

TRUNCATED CONTROL VARIATES FOR WEAK APPROXIMATION SCHEMES *

DENIS BELOMESTNY¹, STEFAN HÄFNER² AND MIKHAIL URUSOV³

Abstract. In this paper we present an enhancement of the regression-based variance reduction approaches recently proposed in Belomestny et al. [1] and [4]. This enhancement is based on a truncation of the control variate and allows for a significant reduction of the computing time, while the complexity stays of the same order. The performances of the proposed truncated algorithms are illustrated by a numerical example.

INTRODUCTION

Let $T > 0$ be a fixed time horizon. Consider a d -dimensional diffusion process $(X_t)_{t \in [0, T]}$ defined by the Itô stochastic differential equation

$$dX_t = \mu(X_t) dt + \sigma(X_t) dW_t, \quad X_0 = x_0 \in \mathbb{R}^d, \quad (0.1)$$

for Lipschitz continuous functions $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$, where $(W_t)_{t \in [0, T]}$ is a standard m -dimensional Brownian motion. Our aim is to compute the expectation

$$u(t, x) := \mathbb{E}[f(X_T^{t,x})], \quad (0.2)$$

for some $f: \mathbb{R}^d \rightarrow \mathbb{R}$, where $X_T^{t,x}$ denotes the solution to (0.1) started at time t in point x . The standard Monte Carlo (SMC) estimate for $u(0, x)$ at a fixed point $x \in \mathbb{R}^d$ has the form

$$V_{N_0} := \frac{1}{N_0} \sum_{i=1}^{N_0} f(\bar{X}_T^{(i)}) \quad (0.3)$$

for some $N_0 \in \mathbb{N}_0$, where \bar{X}_T is an approximation for $X_T^{0,x}$ constructed via a time discretisation of (0.1) (we refer to [7] for a nice overview of various discretisation schemes). In the computation of $u(0, x) = \mathbb{E}[f(X_T^{0,x})]$ by the SMC approach there are two types of error inherent: the (deterministic) discretisation error $\mathbb{E}[f(X_T^{0,x})] - \mathbb{E}[f(\bar{X}_T)]$ and the Monte Carlo (statistical) error, which results from the substitution of $\mathbb{E}[f(\bar{X}_T)]$ with the

* The work of Denis Belomestny is supported by the Russian Science Foundation project 14-50-00150. Stefan Häfner thanks the Faculty of Mathematics of the University of Duisburg-Essen, where this work was carried out.

¹ University of Duisburg-Essen, Essen, Germany and IITP RAS, Moscow, Russia

² PricewaterhouseCoopers GmbH, Frankfurt, Germany

³ University of Duisburg-Essen, Essen, Germany

sample average V_{N_0} . The aim of variance reduction methods is to reduce the latter statistical error. For example, in the so-called control variate variance reduction approach one looks for a random variable ξ with $\mathbb{E}\xi = 0$, which can be simulated, such that the variance of the difference $f(\bar{X}_T) - \xi$ is minimised, that is,

$$\text{Var}[f(\bar{X}_T) - \xi] \rightarrow \min \text{ under } \mathbb{E}\xi = 0.$$

Then one uses the sample average

$$V_{N_0}^{CV} := \frac{1}{N_0} \sum_{i=1}^{N_0} \left[f(\bar{X}_T^{(i)}) - \xi^{(i)} \right] \quad (0.4)$$

instead of (0.3) to approximate $\mathbb{E}[f(\bar{X}_T)]$. The use of control variates for computing expectations of functionals of diffusion processes via Monte Carlo was initiated by Newton [11] and further developed in Milstein and Tretyakov [8]. In Belomestny et al [1] a novel regression-based approach for the construction of control variates, which reduces the variance of the approximated functional $f(\bar{X}_T)$ was proposed. As shown in [1], the ‘‘Monte Carlo approach with the Regression-based Control Variate’’ (abbreviated below as ‘‘RCV approach’’) is able to achieve a higher order convergence of the resulting variance to zero, which in turn leads to a significant complexity reduction as compared to the SMC algorithm. Other prominent examples of algorithms with this property are the multilevel Monte Carlo (MLMC) algorithm of [5] and quadrature-based algorithms of [9] and [10]. The RCV approach becomes especially simple in the case of the so-called weak approximation schemes, i.e., the schemes, where simple random variables are used in place of Brownian increments, and which became quite popular in recent years. However, due to the fact that a lot of computations are required for implementing the RCV approach, its numerical efficiency is not convincing in higher-dimensional examples. The same applies also to the SRCV algorithm of [4]. In this paper we further enhance the performances of the RCV and SRCV algorithms by truncating the control variates, leading to a reduction from $(2^m - 1)$ to m terms at each time point in case of the weak Euler scheme and a reduction from $(3^m 2^{\frac{m(m-1)}{2}} - 1)$ to $m(m+1) = O(m^2)$ terms at each time point in case of the second order weak scheme. It turns out that, while the computing time is reduced significantly, we still have a sufficient variance reduction effect such that the complexity is of the same order as for the original RCV and SRCV approaches.

The paper is organised as follows. In Section 1 we present a smoothness theorem for a general class of discretisation schemes. Section 2 recalls the construction of control variates for weak schemes of the first and the second order. The main truncation results are derived in Section 3. In Section 4 we describe a generic regression algorithm. Section 5 deals with a complexity analysis for the algorithm that is based on the truncated control variate. Section 6 is devoted to a simulation study. Finally, all proofs are collected in Section 7.

1. SMOOTHNESS THEOREM FOR DISCRETISATION SCHEMES

In this section we present a technical result for discretisation schemes, which will be very important in the sequel. To begin with, let $J \in \mathbb{N}$ denote the time discretisation parameter, we set $\Delta := T/J$ and consider discretisation schemes defined on the grid $\{j\Delta : j = 0, \dots, J\}$.

Let us consider a scheme, where d -dimensional approximations $X_{\Delta, j\Delta}$, $j = 0, \dots, J$, satisfy $X_{\Delta, 0} = x_0$ and

$$X_{\Delta, j\Delta} = \Phi_{\Delta}(X_{\Delta, (j-1)\Delta}, \xi_j), \quad j = 1, \dots, J, \quad (1.1)$$

for some Borel measurable functions $\Phi_{\Delta}: \mathbb{R}^{d+\tilde{m}} \rightarrow \mathbb{R}^d$, where $\tilde{m} \geq m$, and for \tilde{m} -dimensional i.i.d. random vectors $\xi_j = (\xi_j^1, \dots, \xi_j^{\tilde{m}})^{\top}$ with independent coordinates satisfying $\mathbb{E}[\xi_j^i] = 0$ and $\text{Var}[\xi_j^i] = 1$ for all $i = 1, \dots, \tilde{m}$, $j = 1, \dots, J$. Moreover, let \mathcal{G}_0 be the trivial σ -field and $\mathcal{G}_j = \sigma(\xi_1, \dots, \xi_j)$, $j = 1, \dots, J$. In the chapters below we will focus on different kinds of discretisation schemes, resulting in different convergence behaviour.

We now define the random function $G_{l,j}(x)$ for $J \geq l \geq j \geq 0$, $x \in \mathbb{R}^d$, as follows

$$\begin{aligned} G_{l,j}(x) &\equiv \Phi_{\Delta,l} \circ \Phi_{\Delta,l-1} \circ \dots \circ \Phi_{\Delta,j+1}(x), \quad l > j, \\ G_{l,j}(x) &\equiv x, \quad l = j, \end{aligned} \quad (1.2)$$

where $\Phi_{\Delta,l}(x) := \Phi_{\Delta}(x, \xi_l)$ for $l = 1, \dots, J$. By $\Phi_{\Delta,l}^k$, $k \in \{1, \dots, d\}$, we denote the k -th component of the function $\Phi_{\Delta,l}$. Note that it holds

$$q_j(x) := \mathbb{E}[f(X_{\Delta,T}) | X_{\Delta,j\Delta} = x] = \mathbb{E}[f(G_{J,j}(x))]. \quad (1.3)$$

Let us define the operator D^α as follows

$$D^\alpha g(x) := \frac{\partial^{|\alpha|} g(x)}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}, \quad (1.4)$$

where g is a real-valued function, $\alpha \in \mathbb{N}_0^d$ and $|\alpha| = \alpha_1 + \dots + \alpha_d$ ($\mathbb{N}_0 := \mathbb{N} \cup \{0\}$).

In the next theorem we present some smoothness conditions on q_j , which will be used several times in the chapters below.

Theorem 1.1. *Let $K \in \{1, 2, 3\}$. Suppose that f is K times continuously differentiable with bounded partial derivatives up to order K , $\Phi_{\Delta}(\cdot, \xi)$ is K times continuously differentiable (for any fixed ξ), and that, for any $n \in \mathbb{N}$, $l \geq j$, $k \in \{1, \dots, d\}$, $\alpha \in \mathbb{N}_0^d$ with $1 \leq |\alpha| \leq K$, it holds*

$$\left| \mathbb{E} \left[\left(D^\alpha \Phi_{\Delta,l+1}^k(G_{l,j}(x)) \right)^n \middle| \mathcal{G}_l \right] \right| \leq \begin{cases} (1 + A_n \Delta), & |\alpha| = \alpha_k = 1 \\ B_n \Delta, & (|\alpha| > 1) \vee (\alpha_k \neq 1) \end{cases} \quad (1.5)$$

with probability one for some constants $A_n > 0$, $B_n > 0$. Moreover, suppose that for any $n_1, n_2 \in \mathbb{N}$, $\alpha, \beta \in \mathbb{N}_0^d$, with $|\alpha| = 1$, $1 \leq |\beta| \leq K$, $\alpha \neq \beta$, it holds

$$\left| \mathbb{E} \left[\left(D^\alpha \Phi_{\Delta,l+1}^k(G_{l,j}(x)) \right)^{n_1} \left(D^\beta \Phi_{\Delta,l+1}^k(G_{l,j}(x)) \right)^{n_2} \middle| \mathcal{G}_l \right] \right| \leq E_{n_1, n_2} \Delta \quad (1.6)$$

for some constants $E_{n_1, n_2} > 0$. Then we obtain for all $j \in \{0, \dots, J\}$ that q_j is K times continuously differentiable with bounded partial derivatives up to order K .

2. REPRESENTATIONS FOR WEAK APPROXIMATION SCHEMES

Below we focus on weak schemes of first and second order.

2.1. Weak Euler scheme

In this subsection we treat weak schemes of order 1. Let us consider a scheme, where d -dimensional approximations $X_{\Delta,j\Delta}$, $j = 0, \dots, J$, satisfy $X_{\Delta,0} = x_0$ and

$$X_{\Delta,j\Delta} = \Phi_{\Delta}(X_{\Delta,(j-1)\Delta}, \xi_j), \quad j = 1, \dots, J, \quad (2.1)$$

for some functions $\Phi_{\Delta}: \mathbb{R}^{d+m} \rightarrow \mathbb{R}^d$, with $\xi_j = (\xi_j^1, \dots, \xi_j^m)$, $j = 1, \dots, J$, being m -dimensional iid random vectors with iid coordinates such that

$$\mathbb{P}(\xi_j^k = \pm 1) = \frac{1}{2}, \quad k = 1, \dots, m.$$

That is, relating to the framework in Section 1, we have $\tilde{m} = m$ and use the discrete increments ξ_j^i , $i = 1, \dots, m$. A particular case is the weak Euler scheme (also called the *simplified weak Euler scheme* in [7, Section 14.1]) of order 1, which is given by

$$\Phi_\Delta(x, y) = x + \mu(x) \Delta + \sigma(x) y \sqrt{\Delta}. \quad (2.2)$$

Let us recall the functions (cf. (1.3))

$$q_j(x) = \mathbb{E}[f(X_{\Delta,T}) | X_{\Delta,j\Delta} = x].$$

The proposition below summarises important representations for the weak Euler scheme, which were derived in [1].

Proposition 2.1. *The following representation holds*

$$f(X_{\Delta,T}) = \mathbb{E}f(X_{\Delta,T}) + \sum_{j=1}^J \sum_{r=1}^m \sum_{1 \leq s_1 < \dots < s_r \leq m} a_{j,r,s}(X_{\Delta,(j-1)\Delta}) \prod_{i=1}^r \xi_j^{s_i}, \quad (2.3)$$

where we use the notation $s = (s_1, \dots, s_r)$. The coefficients $a_{j,r,s}: \mathbb{R}^d \rightarrow \mathbb{R}$ can be computed by the formula

$$a_{j,r,s}(x) = \mathbb{E} \left[f(X_{\Delta,T}) \prod_{i=1}^r \xi_j^{s_i} \mid X_{\Delta,(j-1)\Delta} = x \right] \quad (2.4)$$

for all j, r , and s as in (2.3). Moreover, we have the following recursion formulas

$$\begin{aligned} q_{j-1}(x) &= \mathbb{E}[q_j(X_{\Delta,j\Delta}) | X_{\Delta,(j-1)\Delta} = x] = \frac{1}{2^m} \sum_{y=(y^1, \dots, y^m) \in \{-1, 1\}^m} q_j(\Phi_\Delta(x, y)), \\ a_{j,r,s}(x) &= \frac{1}{2^m} \sum_{y=(y^1, \dots, y^m) \in \{-1, 1\}^m} \left[\prod_{i=1}^r y^{s_i} \right] q_j(\Phi_\Delta(x, y)), \end{aligned} \quad (2.5)$$

for all $j \in \{1, \dots, J\}$, $r \in \{1, \dots, m\}$, $1 \leq s_1 < \dots < s_r \leq m$, where $q_J \equiv f$.

The next proposition (cf. Proposition 3.2 in [1]) shows the properties of the weak Euler scheme combined with the control variate

$$M_{\Delta,T}^{(1)} := \sum_{j=1}^J \sum_{r=1}^m \sum_{1 \leq s_1 < \dots < s_r \leq m} a_{j,r,s}(X_{\Delta,(j-1)\Delta}) \prod_{i=1}^r \xi_j^{s_i}, \quad (2.6)$$

where the coefficients $a_{j,r,s}(x)$ are given by (2.4).

Proposition 2.2. *Assume that μ and σ in (0.1) are Lipschitz continuous with components $\mu^i, \sigma^{i,r}: \mathbb{R}^d \rightarrow \mathbb{R}$, $i = 1, \dots, d$, $r = 1, \dots, m$, being 4 times continuously differentiable with their partial derivatives of order up to 4 having polynomial growth. Let $f: \mathbb{R}^d \rightarrow \mathbb{R}$ be 4 times continuously differentiable with partial derivatives of order up to 4 having polynomial growth. Provided that (2.2) holds and that, for sufficiently large $p \in \mathbb{N}$, the expectations $\mathbb{E}|X_{\Delta,j\Delta}|^{2p}$ are uniformly bounded in J and $j = 0, \dots, J$, we have for this “simplified weak Euler scheme”*

$$|\mathbb{E}[f(X_T) - f(X_{\Delta,T})]| \leq c\Delta,$$

where the constant c does not depend on Δ . Moreover, it holds $\text{Var} [f(X_{\Delta,T}) - M_{\Delta,T}^{(1)}] = 0$.

Discussion. In order to use the control variate $M_{\Delta,T}^{(1)}$ in practice, we need to estimate the unknown coefficients $a_{j,r,s}$. Thus, practically implementable control variates $\widetilde{M}_{\Delta,T}^{(1)}$ have the form (2.6) with some estimated functions $\tilde{a}_{j,r,s}: \mathbb{R}^d \rightarrow \mathbb{R}$. Notice that they remain valid control variates, i.e. we still have $\mathbb{E}[\widetilde{M}_{\Delta,T}^{(1)}] = 0$, which is due to the martingale transform structure¹ in (2.6).

2.2. Second order weak scheme

Now we treat weak schemes of order 2. We consider a scheme, where d -dimensional approximations $X_{\Delta,j\Delta}$, $j = 0, \dots, J$, satisfy $X_{\Delta,0} = x_0$ and

$$X_{\Delta,j\Delta} = \Phi_{\Delta}(X_{\Delta,(j-1)\Delta}, \xi_j, V_j), \quad j = 1, \dots, J, \quad (2.7)$$

for some functions $\Phi_{\Delta}: \mathbb{R}^{d+m+m \times m} \rightarrow \mathbb{R}^d$. Here,

- (S1) $\xi_j = (\xi_j^i)_{i=1}^m$ are m -dimensional random vectors,
- (S2) $V_j = (V_j^{il})_{i,l=1}^m$ are random $m \times m$ -matrices,
- (S3) the pairs (ξ_j, V_j) , $j = 1, \dots, J$, are i.i.d.,
- (S4) for each j , the random elements ξ_j and V_j are independent,
- (S5) for each j , the random variables ξ_j^i , $i = 1, \dots, m$, are i.i.d. with

$$\mathbb{P}(\xi_j^i = \pm\sqrt{3}) = \frac{1}{6}, \quad \mathbb{P}(\xi_j^i = 0) = \frac{2}{3},$$

- (S6) for each j , the random variables V_j^{il} , $1 \leq i < l \leq m$, are i.i.d. with

$$\mathbb{P}(V_j^{il} = \pm 1) = \frac{1}{2},$$

- (S7) $V_j^{li} = -V_j^{il}$, $1 \leq i < l \leq m$, $j = 1, \dots, J$,
- (S8) $V_j^{ii} = -1$, $i = 1, \dots, m$, $j = 1, \dots, J$.

Hence, the matrices V_j can be generated by means of $\frac{m(m-1)}{2}$ i.i.d. random variables. That is, relating to the framework in Section 1, we have \tilde{m} -dimensional random vectors $\tilde{\xi}_j := ((\xi_j^i)_{i=1,\dots,m}, (V_j^{il})_{1 \leq i < l \leq m})$ with $\tilde{m} = m + \frac{m(m-1)}{2} = \frac{m(m+1)}{2}$.

Remark 2.3. In order to obtain a second order weak scheme in the multidimensional case, we need to incorporate additional random elements V_j into the structure of the scheme. This is the reason why we now consider (2.7) instead of (2.1). For instance, to get the second order weak scheme (also called the *simplified order 2 weak Taylor scheme*) of [7, Section 14.2] in the multidimensional case, we need to define the functions $\Phi_{\Delta}(x, y, z)$, $x \in \mathbb{R}^d$, $y \in \mathbb{R}^m$, $z \in \mathbb{R}^{m \times m}$, as explained below. First we define the function $\Sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ by the formula

$$\Sigma(x) = \sigma(x)\sigma(x)^{\top}$$

and recall that the coordinates of vectors and matrices are denoted by superscripts, e.g. $\Sigma(x) = (\Sigma^{kl}(x))_{k,l=1}^d$, $\Phi_{\Delta}(x, y, z) = (\Phi_{\Delta}^k(x, y, z))_{k=1}^d$. Let us introduce the operators \mathcal{L}^r , $r = 0, \dots, m$, that act on sufficiently smooth

¹This phrase means that the discrete-time process $\tilde{M} = (\tilde{M}_l)_{l=0,\dots,J}$, where $\tilde{M}_0 = 0$ and \tilde{M}_l is defined like the right-hand side of (2.6) but with $\sum_{j=1}^J$ being replaced by $\sum_{j=1}^l$ and $a_{j,r,s}$ by $\tilde{a}_{j,r,s}$ is a martingale, which is a straightforward calculation.

functions $g: \mathbb{R}^d \rightarrow \mathbb{R}$ as follows:

$$\begin{aligned}\mathcal{L}^0 g(x) &:= \sum_{k=1}^d \mu^k(x) \frac{\partial g}{\partial x^k}(x) + \frac{1}{2} \sum_{k,l=1}^d \Sigma^{kl}(x) \frac{\partial^2 g}{\partial x^l \partial x^k}(x), \\ \mathcal{L}^r g(x) &:= \sum_{k=1}^d \sigma^{kr}(x) \frac{\partial g}{\partial x^k}(x), \quad r = 1, \dots, m.\end{aligned}$$

The r -th coordinate Φ_Δ^r , $r = 1, \dots, d$, in the simplified order 2 weak Taylor scheme of [7, Section 14.2] is now given by the formula

$$\begin{aligned}\Phi_\Delta^r(x, y, z) &= x^r + \sum_{k=1}^m \sigma^{rk}(x) y^k \sqrt{\Delta} \\ &+ \left[\mu^r(x) + \frac{1}{2} \sum_{k,l=1}^m \mathcal{L}^k \sigma^{rl}(x) (y^k y^l + z^{kl}) \right] \Delta \\ &+ \frac{1}{2} \sum_{k=1}^m [\mathcal{L}^0 \sigma^{rk}(x) + \mathcal{L}^k \mu^r(x)] y^k \Delta^{3/2} + \frac{1}{2} \mathcal{L}^0 \mu^r(x) \Delta^2,\end{aligned}\tag{2.8}$$

provided the coefficients μ and σ of (0.1) are sufficiently smooth. We will need to work explicitly with (2.8) at some point, but all results in this subsection assume structure (2.7) only.

Let us define the index sets

$$\mathcal{I}_1 = \{1, \dots, m\}, \quad \mathcal{I}_2 = \{(k, l) \in \mathcal{I}_1^2 : k < l\}$$

and the system

$$\mathcal{A} = \{(U_1, U_2) \in \mathcal{P}(\mathcal{I}_1) \times \mathcal{P}(\mathcal{I}_2) : U_1 \cup U_2 \neq \emptyset\},$$

where $\mathcal{P}(\mathcal{I})$ denotes the set of all subsets of a set \mathcal{I} . For any $U_1 \subseteq \mathcal{I}_1$ and $o \in \{1, 2\}^{U_1}$, we write o as $o = (o_r)_{r \in U_1}$. Below we use the convention that a product over the empty set is always one.

For $k \in \mathbb{N}_0$, $H_k: \mathbb{R} \rightarrow \mathbb{R}$ stands for the (normalized) k -th Hermite polynomial, i.e.

$$H_k(x) := \frac{(-1)^k}{\sqrt{k!}} e^{\frac{x^2}{2}} \frac{d^k}{dx^k} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

We remark that, in particular, $H_0 \equiv 1$, $H_1(x) = x$ and $H_2(x) = \frac{1}{\sqrt{2}}(x^2 - 1)$.

As in Subsection 2.1, we summarise important representations from [1] below.

Proposition 2.4. *It holds*

$$f(X_{\Delta, T}) = \mathbb{E} f(X_{\Delta, T}) + \sum_{j=1}^J \sum_{(U_1, U_2) \in \mathcal{A}} \sum_{o \in \{1, 2\}^{U_1}} a_{j, o, U_1, U_2}(X_{\Delta, (j-1)\Delta}) \prod_{r \in U_1} H_{o_r}(\xi_j^r) \prod_{(k, l) \in U_2} V_j^{kl}, \tag{2.9}$$

where the coefficients $a_{j, o, U_1, U_2}: \mathbb{R}^d \rightarrow \mathbb{R}$ can be computed by the formula

$$a_{j, o, U_1, U_2}(x) = \mathbb{E} \left[f(X_{\Delta, T}) \prod_{r \in U_1} H_{o_r}(\xi_j^r) \prod_{(k, l) \in U_2} V_j^{kl} \middle| X_{\Delta, (j-1)\Delta} = x \right]. \tag{2.10}$$

Moreover, we have for each $j \in \{1, \dots, J\}$,

$$\begin{aligned} q_{j-1}(x) &= \mathbb{E}[q_j(X_{\Delta, j\Delta}) | X_{\Delta, (j-1)\Delta} = x] \\ &= \frac{1}{2^{\frac{m(m-1)}{2}}} \frac{1}{6^m} \sum_{(y^1, \dots, y^m) \in \{-\sqrt{3}, 0, \sqrt{3}\}^m} \sum_{(z^{uv})_{1 \leq u < v \leq m} \in \{-1, 1\}^{\frac{m(m-1)}{2}}} 4^{\sum_{i=1}^m I(y^i=0)} q_j(\Phi_{\Delta}(x, y, z)), \end{aligned}$$

and, for all $(U_1, U_2) \in \mathcal{A}$, $o \in \{1, 2\}^{U_1}$, it holds

$$\begin{aligned} a_{j,o,U_1,U_2}(x) &= \frac{1}{2^{\frac{m(m-1)}{2}}} \frac{1}{6^m} \sum_{(y^1, \dots, y^m) \in \{-\sqrt{3}, 0, \sqrt{3}\}^m} \sum_{(z^{uv})_{1 \leq u < v \leq m} \in \{-1, 1\}^{\frac{m(m-1)}{2}}} \\ &\quad 4^{\sum_{i=1}^m I(y^i=0)} \prod_{r \in U_1} H_{o_r}(y^r) \prod_{(k,l) \in U_2} z^{kl} q_j(\Phi_{\Delta}(x, y, z)), \end{aligned} \quad (2.11)$$

where $y = (y^1, \dots, y^m)$, $z = (z^{uv})$ is the $m \times m$ -matrix with $z^{vu} = -z^{uv}$, $u < v$, $z^{uu} = -1$ and $q_J \equiv f$.

Using Theorem 2.4, we obtain the following result (see Proposition 3.6 in [1]), which provides a bound for the discretisation error and a perfect control variate for the discretised quantity.

Proposition 2.5. *Assume, that μ and σ in (0.1) are Lipschitz continuous with components $\mu^i, \sigma^{i,r}: \mathbb{R}^d \rightarrow \mathbb{R}$, $i = 1, \dots, d$, $r = 1, \dots, m$, being 6 times continuously differentiable with their partial derivatives of order up to 6 having polynomial growth. Let $f: \mathbb{R}^d \rightarrow \mathbb{R}$ be 6 times continuously differentiable with partial derivatives of order up to 6 having polynomial growth. Provided that (2.8) holds and that, for sufficiently large $p \in \mathbb{N}$, the expectations $\mathbb{E}|X_{\Delta, j\Delta}|^{2p}$ are uniformly bounded in J and $j = 0, \dots, J$, we have for this “simplified second order weak Taylor scheme”*

$$|\mathbb{E}[f(X_T) - f(X_{\Delta, T})]| \leq c\Delta^2,$$

where the constant c does not depend on Δ . Moreover, we have $\text{Var} \left[f(X_{\Delta, T}) - M_{\Delta, T}^{(2)} \right] = 0$ for the control variate

$$M_{\Delta, T}^{(2)} := \sum_{j=1}^J \sum_{(U_1, U_2) \in \mathcal{A}} \sum_{o \in \{1, 2\}^{U_1}} a_{j,o,U_1,U_2}(X_{\Delta, (j-1)\Delta}) \prod_{r \in U_1} H_{o_r}(\xi_j^r) \prod_{(k,l) \in U_2} V_j^{kl}, \quad (2.12)$$

where the coefficients $a_{j,o,U_1,U_2}(x)$ are defined in (2.10).

3. TRUNCATED CONTROL VARIATES FOR WEAK APPROXIMATION SCHEMES

Below we recall the assumptions from [1], suggest sufficient conditions for them in terms of the functions f, μ, σ , and then suggest some stronger conditions that will justify the use of truncated control variates.

3.1. Weak Euler scheme

Note that we considered only the second order weak scheme in terms of the regression and complexity analyses in [1]. However, analogous assumptions for the weak Euler scheme are as follows (cf. Proposition 2.1): fix some $j \in \{1, \dots, J\}$, $r \in \{1, \dots, m\}$, $s = (s_1, \dots, s_r)$ with $1 \leq s_1 < \dots < s_r \leq m$, set $\zeta_{j,r,s} := f(X_{\Delta, T}) \prod_{i=1}^r \xi_j^{s_i}$ and remark that $a_{j,r,s}(x) = \mathbb{E}[\zeta_{j,r,s} | X_{\Delta, (j-1)\Delta} = x]$. We assume that, for some positive constants Σ, A , it holds:

- (A1) $\sup_{x \in \mathbb{R}^d} \text{Var}[\zeta_{j,r,s} | X_{\Delta, (j-1)\Delta} = x] \leq \Sigma < \infty$,
- (A2) $\sup_{x \in \mathbb{R}^d} |a_{j,r,s}(x)| \leq A\sqrt{\Delta} < \infty$.

In the following theorem we suggest sufficient conditions for the above assumptions.

Theorem 3.1. (i) *Let f be bounded. Then (A1) holds.*

(ii) *Let all the functions σ^{ki} , $k \in \{1, \dots, d\}$, $i \in \{1, \dots, m\}$, be bounded and all the functions f, μ^k, σ^{ki} be continuously differentiable with bounded partial derivatives. Then (A2) holds.*

Next we suggest some stronger conditions that give us somewhat more than (A2).

Theorem 3.2. *Let all the functions σ^{ki} , $k \in \{1, \dots, d\}$, $i \in \{1, \dots, m\}$, be bounded and all the functions f, μ^k, σ^{ki} be twice continuously differentiable with bounded partial derivatives up to order 2. Then it holds*

$$(A3) \sup_{x \in \mathbb{R}^d} |a_{j,r,s}(x)| \lesssim \Delta, \text{ whenever } r > 1.$$

Remark 3.3. As a generalisation of Theorem 3.2, it is natural to expect that it holds, under additional smoothness conditions on f, μ, σ ,

$$\sup_{x \in \mathbb{R}^d} |a_{j,r,s}(x)| \lesssim \Delta^{r/2}$$

for all $j \in \{1, \dots, J\}$, $r \in \{1, \dots, m\}$ and $1 \leq s_1 < \dots < s_r \leq m$.

Let us define the “truncated control variate”

$$M_{\Delta,T}^{(1),trunc} := \sum_{j=1}^J \sum_{i=1}^m a_{j,1,e_i}(X_{\Delta,(j-1)\Delta}) \xi_j^i, \quad (3.1)$$

where $e_i \in \mathbb{R}^m$ denotes the i -th unit vector in \mathbb{R}^m and $a_{j,1,e_i}$ is given by (cf. (2.4))

$$a_{j,1,e_i}(x) = \mathbb{E} [f(X_{\Delta,T}) \xi_j^i \mid X_{\Delta,(j-1)\Delta} = x].$$

Note that the superscript “trunc” comes from “truncated”. That is, we consider in $M_{\Delta,T}^{(1),trunc}$ only the terms of the control variate $M_{\Delta,T}^{(1)}$ for which $r = 1$ (cf. (2.6)).

Next we study the truncation error that arises from replacing $M_{\Delta,T}^{(1)}$ by $M_{\Delta,T}^{(1),trunc}$.

Theorem 3.4. *Suppose that all the functions σ^{ki} , $k \in \{1, \dots, d\}$, $i \in \{1, \dots, m\}$ are bounded and all the functions f, μ^k, σ^{ki} are twice continuously differentiable with bounded partial derivatives up to order 2. Then it holds (cf. Proposition 2.2)*

$$\text{Var} [f(X_{\Delta,T}) - M_{\Delta,T}^{(1),trunc}] \lesssim \Delta. \quad (3.2)$$

Notice that under Assumption (A2) alone the variance in (3.2) would have been $O(1)$.

3.2. Second order weak scheme

First we recall some of the required assumptions in [1]: let us fix some $j \in \{1, \dots, J\}$, $(U_1, U_2) \in \mathcal{A}$, $o \in \{1, 2\}^{U_1}$, set

$$\zeta_{j,o,U_1,U_2} := f(X_{\Delta,T}) \prod_{r \in U_1} H_{o_r}(\xi_j^r) \prod_{(k,l) \in U_2} V_j^{kl}$$

and remark that $a_{j,o,U_1,U_2}(x) = \mathbb{E}[\zeta_{j,o,U_1,U_2} \mid X_{\Delta,(j-1)\Delta} = x]$. We assume that, for some positive constants Σ, A , it holds:

$$(B1) \sup_{x \in \mathbb{R}^d} \text{Var}[\zeta_{j,o,U_1,U_2} \mid X_{\Delta,(j-1)\Delta} = x] \leq \Sigma < \infty,$$

$$(B2) \sup_{x \in \mathbb{R}^d} |a_{j,o,U_1,U_2}(x)| \leq A\sqrt{\Delta} < \infty.$$

Below we verify the above assumptions.

Theorem 3.5. (i) Let f be bounded. Then (B1) holds.

(ii) Let all the functions μ^k and σ^{ki} , $k \in \{1, \dots, d\}$, $i \in \{1, \dots, m\}$, be bounded, the function f be continuously differentiable with bounded partial derivatives, and all the functions μ^k, σ^{ki} be three times continuously differentiable with bounded partial derivatives up to order 3. Then (B2) holds.

Let us define the index sets

$$\mathcal{K}_1 := \{r \in U_1 : o_r = 1\}, \quad \mathcal{K}_2 := \{r \in U_1 : o_r = 2\}.$$

In the following theorem we provide some stronger conditions that give us more than (B2).

Theorem 3.6. (i) Let all the functions μ^k and σ^{ki} , $k \in \{1, \dots, d\}$, $i \in \{1, \dots, m\}$, be bounded, the function f be twice continuously differentiable with bounded partial derivatives up to order 2, and all the functions μ^k, σ^{ki} be four times continuously differentiable with bounded partial derivatives up to order 4. Then it holds

$$(B3) \sup_{x \in \mathbb{R}^d} |a_{j,o,U_1,U_2}(x)| \lesssim \Delta, \text{ whenever } |U_2| + |\mathcal{K}_2| + \frac{|\mathcal{K}_1|}{2} \geq 1.$$

(ii) Let all the functions μ^k and σ^{ki} , $k \in \{1, \dots, d\}$, $i \in \{1, \dots, m\}$, be bounded, the function f be three times continuously differentiable with bounded partial derivatives up to order 3, and all the functions μ^k, σ^{ki} be five times continuously differentiable with bounded partial derivatives up to order 5. Then it holds

$$(B4) \sup_{x \in \mathbb{R}^d} |a_{j,o,U_1,U_2}(x)| \lesssim \Delta^{3/2}, \text{ whenever } |U_2| + |\mathcal{K}_2| + \frac{|\mathcal{K}_1|}{2} > 1.$$

Remark 3.7. (i) As a generalisation of Theorem 3.6, it is natural to expect that it holds, under additional smoothness conditions on f, μ, σ ,

$$\sup_{x \in \mathbb{R}^d} |a_{j,o,U_1,U_2}(x)| \lesssim \Delta^{|U_2| + |\mathcal{K}_2| + \frac{|\mathcal{K}_1|}{2}}$$

for all $j \in \{1, \dots, J\}$, $(U_1, U_2) \in \mathcal{A}$ and $o \in \{1, 2\}^{U_1}$.

(ii) Define

$$\Delta_{U_1,U_2} := \begin{cases} \Delta^{|U_2| + |\mathcal{K}_2| + \frac{|\mathcal{K}_1|}{2}} & \text{if } |U_2| + |\mathcal{K}_2| + \frac{|\mathcal{K}_1|}{2} \leq 1, \\ \Delta^{3/2} & \text{otherwise.} \end{cases} \quad (3.3)$$

An equivalent reformulation of assumptions (B2)–(B4) is as follows: there exists some positive constant \tilde{A} such that it holds

$$\sup_{x \in \mathbb{R}^d} |a_{j,o,U_1,U_2}(x)| \leq \tilde{A} \Delta_{U_1,U_2} \quad (3.4)$$

for all j, o, U_1, U_2 .

Similar to Section 3.1, let us define a truncated control variate through

$$M_{\Delta,T}^{(2),trunc} := \sum_{j=1}^J \sum_{\substack{(U_1,U_2) \in \mathcal{A} \\ |U_2| + |\mathcal{K}_2| + \frac{1}{2} |\mathcal{K}_1| \leq 1}} \sum_{o \in \{1,2\}^{U_1}} a_{j,o,U_1,U_2}(X_{\Delta,(j-1)\Delta}) \prod_{r \in U_1} H_{o_r}(\xi_j^r) \prod_{(k,l) \in U_2} V_j^{kl}. \quad (3.5)$$

Next we derive the truncation error that arises from replacing $M_{\Delta,T}^{(2)}$ by $M_{\Delta,T}^{(2),trunc}$.

Theorem 3.8. Suppose that all the functions μ^k and σ^{ki} , $k \in \{1, \dots, d\}$, $i \in \{1, \dots, m\}$ are bounded, the function f is three times continuously differentiable with bounded partial derivatives up to order 3, and all the

functions μ^k, σ^{ki} are five times continuously differentiable with bounded partial derivatives up to order 5. Then it holds (cf. Proposition 2.5)

$$\text{Var} \left[f(X_{\Delta,T}) - M_{\Delta,T}^{(2),trunc} \right] \lesssim \Delta^2. \quad (3.6)$$

4. GENERIC REGRESSION ALGORITHM

In the previous sections we have given several representations for control variates. Now we discuss how to compute the coefficients in these representations via regression. For the sake of clarity, we focus on second order schemes and control variate (3.5) with coefficients given by (2.10).

4.1. Monte Carlo regression

Fix a Q -dimensional vector of real-valued functions $\psi = (\psi^1, \dots, \psi^Q)$ on \mathbb{R}^d . Simulate a big number² N of independent “training paths” of the discretised diffusion $X_{\Delta,j\Delta}$, $j = 0, \dots, J$. In what follows these N training paths are denoted by D_N^{tr} :

$$D_N^{tr} := \left\{ (X_{\Delta,j\Delta}^{tr,(i)})_{j=0,\dots,J} : i = 1, \dots, N \right\}.$$

Let $\alpha_{j,o,U_1,U_2} = (\alpha_{j,o,U_1,U_2}^1, \dots, \alpha_{j,o,U_1,U_2}^Q)$, where $j \in \{1, \dots, J\}$, $(U_1, U_2) \in \mathcal{A}$, $|U_2| + |\mathcal{K}_2| + \frac{1}{2}|\mathcal{K}_1| \leq 1$, $o \in \{1, 2\}^{U_1}$, be a solution of the following least squares optimisation problem:

$$\text{argmin}_{\alpha \in \mathbb{R}^n} \sum_{i=1}^N \left[\zeta_{j,o,U_1,U_2}^{tr,(i)} - \alpha^1 \psi^1(X_{\Delta,(j-1)\Delta}^{tr,(i)}) - \dots - \alpha^Q \psi^Q(X_{\Delta,(j-1)\Delta}^{tr,(i)}) \right]^2$$

with

$$\zeta_{j,o,U_1,U_2}^{tr,(i)} := f(X_{\Delta,T}^{tr,(i)}) \prod_{r \in U_1} H_{o_r} \left((\xi_j^{tr,(i)})^r \right) \prod_{(k,l) \in U_2} (V_j^{tr,(i)})^{kl}.$$

Define an estimate for the coefficient function a_{j,o,U_1,U_2} via

$$\hat{a}_{j,o,U_1,U_2}(x) := \hat{a}_{j,o,U_1,U_2}(x, D_N^{tr}) := \alpha_{j,o,U_1,U_2}^1 \psi^1(x) + \dots + \alpha_{j,o,U_1,U_2}^Q \psi^Q(x), \quad x \in \mathbb{R}^d.$$

The intermediate expression $\hat{a}_{j,o,U_1,U_2}(x, D_N^{tr})$ in the above formula emphasises that the estimates \hat{a}_{j,o,U_1,U_2} of the functions a_{j,o,U_1,U_2} are random in that they depend on the simulated training paths. The cost of computing α_{j,o,U_1,U_2} is of order $O(NQ^2)$, since each α_{j,o,U_1,U_2} is of the form $\alpha_{j,o,U_1,U_2} = B^{-1}b$ with

$$B_{k,l} := \frac{1}{N} \sum_{i=1}^N \psi^k(X_{\Delta,(j-1)\Delta}^{tr,(i)}) \psi^l(X_{\Delta,(j-1)\Delta}^{tr,(i)}) \quad (4.1)$$

and

$$b_k := \frac{1}{N} \sum_{i=1}^N \psi^k(X_{\Delta,(j-1)\Delta}^{tr,(i)}) \zeta_{j,o,U_1,U_2}^{tr,(i)},$$

$k, l \in \{1, \dots, Q\}$. The cost of approximating the family of the coefficient functions a_{j,o,U_1,U_2} , $j \in \{1, \dots, J\}$, $(U_1, U_2) \in \mathcal{A}$, $|U_2| + |\mathcal{K}_2| + \frac{1}{2}|\mathcal{K}_1| \leq 1$, $o \in \{1, 2\}^{U_1}$, is of order $O(Jm(m+1)NQ^2)$.

²In the complexity analysis below we show how large N is required to be in order to provide an estimate within some given tolerance.

4.2. Summary of the algorithm

The algorithm consists of two phases: training phase and testing phase. In the training phase, we simulate N independent training paths D_N^{tr} and construct regression estimates $\hat{a}_{j,o,U_1,U_2}(\cdot, D_N^{tr})$ for the coefficients $a_{j,o,U_1,U_2}(\cdot)$. In the testing phase, independently from D_N^{tr} we simulate N_0 independent testing paths $(X_{\Delta,j\Delta}^{(i)})_{j=0,\dots,J}$, $i = 1, \dots, N_0$, and build the Monte Carlo estimator for $\mathbb{E}[f(X_T)]$ as

$$\mathcal{E} = \frac{1}{N_0} \sum_{i=1}^{N_0} \left(f(X_{\Delta,T}^{(i)}) - \widehat{M}_{\Delta,T}^{(2),trunc,(i)} \right), \quad (4.2)$$

where

$$\widehat{M}_{\Delta,T}^{(2),trunc,(i)} := \sum_{j=1}^J \sum_{\substack{(U_1,U_2) \in \mathcal{A} \\ |U_2| + |\mathcal{K}_2| + \frac{1}{2}|\mathcal{K}_1| \leq 1}} \sum_{o \in \{1,2\}^{U_1}} \hat{a}_{j,o,U_1,U_2}(X_{\Delta,(j-1)\Delta}^{(i)}, D_N^{tr}) \prod_{r \in U_1} H_{o_r}(\xi_j^{r,(i)}) \prod_{(k,l) \in U_2} V_j^{kl,(i)} \quad (4.3)$$

(cf. with (2.12)). Due to the martingale transform structure in (4.3) (recall footnote 1 on page 19), we have $\mathbb{E}[\widehat{M}_{\Delta,T}^{(2),trunc,(i)} | D_N^{tr}] = 0$, hence $\mathbb{E}[\mathcal{E} | D_N^{tr}] = \mathbb{E}[f(X_{\Delta,T}^{(i)}) - \widehat{M}_{\Delta,T}^{(2),trunc,(i)} | D_N^{tr}] = \mathbb{E}[f(X_{\Delta,T})]$, and we obtain (cf. (3.6))

$$\begin{aligned} \text{Var}[\mathcal{E}] &= \mathbb{E}[\text{Var}(\mathcal{E} | D_N^{tr})] + \text{Var}[\mathbb{E}(\mathcal{E} | D_N^{tr})] = \mathbb{E}[\text{Var}(\mathcal{E} | D_N^{tr})] \\ &= \frac{1}{N_0} \mathbb{E} \left[\text{Var} \left(f(X_{\Delta,T}^{(1)}) - \widehat{M}_{\Delta,T}^{(2),trunc,(1)} \mid D_N^{tr} \right) \right] = \frac{1}{N_0} \text{Var} \left[f(X_{\Delta,T}^{(1)}) - \widehat{M}_{\Delta,T}^{(2),trunc,(1)} \right]. \end{aligned}$$

Summarising, we have

$$\begin{aligned} \mathbb{E}[\mathcal{E}] &= \mathbb{E}[f(X_{\Delta,T})], \\ \text{Var}[\mathcal{E}] &= \frac{1}{N_0} \text{Var} \left[f(X_{\Delta,T}^{(1)}) - \widehat{M}_{\Delta,T}^{(2),trunc,(1)} \right]. \end{aligned} \quad (4.4)$$

Notice that the result of (4.4) indeed requires the computations above and cannot be stated right from the outset because the summands in (4.2) are dependent (through D_N^{tr}).

This concludes the description of the generic regression algorithm for constructing the control variate. Further details, such as bounds for the right-hand side of (4.4), depend on a particular implementation, i.e. on the quality of the chosen basis functions.

5. COMPLEXITY ANALYSIS

In this section we extend the complexity analysis presented in [1] to the case of the ‘‘TRCV’’ (truncated RCV) algorithm. Below we only sketch the main results for the second order schemes. We make the following assumption (cf. [2] and [4]):

- (B5) The functions $a_{j,o,U_1,U_2}(x)$ can be well approximated by the functions from $\Psi_Q := \text{span}(\{\psi_1, \dots, \psi_Q\})$, in the sense that there are constants $\kappa > 0$ and $C_\kappa > 0$ such that

$$\inf_{g \in \Psi_Q} \int_{\mathbb{R}^d} (a_{j,o,U_1,U_2}(x) - g(x))^2 \mathbb{P}_{\Delta,j-1}(dx) \leq \frac{C_\kappa}{Q^\kappa},$$

where $\mathbb{P}_{\Delta,j-1}$ denotes the distribution of $X_{\Delta,(j-1)\Delta}$.

Remark 5.1. Note that (B5) is a natural condition to be satisfied for good choices of Ψ_Q . For instance, under appropriate assumptions, in the case of *piecewise polynomial regression* as described in [1], (B5) is satisfied with $\kappa = \frac{2\nu(p+1)}{2d(p+1)+d\nu}$, where the parameters p and ν are explained in [1].

In Lemma 5.2 below we present an L^2 -upper bound for the estimation error of the TRCV algorithm. To this end, we need to describe more precisely, how exactly the regression-based approximations \tilde{a}_{j,o,U_1,U_2} are constructed:

Let functions $\hat{a}_{j,o,U_1,U_2}(x)$ be obtained by regression onto the set of basis functions $\{\psi_1, \dots, \psi_Q\}$, while the approximations $\tilde{a}_{j,o,U_1,U_2}(x)$ of the TRCV algorithm be the truncated estimates, which are defined as follows

$$\tilde{a}_{j,o,U_1,U_2}(x) := T_{\tilde{A}\Delta_{U_1,U_2}} \hat{a}_{j,o,U_1,U_2}(x) := \begin{cases} \hat{a}_{j,o,U_1,U_2}(x) & \text{if } |\hat{a}_{j,o,U_1,U_2}(x)| \leq \tilde{A}\Delta_{U_1,U_2}, \\ \tilde{A}\Delta_{U_1,U_2} \operatorname{sgn} \hat{a}_{j,o,U_1,U_2}(x) & \text{otherwise} \end{cases}, \quad (5.1)$$

where Δ_{U_1,U_2} and \tilde{A} are given in (3.3) and (3.4).

Lemma 5.2. *Under (B1)–(B5), we have*

$$\mathbb{E} \|\tilde{a}_{j,o,U_1,U_2} - a_{j,o,U_1,U_2}\|_{L^2(\mathbb{P}_{\Delta,j-1})}^2 \leq \tilde{c}(\Sigma + \tilde{A}^2 \Delta_{U_1,U_2}^2 (\log N + 1)) \frac{Q}{N} + \frac{8C_\kappa}{Q^\kappa}, \quad (5.2)$$

where \tilde{c} is a universal constant.

Notice that the expectation in the left-hand side of (5.2) means averaging over the randomness in D_N^{tr} . Let $(X_{\Delta,j\Delta})_{j=0,\dots,J}$ be a testing path, which is independent of the training paths D_N^{tr} . We define

$$\tilde{M}_{\Delta,T}^{(2),trunc} := \sum_{j=1}^J \sum_{\substack{(U_1,U_2) \in \mathcal{A} \\ |U_2| + |\mathcal{K}_2| + \frac{1}{2}|\mathcal{K}_1| \leq 1}} \sum_{o \in \{1,2\}^{U_1}} \tilde{a}_{j,o,U_1,U_2}(X_{\Delta,(j-1)\Delta}, D_N^{tr}) \prod_{r \in U_1} H_{o_r}(\xi_j^r) \prod_{(k,l) \in U_2} V_j^{kl} \quad (5.3)$$

(cf. (3.5)). Lemma 5.2 now allows to bound the variance $\operatorname{Var}[f(X_{\Delta,T}) - \tilde{M}_{\Delta,T}^{(2),trunc}]$ from above.³

Theorem 5.3. *Under (B1)–(B5), it holds*

$$\operatorname{Var}[f(X_{\Delta,T}) - \tilde{M}_{\Delta,T}^{(2),trunc}] \lesssim \Delta^2 + Jm(m+1) \left(\tilde{c}(\Sigma + \tilde{A}^2 \Delta (\log N + 1)) \frac{Q}{N} + \frac{8C_\kappa}{Q^\kappa} \right).$$

5.1. Complexity of the TRCV approach

Let us study the complexity of the TRCV approach. The overall cost is of order $JQ \max\{NQ, N_0\}$, provided that we only track the constants which tend to ∞ when $\varepsilon \searrow 0$ with ε being the accuracy to be achieved. That is, the constants, such as d, m, κ, C_κ , are ignored. We have the following constraints

$$\max \left\{ \frac{1}{J^4}, \frac{1}{J^2 N_0}, \frac{JQ}{N N_0}, \frac{J}{Q^\kappa N_0} \right\} \lesssim \varepsilon^2, \quad (5.4)$$

where the first term comes from the squared bias of the estimator and the remaining three ones come from the variance of the estimator (see Theorem 5.3 as well as footnote 3 on page 26). We get the following result.

³Notice that the variance of the TRCV estimate $\frac{1}{N_0} \sum_{i=1}^{N_0} [f(X_{\Delta,T}^{(i)}) - \tilde{M}_{\Delta,T}^{(2),trunc,(i)}]$ with N_0 testing paths is $\frac{1}{N_0} \operatorname{Var}[f(X_{\Delta,T}) - \tilde{M}_{\Delta,T}^{(2),trunc}]$ (cf. (4.4)).

Theorem 5.4. *For the TRCV approach with the second order weak schemes under (B1)–(B5), it is optimal to choose the orders of parameters as follows (cf. [4])*

$$J \asymp \varepsilon^{-\frac{1}{2}}, \quad Q \asymp \varepsilon^{-\frac{5}{4\kappa+4}}, \quad N \asymp \varepsilon^{-\frac{5}{4}}, \quad N_0 \asymp NQ \asymp \varepsilon^{-\frac{5\kappa+10}{4\kappa+4}},$$

provided that $\kappa > 1$.⁴ Thus, we have for the complexity

$$\mathcal{C}_{TRCV} \asymp JNQ^2 \asymp JN_0Q \asymp \varepsilon^{-\frac{7\kappa+17}{4\kappa+4}}. \quad (5.5)$$

Remark 5.5. (i) For the sake of comparison with the SMC and MLMC approaches, we recall at this point that their complexities are

$$\mathcal{C}_{SMC} \asymp \varepsilon^{-2.5} \quad \text{and} \quad \mathcal{C}_{MLMC} \asymp \varepsilon^{-2}$$

at best (we are considering the second order scheme).

(ii) Complexity estimate (5.5) shows that one can go beyond the complexity order ε^{-2} , provided that $\kappa > 9$, and that we can achieve the complexity order $\varepsilon^{-1.75-\delta}$, for arbitrarily small $\delta > 0$, provided κ is large enough.

(iii) The complexity of the TRCV approach is the same that we obtain for the RCV approach (where the “complete” control variate (2.12) is estimated), since the second constraint in (5.4), which does not arise for the RCV approach, is the only inactive one in this case. That is why we truncated $M_{\Delta,T}^{(2),trunc}$ in (3.5) at the level $|U_2| + |\mathcal{K}_2| + \frac{1}{2}|\mathcal{K}_1| \leq 1$. For instance, if we had used a control variate of the form (cf. (3.1))

$$\begin{aligned} & \sum_{j=1}^J \sum_{\substack{(U_1, U_2) \in \mathcal{A} \\ |U_2| + |\mathcal{K}_2| + \frac{1}{2}|\mathcal{K}_1| = \frac{1}{2}}} \sum_{o \in \{1,2\}^{U_1}} a_{j,o,U_1,U_2}(X_{\Delta,(j-1)\Delta}) \prod_{r \in U_1} H_{o_r}(\xi_j^r) \prod_{(k,l) \in U_2} V_j^{kl} \\ &= \sum_{j=1}^J \sum_{i=1}^m a_{j,1,i,\emptyset}(X_{\Delta,(j-1)\Delta}) \xi_j^i \end{aligned}$$

with $a_{j,1,i,\emptyset}(x) = \mathbb{E} [f(X_{\Delta,T}) \xi_j^i \mid X_{\Delta,(j-1)\Delta} = x]$, the bound for the variance in (3.6) would have been of order Δ and due to the resulting constraint $\frac{1}{JN_0} \lesssim \varepsilon^2$, we would have obtained worse complexities than ε^{-2} , since $\mathcal{C}_{TRCV} \gtrsim JN_0$.

6. NUMERICAL RESULTS

The results below are based on program codes written and vectorised in MATLAB and running on a Linux 64-bit operating system.

Let us consider the following SDE for $d = m = 5$ (cf. [1])

$$\begin{aligned} dX_t^i &= -\sin(X_t^i) \cos^3(X_t^i) dt + \cos^2(X_t^i) dW_t^i, \quad X_0^i = 0, \quad i \in \{1, 2, 3, 4\}, \\ dX_t^5 &= \sum_{i=1}^4 \left[-\frac{1}{2} \sin(X_t^i) \cos^2(X_t^i) dt + \cos(X_t^i) dW_t^i \right] + dW_t^5, \quad X_0^5 = 0. \end{aligned} \quad (6.1)$$

⁴Performing the full complexity analysis via Lagrange multipliers one can see that these parameter values are *not* optimal if $\kappa \leq 1$ (a Lagrange multiplier corresponding to a “ ≤ 0 ” constraint is negative). Recall that in the case of piecewise polynomial regression (see [1] and recall Remark 5.1) we have $\kappa = \frac{2\nu(p+1)}{2d(p+1)+d\nu}$. Let us note that in [1] it is required to choose the parameters p and ν according to $p > \frac{d-2}{2}$ and $\nu > \frac{2d(p+1)}{2(p+1)-d}$, which implies that $\kappa > 1$, for κ expressed via p and ν by the above formula.

The solution of (6.1) is given by

$$\begin{aligned} X_t^i &= \arctan(W_t^i), \quad i \in \{1, 2, 3, 4\}, \\ X_t^5 &= \sum_{i=1}^4 \operatorname{arsinh}(W_t^i) + W_t^5. \end{aligned}$$

for $t \in [0, 1]$. Further, we consider the functional

$$f(x) = \cos\left(\sum_{i=1}^5 x^i\right) - 20 \sum_{i=1}^4 \sin(x^i),$$

that is, we have

$$\mathbb{E}[f(X_1)] = (\mathbb{E}[\cos(\arctan(W_1^1) + \operatorname{arsinh}(W_1^1))])^4 \mathbb{E}[\cos(W_1^5)] \approx 0.002069.$$

Here we consider weak schemes of the second order and compare the numerical performances of the SMC, MLMC, RCV, TRCV and TSRCV approaches. The latter one is the truncated version of the SRCV approach of [4]. Like the RCV algorithm, the SRCV one is based on (2.12), the difference is only in how to implement the approximations of the coefficients a_{j,o,U_1,U_2} in practice (while the RCV algorithm is a direct Monte Carlo regression, in the SRCV algorithm the regression is combined with a kind of “stratification”; see [4] for more detail). Therefore, the idea of the truncation (i.e. replacing (2.12) with (3.5)) applies also to the SRCV approach and gives us the TSRCV one.

For simplicity we implemented a global regression for the RCV, TRCV and TSRCV approaches (i.e. the one without considering the truncation operator in (5.1), as a part of the general description in Section 4). More precisely, we use quadratic polynomials (that is $\prod_{i=1}^5 x_i^{l_i}$, where $l_1, \dots, l_5 \in \{0, 1, 2\}$ and $\sum_{i=1}^5 l_i \leq 2$) as well as f as basis functions, hence Ψ_Q consists of $Q = \binom{7}{5} + 1 = 22$ basis functions.

Note that we do not need to consider random variables V_j^{kl} in the second order weak scheme, since $\mathcal{L}^k \sigma^{rl}(x) = 0$ for $k \neq l$ (see (2.8)). This gives us less terms for the RCV approach, namely $3^m - 1 = 242$ rather than $3^m 2^{\frac{m(m-1)}{2}} - 1 = 248831$ terms in (2.12) (the factor $2^{\frac{m(m-1)}{2}} \equiv 1024$ is no longer present). As for the TRCV and TSRCV approaches, this gives us only $\frac{m(m+3)}{2} = 20$ compared to $m(m+1) = 30$ terms in (3.5).

We choose $\kappa = 1.2$, which is related to the piecewise polynomial regression with polynomial degree $p = 2$ (comparable to our setting) and the limiting case $\nu \rightarrow \infty$ (see footnote 4 on page 27). Moreover, for each $\varepsilon = 2^{-i}$, $i \in \{2, 3, 4, 5, 6\}$, we set the parameters J , N and N_0 for the RCV, TRCV and TSRCV approaches as follows (compare with the formulas in Subsection 5.1):

$$J = \lceil \varepsilon^{-0.5} \rceil, \quad N = \begin{cases} 512 \cdot \lceil \varepsilon^{-1.25} \rceil, & \text{RCV, TRCV,} \\ 2048 \cdot \lceil \varepsilon^{-1.25} \rceil, & \text{TSRCV,} \end{cases} \quad N_0 = 512 \cdot \lceil \varepsilon^{-1.82} \rceil.$$

The factors 512 and 2048 are here for stability purposes. For the TRCV and SMC algorithms we additionally consider $\varepsilon = 2^{-7}$, which produces a picture with approximately equal maximal computational time (that is, the time corresponding to the best accuracy) for all algorithms. Next we estimate the numerical complexity for the RCV, TRCV and TSRCV approaches by means of 100 independent simulations and compare it with the one for the SMC and MLMC approach, for which we use the same output as in [1]. As can be seen from Figure 1, the estimated numerical complexity is about $\text{RMSE}^{-1.85}$ for the RCV approach, $\text{RMSE}^{-1.83}$ for the TRCV approach, $\text{RMSE}^{-1.53}$ for the TSRCV approach, $\text{RMSE}^{-2.67}$ for the SMC approach and $\text{RMSE}^{-2.01}$ for the MLMC approach, which we get by regressing the log-time (logarithmic computing time of the whole algorithm in seconds) vs. log-RMSE. Beyond the numerical complexities we observe that the truncation effect from RCV algorithm to its truncated versions is huge. While we have poor results for the RCV approach (as

in [1]), i.e. in this region of ε -values the RCV approach is numerically outperformed by the other ones, the TRCV and TSRCV approaches work best (even better than the SMC and MLMC approaches).

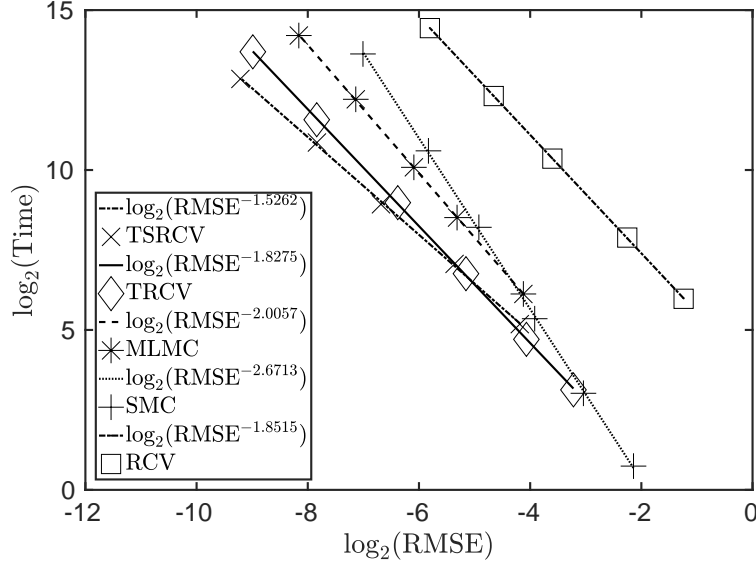


FIGURE 1. Numerical complexities of the RCV, TRCV, TSRCV, SMC and MLMC approaches.

7. PROOFS

Proof of Theorem 1.1

We begin with the following remark. Assumptions (1.5) and (1.6) together with the Cauchy-Schwarz inequality $|\mathbb{E}[XY|\mathcal{G}]| \leq \sqrt{\mathbb{E}[X^2|\mathcal{G}]\mathbb{E}[Y^2|\mathcal{G}]}$ imply that the following generalisation of (1.6) is satisfied: for any $n_1, n_2 \in \mathbb{N}$, $\alpha, \beta \in \mathbb{N}_0^d$, with $1 \leq |\alpha| \leq K$, $1 \leq |\beta| \leq K$, $\alpha \neq \beta$, it holds

$$\left| \mathbb{E} \left[\left(D^\alpha \Phi_{\Delta, l+1}^k(G_{l,j}(x)) \right)^{n_1} \left(D^\beta \Phi_{\Delta, l+1}^k(G_{l,j}(x)) \right)^{n_2} \middle| \mathcal{G}_l \right] \right| \leq C_{n_1, n_2} \Delta \quad (7.1)$$

for some appropriate constants $C_{n_1, n_2} > 0$.

Let us begin with the case $K = 1$. We have for some $k, r \in \{1, \dots, d\}$

$$\frac{\partial}{\partial x^r} G_{l+1, j}^k(x) = \sum_{s=1}^d \frac{\partial}{\partial x^s} \Phi_{\Delta, l+1}^k(G_{l,j}(x)) \frac{\partial}{\partial x^r} G_{l,j}^s(x) =: \sum_{s=1}^d \gamma_s$$

and $\frac{\partial}{\partial x^r} G_{j+1, j}^s(x) = \frac{\partial}{\partial x^r} \Phi_{\Delta}^s(x, \xi_{j+1})$, where $G_{l+1, j}^s$ and Φ_{Δ}^s , $s \in \{1, \dots, d\}$, denote the s -th component of the functions $G_{l+1, j}$ and Φ_{Δ} . Hence

$$\mathbb{E} \left[\left(\frac{\partial}{\partial x^r} G_{l+1, j}^k(x) \right)^2 \right] \leq \mathbb{E} \left[\gamma_k^2 + \sum_{s: s \neq k} (2\gamma_k \gamma_s + (d-1)\gamma_s^2) \right].$$

For an arbitrary $j \in \{0, \dots, J-1\}$, denote

$$\rho_{l+1,n,1}^{r,s} := \mathbb{E} \left[\left(\frac{\partial}{\partial x^r} G_{l+1,j}^s(x) \right)^n \right],$$

then, due to (1.5) and (7.1), we get for $l = j, \dots, J-1$,

$$\rho_{l+1,2,1}^{r,k} \leq (1 + A_2 \Delta) \rho_{l,2,1}^{r,k} + \sum_{s: s \neq k} \left(C_{1,1} \Delta (\rho_{l,2,1}^{r,k} + \rho_{l,2,1}^{r,s}) + (d-1) B_2 \Delta \rho_{l,2,1}^{r,s} \right).$$

Further, denote

$$\rho_{l+1,n,1}^r := \sum_{s=1}^d \rho_{l+1,n,1}^{r,s},$$

then we get

$$\rho_{l+1,2,1}^r \leq (1 + A_2 \Delta) \rho_{l,2,1}^r + 2(d-1) C_{1,1} \Delta \rho_{l,2,1}^r + (d-1)^2 B_2 \Delta \rho_{l,2,1}^r.$$

This gives us

$$\rho_{l+1,2,1}^r \leq (1 + \kappa_1 \Delta) \rho_{l,2,1}^r$$

for some constant $\kappa_1 > 0$, leading to

$$\rho_{l,2,1}^r \leq (1 + \kappa_1 \Delta)^{l-j-1} \rho_{j+1,2,1}^r, \quad l = j+1, \dots, J-1, \quad (7.2)$$

where

$$\rho_{j+1,2,1}^r = \sum_{s=1}^d \mathbb{E} \left[\left(\frac{\partial}{\partial x^r} \Phi_{\Delta}^s(x, \xi_{j+1}) \right)^2 \right],$$

which is bounded due to (1.5). Together with (7.2) we obtain the boundedness of $\{\rho_{J,2,1}^r : J \in \mathbb{N}\}$ and hence the boundedness of

$$\begin{aligned} \left| \frac{\partial}{\partial x^r} q_j(x) \right| &\leq \sum_{s=1}^d \mathbb{E} \left| \frac{\partial}{\partial x^s} f(G_{J,j}(x)) \frac{\partial}{\partial x^r} G_{J,j}^s(x) \right| \\ &\leq \sum_{s=1}^d \sqrt{\mathbb{E} \left[\left(\frac{\partial}{\partial x^s} f(G_{J,j}(x)) \right)^2 \right] \rho_{J,2,1}^{r,s}} \\ &\leq \sqrt{d \sum_{s=1}^d \mathbb{E} \left[\left(\frac{\partial}{\partial x^s} f(G_{J,j}(x)) \right)^2 \right] \rho_{J,2,1}^{r,s}} \leq \text{const} \sqrt{\rho_{J,2,1}^r} \end{aligned}$$

for all $r \in \{1, \dots, d\}$, since f is assumed to be continuously differentiable with bounded partial derivatives.

Let us proceed with the case $K = 2$. We have, due to $(\sum_{k=1}^d a_k)^n \leq d^{n-1} \sum_{k=1}^d a_k^n$,

$$\begin{aligned} &\mathbb{E} \left[\left(\frac{\partial}{\partial x^r} G_{l+1,j}^k(x) \right)^4 \right] \\ &\leq \mathbb{E} \left[\gamma_k^4 + \sum_{s: s \neq k} (4\gamma_k^3 \gamma_s + 6(d-1)\gamma_k^2 \gamma_s^2 + 4(d-1)^2 \gamma_k \gamma_s^3 + (d-1)^3 \gamma_s^4) \right] \end{aligned}$$

and thus, due to $4a^3b \leq 3a^4 + b^4$ and $2a^2b^2 \leq a^4 + b^4$,

$$\begin{aligned} \rho_{l+1,4,1}^{r,k} \leq & (1 + A_4\Delta)\rho_{l,4,1}^{r,k} + \sum_{s: s \neq k} \left(C_{3,1}\Delta(3\rho_{l,4,1}^{r,k} + \rho_{l,4,1}^{r,s}) + 3(d-1)C_{2,2}\Delta(\rho_{l,4,1}^{r,k} + \rho_{l,4,1}^{r,s}) \right. \\ & \left. + (d-1)^2C_{1,3}\Delta(\rho_{l,4,1}^{r,k} + 3\rho_{l,4,1}^{r,s}) + (d-1)^3B_4\Delta\rho_{l,4,1}^{r,s} \right). \end{aligned}$$

This gives us

$$\begin{aligned} \rho_{l+1,4,1}^r \leq & (1 + A_4\Delta)\rho_{l,4,1}^r + 4(d-1)C_{3,1}\Delta\rho_{l,4,1}^r + 6(d-1)^2C_{2,2}\Delta\rho_{l,4,1}^r \\ & + 4(d-1)^3C_{1,3}\Delta\rho_{l,4,1}^r + (d-1)^4B_4\Delta\rho_{l,4,1}^r. \end{aligned}$$

Hence, we obtain

$$\rho_{l+1,4,1}^r \leq (1 + \kappa_2\Delta)\rho_{l,4,1}^r,$$

for some constant $\kappa_2 > 0$, leading to

$$\rho_{l,4,1}^r \leq (1 + \kappa_2\Delta)^{l-j-1}\rho_{j+1,4,1}^r, \quad l = j+1, \dots, J-1,$$

where

$$\rho_{j+1,4,1}^r = \sum_{s=1}^d \mathbb{E} \left[\left(\frac{\partial}{\partial x^r} \Phi_{\Delta}^s(x, \xi_{j+1}) \right)^4 \right].$$

Next, we have for some $k, o, r \in \{1, \dots, d\}$

$$\begin{aligned} \frac{\partial^2}{\partial x^r \partial x^o} G_{l+1,j}^k(x) &= \sum_{s=1}^d \frac{\partial}{\partial x^s} \Phi_{\Delta, l+1}^k(G_{l,j}(x)) \frac{\partial^2}{\partial x^r \partial x^o} G_{l,j}^s(x) \\ &+ \sum_{s,u=1}^d \frac{\partial^2}{\partial x^s \partial x^u} \Phi_{\Delta, l+1}^k(G_{l,j}(x)) \frac{\partial}{\partial x^r} G_{l,j}^s(x) \frac{\partial}{\partial x^o} G_{l,j}^u(x) \\ &=: \sum_{s=1}^d \eta_{1,s} + \sum_{s,u=1}^d \eta_{2,s,u} \end{aligned}$$

and $\frac{\partial^2}{\partial x^r \partial x^o} G_{j+1,j}^s(x) = \frac{\partial^2}{\partial x^r \partial x^o} \Phi_{\Delta}^s(x, \xi_{j+1})$. Hence

$$\begin{aligned} & \mathbb{E} \left[\left(\frac{\partial^2}{\partial x^r \partial x^o} G_{l+1,j}^k(x) \right)^2 \right] \\ \leq & \mathbb{E} \left[\eta_{1,k}^2 + \sum_{s: s \neq k} (2\eta_{1,k}\eta_{1,s} + (d-1)\eta_{1,s}^2) + 2 \sum_{s,u,v=1}^d \eta_{1,v}\eta_{2,s,u} + d^2 \sum_{s,u=1}^d \eta_{2,s,u}^2 \right] \end{aligned}$$

Denote

$$\rho_{l+1,n,2}^{r,o,s} = \mathbb{E} \left[\left(\frac{\partial^2}{\partial x^r \partial x^o} G_{l+1,j}^s(x) \right)^n \right],$$

then we get, due to

$$\begin{aligned} 2\mathbb{E}[XYZ] &\leq 2\sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^4]}\sqrt{\mathbb{E}[Z^4]} \leq \mathbb{E}[X^2] + \sqrt{\mathbb{E}[Y^4]}\sqrt{\mathbb{E}[Z^4]} \\ &\leq \mathbb{E}[X^2] + \frac{1}{2}(\mathbb{E}[Y^4] + \mathbb{E}[Z^4]), \end{aligned}$$

as well as (1.5) and (7.1),

$$\begin{aligned} \rho_{l+1,2,2}^{r,o,k} &\leq (1 + A_2\Delta)\rho_{l,2,2}^{r,o,k} + \sum_{s: s \neq k} \left(C_{1,1}\Delta(\rho_{l,2,2}^{r,o,k} + \rho_{l,2,2}^{r,o,s}) + (d-1)B_2\Delta\rho_{l,2,2}^{r,o,s} \right) \\ &\quad + \sum_{s,u,v=1}^d C_{1,1}\Delta \left(\rho_{l,2,2}^{r,o,v} + \frac{1}{2}(\rho_{l,4,1}^{r,s} + \rho_{l,4,1}^{o,u}) \right) \\ &\quad + d^2 \sum_{s,u=1}^d B_2\Delta \frac{1}{2} (\rho_{l,4,1}^{r,s} + \rho_{l,4,1}^{o,u}). \end{aligned}$$

Further, denote

$$\rho_{l+1,n,2}^{r,o} = \sum_{s=1}^d \rho_{l+1,n,2}^{r,o,s},$$

then we get for $l = j+1, \dots, J-1$,

$$\begin{aligned} \rho_{l+1,2,2}^{r,o} &\leq (1 + A_2\Delta)\rho_{l,2,2}^{r,o} + 2(d-1)C_{1,1}\Delta\rho_{l,2,2}^{r,o} + (d-1)^2B_2\Delta\rho_{l,2,2}^{r,o} \\ &\quad + d^3C_{1,1}\Delta \left(\rho_{l,2,2}^{r,o} + \frac{1}{2}(\rho_{l,4,1}^r + \rho_{l,4,1}^o) \right) + d^4B_2\Delta \frac{1}{2} (\rho_{l,4,1}^r + \rho_{l,4,1}^o). \end{aligned}$$

This gives us

$$\rho_{l+1,2,2}^{r,o} \leq (1 + \kappa_3\Delta)\rho_{l,2,2}^{r,o} + \kappa_4\Delta,$$

for some constants $\kappa_3, \kappa_4 > 0$, leading to

$$\rho_{l,2,2}^{r,o} \leq (1 + \kappa_3\Delta)^{l-j-1} \rho_{j+1,2,2}^{r,o} + \kappa_5, \quad l = j+1, \dots, J-1,$$

where $\kappa_5 > 0$ and

$$\rho_{j+1,2,2}^{r,o} = \sum_{s=1}^d \mathbb{E} \left[\left(\frac{\partial^2}{\partial x^r \partial x^o} \Phi_\Delta^s(x, \xi_{j+1}) \right)^2 \right].$$

Thus, we obtain the boundedness of

$$\begin{aligned} \left| \frac{\partial^2}{\partial x^r \partial x^o} q_j(x) \right| &\leq \sum_{s=1}^d \mathbb{E} \left| \frac{\partial}{\partial x^s} f(G_{J,j}(x)) \frac{\partial^2}{\partial x^r \partial x^o} G_{J,j}^s(x) \right| \\ &\quad + \sum_{s,u=1}^d \mathbb{E} \left| \frac{\partial^2}{\partial x^s \partial x^u} f(G_{J,j}(x)) \frac{\partial}{\partial x^r} G_{J,j}^s(x) \frac{\partial}{\partial x^o} G_{J,j}^u(x) \right| \\ &\leq \sum_{s=1}^d \sqrt{\mathbb{E} \left[\left(\frac{\partial}{\partial x^s} f(G_{J,j}(x)) \right)^2 \right] \rho_{J,2,2}^{r,o,s}} \\ &\quad + \sum_{s,u=1}^d \sqrt{\mathbb{E} \left[\left(\frac{\partial^2}{\partial x^s \partial x^u} f(G_{J,j}(x)) \right)^2 \right] \sqrt[4]{\rho_{J,4,1}^{r,s} \rho_{J,4,1}^{o,u}}} \end{aligned}$$

for all $r, o \in \{1, \dots, d\}$, since f is assumed to be twice continuously differentiable with bounded partial derivatives up to order 2.

Let us proceed with the final case $K = 3$. We have

$$\begin{aligned} & \mathbb{E} \left[\left(\frac{\partial}{\partial x^r} G_{l+1,j}^k(x) \right)^6 \right] \\ \leq & \mathbb{E} \left[\gamma_k^6 + \sum_{s: s \neq k} (6\gamma_k^5 \gamma_s + 15(d-1)\gamma_k^4 \gamma_s^2 + 20(d-1)^2 \gamma_k^3 \gamma_s^3 + 15(d-1)^3 \gamma_k^2 \gamma_s^4 \right. \\ & \left. + 6(d-1)^4 \gamma_k \gamma_s^5 + (d-1)^5 \gamma_s^6) \right] \end{aligned}$$

and thus, due to $6a^5b \leq 5a^6 + b^6$, $3a^4b^2 \leq 2a^6 + b^6$ and $2a^3b^3 \leq a^6 + b^6$,

$$\begin{aligned} \rho_{l+1,6,1}^{r,k} & \leq (1 + A_6 \Delta) \rho_{l,6,1}^{r,k} \\ & + \sum_{s: s \neq k} \left(C_{5,1} \Delta (5\rho_{l,6,1}^{r,k} + \rho_{l,6,1}^{r,s}) + 5(d-1)C_{4,2} \Delta (2\rho_{l,6,1}^{r,k} + \rho_{l,6,1}^{r,s}) \right. \\ & \quad \left. + 10(d-1)^2 C_{3,3} \Delta (\rho_{l,6,1}^{r,k} + \rho_{l,6,1}^{r,s}) + 5(d-1)^3 C_{2,4} \Delta (\rho_{l,6,1}^{r,k} + 2\rho_{l,6,1}^{r,s}) \right. \\ & \quad \left. + (d-1)^4 C_{1,5} \Delta (\rho_{l,6,1}^{r,k} + 5\rho_{l,6,1}^{r,s}) + (d-1)^5 B_6 \Delta \rho_{l,6,1}^{r,s} \right). \end{aligned}$$

This gives us

$$\begin{aligned} \rho_{l+1,6,1}^r & \leq (1 + A_6 \Delta) \rho_{l,6,1}^r + 6(d-1)C_{5,1} \Delta \rho_{l,6,1}^r + 15(d-1)^2 C_{4,2} \Delta \rho_{l,6,1}^r \\ & + 20(d-1)^3 C_{3,3} \Delta \rho_{l,6,1}^r + 15(d-1)^4 C_{2,4} \Delta \rho_{l,6,1}^r + 6(d-1)^5 C_{1,5} \Delta \rho_{l,6,1}^r \\ & + (d-1)^6 B_6 \Delta \rho_{l,6,1}^r. \end{aligned}$$

Hence, we obtain

$$\rho_{l+1,6,1}^r \leq (1 + \kappa_6 \Delta) \rho_{l,6,1}^r$$

for some constant $\kappa_6 > 0$, leading to

$$\rho_{l,6,1}^r \leq (1 + \kappa_6 \Delta)^{l-j-1} \rho_{j+1,6,1}^r, \quad l = j+1, \dots, J-1,$$

where

$$\rho_{j+1,6,1}^r = \sum_{s=1}^d \mathbb{E} \left[\left(\frac{\partial}{\partial x^r} \Phi_{\Delta}^s(x, \xi_{j+1}) \right)^6 \right]$$

Moreover, we have

$$\begin{aligned} & \mathbb{E} \left[\left(\frac{\partial}{\partial x^r} G_{l+1,j}^k(x) \right)^8 \right] \\ \leq & \mathbb{E} \left[\gamma_k^8 + \sum_{s: s \neq k} (8\gamma_k^7 \gamma_s + 28(d-1)\gamma_k^6 \gamma_s^2 + 56(d-1)^2 \gamma_k^5 \gamma_s^3 + 70(d-1)^3 \gamma_k^4 \gamma_s^4 \right. \\ & \left. + 56(d-1)^4 \gamma_k^3 \gamma_s^5 + 28(d-1)^5 \gamma_k^2 \gamma_s^6 + 8(d-1)^6 \gamma_k \gamma_s^7 + (d-1)^7 \gamma_s^8) \right] \end{aligned}$$

and thus, due to $8a^7b \leq 7a^8 + b^8$, $4a^6b^2 \leq 3a^8 + b^8$, $8a^5b^3 \leq 5a^8 + 3b^8$ and $2a^4b^4 \leq a^8 + b^8$,

$$\begin{aligned} \rho_{l+1,8,1}^{r,k} &\leq (1 + A_8\Delta)\rho_{l,8,1}^{r,k} \\ &+ \sum_{s \neq k} \left(C_{7,1}\Delta(7\rho_{l,8,1}^{r,k} + \rho_{l,8,1}^{r,s}) + 7(d-1)C_{6,2}\Delta(3\rho_{l,8,1}^{r,k} + \rho_{l,8,1}^{r,s}) \right. \\ &\quad + 7(d-1)^2C_{5,3}\Delta(5\rho_{l,8,1}^{r,k} + 3\rho_{l,8,1}^{r,s}) + 35(d-1)^3C_{4,4}\Delta(\rho_{l,8,1}^{r,k} + \rho_{l,8,1}^{r,s}) \\ &\quad + 7(d-1)^4C_{3,5}\Delta(3\rho_{l,8,1}^{r,k} + 5\rho_{l,8,1}^{r,s}) + 7(d-1)^5C_{2,6}\Delta(\rho_{l,8,1}^{r,k} + 3\rho_{l,8,1}^{r,s}) \\ &\quad \left. + (d-1)^6C_{1,7}\Delta(\rho_{l,8,1}^{r,k} + 7\rho_{l,8,1}^{r,s}) + (d-1)^7B_8\Delta\rho_{l,8,1}^{r,s} \right). \end{aligned}$$

This gives us

$$\begin{aligned} \rho_{l+1,8,1}^r &\leq (1 + A_8\Delta)\rho_{l,8,1}^r + 8(d-1)C_{7,1}\Delta\rho_{l,8,1}^r + 28(d-1)^2C_{6,2}\Delta\rho_{l,8,1}^r \\ &\quad + 56(d-1)^3C_{5,3}\Delta\rho_{l,8,1}^r + 70(d-1)^4C_{4,4}\Delta\rho_{l,8,1}^r \\ &\quad + 56(d-1)^5C_{3,5}\Delta\rho_{l,8,1}^r + 28(d-1)^6C_{2,6}\Delta\rho_{l,8,1}^r \\ &\quad + 8(d-1)^7C_{1,7}\Delta\rho_{l,8,1}^r + (d-1)^8B_8\Delta\rho_{l,8,1}^r. \end{aligned}$$

Hence, we obtain

$$\rho_{l+1,8,1}^r \leq (1 + \kappa_7\Delta)\rho_{l,8,1}^r,$$

for some constant $\kappa_7 > 0$, leading to

$$\rho_{l,8,1}^r \leq (1 + \kappa_7\Delta)^{l-j-1}\rho_{j+1,8,1}^r, \quad l = j+1, \dots, J-1,$$

where

$$\rho_{j+1,8,1}^r = \sum_{s=1}^d \mathbb{E} \left[\left(\frac{\partial}{\partial x^r} \Phi_{\Delta}^s(x, \xi_{j+1}) \right)^8 \right].$$

Moreover, we have

$$\begin{aligned} &\mathbb{E} \left[\left(\frac{\partial^2}{\partial x^r \partial x^o} G_{l+1,j}^k(x) \right)^4 \right] \\ &\leq \mathbb{E} \left[\eta_{1,k}^4 + \sum_{s \neq k} (4\eta_{1,k}^3\eta_{1,s} + 6(d-1)\eta_{1,k}^2\eta_{1,s}^2 + 4(d-1)^2\eta_{1,k}\eta_{1,s}^3 + (d-1)^3\eta_{1,s}^4) \right. \\ &\quad \left. + \sum_{s,u,v=1}^d (4d^2\eta_{1,v}^3\eta_{2,s,u} + 6d^3\eta_{1,v}^2\eta_{2,s,u}^2 + 4d^4\eta_{1,v}\eta_{2,s,u}^3) + d^6 \sum_{s,u=1}^d \eta_{2,s,u}^4 \right] \end{aligned}$$

and thus, due to $4a^3bc \leq 3a^4 + \frac{1}{2}(b^8 + c^8)$, $2a^2b^2c^2 \leq a^4 + \frac{1}{2}(b^8 + c^8)$ and $4ab^3c^3 \leq a^4 + \frac{3}{2}(b^8 + c^8)$,

$$\begin{aligned}
\rho_{l+1,4,2}^{r,o,k} &\leq (1 + A_4\Delta)\rho_{l,4,2}^{r,o,k} \\
&\quad + \sum_{s:s \neq k} \left(C_{3,1}\Delta(3\rho_{l,4,2}^{r,o,k} + \rho_{l,4,2}^{r,o,s}) + 3(d-1)C_{2,2}\Delta(\rho_{l,4,2}^{r,o,k} + \rho_{l,4,2}^{r,o,s}) \right. \\
&\quad \left. + (d-1)^2C_{1,3}\Delta(\rho_{l,4,2}^{r,o,k} + 3\rho_{l,4,2}^{r,o,s}) + (d-1)^3B_4\Delta\rho_{l,4,2}^{r,o,s} \right) \\
&\quad + \sum_{s,u,v=1}^d \left(d^2C_{3,1}\Delta \left(3\rho_{l,4,2}^{r,o,v} + \frac{1}{2}(\rho_{l,8,1}^{r,s} + \rho_{l,8,1}^{o,u}) \right) \right. \\
&\quad \left. + 3d^3C_{2,2}\Delta \left(\rho_{l,4,2}^{r,o,v} + \frac{1}{2}(\rho_{l,8,1}^{r,s} + \rho_{l,8,1}^{o,u}) \right) \right. \\
&\quad \left. + d^4C_{1,3}\Delta \left(\rho_{l,4,2}^{r,o,v} + \frac{3}{2}(\rho_{l,8,1}^{r,s} + \rho_{l,8,1}^{o,u}) \right) \right) \\
&\quad + d^6 \sum_{s,u=1}^d B_4\Delta \frac{1}{2}(\rho_{l,8,1}^{r,s} + \rho_{l,8,1}^{o,u}).
\end{aligned}$$

This gives us

$$\begin{aligned}
\rho_{l+1,4,2}^{r,o} &\leq (1 + A_4\Delta)\rho_{l,4,2}^{r,o} + 4(d-1)C_{3,1}\Delta\rho_{l,4,2}^{r,o} + 6(d-1)^2C_{2,2}\Delta\rho_{l,4,2}^{r,o} \\
&\quad + 4(d-1)^3C_{1,3}\Delta\rho_{l,4,1}^r + (d-1)^4B_4\Delta\rho_{l,4,2}^{r,o} \\
&\quad + d^5C_{3,1}\Delta \left(3\rho_{l,4,2}^{r,o} + \frac{1}{2}(\rho_{l,8,1}^r + \rho_{l,8,1}^o) \right) \\
&\quad + 3d^6C_{2,2}\Delta \left(\rho_{l,4,2}^{r,o} + \frac{1}{2}(\rho_{l,8,1}^r + \rho_{l,8,1}^o) \right) \\
&\quad + d^7C_{1,3}\Delta \left(\rho_{l,4,2}^{r,o} + \frac{3}{2}(\rho_{l,8,1}^r + \rho_{l,8,1}^o) \right) + d^8B_4\Delta \frac{1}{2}(\rho_{l,8,1}^r + \rho_{l,8,1}^o).
\end{aligned}$$

Hence, we obtain

$$\rho_{l+1,4,2}^{r,o} \leq (1 + \kappa_8\Delta)\rho_{l,4,2}^{r,o} + \kappa_9\Delta,$$

for some constants $\kappa_8, \kappa_9 > 0$, leading to

$$\rho_{l,4,2}^{r,o} \leq (1 + \kappa_8\Delta)^{l-j-1}\rho_{j+1,4,2}^{r,o} + \kappa_{10}, \quad l = j+1, \dots, J-1,$$

where $\kappa_{10} > 0$ and

$$\rho_{j+1,4,2}^{r,o} = \sum_{s=1}^d \mathbb{E} \left[\left(\frac{\partial^2}{\partial x^r \partial x^o} \Phi_{\Delta}^s(x, \xi_{j+1}) \right)^4 \right].$$

Next, we have for some $k, o, r, z \in \{1, \dots, d\}$

$$\begin{aligned}
& \frac{\partial^3}{\partial x^r \partial x^o \partial x^z} G_{l+1,j}^k(x) \\
= & \sum_{s=1}^d \frac{\partial}{\partial x^s} \Phi_{\Delta, l+1}^k(G_{l,j}(x)) \frac{\partial^3}{\partial x^r \partial x^o \partial x^z} G_{l,j}^s(x) \\
& + \sum_{s,u=1}^d \frac{\partial^2}{\partial x^s \partial x^u} \Phi_{\Delta, l+1}^k(G_{l,j}(x)) \left(\frac{\partial^2}{\partial x^r \partial x^o} G_{l,j}^s(x) \frac{\partial}{\partial x^z} G_{l,j}^u(x) + \frac{\partial^2}{\partial x^r \partial x^z} G_{l,j}^s(x) \frac{\partial}{\partial x^o} G_{l,j}^u(x) \right. \\
& \quad \left. + \frac{\partial}{\partial x^r} G_{l,j}^s(x) \frac{\partial^2}{\partial x^o \partial x^z} G_{l,j}^u(x) \right) \\
& + \sum_{s,u,v=1}^d \frac{\partial^3}{\partial x^s \partial x^u \partial x^v} \Phi_{\Delta, l+1}^k(G_{l,j}(x)) \frac{\partial}{\partial x^r} G_{l,j}^s(x) \frac{\partial}{\partial x^o} G_{l,j}^u(x) \frac{\partial}{\partial x^z} G_{l,j}^v(x) \\
= & \sum_{s=1}^d \psi_{1,s} + \sum_{s,u=1}^d \psi_{2,s,u} + \sum_{s,u,v=1}^d \psi_{3,s,u,v}
\end{aligned}$$

and $\frac{\partial^3}{\partial x^r \partial x^o \partial x^z} G_{j+1,j}^s(x) = \frac{\partial^3}{\partial x^r \partial x^o \partial x^z} \Phi_{\Delta}^s(x, \xi_{j+1})$. Hence

$$\begin{aligned}
& \mathbb{E} \left[\left(\frac{\partial^3}{\partial x^r \partial x^o \partial x^z} G_{l+1,j}^k(x) \right)^2 \right] \\
\leq & \mathbb{E} \left[\psi_{1,k}^2 + \sum_{s: s \neq k} (2\psi_{1,k}\psi_{1,s} + (d-1)\psi_{1,s}^2) + 2 \sum_{s,u,v=1}^d \psi_{1,v}\psi_{2,s,u} \right. \\
& \left. + 2 \sum_{s,u,v,w=1}^d \psi_{1,w}\psi_{3,s,u,v} + 2d^2 \sum_{s,u=1}^d \psi_{2,s,u}^2 + 2d^3 \sum_{s,u,v=1}^d \psi_{3,s,u,v}^2 \right]
\end{aligned}$$

Denote

$$\rho_{l+1,n,3}^{r,o,z,s} = \mathbb{E} \left[\left(\frac{\partial^3}{\partial x^r \partial x^o \partial x^z} G_{l+1,j}^s(x) \right)^n \right],$$

then we get, due to $3a^2b^2c^2 \leq a^6 + b^6 + c^6$ and

$$\begin{aligned}
2\mathbb{E}[XYZU] & \leq 2\sqrt{\mathbb{E}[X^2]} \sqrt[6]{\mathbb{E}[Y^6]} \sqrt[6]{\mathbb{E}[Z^6]} \sqrt[6]{\mathbb{E}[U^6]} \leq \mathbb{E}[X^2] + \sqrt[3]{\mathbb{E}[Y^6]} \sqrt[3]{\mathbb{E}[Z^6]} \sqrt[3]{\mathbb{E}[U^6]} \\
& \leq \mathbb{E}[X^2] + \frac{1}{3} (\mathbb{E}[Y^6] + \mathbb{E}[Z^6] + \mathbb{E}[U^6]),
\end{aligned}$$

as well as (1.5) and (7.1),

$$\begin{aligned}
\rho_{l+1,2,3}^{r,o,z,k} &\leq (1 + A_2\Delta)\rho_{l,2,3}^{r,o,z,k} + \sum_{s: s \neq k} \left(C_{1,1}\Delta(\rho_{l,2,3}^{r,o,z,k} + \rho_{l,2,3}^{r,o,z,s}) + (d-1)B_2\Delta\rho_{l,2,3}^{r,o,z,s} \right) \\
&+ \sum_{s,u,v=1}^d C_{1,1}\Delta \left(\rho_{l,2,2}^{r,o,z,v} + \frac{1}{2} \left(\rho_{l,4,1}^{r,s} + \rho_{l,4,1}^{o,u} + \rho_{l,4,1}^{z,u} + \rho_{l,4,2}^{r,o,s} + \rho_{l,4,2}^{r,z,s} + \rho_{l,4,2}^{o,z,u} \right) \right) \\
&+ \sum_{s,u,v,w=1}^d C_{1,1}\Delta \left(\rho_{l,2,2}^{r,o,z,w} + \frac{1}{3} \left(\rho_{l,6,1}^{r,s} + \rho_{l,6,1}^{o,u} + \rho_{l,6,1}^{z,v} \right) \right) \\
&+ 3d^2 \sum_{s,u=1}^d B_2\Delta \left(\rho_{l,4,1}^{r,s} + \rho_{l,4,1}^{o,u} + \rho_{l,4,1}^{z,u} + \rho_{l,4,2}^{r,o,s} + \rho_{l,4,2}^{r,z,s} + \rho_{l,4,2}^{o,z,u} \right) \\
&+ d^3 \sum_{s,u,v=1}^d B_2\Delta \frac{1}{3} \left(\rho_{l,6,1}^{r,s} + \rho_{l,6,1}^{o,u} + \rho_{l,6,1}^{z,v} \right).
\end{aligned}$$

Further, denote

$$\rho_{l+1,2,3}^{r,o,z} = \sum_{s=1}^d \rho_{l+1,2,3}^{r,o,z,s},$$

then we get

$$\begin{aligned}
\rho_{l+1,2,3}^{r,o,z} &\leq (1 + A_2\Delta)\rho_{l,2,2}^{r,o,z} + 2(d-1)C_{1,1}\Delta\rho_{l,2,2}^{r,o,z} + (d-1)^2B_2\Delta\rho_{l,2,2}^{r,o,z} \\
&+ d^3C_{1,1}\Delta \left(\rho_{l,2,2}^{r,o,z} + \frac{1}{2} \left(\rho_{l,4,1}^r + \rho_{l,4,1}^o + \rho_{l,4,1}^z + \rho_{l,4,2}^{r,o} + \rho_{l,4,2}^{r,z} + \rho_{l,4,2}^{o,z} \right) \right) \\
&+ d^4C_{1,1}\Delta \left(\rho_{l,2,2}^{r,o,z} + \frac{1}{3} \left(\rho_{l,6,1}^r + \rho_{l,6,1}^o + \rho_{l,6,1}^z \right) \right) \\
&+ 3d^4B_2\Delta \left(\rho_{l,4,1}^r + \rho_{l,4,1}^o + \rho_{l,4,1}^z + \rho_{l,4,2}^{r,o} + \rho_{l,4,2}^{r,z} + \rho_{l,4,2}^{o,z} \right) \\
&+ d^6B_2\Delta \frac{1}{3} \left(\rho_{l,6,1}^r + \rho_{l,6,1}^o + \rho_{l,6,1}^z \right).
\end{aligned}$$

This gives us

$$\rho_{l+1,2,3}^{r,o,z} \leq (1 + \kappa_{11}\Delta)\rho_{l,2,2}^{r,o,z} + \kappa_{12}\Delta,$$

for some constants $\kappa_{11}, \kappa_{12} > 0$, leading to

$$\rho_{l,2,2}^{r,o,z} \leq (1 + \kappa_{11}\Delta)^{l-j-1} \rho_{j+1,2,3}^{r,o,z} + \kappa_{13}, \quad l = j+1, \dots, J-1,$$

where $\kappa_{13} > 0$ and

$$\rho_{j+1,2,3}^{r,o,z} = \sum_{s=1}^d \mathbb{E} \left[\left(\frac{\partial^3}{\partial x^r \partial x^o \partial x^z} \Phi_{\Delta}^s(x, \xi_{j+1}) \right)^2 \right].$$

Thus, we obtain the boundedness of

$$\begin{aligned}
& \left| \frac{\partial^3}{\partial x^r \partial x^o \partial x^z} q_j(x) \right| \\
\leq & \sum_{s=1}^d \mathbb{E} \left| \frac{\partial}{\partial x^s} f(G_{J,j}(x)) \frac{\partial^3}{\partial x^r \partial x^o \partial x^z} G_{J,j}^s(x) \right| \\
& + \sum_{s,u=1}^d \mathbb{E} \left| \frac{\partial^2}{\partial x^s \partial x^u} f(G_{J,j}(x)) \left(\frac{\partial^2}{\partial x^r \partial x^o} G_{J,j}^s(x) \frac{\partial}{\partial x^z} G_{J,j}^u(x) + \frac{\partial^2}{\partial x^r \partial x^z} G_{J,j}^s(x) \frac{\partial}{\partial x^o} G_{J,j}^u(x) \right. \right. \\
& \quad \left. \left. + \frac{\partial}{\partial x^r} G_{J,j}^s(x) \frac{\partial^2}{\partial x^o \partial x^z} G_{J,j}^u(x) \right) \right| \\
& + \sum_{s,u,v=1}^d \mathbb{E} \left| \frac{\partial^3}{\partial x^s \partial x^u \partial x^v} f(G_{J,j}(x)) \frac{\partial}{\partial x^r} G_{J,j}^s(x) \frac{\partial}{\partial x^o} G_{J,j}^u(x) \frac{\partial}{\partial x^z} G_{J,j}^v(x) \right| \\
\leq & \sum_{s=1}^d \sqrt{\mathbb{E} \left[\left(\frac{\partial}{\partial x^s} f(G_{J,j}(x)) \right)^2 \right] \rho_{J,2,3}^{r,o,z,s}} \\
& + \sum_{s,u=1}^d \sqrt{\mathbb{E} \left[\left(\frac{\partial^2}{\partial x^s \partial x^u} f(G_{J,j}(x)) \right)^2 \right] \left(\sqrt[4]{\rho_{J,4,2}^{r,o,s} \rho_{J,4,1}^{z,u}} + \sqrt[4]{\rho_{J,4,2}^{r,z,s} \rho_{J,4,1}^{o,u}} + \sqrt[4]{\rho_{J,4,1}^{r,s} \rho_{J,4,2}^{o,z,u}} \right)} \\
& + \sum_{s,u,v=1}^d \sqrt{\mathbb{E} \left[\left(\frac{\partial^3}{\partial x^s \partial x^u \partial x^v} f(G_{J,j}(x)) \right)^2 \right] \sqrt[6]{\rho_{J,6,1}^{r,s} \rho_{J,6,1}^{o,u} \rho_{J,6,1}^{z,v}}}
\end{aligned}$$

for all $r, o, z \in \{1, \dots, d\}$, since f is assumed to be three times continuously differentiable with bounded partial derivatives up to order 3.

7.1. Proof of Theorem 3.1

(i) Straightforward.

(ii) Let us define $\mu_\Delta(x) := x + \mu(x)\Delta$. Then we obtain via Taylor's theorem (cf. (2.2))

$$q_j(\Phi_\Delta(x, y)) = q_j(\mu_\Delta(x)) + \sqrt{\Delta} \sum_{k=1}^d \sum_{i=1}^m \sigma^{ki}(x) y^i \int_0^1 \frac{\partial q_j}{\partial x^k}(\mu_\Delta(x) + t\sigma(x)\sqrt{\Delta}y) dt.$$

This gives us (see (2.5))

$$\begin{aligned}
a_{j,r,s}(x) &= \frac{1}{2^m} \sum_{y \in \{-1,1\}^m} q_j(\Phi_\Delta(x, y)) \prod_{o=1}^r y^{s_o} \\
&= \frac{\sqrt{\Delta}}{2^m} \sum_{y \in \{-1,1\}^m} \left(\prod_{o=1}^r y^{s_o} \right) \sum_{k=1}^d \sum_{i=1}^m \sigma^{ki}(x) y^i \int_0^1 \frac{\partial q_j}{\partial x^k}(\mu_\Delta(x) + t\sigma(x)\sqrt{\Delta}y) dt, \tag{7.3}
\end{aligned}$$

since

$$\frac{1}{2^m} \sum_{y \in \{-1,1\}^m} \prod_{o=1}^r y^{s_o} = \mathbb{E} \left[\prod_{o=1}^r \xi_j^{s_o} \right] = 0.$$

Next we apply Theorem 1.1 for the case $K = 1$ to get that all the functions q_j are continuously differentiable with bounded partial derivatives. Clearly, the assumptions in this theorem hold, when all the functions f, μ^k, σ^{ki} , $k \in \{1, \dots, d\}$, $i \in \{1, \dots, m\}$ are continuously differentiable with bounded derivatives. Together with the assumption, that all the functions σ^{ki} are bounded, we get from (7.3) that $a_{j,r,s}$ is of order $\sqrt{\Delta}$ for all j, r, s .

7.2. Proof of Theorem 3.2

Let us consider a higher order Taylor expansion compared to the proof of Theorem 3.1 and recall that $\mu_\Delta(x) = x + \mu(x)\Delta$. We have for any $y \in \{-1, 1\}^m$

$$\begin{aligned} q_j(\Phi_\Delta(x, y)) &= q_j(\mu_\Delta(x)) + \sqrt{\Delta} \sum_{k=1}^d \frac{\partial}{\partial x^k} q_j(\mu_\Delta(x)) \sum_{i=1}^m \sigma^{ki}(x) y^i \\ &\quad + \Delta \sum_{k,l=1}^d (2 - \delta_{k,l}) \int_0^1 (1-t) \frac{\partial^2}{\partial x^k \partial x^l} q_j(\mu_\Delta(x) + t\sqrt{\Delta}\sigma(x)y) dt \sum_{i=1}^m \sigma^{ki}(x) y^i \sum_{i=1}^m \sigma^{li}(x) y^i, \end{aligned} \quad (7.4)$$

where $\delta_{\cdot,\cdot}$ is the Kronecker delta. This gives us for $r \geq 2$ (cf. (2.5))

$$\begin{aligned} a_{j,r,s}(x) &= \frac{1}{2^m} \sum_{y \in \{-1, 1\}^m} q_j(\Phi_\Delta(x, y)) \prod_{o=1}^r y^{s_o} \\ &= \frac{\Delta}{2^m} \sum_{k,l=1}^d \left[(2 - \delta_{k,l}) \sum_{y \in \{-1, 1\}^m} \left(\sum_{i=1}^m \sigma^{ki}(x) y^i \sum_{i=1}^m \sigma^{li}(x) y^i \prod_{o=1}^r y^{s_o} \right. \right. \\ &\quad \left. \left. \cdot \int_0^1 (1-t) \frac{\partial^2}{\partial x^k \partial x^l} q_j(\mu_\Delta(x) + t\sqrt{\Delta}\sigma(x)y) dt \right) \right], \end{aligned} \quad (7.5)$$

due to (cf. (7.4))

$$\frac{1}{2^m} \sum_{y \in \{-1, 1\}^m} y^i \prod_{o=1}^r y^{s_o} = \mathbb{E} \left[\xi_j^i \prod_{o=1}^r \xi_j^{s_o} \right] = 0 \quad (7.6)$$

for all $i \in \{1, \dots, m\}$. (Note that (7.6) does not hold for $r = 1$.) Applying Theorem 1.1 (case $K = 2$), we get that q_j is twice continuously differentiable with bounded partial derivatives up to order 2, provided that all the functions $f, \mu^k, \sigma^{k,i}$ are twice continuously differentiable with bounded partial derivatives up to order 2. Together with the assumption, that all the functions σ^{kl} are bounded, we get from (7.5) that $a_{j,r,s}$ is of order Δ for all j, r, s with $r > 1$.

7.3. Proof of Theorem 3.4

Here we apply Theorem 3.2, which gives us (cf. (2.6))

$$\begin{aligned} \text{Var} \left[f(X_{\Delta,T}) - M_{\Delta,T}^{(1),trunc} \right] &= \text{Var} \left[M_{\Delta,T}^{(1)} - M_{\Delta,T}^{(1),trunc} \right] \\ &= \text{Var} \left[\sum_{j=1}^J \sum_{r=2}^m \sum_{1 \leq s_1 < \dots < s_r \leq m} a_{j,r,s}(X_{\Delta,(j-1)\Delta}) \prod_{i=1}^r \xi_j^{s_i} \right] \\ &\lesssim \sum_{j=1}^J \sum_{r=2}^m \sum_{1 \leq s_1 < \dots < s_r \leq m} \mathbb{E} \left[a_{j,r,s}^2(X_{\Delta,(j-1)\Delta}) \right] \lesssim \Delta, \end{aligned}$$

since $\mathbb{E} \left[a_{j,r,s}^2(X_{\Delta,(j-1)\Delta}) \right] \lesssim \Delta^2$ for all j, r, s with $r > 1$.

7.4. Proof of Theorem 3.5

The proof works similarly to the one of Theorem 3.1. More precisely, here we define (cf. (2.8))

$$\mu_{\Delta}(x) := x + \mu(x)\Delta + \frac{1}{2}\mathcal{L}^0\mu(x)\Delta^2.$$

Then we derive the zero-order Taylor expansion for $q_j(\Phi_{\Delta}(x, y, z))$ around $\mu_{\Delta}(x)$, use that

$$\mathbb{E} \left[\prod_{r \in U_1} H_{o_r}(\xi_j^r) \prod_{(k,l) \in U_2} V_j^{kl} \right] = 0$$

and observe that all components $\tilde{\Phi}_{\Delta}^k(x, y, z) := \Phi_{\Delta}^k(x, y, z) - \mu_{\Delta}^k(x)$, $k \in \{1, \dots, d\}$ (as an analogue of $\sqrt{\Delta} \sum_{i=1}^m \sigma^{ki}(x)y^i$ in case of the weak Euler scheme), are of order $\sqrt{\Delta}$ under less strict assumptions than required in the present theorem. Finally we apply Theorem 1.1 (case $K = 1$) which gives us that q_j is continuously differentiable with bounded partial derivatives under the assumptions, that all functions μ^k and σ^{ki} are bounded and all the functions f, μ^k, σ^{ki} are three-times continuously differentiable with bounded partial derivatives up to order 3. Consequently, all the functions a_{j,o,U_1,U_2} are of order $\sqrt{\Delta}$.

7.5. Proof of Theorem 3.6

(i) The proof works similarly to the one of Theorem 3.2, that is, we consider a Taylor expansion for $q_j(\Phi_{\Delta}(x, y, z))$ of order 1, around the same point $\mu_{\Delta}(x)$ as in the proof of Theorem 3.5. Then we use

$$\mathbb{E} \left[\tilde{\Phi}_{\Delta}^k(x, \xi_j, V_j) \prod_{r \in U_1} H_{o_r}(\xi_j^r) \prod_{(k,l) \in U_2} V_j^{kl} \right] = 0, \quad k \in \{1, \dots, d\},$$

whenever $|U_2| + |\mathcal{K}_2| + \frac{|\mathcal{K}_1|}{2} \geq 1$ (where again $\tilde{\Phi}_{\Delta}^k(x, y, z) = \Phi_{\Delta}^k(x, y, z) - \mu_{\Delta}^k(x)$). Then we apply Theorem 1.1 (case $K = 2$) which gives us that q_j is twice continuously differentiable with bounded partial derivatives up to order 2 under the assumptions, that all functions μ^k and σ^{ki} are bounded and all the functions f, μ^k, σ^{ki} are four-times continuously differentiable with bounded partial derivatives up to order 4. Finally, we get that all the functions a_{j,o,U_1,U_2} are of order Δ , since the product of all functions $\tilde{\Phi}_{\Delta}^k(x, y, z)\tilde{\Phi}_{\Delta}^l(x, y, z)$, $k, l \in \{1, \dots, d\}$, is of order Δ under the above assumptions.

(ii) Here we consider the Taylor expansion of order 2, that is

$$\begin{aligned}
q_j(\Phi_\Delta(x, y, z)) &= q_j(\mu_\Delta(x)) + \sum_{k=1}^d \frac{\partial}{\partial x^k} q_j(\mu_\Delta(x)) \tilde{\Phi}_\Delta^k(x, y, z) \\
&+ \sum_{k,l=1}^d \frac{1}{2} (2 - \delta_{k,l}) \frac{\partial^2}{\partial x^k \partial x^l} q_j(\mu_\Delta(x)) \tilde{\Phi}_\Delta^k(x, y, z) \tilde{\Phi}_\Delta^l(x, y, z) \\
&+ \sum_{k,l,n=1}^d \left[\left(3 - \frac{3}{2} (\delta_{k,l} + \delta_{k,n} + \delta_{l,n}) + 2\delta_{k,l}\delta_{k,n}\delta_{l,n} \right) \tilde{\Phi}_\Delta^k(x, y, z) \tilde{\Phi}_\Delta^l(x, y, z) \tilde{\Phi}_\Delta^n(x, y, z) \right. \\
&\quad \left. \cdot \int_0^1 (1-t)^2 \frac{\partial^3}{\partial x^k \partial x^l \partial x^n} q_j(\mu_\Delta(x) + t\tilde{\Phi}_\Delta(x, y, z)) dt \right].
\end{aligned}$$

Next we use

$$\mathbb{E} \left[\tilde{\Phi}_\Delta^k(x, \xi_j, V_j) \tilde{\Phi}_\Delta^l(x, \xi_j, V_j) \prod_{r \in U_1} H_{o_r}(\xi_j^r) \prod_{(k,l) \in U_2} V_j^{kl} \right] = 0, \quad k, l \in \{1, \dots, d\},$$

whenever $|U_2| + \frac{|\mathcal{K}_1|}{2} > 1$, and thus we obtain (cf. (2.11))

$$\begin{aligned}
&a_{j,o,U_1,U_2}(x) \\
&= \frac{1}{2^{\frac{m(m-1)}{2}}} \frac{1}{6^m} \sum_{y \in \{-\sqrt{3}, 0, \sqrt{3}\}^m} \sum_{z \in \{-1, 1\}^{\frac{m(m-1)}{2}}} 4^{\sum_{i=1}^m I(y^i=0)} \prod_{r \in U_1} H_{o_r}(y^r) \prod_{(k,l) \in U_2} z^{kl} q_j(\Phi_\Delta(x, y, z)) \\
&= \frac{1}{2^{\frac{m(m-1)}{2}}} \frac{1}{6^m} \sum_{y \in \{-\sqrt{3}, 0, \sqrt{3}\}^m} \sum_{z \in \{-1, 1\}^{\frac{m(m-1)}{2}}} 4^{\sum_{i=1}^m I(y^i=0)} \prod_{r \in U_1} H_{o_r}(y^r) \prod_{(k,l) \in U_2} z^{kl} \\
&\quad \cdot \sum_{k,l,n=1}^d \left[\left(3 - \frac{3}{2} (\delta_{k,l} + \delta_{k,n} + \delta_{l,n}) + 2\delta_{k,l}\delta_{k,n}\delta_{l,n} \right) \tilde{\Phi}_\Delta^k(x, y, z) \tilde{\Phi}_\Delta^l(x, y, z) \tilde{\Phi}_\Delta^n(x, y, z) \right. \\
&\quad \left. \cdot \int_0^1 (1-t)^2 \frac{\partial^3}{\partial x^k \partial x^l \partial x^n} q_j(\mu_\Delta(x) + t\tilde{\Phi}_\Delta(x, y, z)) dt \right]
\end{aligned}$$

Then we apply Theorem 1.1 (case $K = 3$) which gives us that q_j is three-times continuously differentiable with bounded partial derivatives up to order 3 under the assumptions, that all functions μ^k and σ^{ki} are bounded and all the functions f, μ^k, σ^{ki} are five-times continuously differentiable with bounded partial derivatives up to order 5. Finally, we get that all the functions a_{j,o,U_1,U_2} are of order $\Delta^{3/2}$, since the product of all functions $\tilde{\Phi}_\Delta^k(x, y, z) \tilde{\Phi}_\Delta^l(x, y, z) \tilde{\Phi}_\Delta^n(x, y, z)$, $k, l, n \in \{1, \dots, d\}$, is of order $\Delta^{3/2}$ under the above assumptions.

7.6. Proof of Theorem 3.8

The proof is similar to the one of Theorem 3.4.

7.7. Proof of Lemma 5.2

We refer to Theorem 11.3 in [6]. When applying it, we obtain actually

$$\mathbb{E}\|\tilde{a}_{j,o,U_1,U_2} - a_{j,o,U_1,U_2}\|_{L^2(\mathbb{P}_{\Delta,j-1})}^2 \leq \tilde{c} \max\left\{\Sigma, \tilde{A}^2 \Delta_{U_1,U_2}^2\right\} (\log N + 1) \frac{Q}{N} + \frac{8C_\kappa}{Q^\kappa}. \quad (7.7)$$

However, the maximum in (7.7) is in fact a sum of two terms Σ and $\tilde{A}^2 \Delta_{U_1,U_2} (\log N_r + 1)$ so that the logarithm is only included in one term (see the proof of Theorem 11.3 in [6]).

7.8. Proof of Theorem 5.3

Using the martingale transform structure in (2.12) and (3.5) (recall footnote 1 on page 19) together with the orthonormality of the system $\prod_{r \in U_1} H_{o_r}(\xi_j^r) \prod_{(k,l) \in U_2} V_j^{kl}$, we get by (3.6) and (5.2)

$$\begin{aligned} \text{Var}\left[f(X_{\Delta,T}) - \widetilde{M}_{\Delta,T}^{(2),trunc}\right] &= \text{Var}\left[f(X_{\Delta,T}) - M_{\Delta,T}^{(2),trunc}\right] + \text{Var}\left[M_{\Delta,T}^{(2),trunc} - \widetilde{M}_{\Delta,T}^{(2),trunc}\right] \\ &\lesssim \Delta^2 + \sum_{j=1}^J \sum_{\substack{(U_1,U_2) \in \mathcal{A} \\ |U_2| + |\mathcal{K}_2| + \frac{1}{2}|\mathcal{K}_1| \leq 1}} \sum_{o \in \{1,2\}^{U_1}} \mathbb{E}\|\tilde{a}_{j,o,U_1,U_2} - a_{j,o,U_1,U_2}\|_{L^2(\mathbb{P}_{\Delta,j-1})}^2 \\ &\leq \Delta^2 + Jm(m+1) \left(\tilde{c}(\Sigma + \tilde{A}^2 \Delta (\log N + 1)) \frac{Q}{N} + \frac{8C_\kappa}{Q^\kappa} \right), \end{aligned}$$

since $\Delta_{U_1,U_2}^2 \leq \Delta$.

7.9. Proof of Theorem 5.4

The proof is similar to the complexity analysis performed in [3].

REFERENCES

- [1] D. Belomestny, S. Häfner, T. Nagapetyan, and M. Urusov. Variance reduction for discretised diffusions via regression. *Preprint, arXiv:1510.03141v3*, 2016.
- [2] D. Belomestny, S. Häfner, and M. Urusov. Regression-based complexity reduction of the dual nested Monte Carlo methods. *Preprint, arXiv:1611.06344*, 2016.
- [3] D. Belomestny, S. Häfner, and M. Urusov. Regression-based variance reduction approach for strong approximation schemes. *Preprint, arXiv:1612.03407*, 2016.
- [4] D. Belomestny, S. Häfner, and M. Urusov. Stratified regression-based variance reduction approach for weak approximation schemes. *Preprint*, 2016.
- [5] M. B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
- [6] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer Series in Statistics. Springer-Verlag, New York, 2002.
- [7] P. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23. Springer, 1992.
- [8] G. N. Milstein and M. V. Tretyakov. Practical variance reduction via regression for simulating diffusions. *SIAM Journal on Numerical Analysis*, 47(2):887–910, 2009.
- [9] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. On the complexity of computing quadrature formulas for marginal distributions of SDEs. *Journal of Complexity*, 31(1):110–145, 2015.
- [10] T. Müller-Gronbach and L. Yaroslavtseva. Deterministic quadrature formulas for SDEs based on simplified weak Itô-Taylor steps. *Foundations of Computational Mathematics*, 16(5):1325–1366, 2016.
- [11] N. J. Newton. Variance reduction for simulated diffusions. *SIAM journal on applied mathematics*, 54(6):1780–1805, 1994.