Multilevel Richardson–Romberg extrapolation

VINCENT LEMAIRE[*] and GILLES PAGÈS[**]

Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, UPMC Paris 6 (Sorbonne Université).
E-mail: [*]vincent.lemaire@upmc.fr; [**]gilles.pages@upmc.fr

We propose and analyze a Multilevel Richardson–Romberg (ML2R) estimator which combines the higher order bias cancellation of the Multistep Richardson–Romberg method introduced in [*Monte Carlo Methods Appl.* **13** (2007) 37–70] and the variance control resulting from Multilevel Monte Carlo (MLMC) paradigm (see [*Ann. Appl. Probab.* **24** (2014) 1585–1620, In *Large-Scale Scientific Computing* (2001) 58–67 Berlin]). Thus, in standard frameworks like discretization schemes of diffusion processes, the root mean squared error (RMSE) $\varepsilon > 0$ can be achieved with our ML2R estimator with a global complexity of $\varepsilon^{-2} \log(1/\varepsilon)$ instead of $\varepsilon^{-2}(\log(1/\varepsilon))^2$ with the standard MLMC method, at least when the weak error $\mathbf{E}[Y_h] - \mathbf{E}[Y_0]$ of the biased implemented estimator Y_h can be expanded at any order in h and $\|Y_h - Y_0\|_2 = O(h^{\frac{1}{2}})$. The ML2R estimator is then halfway between a regular MLMC and a virtual unbiased Monte Carlo. When the strong error $\|Y_h - Y_0\|_2 = O(h^{\frac{\beta}{2}})$, $\beta < 1$, the gain of ML2R over MLMC becomes even more striking. We carry out numerical simulations to compare these estimators in two settings: vanilla and path-dependent option pricing by Monte Carlo simulation and the less classical Nested Monte Carlo simulation.

Keywords: Euler scheme; multilevel Monte Carlo estimator; multistep; nested Monte Carlo method; option pricing; Richardson–Romberg extrapolation

1. Introduction

The aim of this paper is to combine the multilevel Monte Carlo estimator introduced by S. Heinrich in [19] and developed by M. Giles in [12] (see also [23] for the statistical Romberg approach) and the (consistent) Multistep Richardson–Romberg extrapolation (see [27]) in order to minimize the simulation cost of a quantity of interest $I_0 = \mathbf{E}[Y_0]$ when the random variable Y_0 cannot be simulated at a reasonable cost (typically a functional of a generic multidimensional diffusion process or a conditional expectation). Both methods rely on the existence of a family of random variables Y_h, $h > 0$, which strongly approximate Y_0 as h goes to 0 whose bias $\mathbf{E}[Y_h] - \mathbf{E}[Y_0]$ can be expanded as a polynomial function of h (or h^α, $\alpha > 0$).

However, the two methods suffer from opposite but significant drawbacks: the multilevel Monte Carlo estimator does not fully take advantage of the existence of such an expansion beyond the first order whereas the Multistep Richardson–Romberg extrapolation induces an increase of the variance of the resulting estimator. Let us be more precise.

Consider a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and suppose that we have a family $(Y_h)_{h \geq 0}$ of real valued random variables in $\mathbf{L}^2(\mathbf{P})$, associated to Y_0 supposed to be non-degenerate, and satisfying $\lim_{h \to 0} \|Y_h - Y_0\|_2 = 0$ where h takes values in $\mathcal{H} = \{\mathbf{h}/n, n \geq 1\}$ for a fixed $\mathbf{h} \in (0, +\infty)$. Usually, the random variable Y_h appears as a functional of a time discretization scheme of step h

or from an inner approximation in a nested Monte Carlo simulation. The parameter h is called the *bias parameter* in what follows. Furthermore, we assume that, for every admissible $h \in \mathcal{H}$, the random variable Y_h can be simulated at a reasonable computational cost whereas Y_0 can not.

We aim at computing an as good as possible approximation of $I_0 = \mathbf{E}[Y_0]$ by a Monte Carlo type simulation. The starting point is of course to fix a parameter $h \in \mathcal{H}$ to consider a standard Monte Carlo estimator based on Y_h to compute I_0. So, let $(Y_h^{(k)})_{k \geq 1}$ be a sequence of independent copies of Y_h and the estimator $I_N^{(h)} = \frac{1}{N} \sum_{k=1}^{N} Y_h^{(k)}$. By the strong law of numbers and the central limit theorem, we have a control of the renormalized *statistical error* $\sqrt{N}(I_N^{(h)} - \mathbf{E}[Y_h])$ which behaves as a centered Gaussian with variance var(Y_h). On the other hand, there is a *bias error* due to the approximation of I_0 by $I_h = \mathbf{E}[Y_h]$. This bias error is also known as the *weak error* when Y_h is a functional of the time discretization scheme of a stochastic differential equation with step h. In many applications, the bias error can be expanded as

$$\mathbf{E}[Y_h] - \mathbf{E}[Y_0] = c_1 h^{\alpha} + \cdots + c_R h^{\alpha R} + o(h^{\alpha R}), \tag{1}$$

where α is a positive real parameter (usually $\alpha = \frac{1}{2}, 1$ or 2). In this paper, we fully take into account this error expansion and provide a very efficient estimator which can be viewed as a coupling between an MLMC estimator and a Multistep Richardson–Romberg extrapolation. Multilevel methods require the additional strong convergence rate assumption $\|Y_h - Y_0\|_2^2 = \mathcal{O}(h^{\beta})$ involving a parameter $\beta \in (0, 1)$. About the need of this assumption, we refer to [21] for some counterexamples. However, the recent paper [4] shows how this strong error assumption can be relaxed in some situations.

We first present a brief description of the original MLMC estimator as described in [12]. The main idea is to use the following telescopic summation with depth $L \geq 2$,

$$\mathbf{E}[Y_{h_L}] = \mathbf{E}[Y_h] + \sum_{j=2}^{L} \mathbf{E}[Y_{h_j} - Y_{h_{j-1}}],$$

where $(h_j)_{j=1,\ldots,L}$ is a geometrically decreasing sequence of different bias parameters $h_j = M^{-(j-1)}h$, $M \geq 2$. For each level $j \in \{1, \ldots, L\}$, the computation of $\mathbf{E}[Y_{h_j} - Y_{h_{j-1}}]$ is performed by a standard Monte Carlo procedure. The key point is that, at each level j, we consider a random sample of $(Y_{h_{j-1}}, Y_{h_j})$ of size $N_j = \lceil Nq_j \rceil$, where $q = (q_1, \ldots, q_L) \in \mathcal{S}_+(L) = \{q \in (0,1)^L, \sum_{j=1}^{L} q_j = 1\}$ (L-dimensional simplex), with in mind that the marginals $Y_{h_{j-1}}$ and Y_{h_j} are highly correlated since $\lim_{h \to 0} \|Y_h - Y_0\|_2 = 0$ (see Section 5.2 for details). More precisely, we consider L copies of the biased family denoted $Y^{(j)} = (Y_h^{(j)})_{h \in \mathcal{H}}$, $j \in \{1, \ldots, L\}$ attached to *independent* random copies $Y_0^{(j)}$ of Y_0. The MLMC estimator then writes

$$I_{h,L,q}^{N} = \frac{1}{N_1} \sum_{k=1}^{N_1} Y_h^{(1),k} + \sum_{j=2}^{L} \frac{1}{N_j} \sum_{k=1}^{N_j} (Y_{h_j}^{(j),k} - Y_{h_{j-1}}^{(j),k}), \tag{2}$$

where $(Y^{(j),k})_{k \geq 1}$, $j = 1, \ldots, L$ are independent sequences of independent copies of $Y^{(j)}$ and $N_1, \ldots, N_L$ are positive integers. The analysis of the computational complexity and the study

of the bias–variance structure of this estimator will appear as a particular case of a generalized multilevel framework that we will introduce and analyze in Section 3. This framework, following the original MLMC, highly relies on the combination of a strong rate of approximation of Y_0 by Y_h and a first order control of weak error $\mathbf{E}[Y_h] - \mathbf{E}[Y_0]$. This MLMC estimator has been extensively applied to various fields of numerical probability (jump diffusions [8,9], American options [5]), computational statistics and more general numerical analysis problems (high dimensional parabolic SPDEs, see [3], etc.). For more references, we refer to the web page http://people.maths.ox.ac.uk/gilesm/mlmc_community.html and the references therein.

On the other hand, the Multistep Richardson–Romberg extrapolation takes advantage of the full expansion (1). Let us first recall the one-step Richardson–Romberg Monte Carlo estimator. We still consider a biased family denoted $Y = (Y_h)_{h \in \mathcal{H}}$ attached to the random variable Y_0. The one-step Richardson–Romberg Monte Carlo estimator then writes

$$I_{h, \frac{h}{2}}^N = \frac{1}{N} \sum_{k=1}^N \left(2Y_{\frac{h}{2}}^k - Y_h^k\right),$$

where $(Y^k)_{k \geq 1}$ is a sequence of independent copies of Y. It is clear that this linear combination of Monte Carlo estimators satisfies the following bias error expansion (of order 2 in h)

$$\mathbf{E}[2Y_{\frac{h}{2}} - Y_h] - \mathbf{E}[Y_0] = -\frac{c_2}{2}h^2 + o(h^2).$$

Moreover, the asymptotic variance of this estimator satisfies $\mathrm{var}(I_{h, \frac{h}{2}}^N) = \mathrm{var}(2Y_{\frac{h}{2}} - Y_h)/N \approx \mathrm{var}(Y_0)/N$ which is the same as the crude Monte Carlo estimator. Then, it is natural to design a linear estimator with bias error in h^3 by linearly combining Y_h, $Y_{\frac{h}{2}}$ and $Y_{\frac{h}{3}}$ and so on. Such an extension called Multistep Richardson–Romberg extrapolation for Monte Carlo estimator has been introduced and extensively investigated in [27] in the framework of discretization of diffusion processes. More details are given in Section 2.4.

The aim of this paper is to show that an appropriate combination of the MLMC estimator and the Multistep Richardson–Romberg extrapolation outperforms the standard MLMC. More precisely, we will see in Section 3 that an implementation of the Multilevel Richardson Romberg estimator (ML2R) turns out to be a *weighted version of the MLMC* and writes

$$I_{h,R,q}^N = \frac{1}{N_1} \sum_{k=1}^{N_1} Y_h^{(1),k} + \sum_{j=2}^R \frac{\mathbf{W}_j}{N_j} \sum_{k=1}^{N_j} \left(Y_{h_j}^{(j),k} - Y_{h_{j-1}}^{(j),k}\right), \tag{3}$$

where $R \geq 2$ is the *depth level* – similar to L in (2) – and $(Y^{(j),k})_{k \geq 1}$ are like in (2). We denote by $n_j = M^{j-1}$ the jth *refiner coefficient* of the initial bias parameter $h \in \mathcal{H}$ where the integer $M \geq 2$ is called the *root* of the refiners. A strong feature of our approach comes from the fact that the weights $(\mathbf{W}_k)_{k=2,\ldots,R}$ are explicit and only depend on α (given by (1)), M and R. In practice these ML2R weights read $\mathbf{W}_j = \sum_{i=j}^R \mathbf{w}_i$ where $\mathbf{w}_i$ are given further on by (14). These derivative weights $(\mathbf{w}_i)_{i \in \{1,\ldots,R\}}$ have been introduced in [27] to kill the successive bias terms that appear in the expansion (1).

To compare the two methods MLMC and ML2R, we consider the following optimization problem: minimizing the global simulation cost (of one estimator) subject to the constraint that the resulting $\mathbf{L}^2$-error or root mean squared error (RMSE) must be lower than a prescribed $\varepsilon > 0$. We solve the problem step by step for both estimators (in fact, for a more general unifying class of estimators). In the first stage, we minimize the *effort* of the estimator (product of its variance by its complexity) to optimally dispatch the N_j across all levels (it can be viewed as a stratification procedure). Doing so, we are able to specify the initial bias parameter h and the depth level R as functions of ε and the structural parameters (α, β, V_1, var(Y_0)). A light preprocessing makes possible to optimize the choice of the root M of the refiners. Basically (see Theorem 3.12), the numerical cost of the ML2R estimator implemented with these optimal parameters (depending on ε) and denoted Cost(ML2R) satisfies

$$\text{Cost(ML2R)} \lesssim K(\alpha, \beta, M) v(\beta, \varepsilon),$$

where $f \lesssim g$ means that $\limsup_{\varepsilon \to 0} f(\varepsilon)/g(\varepsilon) \leq 1$, $K(\alpha, \beta, M)$ is an explicit bound (see (45) in Theorem 3.12) and

$$v(\beta, \varepsilon) = \begin{cases} \varepsilon^{-2}, & \text{if } \beta > 1, \\ \varepsilon^{-2} \log(1/\varepsilon), & \text{if } \beta = 1, \\ \varepsilon^{-2} e^{\frac{1-\beta}{\sqrt{\alpha}}\sqrt{2\log(1/\varepsilon)\log(M)}}, & \text{if } \beta < 1. \end{cases}$$

Note that $e^{\frac{1-\beta}{\sqrt{\alpha}}\sqrt{2\log(1/\varepsilon)\log(M)}} = o(\varepsilon^{-\eta})$ for all $\eta > 0$. As first established in [12], we prove likewise that the optimal numerical cost of the MLMC estimator denoted Cost(MLMC) satisfies a similar result with $v(\beta, \varepsilon) = \varepsilon^{-2}$ if $\beta > 1$, $v(\beta, \varepsilon) = \varepsilon^{-2} \log(1/\varepsilon)^2$ if $\beta = 1$ and $v(\beta, \varepsilon) = \varepsilon^{-2-\frac{1-\beta}{\alpha}}$ if $\beta < 1$. In the case $\beta = 1$, the gain of $\log(1/\varepsilon)$ may look as a minor improvement but, beyond the fact that it is halfway to a virtual unbiased simulation, this improvement is obtained with respect to an already tremendously efficient method to speed up crude Monte Carlo simulation. In fact, as emphasized in our numerical experiments (see Section 5), this may lead to a significant reduction factor for CPU time, e.g. when $\alpha < 1$: pricing a Black–Scholes Lookback Call option with a prescribed quadratic error $\varepsilon = 2^{-8}$, yields a reduction factor of 3.5 in favor of ML2R. When $\beta < 1$, the above theoretical reduction factor asymptotically goes to $+\infty$ as ε goes to 0 in a very steep way. Thus, the reduction in CPU time factor reaches, *mutatis mutandis*, 22 for a Black–Scholes Up&Out Barrier call option for which $\beta = \frac{1}{2}$ (still using a regular Euler scheme *without Brownian bridge*). When compared on the basis of the resulting empirical RMSE, these factors become even larger (approximately 48 and 61, respectively). In fact, it confirms that $\beta < 1$ is the setting where our ML2R estimator is the most powerful compared to regular MLMC. Additional simulations are available on the web page https://simulations.lpma-paris.fr/multilevel/.

The paper is organized as follows: in Section 2, we propose a general parametrized framework to formalize the optimization of a *biased* Monte Carlo simulation based on the $\mathbf{L}^2$-error minimization. The crude Monte Carlo estimator and the Multistep Richardson–Romberg estimator appear as the first two examples, allowing us to make precise few notations as well as our main assumptions. In Section 3, we first introduce the extended family of multilevel estimators attached to design matrices $\mathbf{T}$. Among them, we specify in more details our proposal: the new

ML2R estimator, but also the standard MLMC estimator. Two typical fields of application are presented in Section 4: the time discretization of stochastic processes (Euler scheme) and the nested Monte Carlo method, for which a weak expansion of the error at any order is established in the regular case. In Section 5, we present and comment on numerical experiments carried out in the above two fields.

NOTATIONS:

- Let $\mathbf{N}^* = \{1, 2, \ldots\}$ denote the set of positive integers and $\mathbf{N} = \mathbf{N}^* \cup \{0\}$.
- If $\underline{n} = (n_1, \ldots, n_R) \in (\mathbf{N}^*)^R$, $|\underline{n}| = n_1 + \cdots + n_R$ and $\underline{n}! = \prod_{1 \leq i \leq R} n_i$.
- Let $(e_1, \ldots, e_R)$ denote the canonical basis of $\mathbf{R}^R$ (viewed as a vector space of column vectors). Thus, $e_i = (\delta_{ij})_{1 \leq + j \leq R}$ where δ_{ij} stands for the classical Kronecker symbol.
- $\langle \cdot, \cdot \rangle$ denotes the canonical inner product on $\mathbf{R}^R$.
- For every $x \in \mathbf{R}_+ = [0, +\infty)$, $\lceil x \rceil$ denotes the unique $n \in \mathbf{N}^*$ satisfying $n - 1 < x \leq n$ and $\lfloor x \rfloor$ denotes the unique $n \in \mathbf{N}$ satisfying $n \leq x < n + 1$.
- If $(a_n)_{n \in \mathbf{N}}$ and $(b_n)_{n \in \mathbf{N}}$ are two sequences of real numbers, $a_n \sim b_n$ if $a_n = \varepsilon_n b_n$ with $\lim_n \varepsilon_n = 1$, $a_n = O(b_n)$ if $(\varepsilon_n)_{n \in \mathbf{N}}$ is bounded and $a_n = o(b_n)$ if $\lim_n \varepsilon_n = 0$.
- $\mathrm{var}(X)$ and $\sigma(X)$ denote the variance and the standard deviation of a random variable X respectively.

2. Preliminaries

2.1. Mixing variance and complexity (effort)

We first introduce some notations and recall basic facts on (possibly biased) *linear* estimators. We consider a family of linear statistical estimator $(I_\pi^N)_{N \geq 1}$ of $I_0 \in \mathbf{R}$ where π lies in a parameter set $\Pi \subset \mathbf{R}^d$. By linear, we mean, on the one hand, that, for every integer $N \geq 1$,

$$\mathbf{E}[I_\pi^N] = \mathbf{E}[I_\pi^1]$$

and, on the other hand, that the numerical cost $\mathrm{Cost}(I_\pi^N)$ induced by the simulation of I_π^N is given by

$$\mathrm{Cost}(I_\pi^N) = N\kappa(\pi),$$

where $\kappa(\pi) = \mathrm{Cost}(I_\pi^1)$ is the cost of a single simulation or *unitary complexity*.

We also assume that our estimator is of *Monte Carlo type* in the sense that its variance is *inverse linear* in the size N of the simulation:

$$\mathrm{var}(I_\pi^N) = \frac{v(\pi)}{N},$$

where $v(\pi) = \mathrm{var}(I_\pi^1)$ denotes the variance of one simulation. For example, in a crude biased Monte Carlo $\pi = h \in \mathcal{H}$, in a Multilevel Monte Carlo $\pi = (h, R, q) \in \mathcal{H} \times \mathbf{N}^* \times \mathcal{S}_+$ and in the Multistep Monte Carlo $\pi = (h, R) \in \mathcal{H} \times \mathbf{N}^*$.

We are looking for the "best" estimator in this family $\{(I_\pi^N)_{N \geq 1}, \pi \in \Pi\}$ *that is*, the estimator minimizing the computational cost for a given error $\varepsilon > 0$. In the sequel, we will often consider

N as a continuous variable lying in $\mathbf{R}_+$. A natural choice for measuring the random error $I_\pi^N - I_0$ is to consider the $\mathbf{L}^2$-error or root mean squared error (RMSE) $\sqrt{\mathbf{E}[(I_\pi^N - I_0)^2]} = \|I_\pi^N - I_0\|_2$. Our aim is to minimize the cost of the simulation for a given target error, say $\varepsilon > 0$. This generic problem reads

$$\big(\pi(\varepsilon), N(\varepsilon)\big) = \underset{\|I_\pi^N - I_0\|_2 \leq \varepsilon}{\operatorname{argmin}} \; \operatorname{Cost}\big(I_\pi^N\big). \tag{4}$$

In order to solve this minimization problem, we introduce the notion of *effort* $\phi(\pi)$ of a linear Monte Carlo type estimator I_π^N.

Definition 2.1. *The* effort *of the estimator I_π^N is defined for every $\pi \in \Pi$ by*

$$\phi(\pi) = \nu(\pi)\kappa(\pi). \tag{5}$$

By definition of a linear estimator I_π^N, we have that

$$\phi(\pi) = \nu(\pi)\kappa(\pi) = \operatorname{var}\big(I_\pi^N\big)\operatorname{Cost}\big(I_\pi^N\big) = \operatorname{var}\big(I_\pi^1\big)\operatorname{Cost}\big(I_\pi^1\big)$$

for every integer $N \geq 1$, so that we obtain the fundamental relation

$$\operatorname{Cost}\big(I_\pi^N\big) = N\frac{\phi(\pi)}{\nu(\pi)}. \tag{6}$$

- If the estimators $(I_\pi^N)_{N \geq 1}$ are *unbiased* that is, $\mathbf{E}[I_\pi^N] = I_0$ for every $\pi \in \Pi$, then $\mathbf{E}[(I_\pi^N - I_0)^2] = \|I_\pi^N - I_0\|_2^2 = \operatorname{var}(I_\pi^N) = \frac{1}{N}\nu(\pi)$. The solution of the generic problem (4) then reads

$$\pi(\varepsilon) = \pi^* = \underset{\pi \in \Pi}{\operatorname{argmin}}\,\phi(\pi), \qquad N(\varepsilon) = \frac{\nu(\pi^*)}{\varepsilon^2} = \frac{\phi(\pi^*)}{\kappa(\pi^*)\varepsilon^2}. \tag{7}$$

  Consequently, the most performing estimator I_π^N is characterized as a minimizer of the effort $\phi(\pi)$ as defined above (and the parameter π does not depend on ε).
- When the estimators $(I_\pi^N)_{N \geq 1}$, $\pi \in \Pi$, are *biased*, the mean squared error writes

$$\mathbf{E}\big[(I_\pi^N - I_0)^2\big] = \mu^2(\pi) + \frac{\nu(\pi)}{N},$$

where

$$\mu(\pi) = \mathbf{E}\big[I_\pi^N\big] - I_0 = \mathbf{E}\big[I_\pi^1\big] - I_0$$

denotes the bias (which does not depend on N). Using that $\nu(\pi) = N(\|I_\pi^N - I_0\|_2^2 - \mu(\pi)^2)$, the solution of the generic problem (4) reads

$$\begin{aligned}
\pi(\varepsilon) &= \underset{\pi \in \Pi,\,|\mu(\pi)| < \varepsilon}{\operatorname{argmin}} \left(\frac{\phi(\pi)}{\varepsilon^2 - \mu^2(\pi)}\right), \\
N(\varepsilon) &= \frac{\nu(\pi(\varepsilon))}{\varepsilon^2 - \mu^2(\pi(\varepsilon))} = \frac{\phi(\pi(\varepsilon))}{\kappa(\pi(\varepsilon))(\varepsilon^2 - \mu^2(\pi(\varepsilon)))}.
\end{aligned} \tag{8}$$

2.2. Assumptions on weak and strong approximation errors

We come back to the framework described in the Introduction: let $(Y_h)_{h \in \mathcal{H}}$ be a family of real valued random variables associated to a random variable $Y_0 \in \mathbf{L}^2$. The index set $\mathcal{H}$ is a set of *bias* parameters (in fact *representative* of a bias) defined by $\mathcal{H} = \{\mathbf{h}/n, n \geq 1\}$ for a fixed $\mathbf{h} \in (0, +\infty)$. All random variables Y_h are defined on the same probability space $(\Omega, \mathcal{A}, \mathbf{P})$. The family satisfies two assumptions which formalize the strong and weak rates of approximation of Y_0 by Y_h when $h \to 0$ in $\mathcal{H}$. These assumptions are the basement of multilevel simulation methods (see [12,19]):

Bias error expansion (weak error rate):

$$\exists \alpha > 0, \bar{R} \geq 1, \qquad \mathbf{E}[Y_h] = \mathbf{E}[Y_0] + \sum_{k=1}^{\bar{R}} c_k h^{\alpha k} + h^{\alpha \bar{R}} \eta_{\bar{R}}(h), \qquad \lim_{h \to 0} \eta_{\bar{R}}(h) = 0, \quad (\mathrm{WE}_{\alpha, \bar{R}})$$

where $c_k, k = 1, \ldots, \bar{R}$, are real coefficients and $\eta_{\bar{R}}$ is a real valued function defined on $\mathcal{H}$.

Strong approximation error assumption:

$$\exists \beta > 0, V_1 \in \mathbf{R}_+, \qquad \|Y_h - Y_0\|_2^2 = \mathbf{E}[|Y_h - Y_0|^2] \leq V_1 h^{\beta}. \qquad (\mathrm{SE}_{\beta})$$

Note that the parameters α, β and $\bar{R}$ are structural parameters which only depend on the family $(Y_h)_{h \in \mathcal{H}}$. When $(Y_h)_{h \in \mathcal{H}}$ satisfies $(\mathrm{WE}_{\alpha, \bar{R}})$ for every integer $\bar{R}$, we will say that $(\mathrm{WE}_{\alpha, \bar{R}})$ is fulfilled. Such a family is said to be *admissible* (at level $\bar{R}$ with parameters β and α).

Note that consistency of weak error $(\mathrm{WE}_{\alpha, \bar{R}})$ (when $c_1 \neq 0$) and strong error (SE_{β}) implies that $\beta \leq 2\alpha$.

In the sequel, we will consider that the depth parameter $R \in \{2, \ldots, \bar{R}\}$ so that it always satisfies $(\mathrm{WE}_{\alpha, \bar{R}})$. This parameter R corresponds to the depth level L used in the multilevel literature.

In what follows, we will use the following ratio

$$\theta = \sqrt{\frac{V_1}{\mathrm{var}(Y_0)}} \qquad (9)$$

which relates the quadratic rate of convergence of Y_h to Y_0 and the variance of Y_0.

Remark 2.2 (Alternative strong approximation error assumptions). Throughout the paper, whenever $(\mathrm{WE}_{\alpha, 1})$ holds (with $c_1 \neq 0$), (SE_{β}) can be replaced by one of the following assumptions

$$\exists \beta > 0, V_1 \in \mathbf{R}_+, \qquad \mathrm{var}(Y_h - Y_0) \leq V_1 h^{\beta}, \qquad h \in \mathcal{H} \qquad (\mathrm{Var}_{\beta})$$

and

$$\exists \beta > 0, V_1 \in \mathbf{R}_+, \qquad \mathrm{var}(Y_h - Y_{h'}) \leq V_1 |h - h'|^{\beta}, h, \qquad h' \in \mathcal{H}. \qquad (\mathrm{Var}'_{\beta})$$

The aim being is to reduce the numerical value of V_1 in view of practical implementation for (Var_{β}) and the estimate of the variance of refined levels for (Var'_{β}). Note that, still if $(\mathrm{WE}_{\alpha, 1})$

holds, ((Var$_\beta$) and $\beta \leq 2\alpha$) $\Longleftrightarrow$ (SE$_\beta$) (the first one with a lower V_1 if $\alpha = 2\beta$) and (Var$'_\beta$) $\Rightarrow$ (Var$_\beta$) by letting $h' \to 0$ (the converse being false).

The theoretical results obtained using one of these two assumptions (combined with (WE$_{\alpha,\bar{R}}$)) may induce slight modifications in the exposition of the results of this paper (in particular Theorem 3.6). These variants are briefly discussed in Remark 3.7 and practical guidelines for the estimation of V_1 are discussed in Section 5.1 (Practitioner's corner).

All estimators considered in this work are based on independent copies $(Y_h^{(j)})_{h \in \mathcal{H}}$ (attached to random variables $Y_0^{(j)}$) of $(Y_h)_{h \in \mathcal{H}}$, $j = 1, \ldots, R$. All random variables are supposed to be defined on the same probability space $(\Omega, \mathcal{A}, \mathbf{P})$. Note that, since the above properties (SE$_\beta$) and (WE$_{\alpha,\bar{R}}$), $\bar{R} \geq 1$, only depend on the distribution of $(Y_h)_{h \in \mathcal{H}}$, all these copies will also satisfy these two properties.

We associate to the family $(Y_h)_{h \in \mathcal{H}}$ and a given bias parameter $h \in \mathcal{H}$, the $\mathbf{R}^R$-valued random vector

$$Y_{h,\underline{n}} = (Y_h, Y_{\frac{h}{n_2}}, \ldots, Y_{\frac{h}{n_R}}),$$

where the R-tuple of integers $\underline{n} := (n_1, n_2, \ldots, n_R) \in \mathbf{N}^R$, called *refiners* in the sequel, satisfy

$$n_1 = 1 < n_2 < \cdots < n_R.$$

One defines likewise $Y_{h,\underline{n}}^{(j)}$ for the (independent) copies of $Y_{h,\underline{n}}$.

$\triangleright$ *Specification of the refiners*: In most applications, we will choose refiners n_i as $n_i = M^{i-1}$ where $M \geq 2$. Indeed, this is the standard choice in the regular Multilevel Monte Carlo method as described in [12]. Other choices like $n_i = i$ are possible (see below the original Multistep Richardson–Romberg estimator in Section 2.4).

2.3. Crude Monte Carlo estimator

In our formalism, a crude Monte Carlo simulation and its cost can be described as follows.

Proposition 2.3. *Assume* (WE$_{\alpha,\bar{R}}$) *with* $c_1 \neq 0$ *and* (SE$_\beta$) *with* $\bar{R} \geq 1$. *The Monte Carlo estimator of* $\mathbf{E}[Y_0]$ *defined by*

$$\forall N \geq 1, h \in \mathcal{H}, \qquad \bar{Y}_h^N = \frac{1}{N} \sum_{k=1}^{N} Y_h^k,$$

where $(Y_h^k)_{k \geq 1}$ *is an i.i.d. sequence of copies of* Y_h, *satisfies*

$$\mu(h) = c_1 h^\alpha (1 + \eta_1(h)), \qquad \kappa(h) = \frac{1}{h}, \qquad \phi(h) = \frac{\text{var}(Y_h)}{h}$$

and, for a prescribed $\mathbf{L}^2$*-error* $\varepsilon > 0$*, the optimal parameters* $h^*(\varepsilon)$ *and* $N^*(\varepsilon)$ *solution to* (4) *are given by*

$$h^*(\varepsilon) = (1 + 2\alpha)^{-\frac{1}{2\alpha}} \left(\frac{\varepsilon}{|c_1|}\right)^{\frac{1}{\alpha}}, \qquad N^*(\varepsilon) = \left(1 + \frac{1}{2\alpha}\right) \frac{\mathrm{var}(Y_0)(1 + \theta(h^*(\varepsilon))^{\frac{\beta}{2}})^2}{\varepsilon^2}. \qquad (10)$$

Furthermore, we have

$$\limsup_{\varepsilon \to 0} \varepsilon^{2 + \frac{1}{\alpha}} \min_{\substack{h \in \mathcal{H} \\ |\mu(h)| < \varepsilon}} \mathrm{Cost}(\bar{Y}_h^N) \le |c_1|^{\frac{1}{\alpha}} \left(1 + \frac{1}{2\alpha}\right)(1 + 2\alpha)^{\frac{1}{2\alpha}} \mathrm{var}(Y_0).$$

Proof. The proof is postponed to Appendix B. $\qquad \square$

We refer to the seminal paper [11] for more details on practical implementation of this estimator.

Remark 2.4. For crude Monte Carlo simulation, Assumption (SE_β) can be replaced by $Y_h \xrightarrow{L^2} Y_0$ (without rate), provided $\mathrm{var}(Y_0)(1 + \theta(h^*(\varepsilon))^{\frac{\beta}{2}})^2$ is replaced by $\mathrm{var}(Y_{h^*(\varepsilon)})$ in (10).

2.4. Background on multistep Richardson–Romberg extrapolation

The so-called *Multistep Richardson–Romberg estimator* has been introduced in [27] in the framework of Brownian diffusions. It relies on R (refined) Euler schemes $\bar{X}^{(\frac{h}{n_i})}$, $1 \le i \le R$, defined on a finite interval $[0, T]$ ($T > 0$), where the bias parameter $h = \frac{T}{n}$, $n \ge 1$. In that case, the refiners are set as $n_i = i$, $i = 1, \ldots, R$ (in order to produce a better control of both the variance and the complexity for the proposed estimator, see Remark 2.6 below). The main results are obtained when all the schemes are *consistent* that is, such that all the Brownian increments are generated from the same underlying Brownian motion. As a consequence, under standard smoothness assumptions on the coefficients of the diffusion, the family $Y_h = \bar{X}^{(h)}$, $h \in \mathcal{H} = \{\frac{T}{n}, n \ge 1\}$, makes up an admissible family in the above sense, as will be seen further on in more details.

For a refiner vector $(n_1, n_2, \ldots, n_R)$, we define the *weight* vector $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_R)$ as the unique solution to the Vandermonde system $V\mathbf{w} = e_1$ where

$$V = V\left(1, n_2^{-\alpha}, \ldots, n_R^{-\alpha}\right) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & n_2^{-\alpha} & \cdots & n_R^{-\alpha} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & n_2^{-\alpha(R-1)} & \cdots & n_R^{-\alpha(R-1)} \end{pmatrix}. \qquad (11)$$

The solution $\mathbf{w}$ of the system has a closed form given by Cramer's rule (see Lemma A.1 in Appendix A):

$$\forall i \in \{1, \ldots, R\}, \qquad \mathbf{w}_i = \frac{(-1)^{R-i} n_i^{\alpha(R-1)}}{\prod_{1 \le j < i}(n_i^\alpha - n_j^\alpha) \prod_{i < j \le R}(n_j^\alpha - n_i^\alpha)}. \qquad (12)$$

We also derive the following identity of interest

$$\widetilde{\mathbf{w}}_{R+1} := \sum_{i=1}^{R} \frac{\mathbf{w}_i}{n_i^{\alpha R}} = \frac{(-1)^{R-1}}{\underline{n}!^{\alpha}}, \tag{13}$$

which will be used in (15) and (17) to control the residual bias.

Note that all coefficients $(\mathbf{w}_i)_{1 \le i \le R}$ depend on the depth R of the combined extrapolations. For the standard choices $n_i = i$ or $n_i = M^{i-1}$, $i = 1, \ldots, R$, we obtain the following expressions:

$$\mathbf{w}_i = \begin{cases} \dfrac{(-1)^{R-i} i^{\alpha(R-1)}}{\prod_{j=1}^{i-1}(i^{\alpha} - j^{\alpha}) \prod_{i+1}^{R}(j^{\alpha} - i^{\alpha})}, & \text{if } n_j = j, j \in \{1, \ldots, R\}, \\[4mm] \dfrac{(-1)^{R-i} M^{-\frac{\alpha}{2}(R-i)(R-i+1)}}{\prod_{j=1}^{i-1}(1 - M^{-j\alpha}) \prod_{j=1}^{R-i}(1 - M^{-j\alpha})}, & \text{if } n_j = M^{j-1}, j \in \{1, \ldots, R\}. \end{cases} \tag{14}$$

Note that when $\alpha = 1$ and $n_i = i$, then $\mathbf{w}_i = \frac{(-1)^{R-i} i^R}{i!(R-i)!}$, $i = 1, \ldots, R$.

Assume now ($\text{WE}_{\alpha,\bar{R}}$) and $R \in \{1, \ldots, \bar{R}\}$. In order to design an estimator which kills the bias up to order R, we focus on the random variable resulting from the linear combination $\langle \mathbf{w}, Y_{h,\underline{n}} \rangle = \sum_{i=1}^{R} \mathbf{w}_i Y_{\frac{h}{n_i}}$.

The first equation of the Vandermonde system $V\mathbf{w} = e_1$, namely $(V\mathbf{w})_1 = \sum_{r=1}^{R} \mathbf{w}_r = 1$, implies

$$\lim_{h \to 0} \mathbf{E}\big[\langle \mathbf{w}, Y_{h,\underline{n}} \rangle\big] = \mathbf{E}[Y_0].$$

Furthermore, when expanding the (weak) error, one checks that the other $R - 1$ equations satisfied by the weight vector $\mathbf{w}$ make all terms in front of the c_{r-1}, $r = 2, \ldots, R$ vanish. Finally, we obtain

$$\mathbf{E}\big[\langle \mathbf{w}, Y_{h,\underline{n}} \rangle\big] = \mathbf{E}[Y_0] + \sum_{r=2}^{R} c_{r-1} h^{\alpha(r-1)} (V\mathbf{w})_r + c_R \widetilde{\mathbf{w}}_{R+1} h^{\alpha R} \big(1 + \eta_{R,\underline{n}}(h)\big) \tag{15}$$

$$= \mathbf{E}[Y_0] + c_R \widetilde{\mathbf{w}}_{R+1} h^{\alpha R} \big(1 + \eta_{R,\underline{n}}(h)\big), \tag{16}$$

where

$$\eta_{R,\underline{n}}(h) = \frac{1}{c_R \widetilde{\mathbf{w}}_{R+1}} \sum_{r=1}^{R} \frac{\mathbf{w}_r}{n_r^{\alpha R}} \eta_R\left(\frac{h}{n_r}\right) \to 0 \qquad \text{as } h \to 0. \tag{17}$$

Proposition 2.5. *Assume* ($\text{WE}_{\alpha,\bar{R}}$) *and* ($\text{SE}_{\beta}$)*. Let* $R \in \{2, \ldots, \bar{R}\}$ *be such that* $c_R \neq 0$*. The Multistep Richardson–Romberg estimator of* $\mathbf{E}[Y_0]$ *defined by*

$$\forall N \ge 1, h \in \mathcal{H}, \qquad \bar{Y}_{h,\underline{n}}^{N} = \frac{1}{N} \sum_{k=1}^{N} \langle \mathbf{w}, Y_{h,\underline{n}}^{k} \rangle = \left\langle \mathbf{w}, \frac{1}{N} \sum_{k=1}^{N} Y_{h,\underline{n}}^{k} \right\rangle, \tag{18}$$

where $(Y_{h,\underline{n}}^k)_{k \geq 1}$ is an i.i.d. sequence of copies of $Y_{h,\underline{n}}$, satisfies

$$\mu(h) = (-1)^{R-1} c_R \left(\frac{h^R}{\underline{n}!} \right)^\alpha \left(1 + \eta_{R,\underline{n}}(h) \right), \qquad \kappa(h) = \frac{|\underline{n}|}{h}, \qquad \phi(h) = \frac{|\underline{n}| \, \mathrm{var}(\langle \mathbf{w}, Y_{h,\underline{n}} \rangle)}{h}$$

and, for a prescribed $\mathbf{L}^2$-error $\varepsilon > 0$, the optimal parameters $h^*(\varepsilon)$ and $N^*(\varepsilon)$ solution of (4) are

$$h^*(\varepsilon) = (1 + 2\alpha R)^{-\frac{1}{2\alpha R}} \left(\frac{\varepsilon}{|c_R|} \right)^{\frac{1}{\alpha R}} \underline{n}!^{\frac{1}{R}},$$

$$N^*(\varepsilon) = \left(1 + \frac{1}{2\alpha R} \right) \frac{\mathrm{var}(Y_0)(1 + \theta(h^*(\varepsilon))^{\frac{\beta}{2}})^2}{\varepsilon^2}.$$

Furthermore,

$$\inf_{\substack{h \in \mathcal{H} \\ |\mu(h)| < \varepsilon}} \mathrm{Cost}(\bar{Y}_h^N) \sim \left(\frac{(1 + 2\alpha R)^{1 + \frac{1}{2\alpha R}}}{2\alpha R} \right) \frac{|c_R|^{\frac{1}{\alpha R}} |\underline{n}| \, \mathrm{var}(Y_0)}{\underline{n}!^{\frac{1}{R}} \varepsilon^{2 + \frac{1}{\alpha R}}} \qquad \text{as } \varepsilon \to 0. \tag{19}$$

Proof. The proof is postponed to Appendix B (but takes advantage of the formalism developed in the next section). □

Remark 2.6.

- As for (SE_β), Remark 2.4 still applies.
- In this approach, the bias reduction suffers from an increase of the simulation cost by the $|\underline{n}|$ factor which appears in the numerator of (19). The choice of the refiners made in [27], namely $n_i = i$, $i = 1, \dots, R$, is justified by the control of the ratio $\frac{|\underline{n}|}{\underline{n}!^{\frac{1}{R}}}$: for such a choice, it behaves linearly, like $\frac{e}{2}(R + 1)$, for large values of R whereas with $n_i = M^{i-1}$ it goes to infinity geometrically in $O(M^{\frac{R-1}{2}})$.

3. A paradigm for multilevel simulation methods

3.1. General framework

Multilevel decomposition

In spite of Proposition 2.5 which shows that the numerical cost of the Multistep method behaves like $\varepsilon^{2 + \frac{1}{\alpha R}}$, one observes in practice that the increase of the ratio $\frac{|\underline{n}|}{\underline{n}!}$ (when R grows) in front of $\mathrm{var}(Y_0)$ in (19) reduces the impact of the bias reduction.

An idea is then to introduce independent linear combination of copies of $\bar{Y}_{h,\underline{n}}$ to reduce the variance taking advantage of the basic fact that if X and X' are independent with the same

distribution then $\mathbf{E}[\frac{X+X'}{2}] = \mathbf{E}[X]$ and $\mathrm{var}(\frac{X+X'}{2}) = \frac{1}{2}\mathrm{var}(X)$, combined with an appropriate allocation policy to control the complexity of the resulting estimator. So, let us consider now R *independent* copies $(Y_{h,\underline{n}}^{(j)})$, $j = 1, \ldots, R$, of the random vector $Y_{h,\underline{n}}$ and the linear combination

$$\sum_{j=1}^{R} \langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} \rangle = \sum_{i,j=1}^{R} \mathbf{T}_i^j Y_{\frac{h}{n_i}}^{(j)},$$

where $\mathbf{T} = [\mathbf{T}^1 \ldots \mathbf{T}^R]$ is an $R \times R$ matrix with column vectors $\mathbf{T}^j \in \mathbf{R}^R$ satisfying the constraint

$$\sum_{1 \leq i,j \leq R} \mathbf{T}_i^j = 1.$$

As emphasized further on, we will also need that each column vector $\mathbf{T}^j$, $j \in 2, \ldots, R$, has zero sum. In turn, this suggests to introduce the notion of Multilevel estimator as a family of "stratified" estimators of $\mathbf{E}[Y_0]$ attached to the random vectors $\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} \rangle$, $j = 1, \ldots, R$. This leads to the following definitions.

Definition 3.1 (Design matrix). *Let* $R \geq 2$. *An* $R \times R$*-matrix* $\mathbf{T} = [\mathbf{T}^1 \ldots \mathbf{T}^R]$ *is an* R*-level design matrix if*

$$\langle \mathbf{T}^j, \mathbf{1} \rangle = \sum_{i=1}^{R} \mathbf{T}_i^j = 0, \qquad j = 2, \ldots, R. \tag{20}$$

Note that such a design matrix always satisfies $\sum_{i,j=1}^{d} \mathbf{T}_i^j = 1$.

Definition 3.2 (General multilevel estimator). *Let* $R \geq 2$ *and let* $(Y_{h,\underline{n}}^{(j),k})_{k \geq 1}$ *be an i.i.d. sequence of copies of* $Y_{h,\underline{n}}^{(j)}$. *A Multilevel estimator of depth* R *attached to an allocation policy* $q = (q_1, \ldots, q_R)$ *with* $q_j > 0$, $j = 1, \ldots, R$, *and* $\sum_j q_j = 1$ *and a design matrix* $\mathbf{T}$, *is defined for every integer* $N \geq 1$ *and* $h \in \mathcal{H}$ *by*

$$\bar{Y}_{h,\underline{n}}^{N,q} = \sum_{j=1}^{R} \frac{1}{N_j} \sum_{k=1}^{N_j} \langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j),k} \rangle, \tag{21}$$

where for all $j \in \{1, \ldots, R\}$, $N_j = \lceil q_j N \rceil$ *(allocated budget to compute* $\mathbf{E}[\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} \rangle]$).

- *If furthermore the* R*-level design matrix* $\mathbf{T}$ *satisfies*

$$\mathbf{T}^1 = e_1 \quad \text{and} \quad \sum_{j=1}^{R} \mathbf{T}^j = e_R, \tag{22}$$

the estimator is called a Multilevel Monte Carlo (MLMC) estimator of order R.

- *If, furthermore, the R-level design matrix* **T** *satisfies*

$$\mathbf{T}^1 = e_1 \quad and \quad \sum_{j=1}^{R} \mathbf{T}^j = \mathbf{w}, \qquad where\ \mathbf{w}\ is\ the\ unique\ solution\ to\ (12), \qquad (23)$$

the estimator is called a Multilevel Richardson–Romberg (ML2R) estimator of order R.

Remark 3.3.

- Note that the assumption $\mathbf{T}^1 = e_1$ is not really necessary. It simply allows for more concise formulas in what follows.
- In this framework, denoting by **0** the null column vector of $\mathbf{R}^R$, the crude Monte Carlo is associated to the design matrix $\mathbf{T} = (e_1, \mathbf{0}, \ldots, \mathbf{0})$ and the Multistep Richardson–Romberg estimator is associated to $\mathbf{T} = (\mathbf{w}, \mathbf{0}, \ldots, \mathbf{0})$.
- Introducing such general families of matrices will allow us to justify the final choice of design matrices. To reduce the complexity of the resulting estimators leads us to choose as sparse as possible design matrices satisfying the constraints (22) or (23).

Within the abstract framework of a parametrized Monte Carlo simulation described in Section 2.1, the structure parameter π of the multilevel estimators $(\bar{Y}_{h,\underline{n}}^{N,q})_{N \geq 1}$ defined by (21) is

$$\pi = (\pi_0, q) \qquad where \begin{cases} q = (q_1, \ldots, q_R) \in (0, 1)^R, & \sum_i q_i = 1, \\ \\ \pi_0 = (h, n_1, \ldots, n_R, R, \mathbf{T}) \in \Pi_0. \end{cases}$$

Cost, complexity and effort of a multilevel estimator

In order to minimize the effort $\phi(\pi)$ of the estimator (21), let us first evaluate its unitary computational complexity. For a simulation size N, the numerical cost induced by the estimators $Y_{h,\underline{n}}^{N,q}$, $N \geq 1$, reads

$$\mathrm{Cost}\big(\bar{Y}_{h,\underline{n}}^{N,q}\big) = \sum_{j=1}^{R} N_j \sum_{i=1}^{R} \frac{1}{h} n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}} = N\kappa(\pi), \qquad (24)$$

where the unitary complexity $\kappa(\pi)$ is given by

$$\kappa(\pi) = \frac{1}{h} \sum_{j=1}^{R} q_j \sum_{i=1}^{R} n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}. \qquad (25)$$

At this stage, it is clear that the design matrix **T** must be *as sparse as* possible to minimize $\kappa(\pi)$. However, it may happen, like for nested Monte Carlo (see Section 4.2 for details), that the unitary complexity writes

$$\kappa(\pi) = \frac{1}{h} \sum_{j=1}^{R} q_j \max_{1 \leq i \leq R} (n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}). \qquad (26)$$

The variance of the Multilevel estimator is inverse linear in N (hence of Monte Carlo type) since, using the independence of the levels, we get

$$\mathrm{var}\big(\bar{Y}_{h,\underline{n}}^{N,q}\big) = \sum_{j=1}^{R} \frac{1}{N_j^2}\, \mathrm{var}\left(\sum_{k=1}^{N_j} \langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j),k}\rangle\right)$$

$$= \frac{1}{N}\sum_{j=1}^{R}\frac{1}{q_j}\,\mathrm{var}\big(\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)}\rangle\big)$$

so that the effort of such a Multilevel estimator is given by

$$\phi(\pi) = \nu(\pi)\kappa(\pi) = \left(\sum_{j=1}^{R}\frac{1}{q_j}\,\mathrm{var}\big(\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)}\rangle\big)\right)\kappa(\pi). \qquad (27)$$

Bias error of a multilevel estimator

We now establish the bias error in this general framework. The proposition below about the bias error follows straightforwardly from the weak error expansion $(\mathrm{WE}_{\alpha,\bar{R}})$ and the definition of a design matrix **T**.

Proposition 3.4. *Assume* $(\mathrm{WE}_{\alpha,\bar{R}})$.

(a) ML2R estimator: *Let* $R \in \{2,\ldots,\bar{R}\}$ *be the depth of an ML2R estimator. For any admissible allocation policy* $q = (q_1,\ldots,q_R)$, *the bias error reads*

$$\mu(\pi_0, q) = (-1)^{R-1} c_R \left(\frac{h^R}{\underline{n}!}\right)^{\alpha}\big(1 + \eta_{R,\underline{n}}(h)\big), \qquad (28)$$

where $\eta_{R,\underline{n}}(h) = (-1)^{R-1}\underline{n}!^{\alpha}\sum_{r=1}^{R}\frac{w_r}{n_r^{\alpha R}}\eta_R(\frac{h}{n_r})$ *(see* (17)*) with* η_R *defined in* $(\mathrm{WE}_{\alpha,\bar{R}})$.

(b) MLMC estimator: *Let* $R \geq 2$ *be the depth of an MLMC estimator. For any admissible allocation policy* $q = (q_1,\ldots,q_R)$, *the bias error reads*

$$\mu(\pi_0, q) = c_1\left(\frac{h}{n_R}\right)^{\alpha}\left(1 + \eta_1\left(\frac{h}{n_R}\right)\right) \qquad (29)$$

with η_1 *defined in* $(\mathrm{WE}_{\alpha,\bar{R}})$.

Toward the optimal parameters

The optimization problem (8) is not attainable directly, so we decompose it into two successive steps:

Step 1: Minimization of the effort ϕ over all allocation policies $q = (q_j)_{1\leq j\leq R}$ (as a function of a fixed bias parameter h). In practice, we will minimize an upper-bound $\bar{\phi}$ of the effort ϕ

$$q^* = \underset{q\in\mathcal{S}_+(R)}{\mathrm{argmin}}\,\bar{\phi}(\pi_0, q), \qquad \text{where } \phi(\pi) \leq \bar{\phi}(\pi), \text{ and } \phi^*(\pi_0) = \phi(\pi_0, q^*). \qquad (30)$$

This phase is solved in Theorem 3.6 below (an explicit expression for $\bar{\phi}$ is provided in (34)). The quantity $\phi^*(\pi_0)$ is called the *optimally allocated effort* (with a slight abuse of terminology since $\bar{\phi}$ is only an upper bound of ϕ).

Step 2: Minimization of the resulting cost as a function of the remaining parameters π_0 for a prescribed $\mathbf{L}^2$-error $\varepsilon > 0$ (and specification of the resulting size of the simulation and its cost):

$$\pi_0(\varepsilon) = \underset{\substack{\pi_0 \in \Pi_0 \\ |\mu(\pi_0, q^*)| < \varepsilon}}{\operatorname{argmin}} \left(\frac{\phi^*(\pi_0)}{\varepsilon^2 - \mu^2(\pi_0, q^*)} \right), \qquad N\big(\pi_0(\varepsilon)\big) = \frac{\phi^*(\pi_0(\varepsilon))}{\kappa(\pi_0(\varepsilon), q^*)(\varepsilon^2 - \mu^2(\pi_0, q^*))}.$$

It will be solved asymptotically when ε goes to 0 in two sub-steps. First, we consider a fixed depth R (with general refiners) in Proposition 3.9 which provides a closed form for $h^*(\varepsilon)$. Secondly, we let R vary as a function of ε (only for geometric refiners $n_i = M^{i-1}$). This leads to the main result of the paper Theorem 3.12 which yields a closed form for $R^*(\varepsilon)$ (and $N^*(\varepsilon)$) and the various asymptotics for the cost, depending on β and other structural parameters.

3.2. Optimally allocated effort (Step 1)

Throughout our investigations on these estimators, we will make extensive use of the following lemma which is a straightforward consequence of Schwarz's Inequality including its equality case.

Lemma 3.5. *For all $j \in \{1, \ldots, R\}$, let $a_j > 0$, $b_j > 0$ and $q_j > 0$ such that $\sum_{j=1}^R q_j = 1$. Then*

$$\left(\sum_{j=1}^R \frac{a_j}{q_j} \right) \left(\sum_{j=1}^R b_j q_j \right) \geq \left(\sum_{j=1}^R \sqrt{a_j b_j} \right)^2$$

and equality holds if and only if $q_j = \mu \sqrt{a_j b_j^{-1}}$, $j = 1, \ldots, R$, with $\mu = (\sum_{k=1}^R \sqrt{a_k b_k^{-1}})^{-1}$.

Theorem 3.6. *Assume (SE$_\beta$) holds and let θ be defined by (9). Then, the optimally allocated effort ϕ^* defined by (30) satisfies*

$$\phi^*(\pi_0) \leq \bar{\phi}\big(\pi_0, q^*\big) = \frac{\operatorname{var}(Y_0)}{h} \left(1 + \theta h^{\frac{\beta}{2}} \sum_{j=1}^R \left(\sum_{i=1}^R |\mathbf{T}_i^j| \, n_i^{-\frac{\beta}{2}} \right) \left(\sum_{i=1}^R n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}} \right)^{\frac{1}{2}} \right)^2,$$

where $q^* = q^*(\pi_0)$ is an optimal policy (with respect to the upper bound $\bar{\phi}$) given by

$$\begin{cases} q_1^*(\pi_0) = \mu_R^* \big(1 + \theta h^{\frac{\beta}{2}} \big), \\[2mm] q_j^*(\pi_0) = \mu_R^* \theta h^{\frac{\beta}{2}} \left(\sum_{i=1}^R |\mathbf{T}_i^j| \, n_i^{-\frac{\beta}{2}} \right) \left(\sum_{i=1}^R n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}} \right)^{-\frac{1}{2}}, \qquad j = 2, \ldots, R, \end{cases} \tag{31}$$

and μ_R^* is the normalizing constant such that $\sum_{j=1}^R q_j^* = 1$.

Proof. Under assumption (20), we have $\langle \mathbf{T}^1, Y_{h,\underline{n}}^{(1)} \rangle = Y_h^{(1)}$ and, for every $j \in \{2, \ldots, R\}$, $\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} \rangle = \langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} - Y_0^{(j)} \mathbf{1} \rangle$ since $\langle \mathbf{T}^j, \mathbf{1} \rangle = 0$. Hence, using the sub-additivity of standard deviation derived from (Minkowski's inequality) and the strong error assumption, we obtain

$$\forall j \geq 2, \qquad \mathrm{var}\big(\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} \rangle\big) = \sigma \left(\sum_{i=1}^R \mathbf{T}_i^j \big(Y_{\frac{h}{n_i}}^{(j)} - Y_0^{(j)}\big) \right)^2 \leq \left(\sum_{i=1}^R |\mathbf{T}_i^j| \sigma \big(Y_{\frac{h}{n_i}}^{(j)} - Y_0^{(j)}\big) \right)^2$$

$$\leq V_1 h^\beta \left(\sum_{i=1}^R |\mathbf{T}_i^j| |n_i|^{-\frac{\beta}{2}} \right)^2. \tag{32}$$

The variance of the Multilevel estimator is then given by

$$\mathrm{var}\big(\bar{Y}_{h,\underline{n}}^{N,q}\big) \leq \frac{1}{N} \left(\frac{\mathrm{var}(Y_h^{(1)})}{q_1} + V_1 h^\beta \sum_{j=2}^R \frac{1}{q_j} \left(\sum_{i=1}^R |\mathbf{T}_i^j| |n_i|^{-\frac{\beta}{2}} \right)^2 \right). \tag{33}$$

On the other hand, we have

$$\mathrm{var}\big(Y_h^{(1)}\big) = \mathrm{var}(Y_h) \leq \mathbf{E}\big[Y_h - \mathbf{E}[Y_0] \big]^2$$

$$\leq \|Y_h - Y_0\|_2^2 + 2\mathbf{E}\big[(Y_h - Y_0)(Y_0 - \mathbf{E}[Y_0]) \big] + \mathrm{var}(Y_0)$$

$$\leq \mathrm{var}(Y_0) + V_1 h^\beta + 2\sqrt{V_1} h^{\beta/2} \sqrt{\mathrm{var}\, Y_0} = \mathrm{var}(Y_0)\big(1 + \theta h^{\frac{\beta}{2}}\big)^2.$$

Combining (25), the above inequality (33) and the above upper-bound for $\mathrm{var}(Y_h^{(1)})$, we derive the following upper bound $\bar{\phi}(\pi)$ for the effort $\phi(\pi)$ defined by

$$\bar{\phi}(\pi) = \frac{\mathrm{var}(Y_0)}{h} \left(\frac{(1 + \theta h^{\frac{\beta}{2}})^2}{q_1} + \theta^2 h^\beta \sum_{j=2}^R \frac{1}{q_j} \left(\sum_{i=1}^R |\mathbf{T}_i^j| |n_i|^{-\frac{\beta}{2}} \right)^2 \right) \left(\sum_{i,j=1}^R q_j n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}} \right). \tag{34}$$

Applying Lemma 3.5 with $a_1 = (1 + \theta h^{\frac{\beta}{2}})^2$, $b_1 = 1$ and $a_j = \theta^2 h^\beta \big(\sum_{i=1}^R |\mathbf{T}_i^j| |n_i|^{-\frac{\beta}{2}}\big)^2$, $b_j = \sum_{i=1}^R n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}$, $j \in \{2, \ldots, R\}$ completes the proof. □

Remark 3.7 (Accuracy of the bound). As announced in Remark 2.2, we can replace the strong error assumption (SE_β) by a slight modified version e.g. $\mathrm{var}(Y_h - Y_{h'}) \leq V_1 |h - h'|^\beta$. Using this assumption, the upper bound of the previous theorem can be improved. For instance, if we make the natural choice $\mathbf{T}^j = \mathbf{W}_j(e_j - e_{j-1})$ corresponding to the ML2R estimator (see Section 3.3.2), we can replace (32) by

$$\sigma\big(\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} \rangle\big) = |\mathbf{W}_j| \sigma\big(Y_{\frac{h}{n_j}} - Y_{\frac{h}{n_{j-1}}}\big) \leq |\mathbf{W}_j| \sqrt{V_1} \left| \frac{h}{n_j} - \frac{h}{n_{j-1}} \right|^{\beta/2}. \tag{35}$$

Note that if the constant V_1 is sharp, the resulting upper bound derived in Theorem 3.6 is tight.

Remark 3.8 (About variance minimization). We established in the above proof that for every allocation policy $q = (q_1, \ldots, q_R)$,

$$\operatorname{var}\big(\bar{Y}_{h,\underline{n}}^{N,q}\big) \leq \frac{\operatorname{var}(Y_0)}{N}\left(\frac{(1+\theta h^{\frac{\beta}{2}})^2}{q_1} + \theta^2 h^\beta \sum_{j=2}^R \frac{1}{q_j}\left(\sum_{i=1}^R |\mathbf{T}_i^j|n_i^{-\frac{\beta}{2}}\right)^2\right).$$

Then, applying Lemma 3.5 with $a_1 = (1 + \theta h^{\frac{\beta}{2}})^2$, $b_1 = 1$ and $a_j = \theta^2 h^\beta (\sum_{i=1}^R |\mathbf{T}_i^j|n_i^{-\frac{\beta}{2}})^2$, $b_j = 1$, $j \in \{2, \ldots, R\}$, we obtain (since $\sum_{j=1}^R q_j b_j = 1$)

$$\inf_{q \in \mathcal{S}_+(R)} \operatorname{var}\big(\bar{Y}_{h,\underline{n}}^{N,q}\big) \leq \operatorname{var}(Y_0)\left(1 + \theta h^{\frac{\beta}{2}}\sum_{j=1}^R \sum_{i=1}^R |\mathbf{T}_i^j|n_i^{-\frac{\beta}{2}}\right)^2$$

with an optimal choice (to minimize the variance):

$$q_1^\dagger = \mu^\dagger\big(1 + \theta h^{\frac{\beta}{2}}\big), \qquad q_j^\dagger = \mu^\dagger \theta h^{\frac{\beta}{2}}\left(\sum_{i=1}^R |\mathbf{T}_i^j|n_i^{-\frac{\beta}{2}}\right)$$

($\mu^\dagger$ normalizing constant such that $\sum_{j=1}^n q_j^\dagger = 1$). Note that this choice $q^\dagger$ differs from the optimal one q^* obtained in Theorem 3.6.

3.3. Resulting cost optimization (Step 2)

3.3.1. *Bias parameter optimization* (*R fixed*)

In this first sub-step, we fix the depth $R \geq 2$, the design matrix $\mathbf{T}$ and the refiners $n_1, \ldots, n_R$ and we only optimize the bias parameter $h \in \mathcal{H}$ with respect to $\varepsilon > 0$, so that

$$\pi_0(\varepsilon) = h(\varepsilon, n_1, \ldots, n_R, R, \mathbf{T}).$$

We recall that $\phi^*(h) \leq \bar{\phi}(h, q^*) =: \bar{\phi}^*(h)$ where

$$\bar{\phi}^*(h) = \frac{\operatorname{var}(Y_0)}{h}\left(1 + \theta h^{\frac{\beta}{2}}\sum_{j=1}^R\left(\sum_{i=1}^R |\mathbf{T}_i^j|n_i^{-\frac{\beta}{2}}\right)\left(\sum_{i=1}^R n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}\right)^{\frac{1}{2}}\right)^2. \tag{36}$$

Proposition 3.9 (Bias parameter optimization). *Assume* ($\mathrm{WE}_{\alpha,\bar{R}}$) *and* ($\mathrm{SE}_\beta$)*. Let $R \geq 2$ and let n_i, $i = 1, \ldots, R$, be fixed refiners.*

(a) ML2R estimator: *Let $R \in \{2, \ldots, \bar{R}\}$ be such that $c_R \neq 0$. A ML2R estimator of depth R obtained with the allocation policy q^* defined by* (31) *and a bias parameter*

$$h^*(\varepsilon, R) = (1 + 2\alpha R)^{-\frac{1}{2\alpha R}}\left(\frac{\varepsilon}{|c_R|}\right)^{\frac{1}{\alpha R}} \underline{n}!^{\frac{1}{R}} \tag{37}$$

achieves the asymptotic minimal cost, namely

$$\inf_{\substack{h \in \mathcal{H} \\ |\mu(h,q^*)| < \varepsilon}} \mathrm{Cost}\big(\bar{Y}_{h,\underline{n}}^{N,q^*}\big) \sim \left(\frac{(1+2\alpha R)^{1+\frac{1}{2\alpha R}}}{2\alpha R} \right) \frac{|c_R|^{\frac{1}{\alpha R}} \mathrm{var}(Y_0)}{\underline{n}!^{\frac{1}{R}} \varepsilon^{2+\frac{1}{\alpha R}}} \qquad as \; \varepsilon \to 0.$$

(b) MLMC estimator: *Assume* $c_1 \neq 0$. *An MLMC estimator of depth* R *obtained with the allocation policy* q^* *defined in* (31) *and a bias parameter*

$$h^*(\varepsilon, R) = (1+2\alpha)^{-\frac{1}{2\alpha}} \left(\frac{\varepsilon}{|c_1|} \right)^{\frac{1}{\alpha}} n_R \tag{38}$$

achieves the asymptotic minimal cost, namely

$$\inf_{\substack{h \in \mathcal{H} \\ |\mu(h,q^*)| < \varepsilon}} \mathrm{Cost}\big(\bar{Y}_{h,\underline{n}}^{N,q^*}\big) \sim \left(\frac{(1+2\alpha)^{1+\frac{1}{2\alpha}}}{2\alpha} \right) \frac{|c_1|^{\frac{1}{\alpha}} \mathrm{var}(Y_0)}{n_R \varepsilon^{2+\frac{1}{\alpha}}} \qquad as \; \varepsilon \to 0.$$

Proof. (a) By definition of the effort ϕ and the bias μ of the estimator, we have (see Section 2.1)

$$\mathrm{Cost}\big(\bar{Y}_{h,\underline{n}}^{N,q^*}\big) = \frac{\phi^*(h)}{\varepsilon^2 - \mu^2(h,q^*)}.$$

It follows from (36) that the cost minimization problem is upper-bounded by the more tractable problem

$$\inf_{h \in \mathcal{H}, |\mu(h,q^*)| < \varepsilon} \frac{h\bar{\phi}^*(h)}{h(\varepsilon^2 - \mu^2(h,q^*))}$$

with a bias $\mu(h, q^*)$ satisfying (28). First, note that $\lim_{h \to 0} h\bar{\phi}(h, q^*) = \mathrm{var}(Y_0)$. We will consider now the denominator $h(\varepsilon^2 - \mu^2(h, q^*))$. Elementary computations show that, for fixed real numbers a, $R' > 0$, the function $g_{a,R'}$ defined by $g_{a,R'}(\xi) = \xi(1 - a^2 \xi^{2R'})$, $\xi > 0$, satisfies

$$\xi(a, R') := \underset{\xi > 0}{\mathrm{argmax}} \, g_{a,R'}(\xi) = \big((2R'+1)^{\frac{1}{2}} a\big)^{-\frac{1}{R'}}$$

and

$$\max_{(0,+\infty)} g_{a,R'} = \frac{2R'}{(2R'+1)^{1+\frac{1}{2R'}}} a^{-\frac{1}{R'}}.$$

Then, set $R' = R\alpha$, $\tilde{a} = \frac{|\widetilde{\mathbf{w}}_{R+1} c_R|}{\varepsilon}$. Inspired by what precedes, we make the sub-optimal choice $h(\varepsilon, R) = h(\varepsilon, R, \alpha) = \xi(\tilde{a}, \alpha R) = (\frac{\varepsilon}{(2\alpha R+1)^{\frac{1}{2}} |c_R|})^{\frac{1}{\alpha R}} \underline{n}!^{\frac{1}{R}}$ corresponding to the case $\eta_{R,\underline{n}} \equiv 0$. It is clear that, at least for small enough ε, $\mu^2(h, q^*) < \varepsilon^2$ which makes this choice admissible.

Hence,

$$
\begin{aligned}
&\inf_{\substack{h\in\mathcal{H}\\ |\mu(h,q^*)|<\varepsilon}} \frac{\phi^*(h)}{\varepsilon^2 - \mu^2(h,q^*)}\\
&\leq \left(1+\frac{1}{2\alpha R}\right)(2\alpha R+1)^{\frac{1}{2\alpha R}}|c_R|^{\frac{1}{\alpha R}}\frac{h(\varepsilon,R)\bar{\phi}^*(h(\varepsilon,R))}{\underline{n}!^{\frac{1}{R}}\varepsilon^{2+\frac{1}{\alpha R}}}\frac{1}{1-\frac{(\eta_{R,\underline{n}}(h(\varepsilon,R))+1)^2-1}{2\alpha R}}.
\end{aligned}
\tag{39}
$$

The "lim sup" side of the result follows since $\lim_{h\to 0}\eta_{R,\underline{n}}(h)=0$.

On the other hand, it follows from the definition (27) of the effort ϕ that

$$
\phi^*(h)=\frac{1}{h}\left(\sum_{j=1}^{R}\frac{1}{q_j^*}\operatorname{var}(\langle\mathbf{T}^j,Y_{h,\underline{n}}^{(j)}\rangle)\right)\left(\sum_{i,j=1}^{R}q_j n_i \mathbf{1}_{\{\mathbf{T}_i^j\neq 0\}}\right).
$$

Then Schwarz's Inequality implies

$$
\begin{aligned}
\phi^*(h) &\geq \frac{1}{h}\left(\sum_{j=1}^{R}\sqrt{\operatorname{var}(\langle\mathbf{T}^j,Y_{h,\underline{n}}^{(j)}\rangle)}\sqrt{\sum_{i=1}^{R}n_i\mathbf{1}_{\{\mathbf{T}_i^j\neq 0\}}}\right)^2\\
&\geq \frac{1}{h}\max_{1\leq j\leq R}\left(\operatorname{var}(\langle\mathbf{T}^j,Y_{h,\underline{n}}^{(j)}\rangle)\sum_{i=1}^{R}n_i\mathbf{1}_{\{\mathbf{T}_i^j\neq 0\}}\right)\\
&\geq \frac{1}{h}\max_{1\leq j\leq R}\operatorname{var}(\langle\mathbf{T}^j,Y_{h,\underline{n}}^{(j)}\rangle)
\end{aligned}
$$

since $n_i\geq n_1=1$, $i=1,\ldots,R$. Denoting $g(h)=\max_{1\leq j\leq R}\operatorname{var}(\langle\mathbf{T}^j,Y_{h,\underline{n}}^{(j)}\rangle)$ one clearly has $\lim_{h\to 0}g(h)=\operatorname{var}(Y_0)$ under the strong assumption (SE$_\beta$) and, as a consequence, $\lim_{h\to 0}h\phi(h)=\operatorname{var}(Y_0)$. Hence, the cost minimization problem is lower bounded by the more explicit problem

$$
\inf_{\substack{h\in\mathcal{H}\\ |\mu(h,q^*)|<\varepsilon}} \frac{g(h)}{h(\varepsilon^2-\mu^2(h,q^*))}.
$$

Let $\eta\in(0,1)$. There exists $\varepsilon_\eta>0$ such that, for every $h\in(0,h(\varepsilon_\eta,R))$,

$$
\left|g(h)-\operatorname{var}(Y_0)\right|\leq \eta\operatorname{var}(Y_0)\quad\text{and}\quad\left|\eta_{R,\underline{n}}(h)\right|\leq\eta.
$$

Let $\varepsilon\in(0,\varepsilon_\eta)$. We derive from equation (28) that

$$
\mu\left(h(\varepsilon_\eta,R),q^*\right)^2\geq\frac{\varepsilon_\eta^2(1-\eta)}{2\alpha R+1}.
$$

Consequently, if $\varepsilon < \frac{\varepsilon_\eta \sqrt{1-\eta}}{\sqrt{2\alpha R+1}}$, for every $h > 0$ such that $\mu^2(h, q^*) < \varepsilon^2$, one has

$$\frac{g(h)}{h(\varepsilon^2 - \mu(h, q^*)^2)} \geq \frac{\text{var}(Y_0)(1 - \eta)}{h(\varepsilon^2 - (1 - \eta)(\widetilde{\mathbf{w}}_{R+1} c_R)^2 h^{2\alpha R})}.$$

Taking advantage of what was done in the "lim sup" part, we get

$$\inf_{\substack{h \in \mathcal{H} \\ \mu(h, q^*) < \varepsilon}} \frac{g(h)}{h(\varepsilon^2 - \mu(h, q^*)^2)} \geq \left(1 + \frac{1}{2\alpha R}\right)(2\alpha R + 1)^{\frac{1}{2\alpha R}} |c_R|^{\frac{1}{\alpha R}} \frac{\text{var}(Y_0)}{\underline{n}!^{\frac{1}{R}} \varepsilon^{2+\frac{1}{\alpha R}}} (1 - \eta)^{1+\frac{1}{2\alpha R}}.$$

Letting ε and η successively go to zero, yields the "lim inf" side.

(b) Owing to (29), the bias $\mu(h, q)$ is now given by

$$\mu(h, q) = \left(\frac{h}{n_R}\right)^\alpha \left(c_1 + \eta_1\left(\frac{h}{n_R}\right)\right) \qquad \text{with } \lim_{h \to 0} \eta_1(h) = 0.$$

Following the lines of the proof of (a) with $R' = \alpha$ completes the proof. $\qquad \square$

Remark 3.10.

- The fact that the function $\lim_{h \to 0} h\phi^*(h) = \text{var}(Y_0)$ follows from the L^2-strong convergence of Y_h toward Y_0. Its rate of convergence plays no explicit role in this asymptotic rate of the cost as $\varepsilon \to 0$. However, this strong rate is important to design a practical allocation across the R levels, which is the key to avoid an explosion of this term.
- When $c_R = 0$, the same reasoning can be carried out by considering any small parameter $\epsilon_0^R > 0$. Anyway in practice c_R is usual not known and the impact of this situation is briefly discussed further on in Section 3.3.3.
- When $c_1 = 0$, specific weights can be computed (see Practitioner's corner in Section 5.1).

Remark 3.11. The asymptotic number N of simulations given by (8) satisfies

$$N(\varepsilon) \sim \left(1 + \frac{1}{2\alpha R}\right) \frac{\text{var}(Y_0)}{\varepsilon^2} \left(\sum_{j=1}^R q_j^* \sum_{i=1}^R n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}\right)^{-1} \qquad \text{as } \varepsilon \to 0$$

for an ML2R estimator and

$$N(\varepsilon) \sim \left(1 + \frac{1}{2\alpha}\right) \frac{\text{var}(Y_0)}{\varepsilon^2} \left(\sum_{j=1}^R q_j^* \sum_{i=1}^R n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}\right)^{-1} \qquad \text{as } \varepsilon \to 0$$

for an MLMC estimator.

3.3.2. *Templates for the design matrix* $\mathbf{T}$

We now specify the design matrices $\mathbf{T}$ in both multilevel settings MLMC defined in (22) and ML2R defined in (23).

MLMC estimator. The standard Multilevel Monte Carlo design matrix used by [12,19] is derived from the telescopic summation

$$\mathbf{E}[Y_{\frac{h}{n_R}}] = \mathbf{E}[Y_h] + \sum_{j=2}^{R} \mathbf{E}[Y_{\frac{h}{n_j}} - Y_{\frac{h}{n_{j-1}}}].$$

This telescopic sum corresponds to the design matrix $\mathbf{T}$ defined by $\mathbf{T}^j = e_j - e_{j-1}$, $j = 2, \ldots, R$ that is,

$$\mathbf{T} = \begin{pmatrix} 1 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -1 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}. \tag{MLMC}$$

In that case, the resulting upper-bound $\bar{\phi}^*$ of ϕ^* writes, with the convention $n_0 = (n_0)^{-1} = 0$,

$$\bar{\phi}^*(\pi_0) = \frac{\mathrm{var}(Y_0)}{h}\left(1 + \theta h^{\frac{\beta}{2}} \sum_{j=1}^{R}\left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1} + n_j}\right)^2. \tag{40}$$

With this design matrix (MLMC) the MLMC estimator writes

$$\bar{Y}_{h,\underline{n}}^{N,q} = \frac{1}{N_1}\sum_{k=1}^{N_1} Y_h^{(1),k} + \sum_{j=2}^{R}\frac{1}{N_j}\sum_{k=1}^{N_j}\left(Y_{\frac{h}{n_j}}^{(j),k} - Y_{\frac{h}{n_{j-1}}}^{(j),k}\right) \tag{41}$$

with $N_j = \lceil q_j N \rceil$.

ML2R estimator. The natural counterpart for the design matrix $\mathbf{T}$ in the ML2R setting appears as $\mathbf{T}^j = -\mathbf{W}_j e_{j-1} + \mathbf{W}_j e_j$, $j = 2, \ldots, R$ with $\mathbf{W}_j = \sum_{k=j}^{R} \mathbf{w}_k$ and $\mathbf{w}$ given by (12) that is,

$$\mathbf{T} = \begin{pmatrix} 1 & -\mathbf{W}_2 & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{W}_2 & -\mathbf{W}_3 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \mathbf{W}_{R-1} & -\mathbf{W}_R \\ 0 & \cdots & \cdots & \cdots & 0 & \mathbf{W}_R \end{pmatrix}. \tag{ML2R}$$

The resulting upper-bound $\bar{\phi}^*$ reads (still with the convention $n_0 = (n_0)^{-1} = 0$),

$$\bar{\phi}^*(\pi_0) = \frac{\mathrm{var}(Y_0)}{h}\left(1 + \theta h^{\frac{\beta}{2}} \sum_{j=1}^{R}|\mathbf{W}_j|\left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1} + n_j}\right)^2. \tag{42}$$

In the sequel, we will focus on the above choice (ML2R) for the design matrix **T** which leads to the ML2R estimator (3) proposed in the Introduction. With this design matrix (ML2R), the ML2R estimator writes as a weighted version of MLMC

$$\bar{Y}_{h,\underline{n}}^{N,q} = \frac{1}{N_1} \sum_{k=1}^{N_1} Y_h^{(1),k} + \sum_{j=2}^{R} \frac{\mathbf{W}_j}{N_j} \sum_{k=1}^{N_j} \left(Y_{\frac{h}{n_j}}^{(j),k} - Y_{\frac{h}{n_{j-1}}}^{(j),k} \right), \tag{43}$$

where $N_j = \lceil q_j N \rceil$. Alternative choices for **T** are proposed in Section 5.1.

3.3.3. *Global optimization with varying depth R for geometric refiners*

In this final sub-step, we consider geometric refiners with *root $M \geq 2$* of the form

$$n_i = M^{i-1}, \qquad i = 1, \ldots, R$$

and we only analyze the ML2R and MLMC estimators defined by (43) and (41), respectively. Note that geometric refiners have already been considered in regular multilevel Monte Carlo framework in [12].

Theorem 3.12. *Assume* (SE$_\beta$) *holds for $\beta > 0$.*

(a) ML2R estimator: *Assume* (WE$_{\alpha,\bar{R}}$), $\sup_{R \in \mathbf{N}} \sup_{h' \in (0,h)} |\eta_R(h')| < +\infty$ *for every $h \in \mathcal{H}$ and $\lim_{R \to +\infty} |c_R|^{\frac{1}{R}} = \tilde{c}_\infty \in (0,+\infty)$. The ML2R estimator* (43) *with design matrix* **T** *in* (ML2R) *satisfies*

$$\limsup_{\varepsilon \to 0} v(\beta, \varepsilon) \times \inf_{\substack{h \in \mathcal{H}, R \geq 2 \\ |\mu(h,R,q^*)| < \varepsilon}} \text{Cost}\left(\bar{Y}_{h,\underline{n}}^{N,q}\right) \leq K_{\text{ML2R}}(\alpha, \beta, M) \tag{44}$$

with

$$v(\beta, \varepsilon) = \begin{cases} \varepsilon^2, & \text{if } \beta > 1, \\ \varepsilon^2 \left(\log(1/\varepsilon)\right)^{-1}, & \text{if } \beta = 1, \\ \varepsilon^2 e^{-\frac{1-\beta}{\sqrt{\alpha}} \sqrt{2\log(1/\varepsilon)\log(M)}}, & \text{if } \beta < 1. \end{cases}$$

When $\beta < 1$, the best rate is achieved with $M = 2$. These rates are achieved with a depth

$$R^*(\varepsilon) = \left\lceil \frac{1}{2} + \frac{\log(\tilde{c}^{\frac{1}{\alpha}}\mathbf{h})}{\log(M)} + \sqrt{\left(\frac{1}{2} + \frac{\log(\tilde{c}^{\frac{1}{\alpha}}\mathbf{h})}{\log(M)}\right)^2 + 2\frac{\log(A/\varepsilon)}{\alpha \log(M)}} \right\rceil, \qquad A = \sqrt{1+4\alpha},$$

with $\tilde{c} > \tilde{c}_\infty$, satisfying $\lim_{\varepsilon \to 0} R^*(\varepsilon) = +\infty$ and a bias parameter $h^*(\varepsilon) = \mathbf{h}/\lceil \mathbf{h}/h^*(\varepsilon, R^*(\varepsilon)) \rceil$ where $h^*(\varepsilon, R)$ is given by (37). The finite real constant $K_{\text{ML2R}}(\alpha, \beta, M)$ depends on M and on

the structural parameters α, β, V_1, $\mathrm{var}(Y_0)$, $\mathbf{h}$, *namely*

$$K_{\mathrm{ML2R}}(\alpha,\beta,M) = \begin{cases} \dfrac{\mathrm{var}(Y_0)M}{\mathbf{h}}\left(1+\theta\mathbf{h}^{\frac{\beta}{2}}\dfrac{\mathbf{W}_\alpha(M)M^{\frac{\beta-1}{2}}\sqrt{1+M}(1+M^{-\frac{\beta}{2}})}{1-M^{\frac{1-\beta}{2}}}\right)^2, \\ \hfill \text{if } \beta > 1, \\[2mm] \dfrac{2V_1}{\alpha}\left(\dfrac{\mathbf{W}_\alpha(M)M(1+M)(1+M^{-\frac{1}{2}})^2}{\log(M)}\right), \qquad \text{if } \beta = 1, \\[2mm] V_1\mathbf{h}^{1-\beta}\tilde{c}_\infty^{\frac{(1-\beta)}{\alpha}}\left(\dfrac{\mathbf{W}_\alpha^2(M)M^{2-\beta}(1+M)(1+M^{-\frac{\beta}{2}})^2}{(M^{\frac{1-\beta}{2}}-1)^2}\right), \\ \hfill \text{if } \beta < 1, \end{cases} \tag{45}$$

where $\mathbf{W}_\alpha(M) = \frac{M^{-\alpha}}{\pi_{\alpha,M}^2}\sum_{k\ge 0}M^{-\alpha\frac{k(k+3)}{2}}+\frac{1}{\pi_{\alpha,M}}$ *with* $\pi_{\alpha,M}=\prod_{k\ge 1}(1-M^{-\alpha k})$.

(b) *MLMC estimator:* *Assume* $(\mathrm{WE}_{\alpha,1})$ *and* $c_1 \ne 0$. *The MLMC estimator* (41) (*with design matrix* $\mathbf{T}$ *defined in* (MLMC)) *satisfies*

$$\limsup_{\varepsilon\to 0}v(\beta,\varepsilon)\times\inf_{\substack{h\in\mathcal{H},R\ge 2\\|\mu(h,R,q^*)|<\varepsilon}}\mathrm{Cost}\big(\bar{Y}_{h,\underline{n}}^{N,q}\big)\le K_{\mathrm{MLMC}}(\alpha,\beta,M)\tag{46}$$

with

$$v(\beta,\varepsilon)=\begin{cases}\varepsilon^2, & \text{if }\beta>1,\\[1mm]\varepsilon^2\big(\log(1/\varepsilon)\big)^{-2}, & \text{if }\beta=1,\\[1mm]\varepsilon^{2+\frac{1-\beta}{\alpha}}, & \text{if }\beta<1.\end{cases}$$

These rates are achieved with a depth

$$R^*(\varepsilon)=\left\lceil 1+\frac{\log(|c_1|^{\frac{1}{\alpha}}\mathbf{h})}{\log(M)}+\frac{\log(A/\varepsilon)}{\alpha\log(M)}\right\rceil,\qquad A=\sqrt{1+2\alpha}$$

satisfying $\lim_{\varepsilon\to 0}R^*(\varepsilon)=+\infty$ *and a bias parameter* $h^*(\varepsilon)=\mathbf{h}/\lceil\mathbf{h}/h^*(\varepsilon,R^*(\varepsilon))\rceil$ *where* $h^*(\varepsilon,R)$ *is given by* (38). *The finite real constant* $K_{\mathrm{MLMC}}(\alpha,\beta,M)$ *depends on* M *and the structural parameters* α, β, V_1, $\mathrm{var}(Y_0)$, $\mathbf{h}$, *namely*

$$K_{\mathrm{MLMC}}(\alpha,\beta,M) = \begin{cases} \left(1+\dfrac{1}{2\alpha}\right)\dfrac{\mathrm{var}(Y_0)M}{\mathbf{h}}\left(1+\theta\mathbf{h}^{\frac{\beta}{2}}\dfrac{M^{\frac{\beta-1}{2}}\sqrt{1+M}(1+M^{-\frac{\beta}{2}})}{1-M^{\frac{1-\beta}{2}}}\right)^2, \\ \hfill \text{if }\beta>1, \\[2mm] \left(1+\dfrac{1}{2\alpha}\right)\dfrac{V_1}{\alpha^2}\left(\dfrac{M(1+M)(1+M^{-\frac{1}{2}})^2}{\log(M)^2}\right), \qquad \text{if }\beta=1, \\[2mm] \dfrac{(1+2\alpha)^{1+\frac{1-\beta}{2\alpha}}}{2\alpha}V_1\mathbf{h}^{1-\beta}|c_1|^{\frac{(1-\beta)}{\alpha}}\left(\dfrac{M^{2-\beta}(1+M)(1+M^{-\frac{\beta}{2}})^2}{(M^{\frac{1-\beta}{2}}-1)^2}\right), \\ \hfill \text{if }\beta<1. \end{cases}$$

Comments

Claim (b) is essentially established in Giles' complexity theorem from [12].

- When $\beta < 1$, ML2R (with $M = 2$) is asymptotically more efficient than MLMC by a factor $\varepsilon^{-\frac{1-\beta}{\sqrt{\alpha}}} e^{-\frac{1-\beta}{\alpha}} \sqrt{2\log(M)\log(1/\varepsilon)}$ which goes to $+\infty$ as $\varepsilon \to 0$ in a very steep way. To be precise the ratio is greater than 1 as soon as

$$\varepsilon \leq 2^{-\frac{2}{\alpha}}.$$

  It is clear that it is for this setting that ML2R is the most powerful compared to regular MLMC.
- When $\beta = 1$, ML2R is asymptotically more efficient than MLMC by a factor $\log(1/\varepsilon) \to +\infty$ as $\varepsilon \to 0$.
- When $\beta > 1$, both estimators achieve the same rate ε^{-2} as a virtual unbiased Monte Carlo method based on the direct simulation of Y_0. Some numerical experiments carried out with the call in Black–Scholes model discretized by a Milstein scheme strongly suggest that the constant of the ML2R estimator is significant lower than the MLMC one.

Remark 3.13.

- It is proved in Appendix B that $\lim_{M \to +\infty} \mathbf{W}_\alpha(M) = 1$ and, to be more precise, that $\mathbf{W}_\alpha(M) - 1 \sim M^{-\alpha}$ as $M \to +\infty$.
- The assumption on the functions η_R and the sequence $(c_R)_{R \geq 2}$ in (a) of the above proposition are reasonable, though almost impossible to check in practice. In particular, note that as soon as the sequence $(c_R)_{R \geq 2}$ has at most a polynomial growth as a function of R, it satisfies the assumption since $\tilde{c}_\infty = 1$.
- When $\tilde{c}_\infty = 0$, the constant $K(\alpha, \beta, M)$ equals to 0 which emphasizes that we are not in the right asymptotic. In practive $\tilde{c}_\infty$ is replaced in this constant by the parameter $\tilde{c} > 0$ used to define the depth.

Proof of Theorem 3.12. We provide a detailed proof of claim (a), that of (b) following the same lines.

Step 1: We start from equation (39) in the proof of Proposition 3.9 which reads

$$\inf_{\substack{h \in \mathcal{H} \\ |\mu(h,q^*)| < \varepsilon}} \mathrm{Cost}\big(\bar{Y}_{h,\underline{n}}^{N,q^*}\big) \leq \left(1 + \frac{1}{2\alpha R}\right) \frac{\bar{\phi}^*(h^*(\varepsilon, R))}{\varepsilon^2} \frac{1}{1 - \frac{(\eta_{R,\underline{n}}(h^*(\varepsilon,R))+1)^2 - 1}{2\alpha R}}$$

with

$$\bar{\phi}^*\big(h^*(\varepsilon, R)\big) = \frac{\mathrm{var}(Y_0)}{h^*(\varepsilon, R)} \left(1 + \theta h^*(\varepsilon, R)^{\frac{\beta}{2}} \sum_{j=1}^{R} |\mathbf{W}_j| \big(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\big) \sqrt{n_{j-1} + n_j}\right)^2$$

(convention $n_0 = (n_0)^{-1} = 0$). The idea is to choose $R = R^*(\varepsilon)$ as large as possible provided the optimal bias parameter lies in $\mathcal{H}$. The form of the refiners $n_i = M^{i-1}$ implies that $\underline{n}! = M^{\frac{R(R-1)}{2}}$

so that

$$h^*(\varepsilon, R) = (1 + 2\alpha R)^{-\frac{1}{2\alpha R}} |c_R|^{-\frac{1}{\alpha R}} \varepsilon^{\frac{1}{\alpha R}} M^{\frac{R-1}{2}}.$$

To determine the dependence of R with respect to ε we consider the auxiliary function

$$\tilde{h}(\varepsilon, R) = (1 + 4\alpha)^{-\frac{1}{2\alpha R}} \tilde{c}^{-\frac{1}{\alpha}} \varepsilon^{\frac{1}{\alpha R}} M^{\frac{R-1}{2}},$$

and the polynomial function P defined by

$$P(R) = \frac{R(R-1)}{2} \log(M) - R \log(K) - \frac{1}{\alpha} \log(\sqrt{1 + 4\alpha}/\varepsilon)$$

with $K = \tilde{c}^{\frac{1}{\alpha}} \mathbf{h}$, so that $\tilde{h}(\varepsilon, R) = \mathbf{h} e^{\frac{P(R)}{R}}$. Note that P has a unique positive root $R_+(\varepsilon)$ given by

$$R_+(\varepsilon) = \frac{1}{2} + \frac{\log(K)}{\log(M)} + \sqrt{\left(\frac{1}{2} + \frac{\log(K)}{\log(M)}\right)^2 + 2\frac{\log(\sqrt{1 + 4\alpha}/\varepsilon)}{\alpha \log M}},$$

hence satisfying $\tilde{h}(\varepsilon, R_+(\varepsilon)) = \mathbf{h}$. We then consider $R^*(\varepsilon) = \lceil R_+(\varepsilon) \rceil$ and define $h^*(\varepsilon)$ as the projection of $h^*(\varepsilon, R^*(\varepsilon))$ on $\mathcal{H}$ so that $h^*(\varepsilon) \leq \mathbf{h}$ and $h^*(\varepsilon)$ is equal to $\mathbf{h}$ for small enough ε.

Let us show that our choice $h^*(\varepsilon)$ is admissible, i.e. $\mu(\varepsilon) := \mu(h^*(\varepsilon), R^*(\varepsilon), q^*)^2 < \varepsilon^2$ at least for small enough ε. Elementary computations show that

$$\mu(\varepsilon)^2 = \left(c_{R^*(\varepsilon)} M^{-\frac{R^*(\varepsilon)(R^*(\varepsilon)-1)}{2}\alpha} h^*(\varepsilon)^{\alpha R^*(\varepsilon)}\right)^2 \left(1 + \eta_{R^*(\varepsilon),\underline{n}}(h^*(\varepsilon))\right)^2$$

$$= (1 + 4\alpha)^{-1} \varepsilon^2 e^{-2\alpha P(R^*(\varepsilon))} \left(\frac{c_{R^*(\varepsilon)}}{\tilde{c}^{R^*(\varepsilon)}}\right)^2 \left(1 + \eta_{R^*(\varepsilon),\underline{n}}(h^*(\varepsilon))\right)^2.$$

First note that we have $\lim_{R \to +\infty} |c_R|^{\frac{1}{R}} = \tilde{c}_\infty$ and $\tilde{c} > \tilde{c}_\infty$. Moreover, Claim 6 of Proposition A.2 in Appendix A and the assumption on η_R imply that

$$\sup_{0 < h' < \mathbf{h}} \left|\eta_{R^*(\varepsilon),\underline{n}}(h')\right| \leq B_\alpha(M) \sup_{h' \in (0, \mathbf{h})} \left|\eta_{R^*(\varepsilon)}(h')\right| \leq B_\alpha(M) \sup_{R \geq 1} \sup_{h' \in (0, \mathbf{h})} \left|\eta_R(h')\right| < +\infty.$$

As a consequence of the assumption made on the functions η_R, it is clear that $\mu(\varepsilon)^2 = o(\varepsilon^2)$ since $R^*(\varepsilon) \to +\infty$ as $\varepsilon \to 0$. Hence, our choice for the bias parameter is admissible, at least for small enough ε.

Likewise, the assumption on the functions η_R implies $\lim_{\varepsilon \to 0} \frac{(\eta_{R^*(\varepsilon),\underline{n}}(h^*(\varepsilon))+1)^2 - 1}{2\alpha R^*(\varepsilon)} = 0$. We have then proved that

$$\limsup_{\varepsilon \to 0} \left(l(\varepsilon, R^*(\varepsilon)) \inf_{\substack{h \in \mathcal{H} \\ |\mu(h, R, q^*)| < \varepsilon}} \times \text{Cost}(\bar{Y}_{h,\underline{n}}^{N,q^*}) \right) \leq \frac{M \operatorname{var}(Y_0)}{\mathbf{h}}$$

with

$$l(\varepsilon, R) = \varepsilon^2 \left(1 + \theta h^*(\varepsilon, R)^{\frac{\beta}{2}} \sum_{j=1}^R |\mathbf{W}_j| \left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right) \sqrt{n_{j-1} + n_j}\right)^{-2}.$$

It follows from Claim 5 of Proposition A.2 in Appendix A that $\max_{j=1,\dots,R} |\mathbf{W}_i| \leq \mathbf{W}_\alpha(M)$. On the other hand, standard computations show that, for every $j = 2, \dots, R$,

$$
\left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1} + n_j} = M^{\beta-1} M^{j\frac{1-\beta}{2}}\left(1 + M^{-\frac{\beta}{2}}\right)(1+M)^{\frac{1}{2}}. \tag{47}
$$

Moreover, with our convention on n_0, it still holds true as an inequality ($\leq$) for $j = 1$. So

$$
l(\varepsilon, R) \geq \varepsilon^2\left(1 + \theta h^*(\varepsilon, R)^{\frac{\beta}{2}}\mathbf{W}_\alpha(M) M^{\beta-1}\sqrt{1+M}\left(1+M^{-\frac{\beta}{2}}\right)\sum_{j=1}^{R} M^{j\frac{1-\beta}{2}}\right)^{-2}.
$$

Step 2: Now we will inspect successively the three cases depending on the strong rate convergence parameter $\beta > 0$.

Case $\beta = 1$. In that case,

$$
\begin{aligned}
l\left(\varepsilon, R^*(\varepsilon)\right) &\geq \varepsilon^2\left(1 + \theta h^*(\varepsilon)^{\frac{\beta}{2}}\mathbf{W}_\alpha(M)\sqrt{1+M}\left(1+M^{-\frac{1}{2}}\right)R^*(\varepsilon)\right)^{-2}\\
&\geq \varepsilon^2\left(1 + \theta\mathbf{h}^{\frac{\beta}{2}}\mathbf{W}_\alpha(M)\sqrt{1+M}\left(1+M^{-\frac{1}{2}}\right)R_+(\varepsilon)\right)^{-2},
\end{aligned}
$$

and, as $R^*(\varepsilon)^2 \sim R_+(\varepsilon)^2 \sim \frac{2}{\alpha \log(M)}\log(1/\varepsilon)$ as $\varepsilon \to 0$, we get (44) with $K_{\mathrm{ML2R}}(\alpha, 1, M)$ given by (45) keeping in mind that $V_1 = \mathrm{var}(Y_0)\theta^2$.

Case $\beta > 1$. Noting that $\sum_{j=1}^{R} M^{j\frac{1-\beta}{2}} \leq \frac{M^{\frac{1-\beta}{2}}}{1-M^{\frac{1-\beta}{2}}}$, we get

$$
l\left(\varepsilon, R^*(\varepsilon)\right) \geq \varepsilon^2\left(1 + \theta\mathbf{h}^{\frac{\beta}{2}}\frac{\mathbf{W}_\alpha(M)M^{\frac{\beta-1}{2}}\sqrt{1+M}\left(1+M^{-\frac{\beta}{2}}\right)}{1-M^{\frac{1-\beta}{2}}}\right)^{-2},
$$

which yields (44) with $K_{\mathrm{ML2R}}(\alpha, \beta, M)$ given by (45).

Case $\beta < 1$. In that setting, we note this time that $\sum_{j=1}^{R} M^{j\frac{1-\beta}{2}} \leq \frac{M^{(R+1)\frac{1-\beta}{2}}}{M^{\frac{1-\beta}{2}}-1}$ so that

$$
l\left(\varepsilon, R^*(\varepsilon)\right) \geq \varepsilon^2\left(1 + \theta\mathbf{h}^{\frac{\beta}{2}}\frac{\mathbf{W}_\alpha(M)\sqrt{1+M}\left(1+M^{-\frac{\beta}{2}}\right)}{M^{\frac{1-\beta}{2}}-1}M^{(R^*(\varepsilon)-1)\frac{1-\beta}{2}}\right)^{-2}.
$$

As $R_+(\varepsilon)$ satisfies $\tilde{h}(\varepsilon, R_+(\varepsilon)) = \mathbf{h}$, we obtain $M^{\frac{R_+(\varepsilon)-1}{2}} = (1+4\alpha)^{\frac{1}{2\alpha R_+(\varepsilon)}}\mathbf{h}\tilde{c}^{\frac{1}{\alpha}}\varepsilon^{-\frac{1}{\alpha R_+(\varepsilon)}}$. We have $\varepsilon^{-\frac{1}{\alpha R_+(\varepsilon)}} \sim e^{\sqrt{\frac{\log(M)}{2\alpha}\log(1/\varepsilon)}}$ as $\varepsilon \to 0$. Elementary, though tedious, computations yield (44) with $K_{\mathrm{ML2R}}(\alpha, \beta, M)$ given by (45).

(b) The choice for $R^*(\varepsilon)$ follows from by considering the auxiliary function

$$
\tilde{h}(\varepsilon, R) = (1+2\alpha)^{-\frac{1}{2\alpha}}|c_1|^{-\frac{1}{\alpha}}\varepsilon^{\frac{1}{\alpha}}M^{R^*(\varepsilon)-1}.
$$

Then, the proof follows the same lines as that of (a). $\square$

Remark 3.14 (On the constraint h). In the proof, we chose to saturate the constraint $h^* \leq \mathbf{h}$. If we consider $h^* = \chi$ where χ is a free parameter in $(0, \mathbf{h}]$, then the asymptotic constants $K(\alpha, \beta, M)$ for the renormalized optimized cost in Theorem 3.12 depends on χ and one verifies the following facts:

- When $\beta < 1$, one can write $K_{\mathrm{ML2R}}(\alpha, \beta, M, \chi) = \chi^{1-\beta} K_{\mathrm{ML2R}}(\alpha, \beta, M, 1)$ which this time suggests to start the simulation with a small upper bias parameter $\chi < \mathbf{h}$.
- When $\beta = 1$, the asymptotic constant $K_{\mathrm{ML2R}}(\alpha, 1, M, \chi)$ *does not depend* on χ. This suggests that the choice of the upper bias parameter is not decisive, at least for high accuracy computations (ε close to 0). The choice $\chi = \mathbf{h}$ remains the most natural.
- When $\beta > 1$, the asymptotic cost of the simulation increases in ε^2 like a (virtual) unbiased one. In that very case, it appears that the asymptotic constant $K_{\mathrm{ML2R}}(\alpha, \beta, M, \chi)$ can itself be optimized as a function of χ. Namely, if we set

$$\kappa_1 = \frac{\mathrm{var}(Y_0)M}{\chi} \quad \text{and} \quad \kappa_2 = \theta^2 \frac{\mathbf{W}_\alpha(M)^2 M^{\beta-1}(1+M)(1+M^{-\beta})}{(1 - M^{\frac{1-\beta}{2}})^2},$$

then

$$\chi_{\mathrm{opt}} = \beta^{-\frac{2}{\beta+1}} \kappa_2^{-\frac{1}{\beta+1}} \quad \text{and} \quad K_{\mathrm{ML2R}}(\alpha, \beta, M, \chi_{\mathrm{opt}}) = (\beta+1)^2 \beta^{-\frac{2}{\beta+1}} \kappa_1 \kappa_2^{\frac{1}{\beta+1}}.$$

4. Examples of applications

4.1. Brownian diffusion approximation

Euler scheme

In fact, the (one-step) Richardson–Romberg extrapolation is well known as an efficient mean to reduce the time discretization error induced by the use of an Euler scheme to simulate a Brownian diffusion. In this field of Numerical Probability, its introduction goes back to Talay and Tubaro in their seminal paper [30] on weak error expansion, followed by the case of non-smooth functions in [2], under an Hörmander hypo-ellipticity assumption.

It relies on the following theorem.

Theorem 4.1. *Let $b : \mathbf{R}^d \to \mathbf{R}^d$, $\sigma : \mathbf{R}^d \to \mathcal{M}(d, q)$ and let $(W_t)_{t \geq 0}$ be a q-dimensional standard Brownian motion defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. Let $X = (X_t)_{t \in [0, T]}$ be a diffusion process, strong solution to the Stochastic Differential equation (SDE)*

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad t \in [0, T], X_0 = x_0 \in \mathbf{R}^d, \tag{48}$$

and its continuous Euler scheme $\bar{X}^h = (\bar{X}_t^h)_{t \in [0,T]}$ with bias (step) parameter $h = T/n$ defined by

$$\bar{X}_t^h = X_0 + \int_0^t b(\bar{X}_{\underline{s}}^h)\,ds + \int_0^t \sigma(\bar{X}_{\underline{s}}^h)\,dW_s,$$

where

$$\underline{s} = kh \ on \ \big[kh, (k+1)h\big), \qquad k = 0, \dots, n.$$

(a) *Smooth setting* (*Talay–Tubaro* [30]): *If b and σ are infinitely differentiable with bounded partial derivatives and if $f : \mathbf{R}^d \to \mathbf{R}$ is an infinitely differentiable function, whose all partial derivatives have polynomial growth, then, for a fixed $T > 0$ and every integer $R \in \mathbf{N}^*$,*

$$\mathbf{E}\big[f\big(\bar{X}_T^h\big)\big] - \mathbf{E}\big[f(X_T)\big] = \sum_{k=1}^{R} c_k h^k + O\big(h^{R+1}\big), \tag{49}$$

where the coefficients c_k depend on b, σ, f, T (but not on h).

(b) (*Hypo-*)*Elliptic setting* (*Bally–Talay* [2]): *If b and σ are infinitely differentiable with bounded partial derivatives and if σ is uniformly elliptic in the sense that*

$$\forall x \in \mathbf{R}^d, \qquad \sigma\sigma^*(x) \geq \varepsilon_0 I_q, \qquad \varepsilon_0 > 0$$

or, more generally, if (b, σ) satisfies the strong Hörmander hypo-ellipticity assumption, then (49) *holds true for every bounded Borel function $f : \mathbf{R}^d \to \mathbf{R}$.*

Other results based on the direct expansion of the density of the Euler scheme allow to deal with a drift b with linear growth (see [24], in a uniformly elliptic setting, see also [18] at order 1 in a tempered distribution framework). It is commonly shared by the "weak error community", relying on an analogy with recent results on the existence of smooth density from the diffusion, that if the hypo-ellipticity assumption is satisfied except at finitely many points that are never visited by the diffusion, then the claim (b) remains true. The boundedness assumption on σ is probably more technical than a mandatory assumption. For a recent review on weak error, we refer to [22].

To deal with our abstract multilevel framework, we consider for a fixed horizon $T > 0$, the family of Euler schemes $\bar{X}^h$ with step $h \in \mathcal{H} = \{\frac{T}{n}, n \geq 1\}$. We set $Y_h = f(\bar{X}_T^h)$ and $Y_0 = f(X_T)$ for a function f either smooth enough with polynomial growth or simply Borel and bounded, depending on the smoothness of b and σ and the (hypo-)ellipticity of σ. The above theorem says that condition $(\mathrm{WE}_{\alpha, \bar{R}})$ is satisfied with $\bar{R} = +\infty$ and $\alpha = 1$. However, for a fixed $\bar{R}$, the differentiability assumption on b, σ and f can be relaxed by simply assuming that these three functions are $\mathcal{C}_b^{\bar{R}+5}$ on $[0, T] \times \mathbf{R}^d$.

On the other hand, as soon as $f : \mathbf{R}^d \to \mathbf{R}$ is Lipschitz continuous, it is classical results that (SE_β) is satisfied with $\beta = 1$ as an easy consequence of the fact that the (continuous) Euler scheme $\bar{X}^h$ converges for the sup-norm toward X in $\mathbf{L}^2$ (in fact in every $\mathbf{L}^p$-space) at rate $\sqrt{h}$ as the step h goes to 0.

In such a setting, we can implement multilevel estimators with $\alpha = \beta = 1$.

Milstein scheme

The Milstein scheme is a second order scheme which satisfies (SE_β) with $\beta = 2$ and $(\mathrm{WE}_{\alpha, \bar{R}})$ still with $\alpha = 1$ (like the Euler scheme). Consequently, provided it can be implemented, the resulting multilevel estimators should be designed with these parameters.

However, the main drawback of the Milstein scheme when the *SDE* is driven by a multidimensional Brownian motion ($q \geq 2$), is that it requires the simulation of Lévy areas, for which there is no known efficient method (except in dimension 2). In a recent work [14], Giles and Szpruch introduce a suitable *antithetic multilevel correction estimator* which avoids the simulation of these Lévy areas. This approach can be easily combined with our weighted version of MLMC.

Note that in the $\beta > 1$ case, Rhee and Glynn introduced in [28] a class of finite-variance optimally randomized multilevel estimators which are unbiased with a square root convergence rate.

Path-dependent functionals

When a functional $F : \mathcal{C}([0, T], \mathbf{R}^d) \to \mathbf{R}$ is Lipschitz continuous for the sup-norm, it is straightforward that $F(\bar{X}^h)$ and $F(X)$ satisfy (SE_β), with $\beta = 1$ and $\mathcal{H} = \{\frac{T}{n}, n \geq 1\}$ (but this is no longer true if one considers the *stepwise constant* Euler scheme since the rate of convergence is then $\sqrt{\log n / n} \asymp \sqrt{-h \log h}$). More generally, if F is β-Hölder, $\beta \in (0, 1]$, then this family satisfies (SE_β). High order expansions of the weak error are not available in the general case, however first order expansion have been established for specific functionals like $F(\mathbf{w}) = f(\int_0^T \mathbf{w}(s)\, ds)$ or $F(\mathbf{w}) = f(\mathbf{w}(T))\mathbf{1}_{\{\tau_D(\mathbf{w}) > T\}}$ where $\tau_D(\mathbf{w})$ is the exit time of a domain D of $\mathbf{R}^d$ showing that ($WE_{\alpha, \bar{R}}$) holds with $\alpha = 1$ and $\bar{R} = 1$ (see, e.g., [16,25]). More recently, new results on first order weak error expansions have been obtained for functionals of the form $F(\mathbf{w}) = f(\mathbf{w}(T), \sup_{t \in [0,T]} \mathbf{w}(t))$ (see [13] and [1]). Thus, for the weak error expansion, it is shown in [1] that, for every $\eta > 0$, there exists a real constant $C_\eta > 0$ such that

$$\left| \mathbf{E}\left[f\left(X_T, \sup_{t \in [0,T]} X_t \right) \right] - \mathbf{E}\left[f\left(\bar{X}_T^n, \sup_{t \in [0,T]} \bar{X}_t^n \right) \right] \right| \leq \frac{C_\eta}{N^{\frac{2}{3} - \eta}}.$$

For a review of recent results on approximation of solutions of SDEs, we again refer to [22].

Remark 4.2. Note that, as concerns the MLMC estimator, in the general setting of the discretization of a Brownian diffusion by an Euler scheme, a central limit theorem (with stable weak convergence) has been obtained in [6]. In fact both the ML2R and MLMC estimators attached to the design matrices (ML2R) and (MLMC) satisfy, under a sharp version of (SE_β), a central limit theorem (see [15]) as $\varepsilon \to 0$. In the case of ML2R, it requires an in-depth analysis of the asymptotic behaviour of the weight vector $(\mathbf{W}_i) = (\mathbf{W}_i^{\alpha, R})_{1 \leq i \leq R}$ as R goes to ∞.

4.2. Nested Monte Carlo

The purpose of the so-called *nested* Monte Carlo method is to compute by simulation quantities of the form

$$\mathbf{E}\left[f\left(\mathbf{E}[X|Y] \right) \right],$$

where (X, Y) is a couple of $\mathbf{R} \times \mathbf{R}^{q_Y}$-valued random variable defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ with $X \in \mathbf{L}^2(\mathbf{P})$ and $f : \mathbf{R} \to \mathbf{R}$ is a Lipschitz continuous function with Lipschitz

coefficient $[f]_{\mathrm{Lip}}$. Such quantities often appear in financial applications, like compound option pricing or risk estimation (see [7]) and in actuarial sciences (see [10]) where nested Monte Carlo is widely implemented. The idea of replacing conditional expectations by Monte Carlo estimates also appears in [5] where the authors derive a multilevel dual Monte Carlo algorithm for pricing American style derivatives.

We make the following more stringent assumption: there exists a Borel function $F : \mathbf{R}^{q_Z} \times \mathbf{R}^{q_Y} \to \mathbf{R}$ and a random variable $Z : (\Omega, \mathcal{A}) \to \mathbf{R}^{q_Z}$ independent of Y such that

$$X = F(Z, Y).$$

Then, if $X \in \mathbf{L}^2$, one has the following representation

$$\mathbf{E}[X|Y](\omega) = \big(\mathbf{E}\big[F(Z, y)\big]\big)_{|y=Y(\omega)} = \int_{\mathbf{R}^{q_Z}} F\big(z, Y(\omega)\big)\mathbf{P}_Z(\mathrm{d}z).$$

To comply with the multilevel framework, we set

$$\mathcal{H} = \{1/K,\, K \geq 1\}, \qquad Y_0 = f\big(\mathbf{E}[X|Y]\big), \qquad Y_{\frac{1}{K}} = f\left(\frac{1}{K}\sum_{k=1}^{K} F(Z_k, Y)\right),$$

where $(Z_k)_{k\geq 1}$ is an i.i.d. sequence of copies of Z defined on $(\Omega, \mathcal{A}, \mathbf{P})$ and independent of Y (up to an enlargement of the probability space if necessary).

The following proposition shows that the nested Monte Carlo method is eligible for multilevel simulation when f is regular enough with the same parameters as the Euler scheme for Brownian diffusions.

Proposition 4.3. *Assume $X \in \mathbf{L}^{2R}$. If f is Lipschitz continuous and $2R$ times differentiable with $f^{(k)}$ bounded, $k = R, \ldots, 2R$, the nested Monte Carlo satisfies (SE_β) with $\beta = 1$ and $(\mathrm{WE}_{\alpha, \bar{R}})$ with $\alpha = 1$ and $\bar{R} = R - 1$.*

Remark 4.4. When f is no longer smooth, typically if it is the indicator function of an interval, it is still possible to show that nested Monte Carlo is eligible for multilevel Richardson–Romberg approach e.g. in the more constrained framework developed in [17,20] where X can be viewed as an additive perturbation of Y. Assuming enough regularity in y on the joint density $g_N(y, z)$ of Y and the renormalized perturbation, yields an expansion of the weak error (but seems in a different scale). However, in this work we focus on the regular case (see [26] for the non-regular case and applications in actuarial sciences).

The proof follows from the two lemmas below.

Lemma 4.5 (Strong approximation error). *Assume f is Lipschitz continuous. For every $h, h' \in \mathcal{H} \cup \{0\}$,*

$$\|Y_{h'} - Y_h\|_2^2 \leq [f]_{\mathrm{Lip}}^2\big(\|X\|_2^2 - \big\|\mathbf{E}[X|Y]\big\|_2^2\big)\big|h' - h\big| \tag{50}$$

so that $(Y_h)_{h \in \mathcal{H}}$ satisfies (SE_β) with $\beta = 1$ and the alternative assumption (Var'_β) from Remark 2.2.

Proof. Let $h = \frac{1}{K}$, $h' = \frac{1}{K'}$, K, $K' \in \mathbf{N}^*$, $K \leq K'$. Now set for convenience $\widetilde{X}_k = F(Z_k, Y) - \mathbf{E}_Y[F(Z_k, Y)]$, $M_k = \sum_{\ell=1}^k \widetilde{X}_k$ and $\mathcal{G}_k = \sigma(Y, Z_1, \ldots, Z_k)$, $k \geq 0$. It is clear that $(M_k)_{k \geq 0}$ is a square integrable martingale (null at time 0) satisfying $\mathbf{E}[(M_k - M_{k-1})^2 | \mathcal{G}_{k-1}](\|X\|_2^2 - \|\mathbf{E}[X|Y]\|_2^2)$. Elementary computations yield for every integers $K' \geq K \geq 1$,

$$
\begin{aligned}
\|Y_h - Y_{h'}\|_2^2 &= \left\| f\left(\frac{1}{K'} \sum_{k=1}^{K'} F(Z_k, Y) \right) - f\left(\frac{1}{K} \sum_{k=1}^{K} F(Z_k, Y) \right) \right\|_2^2 \\
&\leq [f]_{\text{Lip}}^2 \left\| \frac{1}{K'} \sum_{k=1}^{K'} F(Z_k, Y) - \frac{1}{K} \sum_{k=1}^{K} F(Z_k, Y) \right\|_2^2 \\
&= [f]_{\text{Lip}}^2 \left\| \frac{1}{K'} \sum_{k=1}^{K'} \widetilde{X}_k - \frac{1}{K} \sum_{k=1}^{K} \widetilde{X}_k \right\|_2^2 \\
&= [f]_{\text{Lip}}^2 \left\| \frac{M_{K'}}{K'} - \frac{M_K}{K} \right\|_2^2
\end{aligned}
$$

since $\mathbf{E}_Y[F(Z_k, Y)] = \mathbf{E}_Y[F(Z, Y)]$ does not depend on k owing to the independence of $(Z_k)_{k \geq 1}$ and Y. Then elementary computations show that

$$
\left\| \frac{M_{K'}}{K'} - \frac{M_K}{K} \right\|_2^2 = \frac{K' - K}{KK'} \|\widetilde{X}_k\|_2^2 = (h - h')\left(\|X\|_2^2 - \|\mathbf{E}[X|Y]\|_2^2 \right).
$$

The case $h' = 0$ can be treated likewise (or by letting K' go to infinity). $\square$

Lemma 4.6 (Weak error). *Let* $f : \mathbf{R} \to \mathbf{R}$ *be a* $2R$ *times differentiable function with* $f^{(k)}$, $k = R, \ldots, 2R$, *bounded over the real line. Assume* $X \in \mathbf{L}^{2R}(\mathbf{P})$. *Then there exists* $c_1, \ldots, c_{R-1}$ *such that*

$$
\forall h \in \mathcal{H}, \qquad \mathbf{E}[Y_h] = \mathbf{E}[Y_0] + \sum_{r=1}^{R-1} c_r h^r + O(h^R). \tag{51}
$$

Consequently, $(Y_h)_{h \in \mathcal{H}}$ *satisfies* $(WE_{\alpha, \bar{R}})$ *with* $\alpha = 1$ *and* $\bar{R} = R - 1$.

Proof. Let $K \geq 1$ and $\widetilde{X}_k = F(Z_k, Y) - \mathbf{E}_Y[F(Z_k, Y)] = F(Z_k, Y) - Y_0$, $k = 1, \ldots, K$. By the multinomial formula, we get

$$
(\widetilde{X}_1 + \cdots + \widetilde{X}_K)^k = \sum_{k_1 + \cdots + k_K = k} \frac{k!}{k_1! \cdots k_K!} \widetilde{X}_1^{k_1} \cdots \widetilde{X}_K^{k_K}.
$$

Then, taking conditional expectation given Y, yields

$$\mathbf{E}_Y\big[(\widetilde{X}_1 + \cdots + \widetilde{X}_K)^k\big] = k! \sum_{k_1+\cdots+k_K=k} \prod_{i=1}^{K} \frac{\mathbf{E}_Y[\widetilde{X}^{k_i}]}{k_i!}$$

since $\mathbf{E}_Y[\widetilde{X}_i^{k_i}] = \mathbf{E}_Y[\widetilde{X}^{k_i}]$. As $\mathbf{E}_Y[\widetilde{X}_i] = 0$, we obtain

$$\mathbf{E}_Y\big[(\widetilde{X}_1 + \cdots + \widetilde{X}_K)^k\big] = k! \sum_{k_1+\cdots+k_K=k,\,k_i\neq 1} \prod_{i=1}^{K} \frac{\mathbf{E}_Y[\widetilde{X}^{k_i}]}{k_i!}.$$

Let $I = I(k)$ denote the generic set of indices i such that $k_i \neq 0$. It is clear that $1 \le |I| \le k/2$. By symmetry, we have now that

$$\sum_{k_1+\cdots+k_K=k,\,k_i\neq 1} \prod_{i=1}^{K} \frac{\mathbf{E}_Y[\widetilde{X}^{k_i}]}{k_i!}$$

$$= \sum_{1\le \ell \le (k/2)\wedge K} \sum_{\substack{I\subset\{1,\ldots,K\},|I|=\ell, \\ \sum_{i\in I} k_i = k,\,k_i \ge 2}} \prod_{i=1}^{K} \frac{\mathbf{E}_Y[\widetilde{X}^{k_i}]}{k_i!}$$

$$= \sum_{1\le \ell \le k/2} \binom{K}{l} \sum_{\sum_{1\le i\le \ell} k_i = k-2\ell} \prod_{i=1}^{\ell} \frac{\mathbf{E}_Y[\widetilde{X}^{2+k_i}]}{(2+k_i)!}.$$

As a consequence, for every integer $R \ge 1$,

$$\mathbf{E}_Y[Y_h] = \mathbf{E}_Y[Y_0] + \sum_{k=1}^{2R-1} \frac{f^{(k)}(\mathbf{E}_Y[X])}{k!K^k} \mathbf{E}^Y(\widetilde{X}_1 + \cdots + \widetilde{X}_K)^k + \mathbf{R}_{2R-1}(Y)$$

$$= \mathbf{E}_Y[Y_0] + \sum_{k=1}^{2R-1} \frac{f^{(k)}(\mathbf{E}_Y[X])}{k!K^k} \sum_{1\le \ell \le (k/2)\wedge K} \binom{K}{\ell} c_{k,\ell} + \mathbf{R}_{2R-1}(Y),$$

where

$$a_{k,\ell} = \sum_{k_1+\cdots+k_\ell=k-2\ell} \prod_{i=1}^{\ell} \frac{\mathbf{E}_Y[\widetilde{X}^{2+k_i}]}{(2+k_i)!}$$

and

$$\big|\mathbf{R}_{2R-1}(Y)\big| \le \frac{\|f^{(2R)}\|_{\sup}}{(2R)!} \frac{1}{K^{2R}} \mathbf{E}_Y\big[|\widetilde{X}_1 + \cdots + \widetilde{X}_K|^{2R}\big].$$

By the Marcinkiewicz–Zygmund Inequality, we get

$$\left|\mathbf{R}_{2R-1}(Y)\right| \le \left(B_{2R}^{MZ}\right)^{2R} \frac{\|f^{(2R)}\|_{\sup}}{(2R)!} \frac{1}{K^{2R}} \mathbf{E}_Y\left[|\tilde{X}_1^2 + \cdots + \tilde{X}_K^2|^R\right]$$

$$\le \|f^{(2R)}\|_{\sup} \frac{(B_{2R}^{MZ})^{2R}}{(2R)!} \frac{1}{K^R} \mathbf{E}_Y\left[\tilde{X}^{2R}\right],$$

where $B_p^{MZ} = 18 \dfrac{p^{\frac{3}{2}}}{(p-1)^{\frac{1}{2}}}$, $p > 1$ (see [29], page 499). Now, we write the polynomial function $x(x-1)\cdots(x-\ell+1)$ on the canonical basis $1, x, \ldots, x^n, \ldots$ as follows

$$x(x-1)\cdots(x-\ell+1) = \sum_{m=0}^{\ell} b_{\ell,m} x^m \qquad (b_{\ell,\ell} = 1 \text{ and } b_{\ell,0} = 0).$$

Hence,

$$\mathbf{E}_Y[Y_h] = \mathbf{E}_Y[Y_0] + \sum_{k=1}^{2R-1} \sum_{\ell=1}^{\frac{k}{2}} \sum_{m=1}^{\ell} \frac{f^{(k)}(\mathbf{E}_Y[X])}{k!} \frac{1}{K^{k-m}} a_{k,\ell} b_{\ell,m} + O(K^{-R}),$$

where $K^R O(K^{-R})$ is bounded by a deterministic constant. For every $r \in \{1, \ldots, R-1\}$, set

$$J_{R,r} = \left\{(k,l,m) \in \mathbf{N}^3, 1 \le k \le 2R-1, 1 \le \ell \le k/2, 1 \le m \le \ell, k = m+r\right\}$$

(note that one always has $k \ge (2m) \vee 1$ so that $k - m \ge 1$ when k, l, m vary in the admissible index set). We finally get

$$\mathbf{E}_Y[Y_h] = \mathbf{E}_Y[Y_0] + \sum_{r=1}^{2R-1} \left(\sum_{(k,\ell,m)\in J_{R,r}} \frac{f^{(k)}(\mathbf{E}_Y[X])}{k!} a_{k,\ell} b_{\ell,m}\right) \frac{1}{K^r} + O(K^{-R}).$$

$$= \mathbf{E}_Y[Y_0] + \sum_{r=1}^{R-1} \frac{c_r}{K^r} + O(K^{-R}).$$

Taking the expectation in the above equality yields the announced result. □

Remark 4.7. Though it is not the only term included in the final $O(K^{-R})$, it is worth noticing that $\left(\frac{(B_{2R}^{MZ})^{2R}}{(2R)!}\right)^{\frac{1}{R}} \sim (36R)^2 (\frac{2R}{e})^{-2} \sim 18e^2$ as $R \to +\infty$ owing to Stirling's formula. This suggests that, if all the derivatives of f are uniformly bounded, $\limsup_{R\to+\infty} |c_R|^{\frac{1}{R}} < +\infty$.

5. Numerical experiments

5.1. Practitioner's corner

We summarize here the study of the Section 3. We have proved in Theorem 3.6, Proposition 3.9 and Theorem 3.12 that the asymptotic optimal parameters (as ε goes to 0) R, h, q and N depend on structural parameters α, c_1, β, V_1, $\mathrm{var}(Y_0)$ and $\mathbf{h}$ (recall that $\theta = \sqrt{V_1/\mathrm{var}(Y_0)}$). Note that we did not optimize the design matrix $\mathbf{T}$ and the refiners n_i, $i = 2, \ldots, R$.

About structural parameters

Implementing MLMC or ML2R estimator needs to know both the weak and strong rates of convergence of the biased estimator Y_h toward Y_0. The exponents α and β are generally known by a mathematical study of the approximation (see Section 4.1 for Brownian diffusion discretization and Section 4.2 for nested Monte Carlo). The parameter V_1 comes from the strong approximation rate assumption (SE_β) and a natural approximation for V_1 is

$$V_1 \simeq \limsup_{h \to 0} h^{-\beta} \|Y_h - Y_0\|_2^2.$$

Since Y_0 cannot be simulated at a reasonable computational cost, one may proceed as follows to get a good empirical estimator of V_1. First, assume that, in fact, $\|Y_h - Y_0\|_2^2 \sim V_1 h^\beta$ as $h \to 0$ but that this equivalence still holds as an approximation for not too small parameters h. Then, one derives from Minkowski's Inequality that, for every integer $M \geq 1$,

$$\|Y_h - Y_{\frac{h}{M}}\|_2 \leq \|Y_h - Y_0\|_2 + \|Y_0 - Y_{\frac{h}{M}}\|_2$$

so that

$$V_1 \gtrsim \left(1 + M^{-\frac{\beta}{2}}\right)^{-2} h^{-\beta} \|Y_h - Y_{\frac{h}{M}}\|_2^2.$$

As a consequence, if we choose $M = M_{\max}$ large enough (see (56) below), we are led to consider the following estimator

$$\widehat{V}_1(h) = \left(1 + M_{\max}^{-\frac{\beta}{2}}\right)^{-2} h^{-\beta} \|Y_h - Y_{\frac{h}{M_{\max}}}\|_2^2. \tag{52}$$

If the assumption (SE_β) is replaced by one of the alternative assumptions (Var_β) and (Var'_β) proposed in Remark 2.2, the estimator of V_1 must be modified. For instance, if we consider the assumption $\mathrm{var}(Y_h - Y_{h'}) \leq V_1 |h - h'|^\beta$, a standard estimator of V_1 becomes $\hat{V}_1(h) = (1 - M^{-\beta})^{-1} h^{-\beta} \mathrm{var}(Y_h - Y_{\frac{h}{M}})$, M being fixed.

The estimation of the real constants c_i, c_1 for crude Monte Carlo and an MLMC estimators and $\tilde{c} = \lim_{R \to \infty} |c_R|^{\frac{1}{R}}$ for the ML2R estimator is much more challenging. So, these methods are usually implemented in a blind way by considering the coefficients c_1 and $|c_R|^{\frac{1}{R}}$ equal to 1.

Note that, even in a crude Monte Carlo method, such structural parameters are useful (and sometimes necessary) to deal with the bias error (see Proposition 2.3).

Design of the multilevel

The standard design matrix is fixed by the template (ML2R) for the multilevel Richardson–Romberg estimator and by the template (MLMC) for the multilevel Monte Carlo estimator. Alternative choices could be to consider for the ML2R estimator another design matrix $\mathbf{T}$ satisfying (20) like $\mathbf{T}^j = -\mathbf{w}_j e_1 + \mathbf{w}_j e_j$ for $j \in \{2, \ldots, R\}$ which reads

$$\mathbf{T} = \begin{pmatrix} 1 & -\mathbf{w}_2 & -\mathbf{w}_3 & \cdots & -\mathbf{w}_R \\ 0 & \mathbf{w}_2 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{w}_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \mathbf{w}_R \end{pmatrix}. \tag{53}$$

We could also consider a lower triangular design matrix (through it does not satisfy the conventional assumption $T^1 = e_1$)

$$\mathbf{T} = \begin{pmatrix} \widetilde{\mathbf{W}}_1 & 0 & \cdots & \cdots & \cdots & 0 \\ -\widetilde{\mathbf{W}}_1 & \widetilde{\mathbf{W}}_2 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & -\widetilde{\mathbf{W}}_{R-2} & \widetilde{\mathbf{W}}_{R-1} & 0 \\ 0 & \cdots & \cdots & \cdots & -\widetilde{\mathbf{W}}_{R-1} & 1 \end{pmatrix} \quad \text{where } \widetilde{\mathbf{W}}_j = \sum_{k=1}^{j} \mathbf{w}_k. \tag{54}$$

The refiners can be specified by users but it turns out that the parametrized family $n_i = M^{i-1}$, $i = 1, \ldots, R$ ($M \in \mathbf{N}$, $M \geq 2$) seems the best compromise between variance control and implementability. The parameter α being settled, all the related quantities like $(\mathbf{W}_i(R, M))_{1 \leq i \leq M}$ can be tabulated for various values of M and R and can be stored offline.

Taking advantage of $c_1 = 0$

When $c_1 = 0$, only $R - 1$ weights are needed to cancel the (remaining) coefficients up to order R that is, c_r, $r = 2, \ldots, R - 1$ (instead of R). One easily shows that, if $(\mathbf{w}_r^{(R-1)})_{r=1,\ldots,R-1}$ denotes the weight vector at order $R - 1$ associated to refiners $n_1 = 1 < n_2, \ldots, n_{R-1}$ (for a given α), then the weight vector $\widetilde{\mathbf{w}}^{(R)}$ at order R (with size $R - 1$) reads

$$\widetilde{\mathbf{w}}_r^{(R)} = \frac{n_r^\alpha \mathbf{w}_r^{(R-1)}}{\sum_{1 \leq s \leq R-1} n_s^\alpha \mathbf{w}_s^{(R-1)}}, \qquad r = 1, \ldots, R - 1.$$

Optimal parameters

Diffusion approximation. In the case $n_i = M^{i-1}$ (with the convention $n_0 = n_0^{-1} = 0$), we can summarize the asymptotic optimal value of the parameters q, R, h and N in Table 1 for the (ML2R) estimator and in Table 2 for the (MLMC) estimator.

Table 1. Optimal parameters for the ML2R estimator (standard case)

$R(\varepsilon)$	$\left\lceil \frac{1}{2} + \frac{\log(\tilde{c}^{\frac{1}{\alpha}} \mathbf{h})}{\log(M)} + \sqrt{\left(\frac{1}{2} + \frac{\log(\tilde{c}^{\frac{1}{\alpha}} \mathbf{h})}{\log(M)}\right)^2 + 2\frac{\log(A/\varepsilon)}{\alpha \log(M)}} \right\rceil, \quad A = \sqrt{1+4\alpha}$
$h(\varepsilon)$	$\mathbf{h} / \left\lceil (1+2\alpha R)^{\frac{1}{2\alpha R}} \varepsilon^{-\frac{1}{\alpha R}} \tilde{c}^{\frac{1}{\alpha}} M^{-\frac{R-1}{2}} \mathbf{h} \right\rceil$
$q(\varepsilon)$	$q_1 = \mu_R^* \left(1 + \theta h^{\frac{\beta}{2}}\right)$
	$q_j = \mu_R^* \theta h^{\frac{\beta}{2}} \left(|\mathbf{W}_j(R,M)| \frac{n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}}{\sqrt{n_{j-1}+n_j}} \right), \quad j = 2, \ldots, R; \quad \mu_R^* \text{ s.t. } \sum_{1 \le j \le R} q_j = 1$
$N(\varepsilon)$	$\left(1 + \frac{1}{2\alpha R}\right) \frac{\text{var}(Y_0)(1 + \theta h^{\frac{\beta}{2}} \sum_{j=1}^R |\mathbf{W}_j(R,M)|(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}})\sqrt{n_{j-1}+n_j})}{\varepsilon^2 \mu_R^*}$

Nested Monte Carlo. In a Nested Monte Carlo framework, the unitary complexity is given by (26).

- The unitary cost term $n_{j-1} + n_j$ in Tables 1 and 2 must be replaced by n_j.
- The unitary variance term $n_{j-1}^{-\beta/2} + n_j^{-\beta/2}$ must be replaced by $(\frac{1}{n_{j-1}} - \frac{1}{n_j})^{\beta/2}$.

Optimization of the root M. Note that these optimal parameters given in the above Tables only depend on the structural parameters and on the user's choice of the root $M \ge 2$ for the refiners. For a fixed $\varepsilon > 0$, if we emphasize the dependance in $M = M(\varepsilon)$ that is, $R(M), h(M), q(M)$ and $N(M)$ the global cost C_ε as a function of M is given by

$$C_\varepsilon(M) = \text{Cost}\left(\bar{Y}_{h(M),\underline{n}}^{N(M),q(M)}\right) = N(M)\kappa\left(h(M), R(M), q(M)\right), \tag{55}$$

where $\kappa(h, R, q) = \frac{1}{h} \sum_{j=1}^R q_j \sum_{i=1}^R n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}$ (in the framework of Section 4.1) and $\kappa(h, R, q) = \frac{1}{h} \sum_{j=1}^R q_j \max_{1 \le i \le R} n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}$ (in the framework of Section 4.2). This function

Table 2. Optimal parameters for the MLMC estimator (standard case)

$R(\varepsilon)$	$\left\lceil 1 + \frac{\log(|c_1|^{\frac{1}{\alpha}} \mathbf{h})}{\log(M)} + \frac{\log(A/\varepsilon)}{\alpha \log(M)} \right\rceil, \quad A = \sqrt{1+2\alpha}$
$h(\varepsilon)$	$\mathbf{h} / \left\lceil (1+2\alpha)^{\frac{1}{2\alpha}} \varepsilon^{-\frac{1}{\alpha}} |c_1|^{\frac{1}{\alpha}} M^{-(R-1)} \mathbf{h} \right\rceil$
$q(\varepsilon)$	$q_1 = \mu_R^* \left(1 + \theta h^{\frac{\beta}{2}}\right)$
	$q_j = \mu_R^* \theta h^{\frac{\beta}{2}} \left(\frac{n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}}{\sqrt{n_{j-1}+n_j}} \right), \quad j = 2, \ldots, R; \quad \mu_R^* \text{ s.t. } \sum_{1 \le j \le R} q_j = 1$
$N(\varepsilon)$	$\left(1 + \frac{1}{2\alpha}\right) \frac{\text{var}(Y_0)(1 + \theta h^{\frac{\beta}{2}} \sum_{j=1}^R (n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}})\sqrt{n_{j-1}+n_j})}{\varepsilon^2 \mu_R^*}$

can be optimized for likely values of M. In our numerical experiments, we consider

$$M = \underset{M \in \{2,...,M_{\max}\}}{\operatorname{argmin}} C_\varepsilon(M) \qquad \text{with } M_{\max} = 10. \tag{56}$$

5.2. Correlation between $Y_{\frac{h}{n_i}}$ and $Y_{\frac{h}{n_{i-1}}}$

Diffusion approximation

In many situations (like e.g., the numerical experiments carried out below), discretization schemes of Brownian diffusions need to be simulated with various steps (say $\frac{T}{nn_i}$ and $\frac{T}{nn_{i+1}}$ in our case). This requires to simulate consistent Brownian increments over $[0, \frac{T}{n}]$, then $[\frac{(k-1)T}{n}, \frac{kT}{n}]$, $k = 2, \ldots, n$. This can be performed by simulating recursively the Brownian increments over all successive sub-intervals of interest, having in mind that the "quantum" size for the simulation is given by $\frac{T}{nm}$ where $m = gcd(n_1, \ldots, n_R)$. This recursive refinement is also known as the Brownian Bridge simulation procedure. One can also produce once and for all an *abacus* of coefficients to compute by induction the needed increments from smaller subintervals up to the root interval of length $\frac{T}{n}$. This is done for example, in [27] up to $R = 5$ for $\alpha = 1$ and up to $R = 3$ for $\alpha = \frac{1}{2}$.

Nested Monte Carlo

In a Nested Monte Carlo, the relation between $Y_{\frac{h}{n_i}}$ and $Y_{\frac{h}{n_{i-1}}}$ is simply based on the following rule: the n_{i-1}/h first terms of the sequence of copies of Z used to simulate $Y_{\frac{h}{n_{i-1}}}$ must be used to simulate $Y_{\frac{h}{n_i}}$.

5.3. Methodology

We compare the two MLMC and ML2R estimators for different biased problems. In the sequel, we consider the standard design matrix (ML2R) for the ML2R estimator, idem for the MLMC estimator. After a crude evaluation of $\operatorname{var}(Y_0)$ and V_1 (using (52)) we compute the "optimal" parameter M solution to (56). The others parameters are specified according to Tables 1 and 2 with $\tilde{c} = c_1 = 1$.

The empirical bias error $\tilde{\mu}_L$ of the estimator $\bar{Y}_{h,\underline{n}}^{N,q}$ is obtained using $L = 256$ independent replications of the estimator, namely

$$\tilde{\mu}_L = \frac{1}{L} \sum_{\ell=1}^{L} (\bar{Y}_{h,\underline{n}}^{N,q})^{(\ell)} - I_0,$$

where $I_0 = \mathbf{E}[Y_0]$ is the true value.

The empirical $\mathbf{L}^2$-error or empirical root mean squared error (RMSE) $\tilde{\varepsilon}_L$ of the estimator used in our numerical experiments is given by

$$\tilde{\varepsilon}_L = \sqrt{\frac{1}{L} \sum_{\ell=1}^{L} \left((\bar{Y}_{h,\underline{n}}^{N,q})^{(\ell)} - I_0 \right)^2}. \tag{57}$$

The computations were performed on a computer with 4 multithreaded(16) octo-core processors (Intel(R) Xeon(R) CPU E5-4620 @ 2.20GHz). The code of one estimator runs on a single thread (program in `C++11` available on request).

5.4. Euler scheme of a geometric Brownian motion: Pricing of European options

We consider a geometric Brownian motion $(S_t)_{t \in [0,T]}$, representative in a Black–Scholes model of the dynamics of a risky asset price between time $t = 0$ and time $t = T$:

$$S_t = s_0 e^{(r - \frac{\sigma^2}{2})t + \sigma W_t}, \qquad t \in [0, T], S_0 = s_0 > 0,$$

where r denotes the (constant) "riskless" interest rate, σ denotes the volatility and $W = (W_t)_{t \in [0,T]}$ is a standard Brownian motion defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. The price or *premium* of a so-called *vanilla* option with payoff φ is given by $e^{-rT} \mathbf{E}[\varphi(S_T)]$ and the price of a *path dependent* option with functional payoff φ is given by $e^{-rT} \mathbf{E}[\varphi((S_t)_{t \in [0,T]})]$. Since $(S_t)_{t \in [0,T]}$ is solution to the diffusion *SDE*

$$\mathrm{d}S_t = S_t (r \, \mathrm{d}t + \sigma \, \mathrm{d}W_t), \qquad S_0 = s_0 > 0,$$

one can compute the price of an option by a Monte Carlo simulation in which the true process $(S_t)_{t \in [0,T]}$ is replaced by its Euler scheme $(\bar{S}_{kh})_{0 \le k \le n}$, $h = \frac{T}{n}$ (even if we are aware that S_T can be simulated). The bias parameter set $\mathcal{H}$ is then defined by $\mathcal{H} = \{T/n, n \ge 1\}$ and $\mathbf{h} = T$.

Although nobody would adopt any kind of Monte Carlo simulation to compute option price in this model since a standard difference method on the Black–Scholes parabolic PDE is much more efficient to evaluate a vanilla option and many path-dependent ones, it turns out that the time discretization of a Black–Scholes model and its Euler scheme is a very demanding *benchmark* to test and evaluate the performances of Monte Carlo method(s). As a consequence, it is quite appropriate to carry out numerical tests with ML2R and MLMC.

5.4.1. *Vanilla call option ($\alpha = \beta = 1$)*

The Black–Scholes parameters considered here are $s_0 = 100$, $r = 0.06$ and $\sigma = 0.4$. The payoff is a European *Call* with maturity $T = 1$ year and strike $K = 80$.

In such a regular diffusion setting (both drift and diffusion coefficients are $\mathcal{C}_b^\infty$ and the payoff function is Lipschitz continuous), one has $\alpha = \beta = 1$. The parameters $\theta = \sqrt{V_1 / \mathrm{var}(Y_0)}$ and

var(Y_0) have been roughly estimated following the procedure (52) on a sample of size 100 000 described in Section 5.1, leading to the values $V_1 \simeq 56$ and var(Y_0) $\simeq 876$ (so that $\theta \simeq 0.25$). The empirical $\mathbf{L}^2$-error $\tilde{\epsilon}_L$ is estimated using $L = 256$ runs of the algorithm and the bias is computed using the true value of the price $I_0 = 29.4987$ provided by the Black–Scholes formula.

The results are summarized in Table 3 for the ML2R estimator and in Table 4 for the MLMC estimator.

As an example, the third line of the Table 3 reads as follows: for a prescribed RMSE error $\varepsilon = 2^{-3} = 0.125$, the ML2R estimator $\bar{Y}_{h,\underline{n}}^{N,q}$ (with design matrix (ML2R)) is implemented with the parameters $R = 3$, $h = 1$ and refiners $n_i = 4^{i-1}$ (then $n_1 = 1$, $n_2 = 4$ and $n_3 = 16$) and the sample size $N \simeq 319\,000$. The allocation weights q_i (not reported in this table) are such that the numerical cost $\text{Cost}(\bar{Y}_{h,\underline{n}}^{N,q}) \simeq 709\,800$. For such parameters, the empirical RMSE $\tilde{\epsilon}_L \simeq 0.0928$ and the computational time of $Y_{h,\underline{n}}^{N,q} \simeq 0.559$ seconds. The empirical bias error $\tilde{\mu}_L$ is reported in the 5th column (*bias*) and the empirical unitary variance $\tilde{v}_L$ is reported in the 6th column (*variance*). Recall that $\tilde{\epsilon}_L = \sqrt{(\tilde{\mu}_L)^2 + \tilde{v}_L}$.

Note first that, as expected, the depth parameter $R \geq 2$ and the numerical cost $\text{Cost}(\bar{Y}_{h,\underline{n}}^{N,q})$ grow slower for ML2R than for MLMC as ε goes to 0. Consequently, regarding the CPU-time for a prescribed error $\varepsilon = 2^{-k}$, ML2R is about 10% to 100% (twice) faster than MLMC when k goes from 2 to 8. On the other hand, both estimators ML2R and MLMC provide an empirical RMSE close to the prescribed RMSE i.e. $\tilde{\epsilon}_L \leq \varepsilon$. We can conclude that the automatic tuning of the algorithm parameters is satisfactory for both estimators.

In Figure 1(a) is depicted the CPU-time (4th column) as a function of the *empirical* $\mathbf{L}^2$-error (3rd column). It provides a direct comparison of the performance of both estimators. Each point is labeled by the prescribed RMSE $\varepsilon = 2^{-k}$, $k = 1, \ldots, 8$ for easy reading. The plot is in $\log_2$–log scale. The ML2R estimator (blue solid line) is below the MLMC estimator (red dashed line). The ratio of CPU-times for a given $\tilde{\epsilon}_L$ shows that ML2R goes from 1.28 up to 2.8 faster, within the range of our simulations. Figure 1(b) represents the product (CPU-time) $\times \varepsilon^2$ as a function of ε.

5.4.2. *Lookback option* ($\alpha = 0.5$, $\beta = 1$)

We consider a partial Lookback Call option defined by its functional payoff

$$\varphi(x) = e^{-rT}\left(x(T) - \lambda \min_{t \in [0,T]} x(t)\right)_+, \qquad x \in \mathcal{C}([0,T], \mathbf{R}),$$

where $\lambda \geq 1$. The parameters of the Black–Scholes model are $s_0 = 100$, $r = 0.15$, $\sigma = 0.1$ and $T = 1$ and the coefficient λ is set at $\lambda = 1.1$. For these parameters, the price given by a closed-form expression is $I_0 = 8.89343$.

For such payoff with Lipschitz continuous functional, (SE$_\beta$) holds with $\beta = 1$ and (WE$_{\alpha,\bar{R}}$) holds with $\alpha = 0.5$. Note that the full expansion $\bar{R} = +\infty$ is not yet proved to our knowledge. An estimation of structural parameters yields var(Y_0) $\simeq 41$ and $V_1 \simeq 3.58$ (and then $\theta \simeq 0.29$). Both estimators are implemented using the automatic tuning previously exposed.

Table 3. Call option ($\alpha = 1$, $\beta = 1$): Parameters and results of the ML2R estimator

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$-error	Time (s)	Bias	Variance	R	M	h^{-1}	N	Cost
1	$5.00 \cdot 10^{-01}$	$3.91 \cdot 10^{-01}$	$3.02 \cdot 10^{-02}$	$1.47 \cdot 10^{-01}$	$1.31 \cdot 10^{-01}$	2	5	1	$1.50 \cdot 10^{+04}$	$2.47 \cdot 10^{+04}$
2	$2.50 \cdot 10^{-01}$	$2.18 \cdot 10^{-01}$	$1.12 \cdot 10^{-01}$	$8.99 \cdot 10^{-02}$	$3.96 \cdot 10^{-02}$	2	9	1	$5.91 \cdot 10^{+04}$	$1.06 \cdot 10^{+05}$
3	$1.25 \cdot 10^{-01}$	$9.28 \cdot 10^{-02}$	$5.59 \cdot 10^{-01}$	$-5.61 \cdot 10^{-04}$	$8.62 \cdot 10^{-03}$	3	4	1	$3.19 \cdot 10^{+05}$	$7.09 \cdot 10^{+05}$
4	$6.25 \cdot 10^{-02}$	$5.01 \cdot 10^{-02}$	$2.12 \cdot 10^{+00}$	$-1.90 \cdot 10^{-02}$	$2.15 \cdot 10^{-03}$	3	4	1	$1.27 \cdot 10^{+06}$	$2.84 \cdot 10^{+06}$
5	$3.12 \cdot 10^{-02}$	$2.71 \cdot 10^{-02}$	$8.13 \cdot 10^{+00}$	$-1.15 \cdot 10^{-02}$	$6.00 \cdot 10^{-04}$	3	5	1	$4.99 \cdot 10^{+06}$	$1.15 \cdot 10^{+07}$
6	$1.56 \cdot 10^{-02}$	$1.35 \cdot 10^{-02}$	$3.22 \cdot 10^{+01}$	$-4.41 \cdot 10^{-03}$	$1.63 \cdot 10^{-04}$	3	6	1	$1.99 \cdot 10^{+07}$	$4.72 \cdot 10^{+07}$
7	$7.81 \cdot 10^{-03}$	$6.98 \cdot 10^{-03}$	$1.31 \cdot 10^{+02}$	$-2.32 \cdot 10^{-03}$	$4.33 \cdot 10^{-05}$	3	7	1	$7.98 \cdot 10^{+07}$	$1.95 \cdot 10^{+08}$
8	$3.91 \cdot 10^{-03}$	$3.57 \cdot 10^{-03}$	$5.51 \cdot 10^{+02}$	$-9.35 \cdot 10^{-04}$	$1.19 \cdot 10^{-05}$	3	9	1	$3.25 \cdot 10^{+08}$	$8.37 \cdot 10^{+08}$

Table 4. Call option ($\alpha = 1$, $\beta = 1$): Parameters and results of the MLMC estimator

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$-error	Time (s)	Bias	Variance	R	M	h^{-1}	N	Cost
1	$5.00 \cdot 10^{-01}$	$5.02 \cdot 10^{-01}$	$2.53 \cdot 10^{-02}$	$3.87 \cdot 10^{-01}$	$1.02 \cdot 10^{-01}$	2	4	1	$1.57 \cdot 10^{+04}$	$2.32 \cdot 10^{+04}$
2	$2.50 \cdot 10^{-01}$	$2.85 \cdot 10^{-01}$	$1.31 \cdot 10^{-01}$	$2.25 \cdot 10^{-01}$	$3.04 \cdot 10^{-02}$	2	7	1	$6.48 \cdot 10^{+04}$	$1.06 \cdot 10^{+05}$
3	$1.25 \cdot 10^{-01}$	$1.20 \cdot 10^{-01}$	$6.28 \cdot 10^{-01}$	$8.77 \cdot 10^{-02}$	$6.63 \cdot 10^{-03}$	3	4	1	$3.64 \cdot 10^{+05}$	$7.33 \cdot 10^{+05}$
4	$6.25 \cdot 10^{-02}$	$6.31 \cdot 10^{-02}$	$2.44 \cdot 10^{+00}$	$4.45 \cdot 10^{-02}$	$2.00 \cdot 10^{-03}$	3	6	1	$1.49 \cdot 10^{+06}$	$3.32 \cdot 10^{+06}$
5	$3.12 \cdot 10^{-02}$	$3.42 \cdot 10^{-02}$	$1.05 \cdot 10^{+01}$	$2.48 \cdot 10^{-02}$	$5.59 \cdot 10^{-04}$	3	8	1	$6.15 \cdot 10^{+06}$	$1.47 \cdot 10^{+07}$
6	$1.56 \cdot 10^{-02}$	$1.66 \cdot 10^{-02}$	$5.17 \cdot 10^{+01}$	$1.23 \cdot 10^{-02}$	$1.22 \cdot 10^{-04}$	4	5	1	$3.06 \cdot 10^{+07}$	$8.38 \cdot 10^{+07}$
7	$7.81 \cdot 10^{-03}$	$7.83 \cdot 10^{-03}$	$2.20 \cdot 10^{+02}$	$5.06 \cdot 10^{-03}$	$3.57 \cdot 10^{-05}$	4	7	1	$1.27 \cdot 10^{+08}$	$3.82 \cdot 10^{+08}$
8	$3.91 \cdot 10^{-03}$	$4.48 \cdot 10^{-03}$	$9.14 \cdot 10^{+02}$	$3.26 \cdot 10^{-03}$	$9.43 \cdot 10^{-06}$	4	8	1	$5.17 \cdot 10^{+08}$	$1.62 \cdot 10^{+09}$

(a) CPU–time (y–axis, log scale) as a function of $\tilde{\varepsilon}_L$ (x–axis, $\log_2$ scale).

(b) CPU–time $\times \varepsilon^2$ (y–axis) as a function of ε (x–axis, $\log_2$ scale).
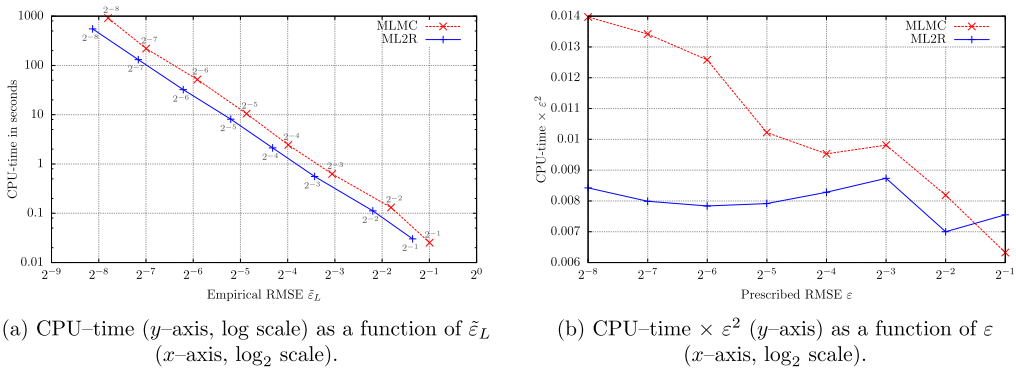
Figure 1. Call option in a Black–Scholes model.

The results are summarized in Table 5 for the ML2R and in Table 6 for the MLMC. Note first that as a function of the prescribed $\varepsilon = 2^{-k}$ the ratio between CPU-times goes from 1.1 ($k = 2$) up to 3.5 ($k = 9$), as does the ratio Cost(MLMC)/Cost(ML2R). However, the empirical RMSE of MLMC is greater than ε (certainly because $c_1 \neq 1$) unlike that of ML2R. One observes that the L^2-error of ML2R has a very small bias $\tilde{\mu}_L$ (5th column) due to the particular choice of the weights $(\mathbf{W}_i)_{1 \leq i \leq R}$.

Figure 2(a) provides a graphical representation of the performance of both estimators, now as a function of the empirical RMSE $\tilde{\varepsilon}$. It shows that ML2R is faster then MLMC by a factor that goes from 18 up to 48 within the range of our simulations.

5.4.3. *Barrier option* ($\alpha = 0.5$, $\beta = 0.5$)

We consider now an up-and-out call option to illustrate the case $\beta = 0.5 < 1$ and $\alpha = 0.5$. This path-dependent option with strike K and barrier $B > K$ is defined by its functional payoff

$$\varphi(x) = e^{-rT} \left(x(T) - K \right)_+ \mathbf{1}_{\{\max_{t \in [0,T]} x(t) \leq B\}}, \qquad x \in \mathcal{C}\left([0, T], \mathbf{R}\right).$$

The parameters of the Black–Scholes model are $s_0 = 100$, $r = 0$, $\sigma = 0.15$ and $T = 1$. With $K = 100$ and $B = 120$, the price computed by closed-form solution is $I_0 = 1.855225$.

We consider here a simple (and highly biased) approximation of $\max_{t \in [0,T]} S_t$ by $\max_{k \in \{1, \ldots, n\}} \bar{S}_{kh}$. This allows us to compare both estimators in the case $\beta = 0.5$. Like in the Lookback option, we assume that $(\text{WE}_{\alpha, \bar{R}})$ holds with $\alpha = 0.5$ and $\bar{R} = +\infty$. A first computational stage gives us $\text{var}(Y_0) \simeq 303$, $V_1 \simeq 5.30$ and $\theta \simeq 0.41$.

The results are summarized in Table 7 for ML2R and in Table 8 for MLMC.

See Figure 3 for a graphical representation. Note that since $\beta = 0.5$, we observe that the function (CPU-time) $\times \varepsilon^2$ increases much faster for MLMC than ML2R as ε goes to 0 which agrees with the theoretical asymptotic rates from Theorem 3.12. In fact, in this highly biased example with slow strong convergence rate, the ratio Cost(MLMC)/Cost(ML2R) as a function of the prescribed $\varepsilon = 2^{-k}$ goes from 1.1 ($k = 2$) up to 22 ($k = 8$), likewise the ratio between

Table 5. Lookback option ($\alpha = 0.5$, $\beta = 1$): Parameters and results of the ML2R estimator

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$-error	Time (s)	Bias	Variance	R	M	h^{-1}	N	Cost
1	$5.00 \cdot 10^{-01}$	$3.54 \cdot 10^{-01}$	$2.42 \cdot 10^{-03}$	$-5.80 \cdot 10^{-02}$	$1.22 \cdot 10^{-01}$	3	6	1	$1.46 \cdot 10^{+03}$	$4.40 \cdot 10^{+03}$
2	$2.50 \cdot 10^{-01}$	$1.80 \cdot 10^{-01}$	$1.04 \cdot 10^{-02}$	$-3.66 \cdot 10^{-02}$	$3.10 \cdot 10^{-02}$	3	6	1	$5.82 \cdot 10^{+03}$	$1.76 \cdot 10^{+04}$
3	$1.25 \cdot 10^{-01}$	$9.95 \cdot 10^{-02}$	$4.17 \cdot 10^{-02}$	$-3.98 \cdot 10^{-02}$	$8.31 \cdot 10^{-03}$	3	7	1	$2.30 \cdot 10^{+04}$	$7.07 \cdot 10^{+04}$
4	$6.25 \cdot 10^{-02}$	$5.45 \cdot 10^{-02}$	$1.53 \cdot 10^{-01}$	$-9.53 \cdot 10^{-03}$	$2.88 \cdot 10^{-03}$	3	10	2	$6.48 \cdot 10^{+04}$	$3.55 \cdot 10^{+05}$
5	$3.12 \cdot 10^{-02}$	$2.31 \cdot 10^{-02}$	$8.69 \cdot 10^{-01}$	$-1.50 \cdot 10^{-03}$	$5.33 \cdot 10^{-04}$	4	5	1	$4.50 \cdot 10^{+05}$	$1.68 \cdot 10^{+06}$
6	$1.56 \cdot 10^{-02}$	$1.22 \cdot 10^{-02}$	$3.43 \cdot 10^{+00}$	$-8.49 \cdot 10^{-04}$	$1.47 \cdot 10^{-04}$	4	6	1	$1.77 \cdot 10^{+06}$	$6.74 \cdot 10^{+06}$
7	$7.81 \cdot 10^{-03}$	$6.31 \cdot 10^{-03}$	$1.39 \cdot 10^{+01}$	$-2.76 \cdot 10^{-04}$	$3.98 \cdot 10^{-05}$	4	7	1	$7.03 \cdot 10^{+06}$	$2.74 \cdot 10^{+07}$
8	$3.91 \cdot 10^{-03}$	$3.34 \cdot 10^{-03}$	$5.74 \cdot 10^{+01}$	$1.19 \cdot 10^{-04}$	$1.11 \cdot 10^{-05}$	4	9	1	$2.83 \cdot 10^{+07}$	$1.16 \cdot 10^{+08}$
9	$1.95 \cdot 10^{-03}$	$1.80 \cdot 10^{-03}$	$2.10 \cdot 10^{+02}$	$1.08 \cdot 10^{-04}$	$3.23 \cdot 10^{-06}$	4	10	2	$7.88 \cdot 10^{+07}$	$5.45 \cdot 10^{+08}$

Table 6. Lookback option ($\alpha = 0.5$, $\beta = 1$): Parameters and results of the MLMC estimator

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$-error	Time (s)	Bias	Variance	R	M	h^{-1}	N	Cost
1	$5.00 \cdot 10^{-01}$	$1.35 \cdot 10^{+00}$	$1.47 \cdot 10^{-03}$	$-1.32 \cdot 10^{+00}$	$6.60 \cdot 10^{-02}$	2	8	1	$1.17 \cdot 10^{+03}$	$2.05 \cdot 10^{+03}$
2	$2.50 \cdot 10^{-01}$	$6.86 \cdot 10^{-01}$	$1.13 \cdot 10^{-02}$	$-6.72 \cdot 10^{-01}$	$1.87 \cdot 10^{-02}$	3	6	1	$6.80 \cdot 10^{+03}$	$1.61 \cdot 10^{+04}$
3	$1.25 \cdot 10^{-01}$	$3.00 \cdot 10^{-01}$	$6.27 \cdot 10^{-02}$	$-2.91 \cdot 10^{-01}$	$5.37 \cdot 10^{-03}$	4	6	1	$3.59 \cdot 10^{+04}$	$1.11 \cdot 10^{+05}$
4	$6.25 \cdot 10^{-02}$	$1.96 \cdot 10^{-01}$	$2.73 \cdot 10^{-01}$	$-1.92 \cdot 10^{-01}$	$1.57 \cdot 10^{-03}$	4	8	1	$1.49 \cdot 10^{+05}$	$5.04 \cdot 10^{+05}$
5	$3.12 \cdot 10^{-02}$	$9.25 \cdot 10^{-02}$	$1.46 \cdot 10^{+00}$	$-9.03 \cdot 10^{-02}$	$4.04 \cdot 10^{-04}$	5	7	1	$7.26 \cdot 10^{+05}$	$2.93 \cdot 10^{+06}$
6	$1.56 \cdot 10^{-02}$	$4.38 \cdot 10^{-02}$	$6.80 \cdot 10^{+00}$	$-4.25 \cdot 10^{-02}$	$1.20 \cdot 10^{-04}$	5	10	1	$3.10 \cdot 10^{+06}$	$1.40 \cdot 10^{+07}$
7	$7.81 \cdot 10^{-03}$	$2.47 \cdot 10^{-02}$	$3.26 \cdot 10^{+01}$	$-2.42 \cdot 10^{-02}$	$2.87 \cdot 10^{-05}$	6	8	1	$1.42 \cdot 10^{+07}$	$7.17 \cdot 10^{+07}$
8	$3.91 \cdot 10^{-03}$	$9.06 \cdot 10^{-03}$	$1.72 \cdot 10^{+02}$	$-8.64 \cdot 10^{-03}$	$7.49 \cdot 10^{-06}$	7	8	1	$6.62 \cdot 10^{+07}$	$3.89 \cdot 10^{+08}$
9	$1.95 \cdot 10^{-03}$	$6.16 \cdot 10^{-03}$	$7.34 \cdot 10^{+02}$	$-6.00 \cdot 10^{-03}$	$1.97 \cdot 10^{-06}$	7	9	1	$2.71 \cdot 10^{+08}$	$1.66 \cdot 10^{+09}$
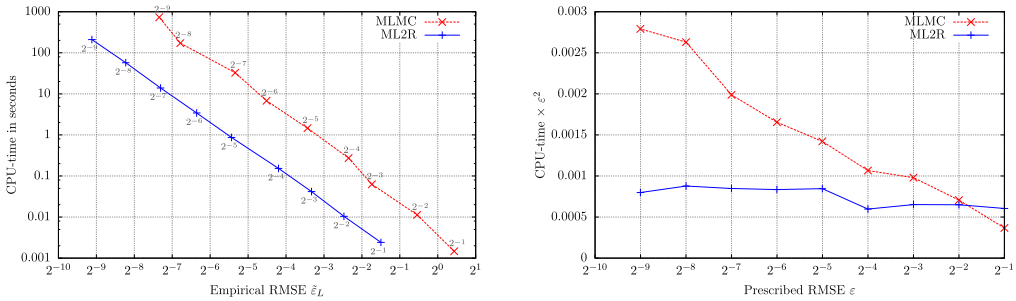
(a) CPU–time (y–axis, log scale) as a function of $\tilde{\varepsilon}_L$ (x–axis, $\log_2$ scale).

(b) CPU–time $\times \varepsilon^2$ (y–axis) as a function of ε (x–axis, $\log_2$ scale).

Figure 2. Lookback option in a Black–Scholes model.

CPU-times behaves. When looking at this ratio as a function of the empirical RMSE, it even goes from 3 up to 61 which is huge having in mind that MLMC provides similar gains with respect to a crude Monte Carlo simulation.

5.5. Nested Monte Carlo: Compound option pricing ($\alpha = \beta = 1$)

A compound option is simply an option on an option. The payoff of a compound option involves the value of another option. A compound option has then two expiration dates $T_1 < T_2$ and two strike prices K_1 and K_2. We consider here the example of a European style *Put* on a *Call* where the underlying risky asset S is still given by a Black–Scholes process with parameters (r, σ). At the first expiration date T_1, the holder has the right to sell a new *Call* option at the strike price K_1. The new *Call* has expiration date T_2 and strike price K_2. The payoff of such a *Put*-on-*Call* option writes

$$\big(K_1 - \mathbf{E}\big[(S_{T_2} - K_2)_+ | S_{T_1}\big]\big)_+.$$

To comply with the multilevel framework, we set $\mathcal{H} = \{1/K, K \geq 1\}$,

$$Y_0 = f\big(\mathbf{E}\big[(S_{T_2} - K_2)_+ | S_{T_1}\big]\big), \qquad Y_{\frac{1}{K}} = f\left(\frac{1}{K} \sum_{k=1}^{K} \big(F\big(Z^k, S_{T_1}\big) - K_2\big)_+\right),$$

where $(Z^k)_{k \geq 1}$ is an i.i.d. sequence of standard Gaussian $\mathcal{N}(0; 1)$, $f(x) = (K_1 - x)_+$ and F is such that

$$S_{T_2} = F(G, S_{T_1}) = S_{T_1} e^{(r - \frac{\sigma^2}{2})(T_2 - T_1) + \sigma \sqrt{T_2 - T_1} Z}.$$

Note that, in this experiments, the underlying process $(S_t)_{t \in [0, T_2]}$ is not discretized in time. The bias error is exclusively due to the inner Monte Carlo estimator of the conditional expectation.

Table 7. Barrier option ($\alpha = 0.5$, $\beta = 0.5$): Parameters and results of the ML2R estimator

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$-error	Time (s)	Bias	Variance	R	M	h^{-1}	N	Cost
1	$5.00 \cdot 10^{-01}$	$3.85 \cdot 10^{-01}$	$6.07 \cdot 10^{-03}$	$-3.92 \cdot 10^{-02}$	$1.46 \cdot 10^{-01}$	3	4	1	$2.65 \cdot 10^{+03}$	$1.17 \cdot 10^{+04}$
2	$2.50 \cdot 10^{-01}$	$1.94 \cdot 10^{-01}$	$2.29 \cdot 10^{-02}$	$-3.82 \cdot 10^{-02}$	$3.62 \cdot 10^{-02}$	3	4	1	$1.06 \cdot 10^{+04}$	$4.66 \cdot 10^{+04}$
3	$1.25 \cdot 10^{-01}$	$1.14 \cdot 10^{-01}$	$9.65 \cdot 10^{-02}$	$-2.00 \cdot 10^{-02}$	$1.26 \cdot 10^{-02}$	3	7	1	$4.02 \cdot 10^{+04}$	$2.07 \cdot 10^{+05}$
4	$6.25 \cdot 10^{-02}$	$6.28 \cdot 10^{-02}$	$5.05 \cdot 10^{-01}$	$-5.45 \cdot 10^{-03}$	$3.92 \cdot 10^{-03}$	3	10	2	$1.34 \cdot 10^{+05}$	$1.44 \cdot 10^{+06}$
5	$3.12 \cdot 10^{-02}$	$2.83 \cdot 10^{-02}$	$3.05 \cdot 10^{+00}$	$1.24 \cdot 10^{-03}$	$8.01 \cdot 10^{-04}$	4	5	1	$1.01 \cdot 10^{+06}$	$7.94 \cdot 10^{+06}$
6	$1.56 \cdot 10^{-02}$	$1.49 \cdot 10^{-02}$	$1.31 \cdot 10^{+01}$	$6.98 \cdot 10^{-04}$	$2.22 \cdot 10^{-04}$	4	6	1	$4.15 \cdot 10^{+06}$	$3.54 \cdot 10^{+07}$
7	$7.81 \cdot 10^{-03}$	$7.81 \cdot 10^{-03}$	$5.79 \cdot 10^{+01}$	$7.82 \cdot 10^{-04}$	$6.03 \cdot 10^{-05}$	4	7	1	$1.71 \cdot 10^{+07}$	$1.58 \cdot 10^{+08}$
8	$3.91 \cdot 10^{-03}$	$4.13 \cdot 10^{-03}$	$2.77 \cdot 10^{+02}$	$-2.01 \cdot 10^{-05}$	$1.71 \cdot 10^{-05}$	4	9	1	$7.39 \cdot 10^{+07}$	$7.81 \cdot 10^{+08}$

Table 8. Barrier option ($\alpha = 0.5$, $\beta = 0.5$): Parameters and results of the MLMC estimator

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$-error	Time (s)	Bias	Variance	R	M	h^{-1}	N	Cost
1	$5.00 \cdot 10^{-01}$	$7.83 \cdot 10^{-01}$	$2.26 \cdot 10^{-03}$	$7.25 \cdot 10^{-01}$	$8.73 \cdot 10^{-02}$	2	8	1	$1.36 \cdot 10^{+03}$	$2.83 \cdot 10^{+03}$
2	$2.50 \cdot 10^{-01}$	$4.03 \cdot 10^{-01}$	$2.05 \cdot 10^{-02}$	$3.67 \cdot 10^{-01}$	$2.75 \cdot 10^{-02}$	3	6	1	$1.03 \cdot 10^{+04}$	$3.57 \cdot 10^{+04}$
3	$1.25 \cdot 10^{-01}$	$1.81 \cdot 10^{-01}$	$1.83 \cdot 10^{-01}$	$1.56 \cdot 10^{-01}$	$8.30 \cdot 10^{-03}$	4	6	1	$7.18 \cdot 10^{+04}$	$4.28 \cdot 10^{+05}$
4	$6.25 \cdot 10^{-02}$	$1.09 \cdot 10^{-01}$	$9.52 \cdot 10^{-01}$	$9.71 \cdot 10^{-02}$	$2.47 \cdot 10^{-03}$	4	8	1	$3.27 \cdot 10^{+05}$	$2.40 \cdot 10^{+06}$
5	$3.12 \cdot 10^{-02}$	$5.33 \cdot 10^{-02}$	$8.38 \cdot 10^{+00}$	$4.70 \cdot 10^{-02}$	$6.27 \cdot 10^{-04}$	5	7	1	$2.11 \cdot 10^{+06}$	$2.40 \cdot 10^{+07}$
6	$1.56 \cdot 10^{-02}$	$2.61 \cdot 10^{-02}$	$6.16 \cdot 10^{+01}$	$2.22 \cdot 10^{-02}$	$1.88 \cdot 10^{-04}$	5	10	1	$1.09 \cdot 10^{+07}$	$1.74 \cdot 10^{+08}$
7	$7.81 \cdot 10^{-03}$	$1.41 \cdot 10^{-02}$	$4.90 \cdot 10^{+02}$	$1.23 \cdot 10^{-02}$	$4.51 \cdot 10^{-05}$	6	8	1	$6.40 \cdot 10^{+07}$	$1.43 \cdot 10^{+09}$
8	$3.91 \cdot 10^{-03}$	$5.58 \cdot 10^{-03}$	$6.05 \cdot 10^{+03}$	$4.43 \cdot 10^{-03}$	$1.15 \cdot 10^{-05}$	7	8	1	$4.37 \cdot 10^{+08}$	$1.67 \cdot 10^{+10}$

(a) CPU–time (y–axis, log scale) as a function of $\tilde{\varepsilon}_L$ (x–axis, $\log_2$ scale).

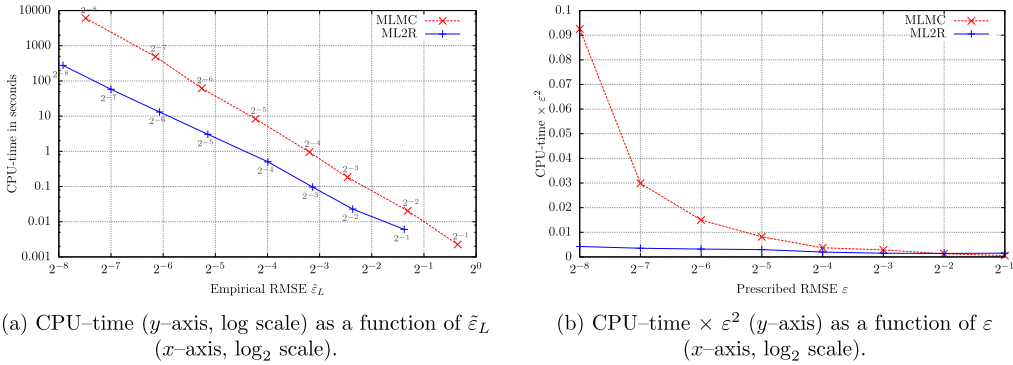(b) CPU–time $\times\ \varepsilon^2$ (y–axis) as a function of ε (x–axis, $\log_2$ scale).

Figure 3. Barrier option in a Black–Scholes model.

The parameters used for the underlying process $(S_t)_{t\in[0,T_2]}$ are $S_0 = 100$, $r = 0.03$ and $\sigma = 0.3$. The parameters of the *Put-on-Call* payoff are $T_1 = 1/12$, $T_2 = 1/2$ and $K_1 = 6.5$, $K_2 = 100$. Section 4.2 strongly suggests that (SE$_\beta$) and (WE$_{\alpha, \bar{R}}$) are satisfied with $\beta = \alpha = 1$. A crude computation of other structural parameters yields $\mathrm{var}(Y_0) \simeq 9.09$, $V_1 \simeq 7.20$ and $\theta \simeq 0.89$.

The results are summarized in Table 9 for ML2R and in Table 10 for MLMC.

Note on Figure 4 that ML2R is faster than MLMC as a function of the empirical RMSE by a factor approximately equal to 5 within the range of our simulations.

Appendix A

Lemma A.1. (a) *The solution of the system* $V\mathbf{w} = e_1$ *where* V *is a Vandermonde matrix*

$$V = V\left(1, n_2^{-\alpha}, \ldots, n_R^{-\alpha}\right) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & n_2^{-\alpha} & \cdots & n_R^{-\alpha} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & n_2^{-\alpha(R-1)} & \cdots & n_R^{-\alpha(R-1)} \end{pmatrix},$$

is given by $\mathbf{w}_i = \dfrac{(-1)^{R-i} n_i^{\alpha(R-1)}}{\prod_{1\leq j<i}(n_i^\alpha - n_j^\alpha) \prod_{i<j\leq R}(n_j^\alpha - n_i^\alpha)}.$

(b) *Furthermore,*

$$\widetilde{\mathbf{w}}_{R+1} = \sum_{i=1}^{R} \frac{\mathbf{w}_i}{n_i^{\alpha R}} = \frac{(-1)^{R-1}}{\prod_{1\leq i\leq R} n_i^\alpha}.$$

Proof. (a) Let $a_i = n_i^{-\alpha}$. Note that by Cramer's rule the solution of this linear system is given by $\mathbf{w}_i = \frac{\det(V_i)}{\det(V)}$ where V_i is the matrix formed by replacing the ith column of V

Table 9. Nested compound option ($\alpha = 1$, $\beta = 1$): Parameters and results of the ML2R estimator

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$-error	Time (s)	Bias	Variance	R	M	h^{-1}	N	Cost
1	$5.00 \cdot 10^{-01}$	$4.36 \cdot 10^{-01}$	$8.82 \cdot 10^{-04}$	$3.17 \cdot 10^{-01}$	$8.95 \cdot 10^{-02}$	2	5	1	$6.53 \cdot 10^{+02}$	$1.37 \cdot 10^{+03}$
2	$2.50 \cdot 10^{-01}$	$2.70 \cdot 10^{-01}$	$4.91 \cdot 10^{-03}$	$2.14 \cdot 10^{-01}$	$2.70 \cdot 10^{-02}$	2	9	1	$2.51 \cdot 10^{+03}$	$6.33 \cdot 10^{+03}$
3	$1.25 \cdot 10^{-01}$	$1.18 \cdot 10^{-01}$	$2.67 \cdot 10^{-02}$	$8.42 \cdot 10^{-02}$	$6.89 \cdot 10^{-03}$	3	3	1	$1.75 \cdot 10^{+04}$	$4.65 \cdot 10^{+04}$
4	$6.25 \cdot 10^{-02}$	$5.94 \cdot 10^{-02}$	$1.05 \cdot 10^{-01}$	$3.79 \cdot 10^{-02}$	$2.09 \cdot 10^{-03}$	3	4	1	$6.27 \cdot 10^{+04}$	$1.87 \cdot 10^{+05}$
5	$3.12 \cdot 10^{-02}$	$3.36 \cdot 10^{-02}$	$4.02 \cdot 10^{-01}$	$2.31 \cdot 10^{-02}$	$5.97 \cdot 10^{-04}$	3	5	1	$2.41 \cdot 10^{+05}$	$7.84 \cdot 10^{+05}$
6	$1.56 \cdot 10^{-02}$	$1.89 \cdot 10^{-02}$	$1.17 \cdot 10^{+00}$	$1.38 \cdot 10^{-02}$	$1.65 \cdot 10^{-04}$	3	6	1	$9.52 \cdot 10^{+05}$	$3.32 \cdot 10^{+06}$
7	$7.81 \cdot 10^{-03}$	$1.20 \cdot 10^{-02}$	$5.13 \cdot 10^{+00}$	$1.00 \cdot 10^{-02}$	$4.45 \cdot 10^{-05}$	3	7	1	$3.80 \cdot 10^{+06}$	$1.41 \cdot 10^{+07}$
8	$3.91 \cdot 10^{-03}$	$6.37 \cdot 10^{-03}$	$2.26 \cdot 10^{+01}$	$5.30 \cdot 10^{-03}$	$1.25 \cdot 10^{-05}$	3	9	1	$1.54 \cdot 10^{+07}$	$6.28 \cdot 10^{+07}$
9	$1.95 \cdot 10^{-03}$	$2.48 \cdot 10^{-03}$	$1.06 \cdot 10^{+02}$	$1.89 \cdot 10^{-03}$	$2.62 \cdot 10^{-06}$	4	4	1	$8.22 \cdot 10^{+07}$	$3.26 \cdot 10^{+08}$

Table 10. Nested compound option ($\alpha = 1$, $\beta = 1$): Parameters and results of the MLMC estimator

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$-error	Time (s)	Bias	Variance	R	M	h^{-1}	N	Cost
1	$5.00 \cdot 10^{-01}$	$8.97 \cdot 10^{-01}$	$5.54 \cdot 10^{-04}$	$8.59 \cdot 10^{-01}$	$6.62 \cdot 10^{-02}$	2	4	1	$6.38 \cdot 10^{+02}$	$1.14 \cdot 10^{+03}$
2	$2.50 \cdot 10^{-01}$	$5.74 \cdot 10^{-01}$	$4.25 \cdot 10^{-03}$	$5.56 \cdot 10^{-01}$	$2.05 \cdot 10^{-02}$	2	7	1	$2.64 \cdot 10^{+03}$	$5.76 \cdot 10^{+03}$
3	$1.25 \cdot 10^{-01}$	$2.69 \cdot 10^{-01}$	$2.37 \cdot 10^{-02}$	$2.58 \cdot 10^{-01}$	$6.08 \cdot 10^{-03}$	3	4	1	$1.72 \cdot 10^{+04}$	$4.57 \cdot 10^{+04}$
4	$6.25 \cdot 10^{-02}$	$1.32 \cdot 10^{-01}$	$1.13 \cdot 10^{-01}$	$1.24 \cdot 10^{-01}$	$1.95 \cdot 10^{-03}$	3	6	1	$6.98 \cdot 10^{+04}$	$2.26 \cdot 10^{+05}$
5	$3.12 \cdot 10^{-02}$	$7.21 \cdot 10^{-02}$	$4.99 \cdot 10^{-01}$	$6.81 \cdot 10^{-02}$	$5.69 \cdot 10^{-04}$	3	8	1	$2.88 \cdot 10^{+05}$	$1.06 \cdot 10^{+06}$
6	$1.56 \cdot 10^{-02}$	$3.78 \cdot 10^{-02}$	$1.57 \cdot 10^{+00}$	$3.59 \cdot 10^{-02}$	$1.40 \cdot 10^{-04}$	4	5	1	$1.53 \cdot 10^{+06}$	$6.21 \cdot 10^{+06}$
7	$7.81 \cdot 10^{-03}$	$1.43 \cdot 10^{-02}$	$8.70 \cdot 10^{+00}$	$1.27 \cdot 10^{-02}$	$4.28 \cdot 10^{-05}$	4	7	1	$6.32 \cdot 10^{+06}$	$3.02 \cdot 10^{+07}$
8	$3.91 \cdot 10^{-03}$	$9.78 \cdot 10^{-03}$	$3.63 \cdot 10^{+01}$	$9.17 \cdot 10^{-03}$	$1.15 \cdot 10^{-05}$	4	8	1	$2.58 \cdot 10^{+07}$	$1.31 \cdot 10^{+08}$
9	$1.95 \cdot 10^{-03}$	$4.95 \cdot 10^{-03}$	$1.68 \cdot 10^{+02}$	$4.61 \cdot 10^{-03}$	$3.21 \cdot 10^{-06}$	4	10	1	$1.07 \cdot 10^{+08}$	$6.06 \cdot 10^{+08}$
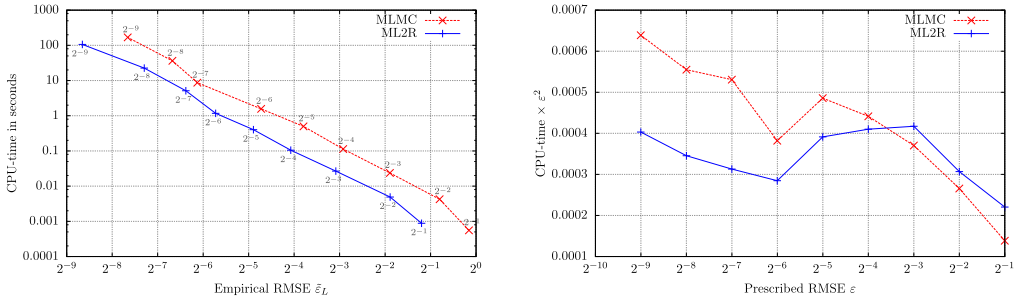
(a) CPU–time (y–axis, log scale) as a function of $\tilde{\varepsilon}_L$ (x–axis, $\log_2$ scale).

(b) CPU–time $\times \varepsilon^2$ (y–axis) as a function of ε (x–axis, $\log_2$ scale).

Figure 4. Nested compound option in a Black–Scholes model.

by the column vector e_1. The first point is that V_i is again a Vandermonde matrix of type $V_i = V(1, \ldots, a_{i-1}, 0, a_{i+1}, \ldots, a_R)$. On the other hand, the determinant of a square Vandermonde matrix can be expressed as $\det(V) = \prod_{1 \leq j < k \leq n}(a_k - a_j)$. We have for every $i \in \{1, \ldots, R\}$

$$
\begin{aligned}
\mathbf{w}_i &= \frac{\prod_{1 \leq j < k \leq R; j, k \neq i}(a_k - a_j) \prod_{1 \leq j < i}(-a_j) \prod_{i < k \leq R} a_k}{\prod_{1 \leq j < k \leq R}(a_k - a_j)} \\
&= \frac{\prod_{1 \leq j < i}(-a_j) \prod_{i < k \leq R} a_k}{\prod_{1 \leq j < i}(a_i - a_j) \prod_{i < k \leq R}(a_k - a_i)}.
\end{aligned}
$$

Using that $a_i = n_i^{-\alpha}$, $i = 1, \ldots, R$, we get

$$
\frac{\prod_{1 \leq j < i}(-a_j)}{\prod_{1 \leq j < i}(a_i - a_j)} = \frac{n_i^{\alpha(i-1)}}{\prod_{1 \leq j < i}(n_i^\alpha - n_j^\alpha)}
$$

and

$$
\frac{\prod_{i < k \leq R} a_k}{\prod_{i < k \leq R}(a_k - a_i)} = \frac{(-1)^{R-i} n_i^{\alpha(R-i)}}{\prod_{i < k \leq R}(n_k^\alpha - n_i^\alpha)}
$$

which completes the proof.

(b) follows by setting $x = 0$ in the decomposition

$$
\frac{1}{\prod_{1 \leq i \leq R}(x - n_i^\alpha)} = \sum_{i=1}^{R} \frac{1}{(x - n_i^\alpha) \prod_{j \neq i}(n_i^\alpha - n_j^\alpha)}. \qquad \square
$$

Proposition A.2. *When $n_i = M^{i-1}$, $i = 1, \ldots, R$, the following holds true for the coefficients* $\mathbf{w}_i = \mathbf{w}_i(R, M)$.

1. *Closed form for* $\mathbf{w}_i$, $i = 1, \ldots, R$:

$$\mathbf{w}_i = \mathbf{w}_i(R, M)$$

$$= (-1)^{R-i} \frac{M^{-\frac{\alpha}{2}(R-i)(R-i+1)}}{\prod_{1 \le j \le i-1}(1 - M^{-j\alpha}) \prod_{1 \le j \le R-i}(1 - M^{-j\alpha})}, \qquad i = 1, \ldots, R.$$

2. *Closed form for* $\widetilde{\mathbf{w}}_{R+1}$:

$$\widetilde{\mathbf{w}}_{R+1} = (-1)^R M^{-\frac{R(R-1)}{2}\alpha}.$$

3. *A useful upper bound*:

$$\sup_{R \in \mathbf{N}^*} \sum_{i=1}^{R-1} \left| \mathbf{w}_i(R, M) \right| \le \frac{M^{-\alpha}}{\pi_{\alpha,M}^2} \sum_{k \ge 0} M^{-\alpha \frac{k(k+3)}{2}} \quad and \quad 1 \le \mathbf{w}_R(R, M) \le \frac{1}{\pi_{\alpha,M}},$$

   where $\pi_{\alpha,M} = \prod_{k \ge 1}(1 - M^{-\alpha k})$.
4. *Asymptotics of the coefficients* $\mathbf{w}_i$ *when* $M \to +\infty$:

$$\lim_{M \to +\infty} \sup_{R \in \mathbf{N}^*} \max_{1 \le i \le R-1} \left| \mathbf{w}_i(R, M) \right| = 0 \quad and \quad \lim_{M \to +\infty} \sup_{R \in \mathbf{N}^*} \left| \mathbf{w}_R(R, M) - 1 \right| = 0.$$

5. *Asymptotics of the coefficients* $\mathbf{W}_i = \mathbf{W}_i(R, M)$ *when* $M \to +\infty$: *the coefficients* $\mathbf{W}_i$ *are defined in* (ML2R). *It follows from what precedes that they satisfy* $\mathbf{W}_1 = 1$,

$$\max_{1 \le i \le R} \left| \mathbf{W}_i(R, M) \right| \le \mathbf{W}_\alpha(M) := \frac{M^{-\alpha}}{\pi_{\alpha,M}^2} \sum_{k \ge 0} M^{-\alpha \frac{k(k+3)}{2}} + \frac{1}{\pi_{\alpha,M}} \qquad (58)$$

   and

$$\max_{1 \le i \le R} \left| \mathbf{W}_i(R, M) - 1 \right| \le \mathbf{W}_\alpha(M) - 1 \sim M^{-\alpha} \to 0 \qquad as \ M \to +\infty.$$

   In particular, the matrix $\mathbf{T} = \mathbf{T}(R, M)$ *in* (ML2R) *converges toward the matrix of the standard Multilevel Monte Carlo* (MLMC) *at level M when* $M \to +\infty$.
6. *One more useful inequality*

$$\forall R \in \mathbf{N}, \qquad \frac{1}{|\widetilde{\mathbf{w}}_{R+1}|} \sum_{r=1}^R \frac{|\mathbf{w}_r(R, M)|}{n_r^{\alpha R}} \le B_\alpha(M) \frac{1}{\pi_{\alpha,M}^2} \sum_{k \ge 0} M^{-\frac{\alpha}{2}k(k+1)}.$$

Proof. Claim 6: For every $r \in \{1, \ldots, R\}$,

$$\frac{|\mathbf{w}_r(R, M)|}{n_r^{\alpha R}} \le \frac{M^{-\frac{\alpha}{2}((R-r)(R-r+1)+2(r-1)R)}}{\pi_{\alpha,M}^2}.$$

Noting that $((R - r)(R - r + 1) + 2(r - 1)R) = R(R - 1) + r(r - 1)$, we derive that

$$\sum_{r=1}^{R} \frac{|\mathbf{w}_r(R, M)|}{n_r^{\alpha R}} \leq \frac{1}{\pi_{\alpha, M}^2} M^{-\alpha \frac{R(R-1)}{2}} \sum_{r=1}^{R} M^{-\alpha \frac{r(r-1)}{2}}$$

which yields the announced inequality since $M^{-\alpha \frac{R(R-1)}{2}} = |\widetilde{\mathbf{w}}_{R+1}|$. $\qquad\square$

Appendix B: Sketch of proof of Propositions 2.3 and 2.5

The multistep Richardson–Romberg estimator with the formal framework of Section 3, is characterized by the design matrix $\mathbf{T} = (\mathbf{w}, \mathbf{0}, \ldots, \mathbf{0})$. Note that the first column is not e_1 but this has no influence on what follows. The expansion of $\mathbf{E}[\bar{Y}_{h,\underline{n}}^N]$ follows from Proposition 2.5. No allocation is needed here since only one Brownian motion is involved. The proof of Proposition 3.9 applies here with $q = (1, 0, \ldots, 0)$. Furthermore,

$$\phi(\bar{Y}_{h,\underline{n}}^N) = \text{var}(\langle \mathbf{w}, Y_{h,\underline{n}}^1 \rangle) \frac{|\underline{n}|}{h} \sim \text{var}(Y_0) \frac{|\underline{n}|}{h} \qquad \text{as } h \to 0$$

since $Y_{h,\underline{n}}^1 \to Y_0 \mathbf{1}$ in $\mathbf{L}^2$ and $\sum_{i=1}^{R} \mathbf{w}_i = 1$. $\qquad\square$

References

[1] Alfonsi, A., Jourdain, B. and Kohatsu-Higa, A. (2014). Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme. *Ann. Appl. Probab.* **24** 1049–1080. MR3199980

[2] Bally, V. and Talay, D. (1996). The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probab. Theory Related Fields* **104** 43–60. MR1367666

[3] Barth, A., Lang, A. and Schwab, C. (2013). Multilevel Monte Carlo method for parabolic stochastic partial differential equations. *BIT* **53** 3–27. MR3029293

[4] Belomestny, D. and Nagapetyan, T. (2014). Multilevel path simulation for weak approximation schemes. Unpublished manuscript.

[5] Belomestny, D., Schoenmakers, J. and Dickmann, F. (2013). Multilevel dual approach for pricing American style derivatives. *Finance Stoch.* **17** 717–742. MR3105931

[6] Ben Alaya, M. and Kebaier, A. (2015). Central limit theorem for the multilevel Monte Carlo Euler method. *Ann. Appl. Probab.* **25** 211–234. MR3297771

[7] Broadie, M., Du, Y. and Moallemi, C.C. (2011). Efficient risk estimation via nested sequential simulation. *Manage. Sci.* **57** 1172–1194.

[8] Dereich, S. (2011). Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *Ann. Appl. Probab.* **21** 283–311. MR2759203

[9] Dereich, S. and Heidenreich, F. (2011). A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stochastic Process. Appl.* **121** 1565–1587. MR2802466

[10] Devineau, L. and Loisel, S. (2009). Construction d'un algorithme d'accélération de la méthode des "simulations dans les simulations" pour le calcul du capital économique solvabilité II. *Bulletin Français D'Actuariat* **10** 188–221.

[11] Duffie, D. and Glynn, P. (1995). Efficient Monte Carlo simulation of security prices. *Ann*. *Appl*. *Probab.* **5** 897–905. MR1384358

[12] Giles, M.B. (2008). Multilevel Monte Carlo path simulation. *Oper*. *Res*. **56** 607–617. MR2436856

[13] 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. MR2519838

[14] Giles, M.B. and Szpruch, L. (2014). Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Ann*. *Appl*. *Probab.* **24** 1585–1620. MR3211005

[15] Giorgi, D., Lemaire, V. and Pagès, G. (2016). Strong and weak asymptotic behaviour of multilevel estimators. Unpublished manuscript.

[16] Gobet, E. (2000). Weak approximation of killed diffusion using Euler schemes. *Stochastic Process*. *Appl*. **87** 167–197. MR1757112

[17] Gordy, M.B. and Juneja, S. (2010). Nested simulation in portfolio risk measurement. *Manage*. *Sci*. **56** 1833–1848.

[18] Guyon, J. (2006). Euler scheme and tempered distributions. *Stochastic Process*. *Appl*. **116** 877–904. MR2254663

[19] Heinrich, S. (2001). Multilevel Monte Carlo methods. In *Large-Scale Scientific Computing* 58–67. Berlin: Springer.

[20] Hong, J.L. and Juneja, S. (2009). Estimating the mean of a non-linear function of conditional expectation. In *Proceedings of the* 2009 *Winter Simulation Conference* (M.D. Rossetti, R.R. Hill, B. Johansson, A. Dunkin and R.G. Ingalls, eds.) 1223–1236. New York: IEEE.

[21] Hutzenthaler, M., Jentzen, A. and Kloeden, P.E. (2013). Divergence of the multilevel Monte Carlo Euler method for nonlinear stochastic differential equations. *Ann*. *Appl*. *Probab.* **23** 1913–1966. MR3134726

[22] Jourdain, B. and Kohatsu-Higa, A. (2011). A review of recent results on approximation of solutions of stochastic differential equations. In *Stochastic Analysis with Financial Applications*. *Progress in Probability* **65** 121–144. Boston, MA: Birkhäuser. MR3050787

[23] Kebaier, A. (2005). Statistical Romberg extrapolation: A new variance reduction method and applications to option pricing. *Ann*. *Appl*. *Probab.* **15** 2681–2705. MR2187308

[24] Konakov, V. and Mammen, E. (2002). Edgeworth type expansions for Euler schemes for stochastic differential equations. *Monte Carlo Methods Appl*. **8** 271–285. MR1931967

[25] Lapeyre, B. and Temam, E. (2001). Competitive Monte Carlo methods for the pricing of Asian options. *J*. *Comput*. *Finance* **5** 39–59.

[26] Lemaire, V. and Pagès, G. (2016). Multilevel estimators for stochastic approximation. Unpublished manuscript.

[27] Pagès, G. (2007). Multi-step Richardson–Romberg extrapolation: Remarks on variance control and complexity. *Monte Carlo Methods Appl*. **13** 37–70. MR2325670

[28] Rhee, C.-h and Glynn, P.W. (2012). A new approach to unbiased estimation for SDEs. In *Proceedings of the Winter Simulation Conference*. Berlin.

[29] Shiryaev, A.N. (1996). *Probability*, 2nd ed. *Graduate Texts in Mathematics* **95**. New York: Springer. MR1368405

[30] Talay, D. and Tubaro, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch*. *Anal*. *Appl*. **8** 483–509 (1991). MR1091544