial z_{n,k}}(y, \mathbf{z}^N) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \right] dy \\ &= - \int_{\mathbb{R}} g(X_T^N(y, \mathbf{z}^N)) \left[- \left(\frac{\partial X_T^N}{\partial y}(y, \mathbf{z}^N) \right)^{-2} \frac{\partial^2 X_T^N}{\partial y^2}(y, \mathbf{z}^N) \frac{\partial X_T^N}{\partial z_{n,k}}(y, \mathbf{z}^N) + \right. \\ &\quad \left. + \left(\frac{\partial X_T^N}{\partial y}(y, \mathbf{z}^N) \right)^{-1} \frac{\partial^2 X_T^N}{\partial z_{n,k} \partial y}(y, \mathbf{z}^N) - y \left(\frac{\partial X_T^N}{\partial y}(y, \mathbf{z}^N) \right)^{-1} \frac{\partial X_T^N}{\partial z_{n,k}}(y, \mathbf{z}^N) \right] \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy. \end{aligned}$$

Hence, Lemmas A.1 and A.2 with Assumption 3.3 (for $p = 2$) imply that

$$\frac{\partial H^N(\mathbf{z}^N)}{\partial z_{n,k}} = \mathcal{O}(2^{-n/2}),$$

with constants independent of n and k .²³ □

For the general case we require the following lemma.

Lemma A.5. *For any $p \in \mathbb{N}$ and indices n_1, \dots, n_p and k_1, \dots, k_p (satisfying $0 \leq k_j < 2^{n_j}$) we have (with constants independent of n_j, k_j)*

$$\frac{\partial^p X_T^N}{\partial z_{n_1, k_1} \cdots \partial z_{n_p, k_p}}(Z_{-1}, \mathbf{Z}^N) = \mathcal{O}\left(2^{-\sum_{j=1}^p n_j/2}\right).$$

The result also holds (mutatis mutandis) if one or several z_{n_j, k_j} are replaced by $y = z_{-1}$ (with n_j set to 0).

Proof. Each $\Delta_{\ell}^N W$ is a linear function of (Z_{-1}, \mathbf{Z}^N) implying that all higher derivatives of $\Delta_{\ell}^N W$ w.r.t. (Z_{-1}, \mathbf{Z}^N) vanish. Hence,

$$\frac{\partial^p X_T^N}{\partial Z_{n_1, k_1} \cdots \partial Z_{n_p, k_p}} = \sum_{\ell_1=2^{N-n_1+1}k_1}^{2^{N-n_1+1}(k_1+1)-1} \cdots \sum_{\ell_p=2^{N-n_p+1}k_p}^{2^{N-n_p+1}(k_p+1)-1} \frac{\partial^p X_T^N}{\partial \Delta_{\ell_1}^N \cdots \partial \Delta_{\ell_p}^N W} \frac{\partial \Delta_{\ell_1}^N W}{\partial Z_{n_1, k_1}} \cdots \frac{\partial \Delta_{\ell_p}^N W}{\partial Z_{n_p, k_p}}.$$

By an argument similar to that made for (A.3), we obtain

$$\frac{\partial^p X_T^N}{\partial \Delta_{\ell_1}^N \cdots \partial \Delta_{\ell_p}^N W} = \mathcal{O}(1).$$

By (A.1), we observe that each summand in the aforementioned sum is of order $\prod_{j=1}^p 2^{-(N-n_j/2)}$. The number of summands in total is $\prod_{j=1}^p 2^{N-n_j+1}$. Therefore, we obtain the desired result. □

²³When $F^N(Z_{-1}, \mathbf{Z}^N) = \mathcal{O}(c)$ for some deterministic constant c , this property is retained when integrating out one of the rdvs, i.e., we still achieve $\int_{\mathbb{R}} F^N(y, \mathbf{Z}^N) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \mathcal{O}(c)$.

Sketch of the proof of Theorem 3.4. We apply integration by parts p times, as performed in the proof of Proposition A.4, which shows that we can again replace the mollified payoff function g_δ by the true, nonsmooth function g . Moreover, using this procedure, we obtain a formula of the form

$$\frac{\partial^p H^N}{\partial z_{n_1, k_1} \cdots \partial z_{n_p, k_p}}(z^N) = \int_{\mathbb{R}} g(X_T^N(y, z^N)) \blacksquare \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy,$$

where \blacksquare represents a long sum of products of various terms. However, when the derivatives w.r.t. y are ignored, each summand contains all derivatives w.r.t. $z_{n_1, k_1}, \dots, z_{n_p, k_p}$ exactly once. (Generally, each summand is a product of the derivatives of X_T^N w.r.t. some z_{n_j, k_j} s, possibly including other terms, such as polynomials in y and derivatives w.r.t. y .) As all other terms are assumed to be of order $\mathcal{O}(1)$ based on Assumptions 3.2 and 3.3, the result suggested by Lemma A.5 is implied, concluding the proof of Theorem 3.4. \square

B Discussion of Assumption 3.3

We present sufficient conditions for Assumption 3.3 to be valid in the one-dimensional setting. Moreover, we discuss its limitation and some multivariate cases in which this assumption holds.

We want to examine the term given by $\left(\frac{\partial X_T^N}{\partial y}(Z_{-1}, \mathbf{Z}^N)\right)^{-p}$ for $p \in \mathbb{N}$. For this, we consider the one-dimensional SDE

$$dX_t = a(X_t)dt + b(X_t)dW_t.$$

For ease of presentation, we set the drift term $a(\cdot)$ to zero. Moreover, using the Brownian bridge construction, we achieve

$$(B.1) \quad dX_t = b(X_t) \left(\frac{y}{\sqrt{T}} dt + dB_t \right),$$

where y is a standard Gaussian rdv and B is the Brownian bridge.

The solution of (B.1), at the final time $T > 0$ is

$$X_T = x_0 + \frac{y}{\sqrt{T}} \int_0^T b(X_s) ds + \int_0^T b(X_s) dB_s,$$

and consequently,

$$\frac{\partial X_T}{\partial y} = \frac{y}{\sqrt{T}} \int_0^T b'(X_s) \frac{\partial X_s}{\partial y} ds + \frac{1}{\sqrt{T}} \int_0^T b(X_s) ds + \int_0^T b'(X_s) \frac{\partial X_s}{\partial y} dB_s.$$

This implies that $\frac{\partial X_T}{\partial y}$ solves

$$\begin{cases} d\left(\frac{\partial X_T}{\partial y}\right) &= \frac{b(X_t)}{\sqrt{T}} dt + b'(X_t) \frac{\partial X_t}{\partial y} dW_t, \\ \frac{\partial X_T}{\partial y} |_{t=0} &= 0. \end{cases}$$

Using Duhamel's principle, we obtain

$$\frac{\partial X_T}{\partial y} = \int_0^T \frac{b(X_s)}{\sqrt{T}} \exp\left(\left(\int_s^T b'(X_u) dW_u\right) - \frac{1}{2} \int_s^T (b')^2(X_u) du\right) ds.$$

If there exists $b_0 \in \mathbb{R}$ such that

$$(B.2) \quad b^2(x) \geq b_0^2, \quad \forall x \in \mathbb{R},$$

then

$$\begin{aligned} \left| \frac{\partial X_T}{\partial y} \right| &\geq \frac{|b_0|}{\sqrt{T}} \int_0^T \exp \left(\left(\int_s^T b'(X_u) dW_u \right) - \frac{1}{2} \int_s^T (b')^2(X_u) du \right) ds, \\ &\geq \frac{|b_0|}{\sqrt{T}} \exp \left(\left(\int_0^T b'(X_u) dW_u \right) - \frac{1}{2} \int_0^T (b')^2(X_u) du \right) ds, \end{aligned}$$

and consequently, for any $p \in \mathbb{N}$, we obtain

$$\left(\left| \frac{\partial X_T}{\partial y} \right| \right)^{-p} \leq \left(\frac{|b_0|}{\sqrt{T}} \right)^{-p} \exp \left(-p \left(\left(\int_0^T b'(X_u) dW_u \right) - \frac{1}{2} \int_0^T (b')^2(X_u) du \right) \right) ds,$$

and the sufficient condition for Assumption 3.3 to be valid is that for any $p \in \mathbb{N}$, there exists a real deterministic constant $D_p > 0$ such that

$$(B.3) \quad \mathbb{E} \left[\exp \left(-p \left(\left(\int_0^T b'(X_u) dW_u \right) - \frac{1}{2} \int_0^T (b')^2(X_u) du \right) \right) \right] \leq D_p.$$

For the particular one-dimensional GBM model, condition (B.3) is clearly satisfied. Moreover, both (i) one-dimensional models with a linear or constant diffusion and (ii) multivariate models with a linear drift and constant diffusion satisfy Assumption 3.3. Interestingly, the multivariate lognormal model can be observed in case (ii) (refer to [8] for further details). However, there may be cases in which Assumption 3.3 is not fulfilled, e.g., $X_T = W_T^2$, corresponding to a system of SDEs where the diffusion coefficient does not satisfy condition (B.2). Nevertheless, the proposed method works well in such cases because (using notation of Section 2.2) $g(X_T) = G(y_1^2)$, and then we can apply our numerical smoothing technique to obtain a highly smooth integrand. Finally, an additional investigation on the sufficient conditions for our smoothness Theorem 3.4 to be valid in high dimensions is an open problem and is not within the scope of this work.

C More Details on the Work Discussion of the ASGQ Method

Under certain conditions of the regularity parameters p and s , we can achieve $\text{Work}_{\text{ASGQ}} = \mathcal{O}(\text{TOL}^{-1})$ under the best scenario ($p, s \gg 1$). In fact, let $q=p/2$, then using the method of Lagrange multipliers, we obtain

$$M_{\text{ASGQ}} \propto \Delta t^{\frac{2q+s-qs}{q(s+2q+s)}}, \quad \text{and} \quad M_{\text{Lag}} \propto \Delta t^{\frac{2q+s-qs}{\frac{3}{2}(q+s+2q+s)}}.$$

Using the constraint in (3.12), we can easily demonstrate that, for an error tolerance TOL, we achieve $\Delta t = \mathcal{O}\left(\text{TOL}^{\frac{qs+2q+s}{qs-2q-s}}\right)$. Therefore, the optimal work, $\text{Work}_{\text{ASGQ}}$, solution of (3.12) satisfies

$$\begin{aligned} \text{Work}_{\text{ASGQ}} &\propto \Delta t^{-1} \times M_{\text{ASGQ}} \times M_{\text{Lag}} \propto \Delta t^{-1} \Delta t^{\frac{2q+s-qs}{q(s+2q+s)}} \Delta t^{\frac{2q+s-qs}{\frac{3}{2}(q+s+2q+s)}} \\ &\propto \text{TOL}^{-1 - \frac{2(2q+s)}{qs-2q-s} - \frac{1}{q} - \frac{2}{s}} \\ &= \mathcal{O}(\text{TOL}^{-1}), \quad \text{because } p, s \gg 1. \end{aligned}$$

D Simulation Schemes for the Heston Dynamics

D.1 Modified Euler scheme

The forward Euler scheme can be used to simulate the Heston model. The literature has reported many solutions to avoid the problems arising from the use of negative values of the volatility process v_t in (4.1) [28]. Table D.1 introduces f_1, f_2 , and f_3 , which imply various schemes when different choices are adopted. The forward Euler scheme to discretize (4.1) yields

$$\begin{aligned}\widehat{S}_{t+\Delta t} &= \widehat{S}_t + \mu\widehat{S}_t\Delta t + \sqrt{\widehat{V}_t\Delta t}\widehat{S}_tZ_s \\ \widehat{V}_{t+\Delta t} &= f_1(\widehat{V}_t) + \kappa(\theta - f_2(\widehat{V}_t))\Delta t + \xi\sqrt{f_3(\widehat{V}_t)\Delta t}Z_V \\ \widehat{V}_{t+\Delta t} &= f_3(\widehat{V}_{t+\Delta t}),\end{aligned}$$

where Z_s and Z_V are two correlated standard normal rdvs with correlation ρ .

Scheme	f_1	f_2	f_3
Full truncation scheme	\widehat{V}_t	\widehat{V}_t^+	\widehat{V}_t^+
Partial truncation scheme	\widehat{V}_t	\widehat{V}_t	\widehat{V}_t^+
Reflection scheme	$ \widehat{V}_t $	$ \widehat{V}_t $	$ \widehat{V}_t $

Table D.1: Different variants for the forward Euler scheme for the Heston model. $\widehat{V}_t^+ = \max(0, \widehat{V}_t)$.

Lord et al. [28] suggested that the full truncation scheme is an optimal option in terms of the weak error convergence. Therefore, we used this variant of the forward Euler scheme.

D.2 Moment-matching scheme

We consider the moment-matching scheme suggested by Andersen and Brotherton-Ratcliffe [3] (the ABR scheme). This scheme assumes that the variance v_t is locally lognormal, and the parameters are determined such that the first two moments of the discretization coincide with the theoretical moments:

$$\begin{aligned}\widehat{V}(t + \Delta t) &= \left(e^{-\kappa\Delta t}\widehat{V}(t) + (1 - e^{-\kappa\Delta t})\theta \right) e^{-\frac{1}{2}\Gamma(t)^2\Delta t + \Gamma(t)\Delta W_v(t)} \\ \Gamma^2(t) &= \Delta t^{-1} \log \left(1 + \frac{\frac{1}{2}\xi^2\kappa^{-1}\widehat{V}(t)(1 - e^{-2\kappa\Delta t})}{\left(e^{-\kappa\Delta t}\widehat{V}(t) + (1 - e^{-\kappa\Delta t})\theta \right)^2} \right).\end{aligned}$$

As reported in [28], the scheme is easy to implement and more effective than many of the Euler variants presented in Section D.1; however, this scheme exhibits a nonrobust weak error behavior w.r.t. the model parameters.

D.3 Heston OU-based scheme

Because any OU process is normally distributed, the sum of n squared OU processes is chi-squared distributed with n degrees of freedom, where $n \in \mathbb{N}_+$. We define \mathbf{X} as a n -dimensional vector-valued

OU process with

$$(D.1) \quad dX_t^i = \alpha X_t^i dt + \beta dW_t^i,$$

where \mathbf{W} is an n -dimensional vector of independent Brownian motions.

We also define the process Y_t as follows:

$$Y_t = \sum_{i=1}^n (X_t^i)^2.$$

Then, because

$$d(X_t^i)^2 = 2X_t^i dX_t^i + 2d\langle X^i \rangle_t = (2\alpha (X_t^i)^2 + \beta^2) dt + 2\beta X_t^i dW_t^i,$$

we can write (using the independence of the Brownian motions):

$$(D.2) \quad dY_t = d\left(\sum_{i=1}^n (X_t^i)^2\right) = \sum_{i=1}^n d(X_t^i)^2 = (2\alpha Y_t + n\beta^2) dt + 2\beta \sum_{i=1}^n X_t^i dW_t^i.$$

Furthermore, the process $Z_t = \int_0^t \sum_{i=1}^n X_u^i dW_u^i$ is a martingale with quadratic variations

$$\langle Z \rangle_t = \int_0^t \sum_{i=1}^n (X_u^i)^2 du = \int_0^t Y_u du.$$

Consequently, using the Lévy characterization theorem, the process $\widetilde{W}_t = \int_0^t \frac{1}{\sqrt{Y_u}} \sum_{i=1}^n X_u^i dW_u^i$ is a Brownian motion. Finally, we obtain

$$(D.3) \quad \begin{aligned} dY_t &= (2\alpha Y_t + n\beta^2) dt + 2\beta \sqrt{Y_t} d\widetilde{W}_t \\ &= \kappa(\theta - Y_t) dt + \xi \sqrt{Y_t} dW_t, \end{aligned}$$

where $\kappa = -2\alpha$, $\theta = -n\beta^2/2\alpha$ and $\xi = 2\beta$.

Equations (D.1), (D.2), and (D.3) indicate that, to simulate the process Y_t given by (D.3), we can simulate the OU process \mathbf{X} with dynamics (D.1) such that its parameters (α, β) are expressed in terms of those of the process Y_t :

$$\alpha = -\frac{\kappa}{2}, \quad \beta = \frac{\xi}{2}, \quad n = \frac{-2\theta\alpha}{\beta^2} = \frac{4\theta\kappa}{\xi^2}.$$

Consequently, we can simulate the volatility of the Heston model using a sum of OU processes.

Remark D.1. The previous derivation can be generalized to cases where n^* is not an integer, by considering the time-change of the squared Bessel process (see Chapter 6 in [26] for details). An alternative method to generalize the scheme for any noninteger n^* is to consider $n^* = n + p$, $p \in (0, 1)$, and compute $E[g(X_{n^*})]$ for any observable g as follows:

$$E[g(X_{n^*})] \approx (1 - p)E[g(X_n)] + pE[g(X_{n+1})].$$

D.4 On the choice of the simulation scheme of the Heston model

We determine the optimal scheme for simulating the Heston model defined in (4.1). In our setting, an optimal scheme is characterized by two properties: (i) the behavior of mixed rate convergence (Section D.4.1), which is a critical requirement for the optimal performance of ASGQ and (ii) the weak error behavior (Section D.4.2) to apply the Richardson extrapolation when necessary.

Although we tested many parameter sets and obtained consistent numerical observations; for illustration, we only present the results for the single call option based on the Heston model with parameters listed in Table 4.1. This set corresponds to $n = 1$, where n represents the number of OU processes used in the Heston OU-based scheme (Section D.3).

D.4.1 Comparison of different schemes in terms of mixed difference rates

As emphasized in [24, 6], one crucial requirement to achieve the optimal performance of the ASGQ is to check the error convergence of the first and mixed difference operators, as expressed by the error contribution ΔE_β in (2.22). This is a measure of how much the quadrature error would decrease after the addition of a new multi-index β to the constructed index set of the ASGQ estimator, $\mathcal{I}_{\text{ASGQ}}$. The ASGQ method exhibits optimal behavior if (i) ΔE_β decreases exponentially fast w.r.t. β_i and (ii) ΔE_β has a product structure so that a faster error decay is observed for second differences compared to the corresponding first difference operators.

In this section, we compare the three approaches of simulating Heston dynamics: (i) the full truncation scheme (Section D.1), (ii) the ABR scheme (Section D.2), and (iii) the Heston OU-based scheme (Section D.3) in terms of the mixed difference convergences. In our numerical experiments, we only observe the differences in the mixed difference rates related to the volatility coordinates because we apply schemes that only differ in the way they simulate the volatility process. Figure D.1 illustrates a comparison of the first difference rates related to the volatility coordinates for the various schemes. The figure reveals that the full truncation scheme is the worst scheme and that the Heston OU-based and the ABR schemes perform very well in terms of the speed of the mixed rate convergence.

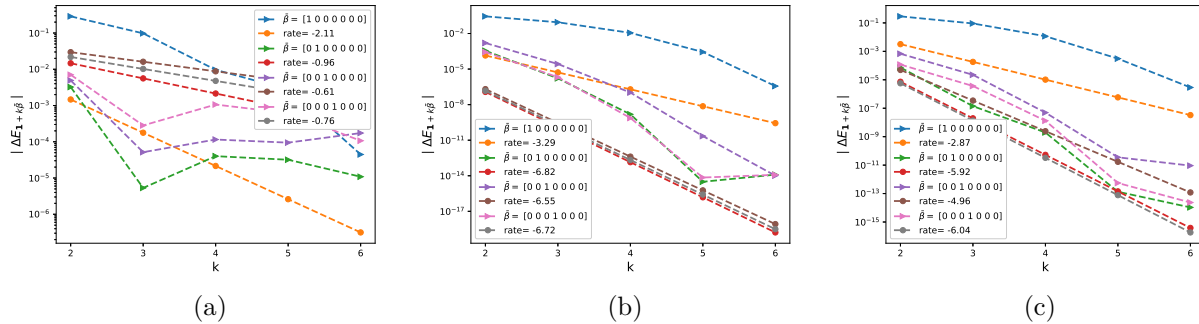


Figure D.1: Rate of error convergence of the first-order differences $|\Delta E_\beta|$, defined in (2.22), ($\beta = \mathbf{1} + k\bar{\beta}$) for the single call option under the Heston model. The parameters are given in Set 1 in Table 4.1, and the number of time steps $N = 4$. We only present the first four dimensions used for the volatility noise (mainly dW_v in (4.1)). (a) Full truncation scheme, (b) ABR scheme, and (c) Heston OU-based scheme.

D.4.2 Comparison in terms of the weak error behavior

We compare the three schemes of simulating Heston dynamics: (i) the full truncation scheme (Section D.1), (ii) the ABR scheme (Section D.2), and (iii) the Heston OU-based scheme (Section D.3 in terms of the weak error convergence. We select the scheme with weak error rate of order 1 (*i.e.*, $\mathcal{O}(\Delta t)$) in the preasymptotic regime to efficiently employ the Richardson extrapolation in our proposed methods. Figure D.2 compares the weak error rates for the different schemes. This figure reveals that the Heston OU-based scheme exhibits a better weak convergence rate closer to 1 than the ABR scheme, which exhibits a weak error rate of 0.7.

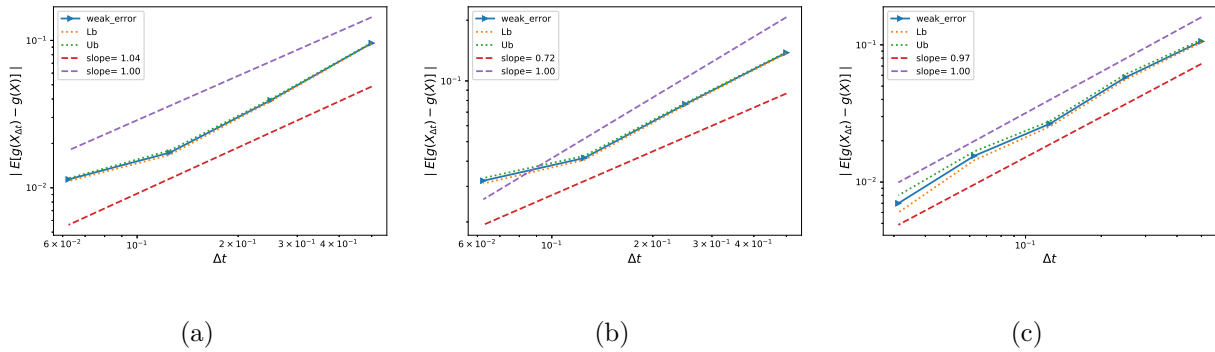


Figure D.2: Weak error convergence for the single call option under the Heston model for the parameters listed in Table 4.1. (a) Full truncation scheme, (b) ABR scheme, and (c) Heston OU-based scheme. The upper and lower bounds are 95% confidence intervals.