

ON THE ACCELERATION OF THE MULTI-LEVEL MONTE CARLO METHOD

KRISTIAN DEBRABANT,* *University of Southern Denmark*

ANDREAS RÖBLER,** *Universität zu Lübeck*

Abstract

The multi-level Monte Carlo method proposed by Giles (2008) approximates the expectation of some functionals applied to a stochastic process with optimal order of convergence for the mean-square error. In this paper a modified multi-level Monte Carlo estimator is proposed with significantly reduced computational costs. As the main result, it is proved that the modified estimator reduces the computational costs asymptotically by a factor $(p/\alpha)^2$ if weak approximation methods of orders α and p are applied in the case of computational costs growing with the same order as the variances decay.

Keywords: Multi-level Monte Carlo; Monte Carlo; weak approximation; variance reduction; stochastic differential equation

2010 Mathematics Subject Classification: Primary 65C05

Secondary 60H35; 65C20; 68U20

1. Introduction

The multi-level Monte Carlo (MLMC) method proposed by Giles [6] approximates the expectation of some functional applied to some stochastic processes, e.g. solutions of stochastic differential equations (SDEs), at a lower computational complexity than classical Monte Carlo simulations; see also [5], [8], and [9]. The MLMC approximation finds application in many fields such as mathematical finance [1], [7], for SDEs driven by a Lévy process [3], by fractional Brownian motion [11], or stochastic PDEs [13]. The main idea of this paper is to reduce the computational costs additionally by applying the MLMC method as a variance reduction technique for some higher-order weak approximation method. As a result, the computational effort can be significantly reduced while the optimal order of convergence for the root-mean-square (RMS) error is preserved.

The outline of this paper is as follows. In Section 2 we provide a brief introduction to the main ideas and results of the MLMC method. In Section 3, based on these results, we present as the main result a modified MLMC algorithm that allows the computational costs to be significantly reduced. Depending on the relationship between the orders of the variance reduction and of the growth of the costs, there exists a reduction of the computational costs by a factor depending on the weak order of the underlying numerical method. As an example, in Section 4 the modified MLMC algorithm is applied to the problem of weak approximation for SDEs driven by Brownian motion.

Received 6 February 2013; revision received 28 April 2014.

* Postal address: Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark.

** Postal address: Institut für Mathematik, Universität zu Lübeck, Ratzeburger Allee 160, D-23562 Lübeck, Germany.
Email address: roessler@math.uni-luebeck.de

2. MLMC simulation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with some filtration $(\mathcal{F}_t)_{t \geq 0}$ and let $X = (X_t)_{t \in I}$ denote an adapted stochastic process on the interval $I = [t_0, T]$ that belongs to a space \mathbb{X} that may be infinite-dimensional. In the following, we are interested in the approximation of $\mathbb{E}_{\mathbb{P}}(f(X))$ for some functional $f \in \mathbb{F}$, where \mathbb{F} denotes a suitable class of functionals that are of interest. Furthermore, let an equidistant discretization $I_h = \{t_0, t_1, \dots, t_N\}$ with $0 \leq t_0 < t_1 < \dots < t_N = T$ of the time interval I with step size h be given. Then, we consider a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ with some filtration $(\tilde{\mathcal{F}}_t)_{t \in I_h}$ and we denote by $Y = (Y_t)_{t \in I_h}$ a discrete-time approximation of X on the grid I_h , adapted to $(\tilde{\mathcal{F}}_t)_{t \in I_h}$. Thus, we consider the approximation $Y \in \mathbb{X}_h$ of $X \in \mathbb{X}$ on a finite-dimensional space \mathbb{X}_h . Here, the probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ may be but do not have to be equal and we assume that Y approximates X in the weak sense with some order $p > 0$, i.e.

$$|\mathbb{E}_{\tilde{\mathbb{P}}}(f(Y)) - \mathbb{E}_{\mathbb{P}}(f(X))| = O(h^p)$$

for all $f \in \mathbb{F}$.

In order to approximate the expectation of $f(X)$ we apply the MLMC estimator introduced in [6]. For some fixed $M \in \mathbb{N}$ with $M \geq 2$ and some $L \in \mathbb{N}$ we define the step sizes $h_l = T/M^l$, and let $Y^l = (Y_t)_{t \in I_{h_l}}$ denote the discrete-time approximation process on the grid I_{h_l} based on step size h_l for $l = 0, 1, \dots, L$. Here, we consider the approximations $Y^l \in \mathbb{X}_{h_l}$ for $l = 0, 1, \dots, L$ of $X \in \mathbb{X}$ on a sequence $\mathbb{X}_{h_0} \subset \mathbb{X}_{h_1} \subset \dots \subset \mathbb{X}_{h_L}$ of finite-dimensional subspaces. Then the MLMC estimator is defined by

$$\hat{Y}_{ML} = \sum_{l=0}^L \hat{Y}^l \tag{2.1}$$

for some $L \in \mathbb{N}$ using the estimators

$$\hat{Y}^0 = \frac{1}{N_0} \sum_{i=1}^{N_0} f(Y^{0(i)}), \quad \hat{Y}^l = \frac{1}{N_l} \sum_{i=1}^{N_l} (f(Y^{l(i)}) - f(Y^{l-1(i)}))$$

for $l = 1, \dots, L$. Then, we obtain

$$\mathbb{E}_{\tilde{\mathbb{P}}}(\hat{Y}_{ML}) = \mathbb{E}_{\tilde{\mathbb{P}}}(f(Y^0)) + \sum_{l=1}^L \mathbb{E}_{\tilde{\mathbb{P}}}(f(Y^l) - f(Y^{l-1})).$$

Here, we have to point out that both approximations $Y^{l(i)}$ and $Y^{l-1(i)}$ are simulated simultaneously based on the same realization of the underlying driving random process, whereas $(Y^{l(i)}, Y^{l-1(i)})$ and $(Y^{l(j)}, Y^{l-1(j)})$ are independent realizations for $i \neq j$.

Now, there are two sources of errors for the approximation. On the one hand, we have a systematical error that depends on the dimension of \mathbb{X}_{h_l} due to the discrete-time approximation $Y^l \in \mathbb{X}_{h_l}$ based on step size h_l which is given by the bias of the method. On the other hand, there is a statistical error from the estimator for the expectation of $f(Y^l)$ by the Monte Carlo simulation. Therefore, in what follows we consider the RMS error

$$e(\hat{Y}_{ML}) = (\mathbb{E}_{\tilde{\mathbb{P}}}(|\hat{Y}_{ML} - \mathbb{E}_{\mathbb{P}}(f(X))|^2))^{1/2} \tag{2.2}$$

of the MLMC method. In order to rate the performance of an approximation method, we will analyse the RMS error of the method compared to the computational costs. Therefore, we

denote by $C(Y)$ the computational costs of the approximation method Y . In order to determine $C(Y)$, one may use a cost model where, e.g., each operation or evaluation of some function is charged with the price of one unit, i.e. one counts the number of required mathematical operations or function evaluations. Furthermore, each random number that has to be generated to compute Y may also be charged with the price of one unit.

It is well known that the optimal order of convergence for the classical Monte Carlo estimator $\hat{Y}_{MC} = (1/N)\sum_{i=1}^N f(Y^{(i)})$ is given by $e(\hat{Y}_{MC}) = O((1/C(\hat{Y}_{MC}))^{p/(2p+1)})$, where p is the weak order of convergence of the approximations Y ; see [4]. Thus, higher-order weak approximation methods result in a higher order of convergence with respect to the RMS error. Clearly, the best RMS order of convergence that can be achieved is at most $\frac{1}{2}$. However, the order bound $\frac{1}{2}$ cannot be reached by any weak order p approximation method in the case of the classical Monte Carlo simulation. Therefore, in order to attain the optimal order of convergence for the RMS error, we apply the MLMC estimator (2.1). The following theorem due to Giles [6] is presented in a slightly generalized version suitable for our considerations.

Theorem 2.1. *For some $L \in \mathbb{N}$, let Y^l denote the approximation process on the grid I_{h_l} with respect to step size $h_l = T/M^l$ for each $l = 0, 1, \dots, L$, respectively. Suppose that there exist some constants $\alpha > 0$ and $c_{1,\alpha}, c_{2,0}, c_2, c_{2,L} > 0$ and $\beta, \beta_L > 0$ such that for the bias*

$$(i) \quad |\mathbb{E}_{\mathbb{P}}(f(X)) - \mathbb{E}_{\mathbb{P}}(f(Y^L))| \leq c_{1,\alpha} h_L^\alpha$$

and for the variances

$$(ii) \quad \text{var}_{\mathbb{P}}(f(Y^0)) \leq c_{2,0} h_0^\beta,$$

$$(iii) \quad \text{var}_{\mathbb{P}}(f(Y^l) - f(Y^{l-1})) \leq c_2 h_l^\beta \text{ for } l = 1, \dots, L - 1,$$

$$(iv) \quad \text{var}_{\mathbb{P}}(f(Y^L) - f(Y^{L-1})) \leq c_{2,L} h_L^{\beta_L}.$$

Furthermore, assume that there exist constants $c_{3,0}, c_3, c_{3,L} > 0$ and $\gamma, \gamma_L \geq 1$ such that for the computational costs

$$(v) \quad C(Y^0) \leq c_{3,0} T h_0^{-\gamma},$$

$$(vi) \quad C(Y^l, Y^{l-1}) \leq c_3 T h_l^{-\gamma} \text{ for } l = 1, \dots, L - 1,$$

$$(vii) \quad C(Y^L, Y^{L-1}) \leq c_{3,L} T h_L^{-\gamma_L}.$$

Then for some arbitrarily prescribed error bound $\varepsilon > 0$ there exist values L and N_l for $l = 0, 1, \dots, L$, such that the RMS error of the MLMC estimator \hat{Y}_{ML} has the bound

$$e(\hat{Y}_{ML}) < \varepsilon$$

with computational costs bounded by

$$C(\hat{Y}_{ML}) \leq \begin{cases} c_4 \varepsilon^{-2} & \text{if } \beta > \gamma, \beta_L \geq \gamma_L, \alpha \geq \frac{1}{2} \max\{\gamma, \gamma_L\}, \\ c_4 \varepsilon^{-2} (\log(\varepsilon))^2 & \text{if } \beta = \gamma, \beta_L \geq \gamma_L, \alpha \geq \frac{1}{2} \max\{\gamma, \gamma_L\}, \\ c_4 \varepsilon^{-2 - \max\{\gamma - \beta, \gamma_L - \beta_L\}/\alpha} & \text{if } \beta < \gamma, \alpha \geq \frac{1}{2} (\max\{\gamma, \gamma_L\} - \max\{\gamma - \beta, \gamma_L - \beta_L\}), \end{cases}$$

for some positive constant c_4 .

In order to apply Theorem 2.1 and the MLMC method, one has to determine the values $\alpha, \beta, \beta_L > 0$ as well as $\gamma, \gamma_L \geq 1$. First, α denotes the weak order of convergence for the bias of the finite-dimensional approximation $Y^L \in \mathbb{X}_{h_L}$ as the dimension of the approximation subspace increases. This value is well known for commonly applied approximations Y^L . Because the approximations $(Y^l)_{l \geq 0}$ converge to X in the weak sense, the differences of two successive approximations $(f(Y^l) - f(Y^{l-1}))_{l \geq 1}$ converge to 0 as the dimensions of the subspaces increase. Then, usually their variances will also tend to 0 with some order β and β_L for the approximations applied on levels $0, 1, \dots, L - 1$ and on level L , respectively. Here, we want to point out that estimates of type (i)–(iv) in Theorem 2.1 are rather natural and turn out to be no considerable restriction for typical applications. Finally, the computational costs to evaluate two correlated approximations Y^l and Y^{l-1} on the finite-dimensional subspaces \mathbb{X}_{h_l} and $\mathbb{X}_{h_{l-1}}$ depend on the dimensions of the subspaces that are proportional to h_l^{-1} . For commonly used discrete-time approximations, one typically has $\gamma = \gamma_L = 1$.

The calculations for the proof follow the lines of the original proof of Giles [6]. Considering the mean-square error

$$e(\hat{Y}_{ML}) = (|\mathbb{E}_{\mathbb{P}}(f(X)) - \mathbb{E}_{\hat{\mathbb{P}}}(f(Y^L))|^2 + \text{var}_{\hat{\mathbb{P}}}(\hat{Y}_{ML}))^{1/2} < \varepsilon,$$

we make use of the weight $q \in (0, 1)$, and claim that

$$|\mathbb{E}_{\mathbb{P}}(f(X)) - \mathbb{E}_{\hat{\mathbb{P}}}(f(Y^L))|^2 < q\varepsilon^2 \quad \text{and} \quad \text{var}_{\hat{\mathbb{P}}}(\hat{Y}_{ML}) < (1 - q)\varepsilon^2.$$

Then, we can calculate L from the bias and we have to solve the minimization problem

$$\min_{\{N_l: 0 \leq l \leq L\}} C(\hat{Y}_{ML})$$

under the constraint that $\text{var}_{\hat{\mathbb{P}}}(\hat{Y}_{ML}) < (1 - q)\varepsilon^2$. As a result of this, we obtain the following values for L and N_l .

Let

$$L = \left\lceil \frac{\log(q^{-1/2} c_{1,\alpha} \varepsilon^{-1} T^\alpha)}{\alpha \log(M)} \right\rceil$$

and

$$\begin{aligned} N_0 &= \left\lceil \frac{1}{1 - q} \varepsilon^{-2} h_0^{(\beta+\gamma)/2} \left(\frac{c_{2,0}}{c_{3,0}} \right)^{1/2} \kappa \right\rceil, \\ N_l &= \left\lceil \frac{1}{1 - q} \varepsilon^{-2} h_l^{(\beta+\gamma)/2} \left(\frac{c_2}{c_3} \right)^{1/2} \kappa \right\rceil \end{aligned} \tag{2.3}$$

for $l = 1, \dots, L - 1$ and

$$N_L = \left\lceil \frac{1}{1 - q} \varepsilon^{-2} h_L^{(\beta_L+\gamma_L)/2} \left(\frac{c_{2,L}}{c_{3,L}} \right)^{1/2} \kappa \right\rceil$$

for some $q \in (0, 1)$, with

- in the $\beta > \gamma$ and $\beta_L \geq \gamma_L$ case or in the $\beta < \gamma$ and $\gamma_L - \beta_L \leq \gamma - \beta$ case

$$\begin{aligned} \kappa &= (c_{2,0} c_{3,0})^{1/2} T^{(\beta-\gamma)/2} + (c_2 c_3)^{1/2} \frac{(M^{-1} T)^{(\beta-\gamma)/2} - h_L^{(\beta-\gamma)/2}}{1 - M^{(\gamma-\beta)/2}} \\ &\quad + (c_{2,L} c_{3,L})^{1/2} h_L^{(\beta_L-\gamma_L)/2}, \end{aligned}$$

- in the $\beta = \gamma$ and $\beta_L \geq \gamma_L$ case

$$\kappa = (c_{2,0} c_{3,0})^{1/2} + (L - 1)(c_2 c_3)^{1/2} + (c_{2,L} c_{3,L})^{1/2} h_L^{(\beta_L-\gamma_L)/2}.$$

3. The improved MLMC estimator

The order of convergence of the MLMC estimator \hat{Y}_{ML} given in (2.1) is optimal in the given framework. However, the computational costs can be reduced if a modified estimator is applied. As yet, the estimator \hat{Y}_{ML} is based on some weak order α approximations Y^l for $l = 0, 1, \dots, L$ on each level. Now, let us apply some cheap low-order weak approximation Y^l on levels $l = 0, 1, \dots, L - 1$ combined with some probably expansive high-order weak approximation \check{Y}^L on the finest level L . The idea is that the approximations Y^l contribute a variance reduction, while the approximation \check{Y}^L results in a small bias of the MLMC estimator, thus reducing the number of levels needed to attain a prescribed accuracy.

Let Y be an order α weak approximation method and let \check{Y} be an order p weak approximation method applied on the finest level. Furthermore, let $L = L_p$ with

$$L_p = \left\lceil \frac{\log(q^{-1/2}c_{1,p}\varepsilon^{-1}T^p)}{p \log(M)} \right\rceil$$

denote the number of levels in order to indicate the dependence on the weak order p . Then, we define the modified MLMC estimator by

$$\hat{Y}_{ML(\alpha,p)} = \sum_{l=0}^{L_p} \hat{Y}^l$$

with the estimators \hat{Y}^l for $l = 0, 1, \dots, L_p - 1$ based on the order α weak approximations Y^l as defined in Section 2. However, now applying the modified estimator

$$\hat{Y}^{L_p} = \frac{1}{N_{L_p}} \sum_{i=1}^{N_{L_p}} (f(\check{Y}^{L_p})^{(i)} - f(Y^{L_p-1})^{(i)}),$$

which combines the weak order α approximations Y^{L_p-1} with the weak order p approximations \check{Y}^{L_p} . Clearly, all conditions of Theorem 2.1 have to be fulfilled for Y^L replaced by \check{Y}^L . Then, in the case of $p > \alpha$, the improved MLMC estimator $\hat{Y}_{ML(\alpha,p)}$ features significantly reduced computational costs compared to the originally proposed estimator $\hat{Y}_{ML} = \hat{Y}_{ML(\alpha,\alpha)}$.

Proposition 3.1. *Let conditions (i)–(vii) of Theorem 2.1 be fulfilled and suppose that there exist constants $\hat{c}_{3,0}, \hat{c}_3, \hat{c}_{3,L_p}, \delta_i > 0$ and $\hat{c}_{3,0}^{(i)}, \hat{c}_3^{(i)}, \hat{c}_{3,L_p}^{(i)} \geq 0$ such that for the computational costs*

- (v') $C(Y^0) = \hat{c}_{3,0}Th_0^{-\gamma} + \sum_{i=1}^k \hat{c}_{3,0}^{(i)}Th_0^{-\gamma+\delta_i}$,
- (vi') $C(Y^l, Y^{l-1}) = \hat{c}_3Th_l^{-\gamma} + \sum_{i=1}^k \hat{c}_3^{(i)}Th_l^{-\gamma+\delta_i}$ for $l = 1, \dots, L_p - 1$,
- (vii') $C(\check{Y}^{L_p}, Y^{L_p-1}) = \hat{c}_{3,L_p}Th_{L_p}^{-\gamma_{L_p}} + \sum_{i=1}^k \hat{c}_{3,L_p}^{(i)}Th_{L_p}^{-\gamma_{L_p}+\delta_i}$

with some $\gamma, \gamma_{L_p} \geq 1$ such that $\gamma - \delta_i \geq 1$ and $\gamma_{L_p} - \delta_i \geq 1$. Then the MLMC estimator $\hat{Y}_{ML(\alpha,p)}$ based on a weak order $\alpha > 0$ approximation scheme on levels $0, 1, \dots, L_p - 1$ and some weak order $p > \alpha$ approximation scheme on level L_p has reduced computational costs.

- (i) *In the $\beta > \gamma$ and $\beta - \gamma < \beta_{L_p} - \gamma_{L_p}$ case, there exists some $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0]$ it holds that*

$$\frac{C(\hat{Y}_{ML(\alpha,\alpha)}(\varepsilon))}{C(\hat{Y}_{ML(\alpha,p)}(\varepsilon))} > 1, \tag{3.1}$$

provided that $\alpha \geq \gamma/2$, $p \geq \frac{1}{2} \max\{\gamma, \gamma_{L_p}\}$, and $p > \frac{1}{4} \max\{\beta + \gamma, \beta - \gamma + 2\gamma_{L_p}\}$. In the $\beta > \gamma$ and $\beta - \gamma = \beta_{L_p} - \gamma_{L_p}$ case then (3.1) holds if, in addition, $c_2 c_3 > (1 - M^{(\gamma-\beta)/2})^2 c_{2,L_p} c_{3,L_p}$ and $\hat{c}_3^2 c_2 / c_3 > (1 - M^{(\gamma-\beta)/2})^2 \hat{c}_{3,L_p}^2 c_{2,L_p} / c_{3,L_p}$. Furthermore, for $0 < \beta - \gamma \leq \beta_{L_p} - \gamma_{L_p}$ it holds that $C(\hat{Y}_{ML(\alpha,p)})(\varepsilon) = O(\varepsilon^{-2})$ if $\alpha > 0$ and $p \geq \frac{1}{2} \max\{\gamma, \gamma_{L_p}\}$.

(ii) In the $\beta = \gamma$ and $\beta_{L_p} \geq \gamma_{L_p}$ case and if $p \geq \frac{1}{2} \max\{\gamma, \gamma_{L_p}\}$, $\alpha \geq \gamma/2$, it holds that

$$\lim_{\varepsilon \rightarrow 0} \frac{C(\hat{Y}_{ML(\alpha,\alpha)})(\varepsilon)}{C(\hat{Y}_{ML(\alpha,p)})(\varepsilon)} \geq \left(\frac{p}{\alpha}\right)^2 \tag{3.2}$$

and $C(\hat{Y}_{ML(\alpha,p)})(\varepsilon) = O(\varepsilon^{-2}(\log(\varepsilon))^2)$ if $\alpha > 0$ and $p \geq \frac{1}{2} \max\{\gamma, \gamma_{L_p}\}$.

(iii) In the $\beta < \gamma$ and $\gamma - \beta = \gamma_{L_p} - \beta_{L_p}$ case it holds that

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \frac{C(\hat{Y}_{ML(p,p)})(\varepsilon)}{C(\hat{Y}_{ML(\alpha,p)})(\varepsilon)} & \tag{3.3} \\ & \geq M^{2(\gamma-\beta)} \left(\frac{\hat{c}_3 c_2}{\hat{c}_{3,L_p} c_{2,L_p}} + \frac{\hat{c}_3 (c_2 c_{3,L_p})^{1/2}}{\hat{c}_{3,L_p} (c_{2,L_p} c_3)^{1/2}} (M^{(\gamma-\beta)/2} - 1) \right. \\ & \quad \left. + \left(\frac{c_2 c_3}{c_{2,L_p} c_{3,L_p}} \right)^{1/2} (M^{(\gamma-\beta)/2} - 1) + (M^{(\gamma-\beta)/2} - 1)^2 \right)^{-1} \tag{3.4} \end{aligned}$$

if $p > \frac{1}{2}(\max\{\gamma, \gamma_{L_p}\} - \gamma + \beta)$. If the parameter $q \in (0, 1)$ is chosen as

$$q = \frac{\gamma - \beta}{\gamma - \beta + 2p}$$

then the computational costs $C(\hat{Y}_{ML(\alpha,p)})$ are asymptotically minimal. In general, if $\beta < \gamma$ or if $\beta_{L_p} < \gamma_{L_p}$ then it holds that $C(\hat{Y}_{ML(\alpha,p)})(\varepsilon) = O(\varepsilon^{-2 - \max\{\gamma-\beta, \gamma_{L_p} - \beta_{L_p}\}/p})$ for $p \geq \frac{1}{2}(\max\{\gamma, \gamma_{L_p}\} - \min\{\gamma - \beta, \gamma_{L_p} - \beta_{L_p}\})$.

We note that in relations (v')–(vii') of Proposition 3.1 a more detailed polynomial dependence of the computational costs from the dimension of the approximation subspaces has to be taken into account, e.g. standard discrete-time approximation methods possess polynomial computational costs and the constants are known explicitly.

Proof. In the following, we will first state some basic formulae and conditions used in the remaining part of the proof. Then, we will calculate lower and upper bounds for the computational costs in the $\beta \neq \gamma$ case. We will then use these to prove first (i) and then (iii). Finally, case (ii) with $\beta = \gamma$ is considered.

Basic formulae. Assume that $\varepsilon < 1$. Let $\delta_0 = 0$, $\hat{c}_{3,0}^{(0)} = \hat{c}_{3,0}$, $\hat{c}_3^{(0)} = \hat{c}_3$, and $\hat{c}_{3,L_p}^{(0)} = \hat{c}_{3,L_p}$. Then the computational costs for $\hat{Y}_{ML(\alpha,p)}$ are

$$\begin{aligned} C(\hat{Y}_{ML(\alpha,p)}) & = \sum_{i=0}^k \hat{c}_{3,0}^{(i)} Th_0^{-\gamma+\delta_i} N_0 + \sum_{i=0}^k \sum_{l=1}^{L_p-1} \hat{c}_3^{(i)} Th_l^{-\gamma+\delta_i} N_l \\ & \quad + \sum_{i=0}^k \hat{c}_{3,L_p}^{(i)} Th_{L_p}^{-\gamma_{L_p}+\delta_i} N_{L_p} \end{aligned}$$

with $L = L_p = \lceil \log(q^{-1/2}c_{1,p}\varepsilon^{-1}T^p)/(p \log(M)) \rceil$ and N_l for $l = 0, 1, \dots, L_p$ given in (2.3). Without loss of generality, suppose that $\delta_i \neq \delta_j$ for $i \neq j$ and that $\delta_k = (\gamma - \beta)/2$ with $\hat{c}_{3,0}^{(k)} = \hat{c}_3^{(k)} = \hat{c}_{3,L_p}^{(k)} = 0$ in the $\beta \geq \gamma$ case. In the following, we make use of the two estimates:

$$\begin{aligned} L_\alpha &\geq \frac{\log(\varepsilon^{-1})}{\alpha \log(M)} + \frac{\log(q^{-1/2}c_{1,\alpha}T^\alpha)}{\alpha \log(M)}, \\ L_p - 1 &\leq \frac{\log(\varepsilon^{-1})}{p \log(M)} + \frac{\log(q^{-1/2}c_{1,p}T^p)}{p \log(M)}. \end{aligned} \tag{3.5}$$

Lower bound for $\beta \neq \gamma$. Let $\beta \neq \gamma$. Then, we obtain the lower bound

$$\begin{aligned} C(\hat{Y}_{ML(\alpha,\alpha)})(\varepsilon) &\tag{3.6} \\ &\geq \frac{T\kappa\varepsilon^{-2}}{1-q} \sum_{i=0}^k \left(h_0^{(\beta-\gamma)/2+\delta_i} \hat{c}_{3,0}^{(i)} \left(\frac{c_{2,0}}{c_{3,0}} \right)^{1/2} + \sum_{l=1}^{L_\alpha} h_l^{(\beta-\gamma)/2+\delta_i} \hat{c}_3^{(i)} \left(\frac{c_2}{c_3} \right)^{1/2} \right) \\ &\geq \frac{T}{1-q} \varepsilon^{-2} \left[\sum_{i=0}^k T^{\beta-\gamma+\delta_i} \hat{c}_{3,0}^{(i)} c_{2,0} \right. \\ &\quad + \sum_{i=0}^k T^{(\beta-\gamma)/2+\delta_i} \hat{c}_{3,0}^{(i)} \left(\frac{c_{2,0}c_{2,0}c_{3,0}}{c_{3,0}} \right)^{1/2} \frac{T^{(\beta-\gamma)/2} - h_{L_\alpha}^{(\beta-\gamma)/2}}{M^{(\beta-\gamma)/2} - 1} \\ &\quad + \sum_{i=0}^{k-1} \hat{c}_3^{(i)} \left(\frac{c_2c_{2,0}c_{3,0}}{c_3} \right)^{1/2} T^{(\beta-\gamma)/2} \frac{T^{(\beta-\gamma)/2+\delta_i} - h_{L_\alpha}^{(\beta-\gamma)/2+\delta_i}}{M^{(\beta-\gamma)/2+\delta_i} - 1} \\ &\quad + \hat{c}_3^{(k)} \left(\frac{c_2c_{2,0}c_{3,0}}{c_3} \right)^{1/2} T^{(\beta-\gamma)/2} \left(\frac{\log(\varepsilon^{-1})}{\alpha \log(M)} + \frac{\log(q^{-1/2}c_{1,\alpha}T^\alpha)}{\alpha \log(M)} \right) \\ &\quad + \sum_{i=0}^{k-1} \hat{c}_3^{(i)} c_2 \frac{T^{(\beta-\gamma)/2+\delta_i} - h_{L_\alpha}^{(\beta-\gamma)/2+\delta_i}}{M^{(\beta-\gamma)/2+\delta_i} - 1} \frac{T^{(\beta-\gamma)/2} - h_{L_\alpha}^{(\beta-\gamma)/2}}{M^{(\beta-\gamma)/2} - 1} \\ &\quad \left. + \hat{c}_3^{(k)} c_2 \frac{T^{(\beta-\gamma)/2} - h_{L_\alpha}^{(\beta-\gamma)/2}}{M^{(\beta-\gamma)/2} - 1} \left(\frac{\log(\varepsilon^{-1})}{\alpha \log(M)} + \frac{\log(q^{-1/2}c_{1,\alpha}T^\alpha)}{\alpha \log(M)} \right) \right], \end{aligned} \tag{3.7}$$

where $\hat{c}_{3,L_\alpha}^{(i)} = \hat{c}_3^{(i)}$, $c_{2,L_\alpha} = c_2$, $c_{3,L_\alpha} = c_3$, $\beta_{L_\alpha} = \beta$, and $\gamma_{L_\alpha} = \gamma$ for $\hat{Y}_{ML(\alpha,\alpha)}$.

Upper bound for $\beta \neq \gamma$. Next, we calculate the upper bound for the $\beta \neq \gamma$ case. Thus,

$$\begin{aligned} C(\hat{Y}_{ML(\alpha,p)})(\varepsilon) &\tag{3.6} \\ &\leq \frac{T\kappa\varepsilon^{-2}}{1-q} \sum_{i=0}^k \left(h_0^{(\beta-\gamma)/2+\delta_i} \hat{c}_{3,0}^{(i)} \left(\frac{c_{2,0}}{c_{3,0}} \right)^{1/2} + \sum_{l=1}^{L_p-1} h_l^{(\beta-\gamma)/2+\delta_i} \hat{c}_3^{(i)} \left(\frac{c_2}{c_3} \right)^{1/2} \right. \\ &\quad \left. + h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2+\delta_i} \hat{c}_{3,L_p}^{(i)} \left(\frac{c_{2,L_p}}{c_{3,L_p}} \right)^{1/2} \right) \\ &\quad + T \sum_{i=0}^k \left(\hat{c}_{3,0}^{(i)} h_0^{-\gamma+\delta_i} + \hat{c}_3^{(i)} \sum_{l=1}^{L_p-1} h_l^{-\gamma+\delta_i} + \hat{c}_{3,L_p}^{(i)} h_{L_p}^{-\gamma_{L_p}+\delta_i} \right) \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{T}{1-q} \varepsilon^{-2} \left[\sum_{i=0}^k \hat{c}_{3,0}^{(i)} \left(c_{2,0} T^{\beta-\gamma+\delta_i} + \left(\frac{c_{2,0} c_{2,0} c_{3,0}}{c_{3,0}} \right)^{1/2} \Lambda_0 T^{(\beta-\gamma)/2+\delta_i} \right. \right. \\
 &\quad \left. \left. + \left(\frac{c_{2,0} c_{2,L_p} c_{3,L_p}}{c_{3,0}} \right)^{1/2} T^{(\beta-\gamma)/2+\delta_i} h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2} \right) \right. \\
 &\quad + \sum_{i=0}^{k-1} \hat{c}_3^{(i)} \left(\frac{c_{2,0} c_{3,0} c_2}{c_3} \right)^{1/2} T^{(\beta-\gamma)/2} \Lambda_i + \sum_{i=0}^{k-1} \hat{c}_3^{(i)} c_2 \Lambda_i \Lambda_0 \\
 &\quad + \left(\hat{c}_3^{(k)} c_2 \Lambda_0 + \hat{c}_3^{(k)} \left(\frac{c_2 c_{2,L_p} c_{3,L_p}}{c_3} \right)^{1/2} h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2} \right. \\
 &\quad \left. + \hat{c}_3^{(k)} \left(\frac{c_{2,0} c_{3,0} c_2}{c_3} \right)^{1/2} T^{(\beta-\gamma)/2} \right) \\
 &\quad \times \left(\frac{\log(\varepsilon^{-1})}{p \log(M)} + \frac{\log(q^{-1/2} c_{1,p} T^p)}{p \log(M)} \right) \\
 &\quad + \sum_{i=0}^{k-1} \hat{c}_3^{(i)} \left(\frac{c_2 c_{2,L_p} c_{3,L_p}}{c_3} \right)^{1/2} \Lambda_i h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2} \\
 &\quad + \sum_{i=0}^k \hat{c}_{3,L_p}^{(i)} \left(\frac{c_{2,0} c_{3,0} c_{2,L_p}}{c_{3,L_p}} \right)^{1/2} T^{(\beta-\gamma)/2} h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2+\delta_i} \\
 &\quad + \sum_{i=0}^k \hat{c}_{3,L_p}^{(i)} \left(\left(\frac{c_2 c_{3,L_p} c_{2,L_p}}{c_{3,L_p}} \right)^{1/2} \Lambda_0 h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2+\delta_i} \right. \\
 &\quad \left. + c_{2,L_p} h_{L_p}^{\beta_{L_p}-\gamma_{L_p}+\delta_i} \right) \Big] \\
 &\quad + T \sum_{i=0}^k \left(\hat{c}_{3,0}^{(i)} T^{\delta_i-\gamma} + \hat{c}_3^{(i)} \frac{(M^{-1} T)^{\delta_i-\gamma} - h_{L_p}^{\delta_i-\gamma}}{1 - M^{\gamma-\delta_i}} + \hat{c}_{3,L_p}^{(i)} h_{L_p}^{\delta_i-\gamma_{L_p}} \right) \tag{3.8}
 \end{aligned}$$

with $\Lambda_i = ((M^{-1} T)^{(\beta-\gamma)/2+\delta_i} - h_{L_p}^{(\beta-\gamma)/2+\delta_i}) / (1 - M^{(\gamma-\beta)/2-\delta_i})$ for $i = 0, \dots, k - 1$.

Proof of (i). In the $\beta > \gamma$ and $\beta_{L_p} > \gamma_{L_p}$ case, we prove that there exists some $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0]$ it follows that $C(\hat{Y}_{ML(\alpha,\alpha)})(\varepsilon) > C(\hat{Y}_{ML(\alpha,p)})(\varepsilon)$. From the lower bound (3.7) for $C(\hat{Y}_{ML(\alpha,\alpha)})(\varepsilon)$ and the upper bound (3.8) for $C(\hat{Y}_{ML(\alpha,p)})(\varepsilon)$ we obtain the following estimate:

$$\begin{aligned}
 &C(\hat{Y}_{ML(\alpha,\alpha)})(\varepsilon) - C(\hat{Y}_{ML(\alpha,p)})(\varepsilon) \\
 &\geq \frac{T}{1-q} \varepsilon^{-2} \\
 &\quad \times \left(\sum_{i=0}^{k-1} T^{(\beta-\gamma)/2+\delta_i} \hat{c}_{3,0}^{(i)} \left(\frac{c_{2,0} c_{2,0} c_{3,0}}{c_{3,0}} \right)^{1/2} h_{L_p}^{(\beta-\gamma)/2} - \frac{M^{(\gamma-\beta)/2} h_{L_\alpha}^{(\beta-\gamma)/2}}{1 - M^{(\gamma-\beta)/2}} \right. \\
 &\quad \left. + \sum_{i=0}^{k-1} \hat{c}_3^{(i)} \left(\frac{c_2 c_{2,0} c_{3,0}}{c_3} \right)^{1/2} T^{(\beta-\gamma)/2} h_{L_p}^{(\beta-\gamma)/2+\delta_i} - \frac{M^{(\gamma-\beta)/2-\delta_i} h_{L_\alpha}^{(\beta-\gamma)/2+\delta_i}}{1 - M^{(\gamma-\beta)/2-\delta_i}} \right)
 \end{aligned}$$

$$\begin{aligned}
 & + \sum_{i=0}^{k-1} \hat{c}_3^{(i)} c_2 \left(\frac{(M^{-1}T)^{(\beta-\gamma)/2+\delta_i} (h_{L_p}^{(\beta-\gamma)/2} - M^{(\gamma-\beta)/2} h_{L_\alpha}^{(\beta-\gamma)/2})}{(1 - M^{(\gamma-\beta)/2-\delta_i})(1 - M^{(\gamma-\beta)/2})} \right) \\
 & + \frac{(M^{-1}T)^{(\beta-\gamma)/2} (h_{L_p}^{(\beta-\gamma)/2+\delta_i} - M^{(\gamma-\beta)/2-\delta_i} h_{L_\alpha}^{(\beta-\gamma)/2+\delta_i}) - h_{L_p}^{\beta-\gamma+\delta_i} + M^{\gamma-\beta-\delta_i} h_{L_\alpha}^{\beta-\gamma+\delta_i}}{(1 - M^{(\gamma-\beta)/2-\delta_i})(1 - M^{(\gamma-\beta)/2})} \\
 & - \sum_{i=0}^{k-1} \hat{c}_{3,0}^{(i)} \left(\frac{c_{2,0} c_{2,L_p} c_{3,L_p}}{c_{3,0}} \right)^{1/2} T^{(\beta-\gamma)/2+\delta_i} h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2} \\
 & - \sum_{i=0}^{k-1} \hat{c}_3^{(i)} \left(\frac{c_2 c_{2,L_p} c_{3,L_p}}{c_3} \right)^{1/2} \Lambda_i h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2} \\
 & - \sum_{i=0}^{k-1} \hat{c}_{3,L_p}^{(i)} \left(\frac{c_{2,0} c_{3,0} c_{2,L_p}}{c_{3,L_p}} \right)^{1/2} T^{(\beta-\gamma)/2} h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2+\delta_i} \\
 & - \sum_{i=0}^{k-1} \hat{c}_{3,L_p}^{(i)} \left(\left(\frac{c_2 c_{3,L_p}}{c_{3,L_p}} \right)^{1/2} \Lambda_0 h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2+\delta_i} + c_{2,L_p} h_{L_p}^{\beta_{L_p}-\gamma_{L_p}+\delta_i} \right) \\
 & - T \sum_{i=0}^{k-1} \left(\hat{c}_{3,0}^{(i)} T^{\delta_i-\gamma} + \hat{c}_3^{(i)} \frac{(M^{-1}T)^{\delta_i-\gamma} - h_{L_p}^{\delta_i-\gamma}}{1 - M^{\gamma-\delta_i}} + \hat{c}_{3,L_p}^{(i)} h_{L_p}^{\delta_i-\gamma_{L_p}} \right). \tag{3.9}
 \end{aligned}$$

In the following, we make use of the estimates $M^{-1} c_{1,\alpha}^{-1/\alpha} q^{1/(2\alpha)} \varepsilon^{1/\alpha} \leq h_{L_\alpha} \leq c_{1,\alpha}^{-1/\alpha} q^{1/(2\alpha)} \varepsilon^{1/\alpha}$ and $M^{-1} c_{1,p}^{-1/p} q^{1/(2p)} \varepsilon^{1/p} \leq h_{L_p} \leq c_{1,p}^{-1/p} q^{1/(2p)} \varepsilon^{1/p}$, i.e. we have $h_{L_p} \rightarrow 0$ and $h_{L_\alpha} \rightarrow 0$ as $\varepsilon \rightarrow 0$.

Multiplying both sides of (3.9) with $(1 - q)/T \varepsilon^2 h_{L_p}^{-\min\{\beta-\gamma, \beta_{L_p}-\gamma_{L_p}\}/2}$ and taking into account the assumptions $4p > \beta + \gamma$ and $4p > \beta - \gamma + 2\gamma_{L_p}$ results in

$$\begin{aligned}
 & \frac{1 - q}{T} \varepsilon^2 h_{L_p}^{-\min\{\beta-\gamma, \beta_{L_p}-\gamma_{L_p}\}/2} (C(\hat{Y}_{ML(\alpha,\alpha)})(\varepsilon) - C(\hat{Y}_{ML(\alpha,p)})(\varepsilon)) \\
 & \geq \left[\sum_{i=0}^{k-1} T^{(\beta-\gamma)/2+\delta_i} \hat{c}_{3,0}^{(i)} \left(\frac{c_{2,0}}{c_{3,0}} \right)^{1/2} \right. \\
 & \quad \times \left((c_2 c_3)^{1/2} \frac{h_{L_p}^{(\beta-\gamma)/2}}{1 - M^{(\gamma-\beta)/2}} - (c_{2,L_p} c_{3,L_p})^{1/2} h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2} \right) \\
 & \quad + T^{(\beta-\gamma)/2} (c_{2,0} c_{3,0})^{1/2} \left(\hat{c}_3^{(0)} \left(\frac{c_2}{c_3} \right)^{1/2} \frac{h_{L_p}^{(\beta-\gamma)/2}}{1 - M^{(\gamma-\beta)/2}} \right. \\
 & \quad \quad \left. \left. - \hat{c}_{3,L_p}^{(0)} \left(\frac{c_{2,L_p}}{c_{3,L_p}} \right)^{1/2} h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2} \right) \right. \\
 & \quad \left. + \sum_{i=0}^{k-1} \hat{c}_3^{(i)} c_2^{1/2} \frac{(M^{-1}T)^{(\beta-\gamma)/2+\delta_i}}{1 - M^{(\gamma-\beta)/2-\delta_i}} \left(c_2^{1/2} \frac{h_{L_p}^{(\beta-\gamma)/2}}{1 - M^{(\gamma-\beta)/2}} - \left(\frac{c_{2,L_p} c_{3,L_p}}{c_3} \right)^{1/2} \right. \right. \\
 & \quad \quad \left. \left. \times h_{L_p}^{(\beta_{L_p}-\gamma_{L_p})/2} \right) \right]
 \end{aligned}$$

$$\begin{aligned}
 &+ \frac{(M^{-1}T)^{(\beta-\gamma)/2}}{1 - M^{(\gamma-\beta)/2}} c_2^{1/2} \left(\hat{c}_3^{(0)} c_2^{1/2} \frac{h_{L_p}^{(\beta-\gamma)/2}}{1 - M^{(\gamma-\beta)/2}} - \hat{c}_{3,L_p}^{(0)} \left(\frac{c_3 c_{2,L_p}}{c_{3,L_p}} \right)^{1/2} \right. \\
 &\quad \left. \times h_{L_p}^{(\beta_{L_p} - \gamma_{L_p})/2} \right) \\
 &+ o(h_{L_p}^{\min\{\beta_{L_p} - \gamma_{L_p}, \beta - \gamma\}/2}) \Big] h_{L_p}^{-\min\{\beta - \gamma, \beta_{L_p} - \gamma_{L_p}\}/2}. \tag{3.10}
 \end{aligned}$$

As a result of (3.10), it follows that in the $\beta - \gamma < \beta_{L_p} - \gamma_{L_p}$ case there exists some $\varepsilon_0 > 0$ such that

$$\frac{C(\hat{Y}_{ML(\alpha,\alpha)})(\varepsilon)}{C(\hat{Y}_{ML(\alpha,p)})(\varepsilon)} > 1 \tag{3.11}$$

for all $\varepsilon \in (0, \varepsilon_0]$. In the $\beta - \gamma = \beta_{L_p} - \gamma_{L_p}$ case there exists some $\varepsilon_0 > 0$ such that (3.11) holds for all $\varepsilon \in (0, \varepsilon_0]$ if $c_2 c_3 > (1 - M^{(\gamma-\beta)/2})^2 c_{2,L_p} c_{3,L_p}$ and $(\hat{c}_3^{(0)})^2 c_2 / c_3 > (1 - M^{(\gamma-\beta)/2})^2 (\hat{c}_{3,L_p}^{(0)})^2 c_{2,L_p} / c_{3,L_p}$. Finally, $C(\hat{Y}_{ML(\alpha,p)})(\varepsilon) = O(\varepsilon^{-2})$ follows from (3.8).

Proof of (iii). In the $\beta < \gamma$ and $\beta < 2p$ case, we have to compare the dominating terms as $\varepsilon \rightarrow 0$. Therefore, from the lower bound we obtain

$$\begin{aligned}
 C(\hat{Y}_{ML(p,p)})(\varepsilon) &\geq \frac{q^{(\beta-\gamma)/(2p)}}{1 - q} \varepsilon^{-2-(\gamma-\beta)/p} T \hat{c}_{3,L_p}^{(0)} c_{2,L_p} c_{1,p}^{(\gamma-\beta)/p} M^{\gamma-\beta} (M^{(\beta-\gamma)/2} - 1)^{-2} \\
 &+ o(\varepsilon^{-2-(\gamma-\beta)/p}) \tag{3.12}
 \end{aligned}$$

and from the upper bound,

$$\begin{aligned}
 C(\hat{Y}_{ML(\alpha,p)})(\varepsilon) &\leq \frac{q^{(\beta-\gamma)/(2p)}}{1 - q} \varepsilon^{-2-(\gamma-\beta)/p} T c_{1,p}^{(\gamma-\beta)/p} \\
 &\times \left(\frac{\hat{c}_3^{(0)} c_2}{(1 - M^{(\gamma-\beta)/2})^2} - \frac{\hat{c}_3^{(0)} (c_2 c_{2,L_p} c_{3,L_p})^{1/2}}{c_3^{1/2} (1 - M^{(\gamma-\beta)/2})} \right. \\
 &\quad \left. - \frac{\hat{c}_{3,L_p}^{(0)} (c_2 c_3 c_{2,L_p})^{1/2}}{c_{3,L_p}^{1/2} (1 - M^{(\gamma-\beta)/2})} + \hat{c}_{3,L_p}^{(0)} c_{2,L_p} \right) + o(\varepsilon^{-2-(\gamma-\beta)/p}). \tag{3.13}
 \end{aligned}$$

Making use of estimates (3.12) and (3.13), this results in estimate (3.4) where $\beta_{L_p} < \gamma_{L_p}$ because we require that $\beta_{L_p} - \gamma_{L_p} = \beta - \gamma < 0$.

In general, it follows that $C(\hat{Y}_{ML(\alpha,p)})(\varepsilon) = O(\varepsilon^{-2-\max\{\gamma-\beta, \gamma_{L_p} - \beta_{L_p}\}/p})$ due to the upper bound (3.8) for $\beta < \gamma$ and any $\beta_{L_p} > 0, \gamma_{L_p} \geq 1$. Furthermore, there is an asymptotically optimal choice for the parameter $q \in (0, 1)$, such that the computational costs are asymptotically minimal. Calculating a lower bound for $C(\hat{Y}_{ML(\alpha,p)})(\varepsilon)$ and taking into account the upper bound (3.13), we obtain

$$C(\hat{Y}_{ML(\alpha,p)})(\varepsilon) = \frac{1}{1 - q} \varepsilon^{-2-(\gamma-\beta)/p} q^{(\beta-\gamma)/(2p)} C + o(\varepsilon^{-2-(\gamma-\beta)/p})$$

with some constant $C > 0$ independent of q and ε . Now, we have to find some $\hat{q} \in (0, 1)$ such that

$$C \varepsilon^{-2-(\gamma-\beta)/p} \frac{\hat{q}^{(\beta-\gamma)/(2p)}}{1 - \hat{q}} = \min_{q \in (0,1)} C \varepsilon^{-2-(\gamma-\beta)/p} \frac{q^{(\beta-\gamma)/(2p)}}{1 - q}$$

for all $0 < \varepsilon < 1$. Solving this minimization problem leads to

$$\hat{q} = \frac{\gamma - \beta}{\gamma - \beta + 2p},$$

which is asymptotically the optimal choice for $q \in (0, 1)$ in the $\beta < \gamma$ case.

Lower bound for $\beta = \gamma$. In the $\beta = \gamma$ case, we obtain the following lower bound:

$$\begin{aligned} & C(\hat{Y}_{ML(\alpha,\alpha)})(\varepsilon) \\ & \geq \frac{T}{1-q} \varepsilon^{-2} \left(\sum_{i=0}^k \hat{c}_{3,0}^{(i)} c_{2,0} h_0^{\delta_i} + \sum_{i=0}^k \hat{c}_{3,0}^{(i)} \left(\frac{c_{2,0} c_2 c_3}{c_{3,0}} \right)^{1/2} L_\alpha h_0^{\delta_i} \right. \\ & \quad + \hat{c}_3^{(0)} \left(\frac{c_2 c_{2,0} c_{3,0}}{c_3} \right)^{1/2} L_\alpha + \sum_{i=1}^k \hat{c}_3^{(i)} \left(\frac{c_2 c_{2,0} c_{3,0}}{c_3} \right)^{1/2} \\ & \quad \times \frac{T^{\delta_i} - h_{L_\alpha}^{\delta_i}}{M^{\delta_i} - 1} + \hat{c}_3^{(0)} c_2 L_\alpha^2 + \sum_{i=1}^k \hat{c}_3^{(i)} c_2 L_\alpha \frac{T^{\delta_i} - h_{L_\alpha}^{\delta_i}}{M^{\delta_i} - 1} \Big) \\ & \geq \frac{T}{1-q} \varepsilon^{-2} \left(\sum_{i=0}^k \hat{c}_{3,0}^{(i)} c_{2,0} T^{\delta_i} + \left(\sum_{i=0}^k \hat{c}_{3,0}^{(i)} \left(\frac{c_{2,0} c_2 c_3}{c_{3,0}} \right)^{1/2} T^{\delta_i} + \hat{c}_3^{(0)} \left(\frac{c_2 c_{2,0} c_{3,0}}{c_3} \right)^{1/2} \right. \right. \\ & \quad \left. \left. + \sum_{i=1}^k \hat{c}_3^{(i)} c_2 \frac{T^{\delta_i} - h_{L_\alpha}^{\delta_i}}{M^{\delta_i} - 1} \right) \right. \\ & \quad \times \left(\frac{\log(\varepsilon^{-1})}{\alpha \log(M)} + \frac{\log(q^{-1/2} c_{1,\alpha} T^\alpha)}{\alpha \log(M)} \right) + \hat{c}_3^{(0)} c_2 \left(\frac{\log(\varepsilon^{-1})}{\alpha \log(M)} \right)^2 \\ & \quad + 2 \hat{c}_3^{(0)} c_2 \frac{\log(\varepsilon^{-1}) \log(q^{-1/2} c_{1,\alpha} T^\alpha)}{\alpha^2 (\log(M))^2} + \hat{c}_3^{(0)} c_2 \left(\frac{\log(q^{-1/2} c_{1,\alpha} T^\alpha)}{\alpha \log(M)} \right)^2 \\ & \quad \left. + \sum_{i=1}^k \hat{c}_3^{(i)} \left(\frac{c_2 c_{2,0} c_{3,0}}{c_3} \right)^{1/2} \frac{T^{\delta_i} - h_{L_\alpha}^{\delta_i}}{M^{\delta_i} - 1} \right), \tag{3.14} \end{aligned}$$

where $\hat{c}_{3,L_\alpha}^{(i)} = \hat{c}_3^{(i)}$, $c_{2,L_\alpha} = c_2$, $c_{3,L_\alpha} = c_3$, $\beta_{L_\alpha} = \beta$, and $\gamma_{L_\alpha} = \gamma$ for $\hat{Y}_{ML(\alpha,\alpha)}$.

Upper bound for $\beta = \gamma$. Next, we calculate the upper bound for $\beta = \gamma$,

$$\begin{aligned} & C(\hat{Y}_{ML(\alpha,p)})(\varepsilon) \\ & \leq \frac{T}{1-q} \varepsilon^{-2} \left[\sum_{i=0}^k \hat{c}_{3,0}^{(i)} c_{2,0} T^{\delta_i} + \sum_{i=1}^k \hat{c}_3^{(i)} \left(\frac{c_{2,0} c_3 c_2}{c_3} \right)^{1/2} \Lambda_i \right. \\ & \quad + \sum_{i=0}^k \hat{c}_{3,0}^{(i)} \left(\frac{c_{2,0} c_{2,L_p} c_{3,L_p}}{c_{3,0}} \right)^{1/2} T^{\delta_i} h_{L_p}^{(\beta_{L_p} - \gamma_{L_p})/2} \\ & \quad + \left(\hat{c}_3^{(0)} \left(\frac{c_{2,0} c_3 c_2}{c_3} \right)^{1/2} + \sum_{i=0}^k \hat{c}_{3,0}^{(i)} \left(\frac{c_{2,0} c_2 c_3}{c_{3,0}} \right)^{1/2} T^{\delta_i} \right. \\ & \quad \left. \left. + \hat{c}_3^{(0)} \left(\frac{c_2 c_{2,L_p} c_{3,L_p}}{c_3} \right)^{1/2} h_{L_p}^{(\beta_{L_p} - \gamma_{L_p})/2} \right) \right] \end{aligned}$$

$$\begin{aligned}
 & + \sum_{i=0}^k \hat{c}_{3,L_p}^{(i)} \left(\frac{c_2 c_3 c_{2,L_p}}{c_{3,L_p}} \right)^{1/2} h_{L_p}^{(\beta_{L_p} - \gamma_{L_p})/2 + \delta_i} + \sum_{i=1}^k \hat{c}_3^{(i)} c_2 \Lambda_i \Big) \\
 & \times \left(\frac{\log(\varepsilon^{-1})}{p \log(M)} + \frac{\log(q^{-1/2} c_{1,p} T^p)}{p \log(M)} \right) \\
 & + \hat{c}_3^{(0)} c_2 \left(\frac{\log(\varepsilon^{-1})}{p \log(M)} + \frac{\log(q^{-1/2} c_{1,p} T^p)}{p \log(M)} \right)^2 \\
 & + \sum_{i=1}^k \hat{c}_3^{(i)} \left(\frac{c_2 c_{2,L_p} c_{3,L_p}}{c_3} \right)^{1/2} \Lambda_i h_{L_p}^{(\beta_{L_p} - \gamma_{L_p})/2} \\
 & + \sum_{i=0}^k \hat{c}_{3,L_p}^{(i)} \left(\left(\frac{c_{2,0} c_{3,0} c_{2,L_p}}{c_{3,L_p}} \right)^{1/2} h_{L_p}^{(\beta_{L_p} - \gamma_{L_p})/2 + \delta_i} + c_{2,L_p} h_{L_p}^{\beta_{L_p} - \gamma_{L_p} + \delta_i} \right) \Big] \\
 & + T \sum_{i=0}^k \left(\hat{c}_{3,0}^{(i)} T^{\delta_i - \gamma} + \hat{c}_3^{(i)} \frac{(M^{-1} T)^{\delta_i - \gamma} - h_{L_p}^{\delta_i - \gamma}}{1 - M^{\gamma - \delta_i}} + \hat{c}_{3,L_p}^{(i)} h_{L_p}^{\delta_i - \gamma_{L_p}} \right), \tag{3.15}
 \end{aligned}$$

where we applied the relation (3.5).

Proof of (ii). Suppose that $\beta_{L_p} \geq \gamma_{L_p}$ and $\gamma, \gamma_{L_p} \leq 2p$. Then from the upper bound (3.15) we obtain $C(\hat{Y}_{ML(\alpha,p)})(\varepsilon) = O(\varepsilon^{-2}(\log(\varepsilon))^2)$. Furthermore, comparing the lower and the upper bounds, (3.14) and (3.15), we asymptotically obtain

$$\begin{aligned}
 \lim_{\varepsilon \rightarrow 0} \frac{C(\hat{Y}_{ML(\alpha,\alpha)})(\varepsilon)}{C(\hat{Y}_{ML(\alpha,p)})(\varepsilon)} & \geq \lim_{\varepsilon \rightarrow 0} \frac{(T/(1-q))\varepsilon^{-2}\hat{c}_3^{(0)}c_2(\log(\varepsilon^{-1})/(\alpha \log(M)))^2 + o(\varepsilon^{-2}(\log(\varepsilon))^2)}{(T/(1-q))\varepsilon^{-2}\hat{c}_3^{(0)}c_2(\log(\varepsilon^{-1})/(p \log(M)))^2 + o(\varepsilon^{-2}(\log(\varepsilon))^2)} \\
 & = \frac{p^2}{\alpha^2},
 \end{aligned}$$

which proves (3.2). This completes the proof.

Remark 3.1. Especially, if $c_3 = \hat{c}_3$ and $c_{3,L_p} = \hat{c}_{3,L_p}$, then it follows in the $\beta < \gamma$ and $\beta < 2p$ case that

$$\lim_{\varepsilon \rightarrow 0} \frac{C(\hat{Y}_{ML(p,p)})(\varepsilon)}{C(\hat{Y}_{ML(\alpha,p)})(\varepsilon)} \geq M^{\gamma - \beta} \left(1 - M^{(\beta - \gamma)/2} \left(1 - \left(\frac{c_2 c_3}{c_{2,L_p} c_{3,L_p}} \right)^{1/2} \right) \right)^{-2}.$$

Thus, if $c_2 c_3 < c_{2,L_p} c_{3,L_p}$ it follows directly that

$$\lim_{\varepsilon \rightarrow 0} \frac{C(\hat{Y}_{ML(p,p)})(\varepsilon)}{C(\hat{Y}_{ML(\alpha,p)})(\varepsilon)} > 1.$$

4. Numerical examples in the case of SDEs

To illustrate the improvement that can be realized with the proposed modified MLMC estimator, we consider the problem of weak approximation for SDEs

$$dX_t = a(X_t) dt + \sum_{j=1}^m b^j(X_t) dB_t^j \tag{4.1}$$

with initial value $X_{t_0} = x_0 \in \mathbb{R}^d$ driven by an m -dimensional Brownian motion.

In the following, we compare for several numerical examples the RMS errors (2.2) versus the computational costs for the MLMC estimator \hat{Y}_{ML} , proposed in [5], [7], and [6] and described in Section 2, with the proposed modified MLMC estimator $\hat{Y}_{ML(\alpha,p)}$ described in Section 3. As a measure for the computational costs, we count the number of evaluations of the drift and diffusion functions, taking into account the dimension d of the solution process as well as the dimension m of the driving Brownian motion.

In the following, we consider on each level $l = 0, 1, \dots, L$ an equidistant discretization $I_{h_l} = \{t_0, \dots, t_{T/2^l}\}$ of $[t_0, T]$ with step size $h_l = T/2^l$. Furthermore, we denote by $Y_n = Y_{t_n}$ the approximation at time t_n . In case of the MLMC estimator \hat{Y}_{ML} , we apply on each level $l = 0, 1, \dots, L$ the Euler–Maruyama scheme on the grid I_{h_l} given by $Y_0 = x_0$ and

$$Y_{n+1} = Y_n + a(Y_n)h_n + \sum_{j=1}^m b^j(Y_n)I_{(j),n},$$

where $h_n = h_l$ and $I_{(j),n} = B_{t_{n+1}}^j - B_{t_n}^j$ for $n = 0, 1, \dots, T/2^l - 1$. The Euler–Maruyama scheme converges with order $\frac{1}{2}$ in the RMS sense and with order $\alpha = 1$ in the weak sense to the solution of the SDE (4.1) at time T [10].

On the other hand, for the modified MLMC estimator $\hat{Y}_{ML(\alpha,p)}$ the Euler–Maruyama scheme is applied on levels $0, 1, \dots, L_p - 1$, whereas on level L_p a second-order weak stochastic Runge–Kutta (SRK) scheme RI6 proposed in [12] is applied. The SRK scheme RI6 on level L_p is defined on the grid $I_{h_{L_p}}$ by $Y_0 = x_0$,

$$\begin{aligned} \check{Y}_{n+1} &= \check{Y}_n + \frac{1}{2}(a(\check{Y}_n) + a(\Upsilon))h_n + \frac{1}{2} \sum_{k=1}^m (b^k(\Upsilon_+^{(k)}) - b^k(\Upsilon_-^{(k)})) \frac{\hat{I}_{(k,k),n}}{\sqrt{h_n}} \\ &+ \sum_{k=1}^m \left(\frac{1}{2}b^k(\check{Y}_n) + \frac{1}{4}b^k(\Upsilon_+^{(k)}) + \frac{1}{4}b^k(\Upsilon_-^{(k)}) \right) I_{(k),n} \\ &+ \frac{1}{2} \sum_{k=1}^m (b^k(\hat{\Upsilon}_+^{(k)}) - b^k(\hat{\Upsilon}_-^{(k)}))\sqrt{h_n} \\ &- \sum_{k=1}^m \left(\frac{1}{2}b^k(\check{Y}_n) - \frac{1}{4}b^k(\hat{\Upsilon}_+^{(k)}) - \frac{1}{4}b^k(\hat{\Upsilon}_-^{(k)}) \right) I_{(k),n}, \end{aligned}$$

where $h_n = h_{L_p}$ and $I_{(k),n} = B_{t_{n+1}}^k - B_{t_n}^k$ for $n = 0, 1, \dots, T/2^{L_p} - 1$ with stages

$$\Upsilon = \check{Y}_n + a(\check{Y}_n)h_n + \sum_{j=1}^m b^j(\check{Y}_n)I_{(j),n},$$

$$\Upsilon_{\pm}^{(k)} = \check{Y}_n + a(\check{Y}_n)h_n \pm b^k(\check{Y}_n)\sqrt{h_n}, \quad \text{and} \quad \hat{\Upsilon}_{\pm}^{(k)} = \check{Y}_n \pm \sum_{j=1, j \neq k}^m b^j(\check{Y}_n) \frac{\hat{I}_{(k,j),n}}{\sqrt{h_n}},$$

where $\hat{I}_{(k,k),n} = \frac{1}{2}(I_{(k),n}^2 - h_n)$ and

$$\hat{I}_{(k,j),n} = \begin{cases} \frac{1}{2}(I_{(k),n}I_{(j),n} - \sqrt{h_n}\tilde{I}_{(k),n}) & \text{if } k < j, \\ \frac{1}{2}(I_{(k),n}I_{(j),n} + \sqrt{h_n}\tilde{I}_{(j),n}) & \text{if } j < k, \end{cases}$$

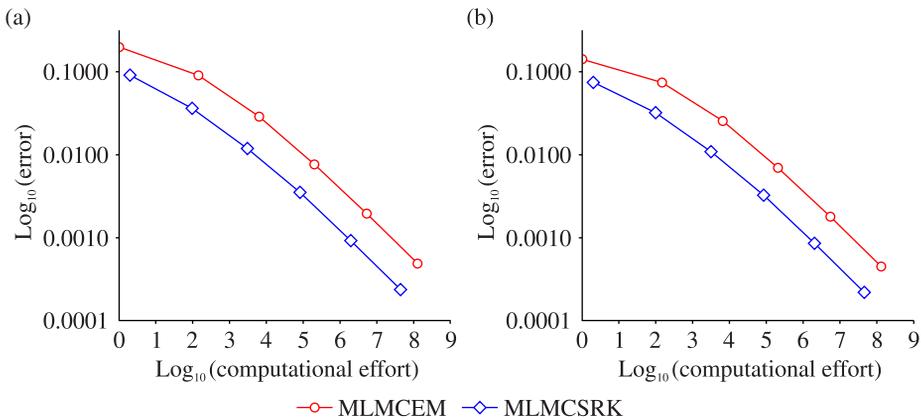


FIGURE 1: Error versus computational effort for (a) SDE (4.2) using $f(x) = x$ and (b) $f(x) = x^2$.

based on independent random variables $\tilde{I}_{(k),n}$ with $\mathbb{P}(\tilde{I}_{(k),n} = \pm\sqrt{h_n}) = \frac{1}{2}$. Thus, we have $\alpha = 1$ and $p = 2$ for the modified MLMC estimator $\hat{Y}_{ML(\alpha,p)}$ in the following. Furthermore, for both schemes the variance decays with the same order as the computational costs increase, i.e. $\beta = \beta_{L_p} = \gamma = \gamma_{L_p} = 1$. Then the optimal order of convergence attained by the MLMC method is $O(\varepsilon^{-2}(\log(\varepsilon))^2)$ due to Theorem 2.1. For the presented simulations, we denote by MLMCEM the numerical results for \hat{Y}_{ML} based on the Euler–Maruyama scheme only and by MLMCSRK the results for $\hat{Y}_{ML(\alpha,p)}$ based on the combination of the Euler–Maruyama scheme and the SRK scheme RI6.

As a first example, we consider the scalar linear SDE with $d = m = 1$ given by

$$dX_t = rX_t dt + \sigma X_t dB_t, \quad X_0 = 0.1, \tag{4.2}$$

using the parameters $r = 1.5$ and $\sigma = 0.1$. We choose $T = 1$ and apply the functionals $f(x) = x$ and $f(x) = x^2$; see Figure 1. The presented simulations are calculated using the prescribed error bounds $\varepsilon = 4^{-j}$ for $j = 0, 1, \dots, 5$. In Figure 1 we can see the significantly reduced computational effort for the estimator $\hat{Y}_{ML(1,2)}$ (MLMCSRK) compared to the estimator \hat{Y}_{ML} (MLMCEM) in case of a linear and a nonlinear functional.

The second example is a nonlinear scalar SDE with $d = m = 1$ given by

$$dX_t = \frac{1}{2}X_t + \sqrt{X_t^2 + 1} dt + \sqrt{X_t^2 + 1} dB_t, \quad X_0 = 0. \tag{4.3}$$

We apply the functional $f(x) = (\log(x + \sqrt{x^2 + 1}))^3 - 6(\log(x + \sqrt{x^2 + 1}))^2 + 8\log(x + \sqrt{x^2 + 1})$. Then the approximated expectation is given by $\mathbb{E}(f(X_t)) = t^3 - 3t^2 + 2t$. Here, the results presented in Figure 2(a) are calculated for $T = 2$, applying the prescribed error bounds $\varepsilon = 4^{-j}$ for $j = 0, 1, \dots, 6$. Here, the improved estimator $\hat{Y}_{ML(1,2)}$ performs much better than \hat{Y}_{ML} also for nonlinear functionals and a nonlinear SDE. Finally, we consider a nonlinear multi-dimensional SDE with a $d = 4$ -dimensional solution process driven by an

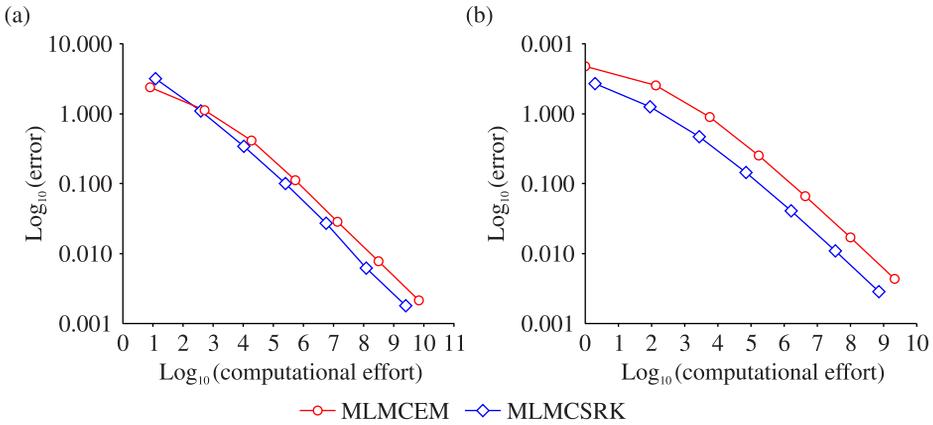


FIGURE 2: Error versus computational effort for (a) the nonlinear SDE (4.3) and (b) SDE (4.4) with noncommutative noise.

$m = 6$ -dimensional Brownian motion with noncommutative noise. Thus,

$$\begin{aligned}
 d \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \\ X_t^4 \end{pmatrix} &= \begin{pmatrix} \frac{243}{154} X_t^1 - \frac{27}{77} X_t^2 + \frac{23}{154} X_t^3 - \frac{65}{154} X_t^4 \\ \frac{27}{77} X_t^1 - \frac{243}{154} X_t^2 + \frac{65}{154} X_t^3 - \frac{23}{154} X_t^4 \\ \frac{5}{154} X_t^1 - \frac{61}{154} X_t^2 + \frac{162}{77} X_t^3 - \frac{36}{77} X_t^4 \\ \frac{61}{154} X_t^1 - \frac{5}{154} X_t^2 + \frac{36}{77} X_t^3 - \frac{162}{77} X_t^4 \end{pmatrix} dt \\
 &+ \frac{1}{9} \sqrt{(X_t^2)^2 + (X_t^3)^2 + \frac{2}{23} \begin{pmatrix} \frac{1}{13} \\ \frac{1}{14} \\ \frac{1}{13} \\ \frac{1}{15} \end{pmatrix}} dB_t^1 + \frac{1}{8} \sqrt{(X_t^4)^2 + (X_t^1)^2 + \frac{1}{11} \begin{pmatrix} \frac{1}{14} \\ \frac{1}{16} \\ \frac{1}{16} \\ \frac{1}{12} \end{pmatrix}} dB_t^2 \\
 &+ \frac{1}{12} \sqrt{(X_t^1)^2 + (X_t^2)^2 + \frac{1}{9} \begin{pmatrix} \frac{1}{6} \\ \frac{1}{5} \\ \frac{1}{5} \\ \frac{1}{6} \end{pmatrix}} dB_t^3 + \frac{1}{14} \sqrt{(X_t^3)^2 + (X_t^4)^2 + \frac{3}{29} \begin{pmatrix} \frac{1}{8} \\ \frac{1}{9} \\ \frac{1}{8} \\ \frac{1}{9} \end{pmatrix}} dB_t^4 \\
 &+ \frac{1}{10} \sqrt{(X_t^1)^2 + (X_t^3)^2 + \frac{1}{13} \begin{pmatrix} \frac{1}{11} \\ \frac{1}{15} \\ \frac{1}{13} \\ \frac{1}{11} \end{pmatrix}} dB_t^5 + \frac{1}{11} \sqrt{(X_t^2)^2 + (X_t^4)^2 + \frac{2}{25} \begin{pmatrix} \frac{1}{12} \\ \frac{1}{13} \\ \frac{1}{16} \\ \frac{1}{13} \end{pmatrix}} dB_t^6
 \end{aligned} \tag{4.4}$$

with initial condition $X_0 = (\frac{1}{8}, \frac{1}{8}, 1, \frac{1}{8})^T$. Then the approximated first moment of the solution is given by $\mathbb{E}(X_T^i) = X_0^i \exp(2T)$ for $i = 1, 2, 3, 4$. The simulation results calculated at $T = 1$ for the error bounds $\varepsilon = 4^{-j}$ for $j = 0, 1, \dots, 6$ are presented in Figure 2(b). Again, in the multi-dimensional noncommutative noise case the proposed estimator $\hat{Y}_{ML(1,2)}$ needs significantly less computational effort compared to the estimator \hat{Y}_{ML} , which reveals the theoretical results (3.2) in Proposition 3.1.

5. Conclusions

In this paper we proposed a modification of the MLMC method introduced by Giles which combines approximation methods of different orders of weak convergence. This modified MLMC method attains the same mean-square order of convergence as the originally proposed method that is in some sense optimal. However, the newly proposed MLMC estimator can attain significantly reduced computational costs. As an example, there is a reduction of costs by a factor $(p/\alpha)^2$ for the problem of weak approximation for SDEs driven by Brownian motion in the $\beta = \gamma$ case. This has been confirmed by some numerical examples for the $p = 2$ and $\alpha = 1$ case, where the calculation cost is a quarter of that required for the standard MLMC estimator. Here, we want to point out that there also exist higher-order weak approximation schemes, e.g. $p = 3$ in the case of SDEs with additive noise [2], that may further improve the benefit of the modified MLMC estimator. Future research will consider the application of this approach to more general SDEs, e.g. SDEs driven by Lévy processes [3] or fractional Brownian motion [11], and to the numerical solution of stochastic partial differential equations [13]. Furthermore, the focus will be on numerical schemes that feature not only high orders of convergence but also minimized constants for the variance estimates.

References

- [1] AVIKAINEN, R. (2009). On irregular functionals of SDEs and the Euler scheme. *Finance Stoch.* **13**, 381–401.
- [2] DEBRABANT, K. (2010). Runge–Kutta methods for third order weak approximation of SDEs with multidimensional additive noise. *BIT* **50**, 541–558.
- [3] DEREICH, S. (2011). Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *Ann. Appl. Prob.* **21**, 283–311.
- [4] DUFFIE, D. AND GLYNN, P. (1995). Efficient Monte Carlo simulation of security prices. *Ann. Appl. Prob.* **5**, 897–905.
- [5] GILES, M. (2008). Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and Quasi-Monte Carlo Methods 2006*, Springer, Berlin, pp. 343–358.
- [6] GILES, M. B. (2008). Multilevel Monte Carlo path simulation. *Operat. Res.* **56**, 607–617.
- [7] GILES, M. B., HIGHAM, D. J. AND MAO, X. (2009). Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff. *Finance Stoch.* **13**, 403–413.
- [8] HEINRICH, S. (2001). Multilevel Monte Carlo methods. In *Large-Scale Scientific Computing* (Lecture Notes Comput. Sci. **2179**), Springer, Berlin, pp. 58–67.
- [9] KEBAIER, A. (2005). Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. *Ann. Appl. Prob.* **15**, 2681–2705.
- [10] KLOEDEN, P. E. AND PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations* (Appl. Math. (New York) **23**). Springer, Berlin.
- [11] KLOEDEN, P. E., NEUENKIRCH, A. AND PAVANI, R. (2011). Multilevel Monte Carlo for stochastic differential equations with additive fractional noise. *Ann. Operat. Res.* **189**, 255–276.
- [12] RÖSSLER, A. (2009). Second order Runge–Kutta methods for Itô stochastic differential equations. *SIAM J. Numerical Anal.* **47**, 1713–1738.
- [13] SCHWAB, C. AND GITTELSON, C. J. (2011). Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numerica* **20**, 291–467.