Multilevel Richardson-Romberg extrapolation

Vincent Lemaire, Gilles Pagès

Multilevel Richardson-Romberg extrapolation

Vincent Lemaire*, Gilles Pagès†

December 19, 2014

Abstract

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 [Pag07] and the variance control resulting from the stratification introduced in the Multilevel Monte Carlo (MLMC) method (see [Gil08, Hei01]). 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: Multilevel Monte Carlo estimator; Richardson-Romberg extrapolation; Multistep; Euler scheme; Nested Monte Carlo method; Stratification; Option pricing.

MSC 2010: primary 65C05, secondary 65C30, 62P05.

1 Introduction

The aim of this paper is to combine the multilevel Monte Carlo estimator introduced by S. Heinrich in [Hei01] and developed by M. Giles in [Gil08] (see also [Keb05] for the statistical Romberg approach) and the (consistent) Multistep Richardson-Romberg extrapolation (see [Pag07]) in order to minimize the simulation cost of a quantity of interest $I_0 = \mathbf{E}[Y_0]$ where the random variable Y_0 cannot be simulated at a reasonable cost (typically a generic multidimensional diffusion process or a conditional expectation). Both methods rely on the existence of a family of simulatable 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 cannot prevent 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 $(Y_h)_{h\geqslant 0}$ 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 an admissible subset of parameters $\mathcal{H} \subset (0, \mathbf{h}]$ such that $0 \in \mathcal{H}$, $\mathbf{h} \in \mathcal{H}$ and $\frac{\mathcal{H}}{n} \subset \mathcal{H}$ for every integer $n \geqslant 1$. We also assume that $\mathbf{h} \in \mathcal{H}$. 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 make the pseudo-assumption, that for every admissible $h \in \mathcal{H}$, the random variable Y_h is *simulatable* whereas Y_0 *is not* (at a reasonable cost). For this reason, the specification of h will be often made in connection with the complexity of the simulation Y_h with in mind to make it inverse linear in h.

*Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, UPMC-Paris 06, E-mail: vincent.lemaire@upmc.fr

†Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, UPMC-Paris 06, E-mail: gilles.pages@upmc.fr

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 $(Y_h^{(k)})_{k\geqslant 1}$ 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 standard control of the renormalized *statistical error* $\sqrt{N}(I_N^{(h)} - \mathbf{E}[Y_h])$ which behaves as a centered Gaussian with variance $\mathrm{var}(Y_h)$. 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.

We first present a brief description of the original MLMC estimator as described in [Gil08]. The main idea is to use the following telescopic summation with depth $L \geqslant 2$

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

where $(h_j)_{j=0,\dots,L}$ is a geometrically decreasing sequence of different bias parameters $h_j = M^{-(j-1)}h$. For each level $j \in \{1,\dots,L\}$, the computation of $\mathbf{E}\left[Y_{h_j} - Y_{h_{j-1}}\right]$ is performed by a standard Monte Carlo procedure. The key point is that, at each level, we consider a number $N_j = \lceil N q_j \rceil$ of scenarios where $q = (q_1,\dots,q_L) \in \mathcal{S}_+(L) = \left\{q \in (0,1)^L, \sum_{j=1}^{L} q_j = 1\right\}$ (L-dimensional simplex) and a random sample of Y_{h_j} and $Y_{h_{j-1}}$ supposed to be perfectly correlated. More precisely, we consider L copies of the biased family denoted $Y^{(j)} = (Y_h^{(j)})_{h\in\mathcal{H}}$, $j \in \{1,\dots,L\}$ attached to *independant* random copies $Y_0^{(j)}$ of Y_0. The MLMC estimator then writes

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

where $(Y^{(j),k})_{k\geqslant 1}$, $j = 1,\dots,L$ are independent sequences of independent copies of $Y^{(j)}$ and $N_1,\dots,N_L$ 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 [DH11, Der11], computational statistics and more general numerical analysis problems (high dimensional parabolic SPDEs, see [BLS13], etc). For more references, we refer to the webpage `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 consider one 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 $(Y^k)_{k\geqslant 1}$ 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)

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

Moreover, the asymptotic variance of this estimator satisfies $\lim_{h\to 0} \mathrm{var}(I^N_{h,\frac{h}{2}}) = \mathrm{var}(Y_0)/N$ which is the same as the crude Monte Carlo estimator. It is natural to reiterate this extrapolation to obtain a linear estimator with bias error of order 3 in h and so on. This extension called Multistep Richardson-Romberg extrapolation for Monte Carlo estimator has been introduced and extensively investigated in [Pag07] 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 MLMC and writes

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

where $R \geqslant 2$ is the depth level – similar to L in (2) – and $(Y^{(j),k})_{k\geqslant 1}$ are like in (2). We denote by $n_j = M^{j-1}$ the j–th *refiner coefficient* of the initial bias parameter $h \in \mathcal{H}$ and call $M \geqslant 2$ the *root* of the refiners. A strong feature of our approach comes from the fact that the weights $(\mathbf{W}_k)_{k=2,\dots,R}$ are explicit and only depend on α (given by (1)), $M \geqslant 2$ and $R \geqslant 2$. In practice these ML2R weights read $\mathbf{W}_j = \sum_{i=j}^{R} \mathbf{w}_i$ where $\mathbf{w}_i$ are given by (12). These derivative weights $(\mathbf{w}_i)_{i\in\{1,\dots,R\}}$ have been introduced in [Pag07] 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). The first stage is a stratification procedure to minimize the *effort estimator* (product of the variance by the complexity) to optimally dispatch the N_j across all level. Doing so we are able to specify the initial bias parameter h and the depth level parameter R as a function of ε and structural fixed parameters $(\alpha, \beta, V_1, \mathrm{var}(Y_0))$. A light preprocessing makes possible to optimize the choice of the root $M \geqslant 2$ of the refiners n_j, $j = 1,\dots,R$. Basically (see Theorem 3.11), the numerical cost of the ML2R estimator implemented with these optimal parameters (depending on ε) and denoted Cost(ML2R) satisfies

$$\limsup_{\varepsilon\to 0} v(\beta,\varepsilon) \times \mathrm{Cost(ML2R)} \leqslant K(\alpha,\beta,M)$$

where $K(\alpha,\beta,M)$ is an explicit bound and

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

As first established in [Gil08], we prove in a similar way that the optimal numerical cost of the MLMC estimator denoted Cost(MLMC) satisfies a similar result with $v(\beta,\varepsilon) = \varepsilon^2\log(1/\varepsilon)^{-2}$ 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 a 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 MLMC. When $\beta < 1$, ML2R 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 Multilevel Richardson-Romberg estimator is the most powerful compared to regular MLMC.

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 allocations matrix $\mathbf{T}$. Among them, we describe 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.
• If $\underline{n} = (n_1, \ldots, n_R) \in (\mathbf{N}^*)^R$, $|\underline{n}| = n_1 + \cdots + n_R$ and $\underline{n}! = \prod_{1 \leqslant i \leqslant R} n_i$.
• Let $(e_1, \ldots, e_R)$ denote the canonical basis of $\mathbf{R}^R$ (viewed as a vector space of column vectors). Thus $e_i = (\delta_{ij})_{1 \leqslant j \leqslant r}$ where δ_{ij} stands for the classical Kronecker symbol.
• $\langle .,. \rangle$ 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 $n - 1 < x \leqslant n$ and $\lfloor x \rfloor$ denotes the unique $n \in \mathbf{N}$ satisfying $n \leqslant x < n + 1$.
• For every $x \in \mathbf{R}_+ = [0, +\infty)$, and $y \in \mathbf{R}_+^* = (0, +\infty)$, $\lceil x \rceil_y$ denotes the unique $n \in \mathbf{N}^*$ satisfying $(n-1)y < x \leqslant ny$ and $\lfloor x \rfloor_y$ denotes the unique $n \in \mathbf{N}$ satisfying $ny \leqslant x < (n+1)y$.
• 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 $a_n = o(b_n)$ is $\lim_n \varepsilon_n = 0$.

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 $(I_\pi^N)_{N \geqslant 1}$ 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

$$\mathbf{E}\left[I_\pi^N\right] = \mathbf{E}\left[I_\pi^1\right], \; N \geqslant 1,$$

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:

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

where $\nu(\pi) = \mathrm{var}(I_\pi^1)$ 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 $\left\{(I_\pi^N)_{N \geqslant 1}, \pi \in \Pi\right\}$ *i.e.* the estimator minimizing the computational cost for a given error $\varepsilon > 0$. In the sequel, we will 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) $\sqrt{\mathbf{E}\left[(I_\pi^N - I_0)^2\right]} = \left\|I_\pi^N - I_0\right\|_2$. Our aim is to minimize the cost of the simulation for a given target error, say $\varepsilon > 0$. This generic problem reads

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

In order to solve with 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

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

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

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

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

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

  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 $(I_\pi^N)_{N\geqslant 1}$, $\pi \in \Pi$, 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 $\nu(\pi) = N\big(\big\|I_\pi^N - I_0\big\|_2^2 - \mu(\pi)^2\big)$, the solution of the generic problem (4) reads

$$\pi(\varepsilon) = \operatorname*{argmin}_{\pi \in \Pi,\ |\mu(\pi)|<\varepsilon} \left(\frac{\phi(\pi)}{\varepsilon^2 - \mu^2(\pi)}\right), \quad N(\varepsilon) = \frac{\nu(\pi(\varepsilon))}{\varepsilon^2 - \mu^2(\pi(\varepsilon))} = \frac{\phi(\pi(\varepsilon))}{\kappa(\pi(\varepsilon))(\varepsilon^2 - \mu^2(\pi(\varepsilon)))}. \tag{8}$$

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 *consistent* set of *step* parameters in the sense that $\mathcal{H} \subset (0, \mathbf{h}]$, $\mathbf{h} \in \mathcal{H}$ and, for every integer $n \geqslant 1$, $\frac{\mathcal{H}}{n} \subset \mathcal{H}$ (hence 0 is a limiting value of $\mathcal{H}$). 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 [Gil08, Hei01]):

Bias error expansion (weak error rate):

$$\exists\, \alpha > 0, \bar{R} \geqslant 1, \quad \mathbf{E}\,[Y_h] = \mathbf{E}\,[Y_0] + \sum_{k=1}^{\bar{R}} c_k h^{\alpha k} + h^{\alpha \bar{R}}\eta_{\bar{R}}(h), \quad \lim_{h\to 0} \eta_{\bar{R}}(h) = 0, \tag{$WE_{\alpha,\bar{R}}$}$$

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

Strong approximation error assumption:

$$\exists\, \beta > 0,\ V_1 \in \mathbf{R}_+, \quad \big\|Y_h - Y_0\big\|_2^2 = \mathbf{E}\left[\big|Y_h - Y_0\big|^2\right] \leqslant V_1 h^\beta. \tag{SE_β}$$

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

Note that when $c_1 \neq 0$, consistency of strong and weak errors implies in what follows that $\beta \leqslant 2\alpha$. In the sequel we will consider a free parameter $R \in \{2,\dots,\bar{R}\}$ for which $(WE_{\alpha,\bar{R}})$ is always satisfied (with the same coefficients c_r up to $r = R$). This parameter R corresponds to the depth level L used in the multilevel literature and will also be referred to as the depth from now on.

In what follows, we will use the following ratio

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

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

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 $(Y_h)_{h\in\mathcal{H}}$, $j = 1,\dots,R$. All random variables are supposed to be defined on the same probability space. Note that, since the above properties (SE_β) and $(WE_{\alpha,\bar{R}})$, $\bar{R} \geqslant 1$, 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

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

where the R-tuple of integers $\underline{n} := (n_1, n_2, \dots, n_R) \in \mathbf{N}^R$, 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 $(Y_h^{(j)})_{h\in\mathcal{H}}$.

$\triangleright$ *Specification of the refiners*: In most applications, we will choose refiners n_i as $n_i = M^{i-1}$ where $M \geqslant 2$. Indeed, this is the standard choice in the regular Multilevel Monte Carlo method as described in [Gil08]. Other choices like $n_i = i$ are possible (see below).

2.3 Crude Monte Carlo estimator

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

Proposition 2.2. *Assume $(WE_{\alpha,\bar{R}})$ with $\bar{R} \geqslant 1$ and $c_1 \neq 0$. The Monte Carlo estimator of $\mathbf{E}[Y_0]$ defined by*

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

where $\left(Y_h^k\right)_{k\geqslant 1}$ is an i.i.d. *sequence of copies of Y_h, satisfy*

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

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

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

Furthermore, we have

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

Proof. The proof is postponed to Annex B. □

We refer to the seminal paper [DG95] for more details on practical implementation of this estimator.

Remark 2.3. *For crude Monte Carlo simulation, Assumption* (SE_β) *is not necessary. Note that, at order* 1*, one can always assume* $c_1 \neq 0$ *considering the first non-zero term* h^α *in the expansion (if any).*

2.4 Background on Multistep Richardson-Romberg extrapolation

The so-called *Multistep Richardson-Romberg estimator* has been introduced in [Pag07] in the framework of Brownian diffusions. It relies on R (refined) Euler schemes $\bar{X}^{(\frac{h}{n_i})}$, $1 \leqslant i \leqslant R$, defined on a finite interval $[0, T]$ ($T > 0$) where the bias parameter $h = \frac{T}{n}$, $n \geqslant 1$. 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 2.5 below). The main results are obtained when all the schemes are *consistent i.e.* 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 $Y_h = \bar{X}^{(h)}$, $h \in \mathcal{H}$, 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

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

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

$$\forall i \in \{1, \ldots, R\}, \quad \mathbf{w}_i = \frac{(-1)^{R-i} n_i^{\alpha(R-1)}}{\prod_{1 \leqslant j < i} (n_i^\alpha - n_j^\alpha) \prod_{i < j \leqslant R} (n_j^\alpha - n_i^\alpha)}. \tag{10}$$

We also derive the following identity of interest

$$\widetilde{\mathbf{w}}_{R+1} := \sum_{i=1}^{R} \frac{\mathbf{w}_i}{n_i^{\alpha R}} = \frac{(-1)^{R-1}}{\underline{n}!^\alpha}. \tag{11}$$

Note that all coefficients $(\mathbf{w}_i)_{1 \leqslant i \leqslant R}$ depend on the depth level 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:

$$\mathbf{w}_i = \begin{cases} \dfrac{(-1)^{R-i} i^{\alpha R}}{\prod_{j=1}^{i-1}(i^\alpha - j^\alpha) \prod_{i+1}^{R}(j^\alpha - i^\alpha)} & \text{if } n_j = j,\ j \in \{1, \ldots, R\}, \\[2ex] \dfrac{(-1)^{R-i} M^{-\frac{\alpha}{2}(R-i)(R-i+1)}}{\prod_{j=1}^{i-1}(1 - M^{-j\alpha}) \prod_{j=1}^{R-i}(1 - M^{-j\alpha})} & \text{if } n_j = M^{j-1},\ j \in \{1, \ldots, R\}. \end{cases} \tag{12}$$

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

Assume now ($WE_{\alpha,\bar{R}}$) 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 $\sum_{r=1}^{R} \mathbf{w}_r = 1$, implies that

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

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, $r=1,\dots,R-1$ vanish. Finally, we obtain

$$\mathbf{E}\left[\langle \mathbf{w}, Y_{h,\underline{n}}\rangle\right] = \mathbf{E}\left[Y_0\right] + c_R \widetilde{\mathbf{w}}_{R+1} h^{\alpha R}\left(1+\eta_{R,\underline{n}}(h)\right) \tag{13}$$

where

$$\eta_{R,\underline{n}}(h) = \frac{1}{c_R \widetilde{\mathbf{w}}_{R+1}} \sum_{r=1}^{R} \frac{\mathbf{w}_r}{n_r^{\alpha R}} \eta_R\left(\frac{h}{n_r}\right) \to 0 \quad \text{as} \quad h \to 0. \tag{14}$$

Proposition 2.4. *Assume* ($WE_{\alpha,\bar{R}}$) *and* $R \in \{2,\dots,\bar{R}\}$. *The Multistep Richardson-Romberg estimator of* $\mathbf{E}[Y_0]$ *defined by*

$$\forall N \geqslant 1,\ h \in \mathcal{H}, \quad \bar{Y}_{h,\underline{n}}^N = \frac{1}{N}\sum_{k=1}^{N} \langle \mathbf{w}, Y_{h,\underline{n}}^k\rangle = \langle \mathbf{w}, \frac{1}{N}\sum_{k=1}^{N} Y_{h,\underline{n}}^k\rangle \tag{15}$$

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

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

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

$$h^*(\varepsilon) = (1+2\alpha R)^{-\frac{1}{2\alpha R}}\left(\frac{\varepsilon}{|c_R|}\right)^{\frac{1}{\alpha R}} \underline{n}!^{\frac{1}{R}}, \quad N^*(\varepsilon) = \left(1+\frac{1}{2\alpha R}\right)\frac{\operatorname{var}(Y_0)(1+\theta(h^*(\varepsilon))^{\frac{\beta}{2}})^2}{\varepsilon^2}.$$

Furthermore,

$$\inf_{\substack{h\in\mathcal{H}\\ |\mu(h)|<\varepsilon}} \operatorname{Cost}(\bar{Y}_h^N) \sim \left(\frac{(1+2\alpha R)^{1+\frac{1}{2\alpha R}}}{2\alpha R}\right)\frac{|c_R|^{\frac{1}{\alpha R}} |\underline{n}| \operatorname{var}(Y_0)}{\underline{n}!^{\frac{1}{R}} \varepsilon^{2+\frac{1}{\alpha R}}} \quad \text{as } \varepsilon \to 0. \tag{16}$$

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

Remark 2.5. *In this approach the bias reduction suffers from an increase of the simulation cost by a* $|\underline{n}|$ *factor which appears in the numerator of* (16). *The choice of the refiners in [Pag07], 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 in* R *– like* $\frac{e}{2}(R+1)$ *– for large values of* R, *whereas with* $n_i = M^{i-1}$ *it goes to infinity like* $M^{\frac{R-1}{2}}$.

3 A paradigm for Multilevel simulation methods

3.1 General framework

Multilevel decomposition

In spite of Proposition 2.4 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 ot the ratio $\frac{|\underline{n}|}{\underline{n}!}$ (when R grows) in front of $\operatorname{var}(Y_0)$ in (16) 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 $\mathbf{E}\left[\frac{X+X'}{2}\right] = \mathbf{E}[X]$ and $\operatorname{var}(\frac{X+X'}{2}) = \frac{1}{2}\operatorname{var}(X)$, combined with an appropriate stratification strategy to control the complexity of the resulting estimator. So, let us consider now R *independant* copies $(Y_{h,\underline{n}}^{(j)})$, $j=1,\dots,R$, 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 $\mathbf{T} = [\mathbf{T}^1 \dots \mathbf{T}^R]$ is an $R \times R$ matrix with column vectors $\mathbf{T}^j \in \mathbf{R}^R$ satisfying the constraint

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

Under Assumption ($WE_{\alpha,\bar{R}}$) and $R \in \{2, \dots, \bar{R}\}$,

$$\begin{aligned}
\mathbf{E}\Big[\sum_{j=1}^{R} \langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} \rangle\Big] - \mathbf{E}[Y_0] &= \sum_{j=1}^{R} \langle \mathbf{T}^j, \mathbf{E}\left[Y_{h,\underline{n}}^{(j)}\right] \rangle - \mathbf{E}[Y_0] \\
&= \sum_{i=1}^{R} \Big(\sum_{j=1}^{R} \mathbf{T}_i^j\Big) \mathbf{E}\Big[Y_{\frac{h}{n_i}}\Big] - \mathbf{E}[Y_0] \\
&= \sum_{i=1}^{R} \Big(\sum_{j=1}^{R} \mathbf{T}_i^j\Big) \Big(\mathbf{E}\Big[Y_{\frac{h}{n_i}}\Big] - \mathbf{E}[Y_0]\Big) = o(h^\alpha).
\end{aligned}$$

The strong $\mathbf{L}^2$–convergence of the estimator also holds (without rate) as soon as Y_h strongly converges toward Y_0 (in $\mathbf{L}^2$).

As emphasized further on, we will also need that each column vector $\mathbf{T}^j$, $j \in 2, \dots, R$, 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 $\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} \rangle$, $j = 1, \dots, R$. This leads to the following definitions.

Definition 3.1 (Allocation matrix)**.** *Let $R \geqslant 2$. An $R \times R$-matrix $\mathbf{T} = [\mathbf{T}^1 \dots \mathbf{T}^R]$ is an R-level allocation matrix if*

$$\langle \mathbf{T}^j, \mathbf{1} \rangle = \sum_{i=1}^{R} \mathbf{T}_i^j = 0, \; j = 2, \dots, R. \tag{17}$$

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

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

$$\bar{Y}_{h,\underline{n}}^{N,q} = \sum_{j=1}^{R} \frac{1}{N_j} \sum_{k=1}^{N_j} \langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j),k} \rangle \tag{18}$$

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

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

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

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

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

$$\mathbf{T}^1 = e_1 \quad \textit{and} \quad \sum_{j=1}^{R} \mathbf{T}^j = \mathbf{w}, \; \textit{where } \mathbf{w} \textit{ is the unique solution to (10)},$$

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 $\mathbf{0}$ the null column vector of $\mathbf{R}^R$, the crude Monte Carlo is associated to the allocation matrix $\mathbf{T} = (e_1, \mathbf{0}, \dots, \mathbf{0})$ an the multistep Richardson-Romberg estimator is associated to $\mathbf{T} = (\mathbf{w}, \mathbf{0}, \dots, \mathbf{0})$.*

Within the abstract framework of a parametrized Monte Carlo simulation described in Section 2.1, the structure parameter π of the multilevel estimators $(\bar{Y}^{N,q}_{h,\underline{n}})_{N\geqslant 1}$ defined by (18) is

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

Cost, complexity and effort of a Multilevel estimator

In order to optimize (minimize) the effort $\phi(\pi)$ of the estimator (18), let us evaluate its unitary computational complexity. For a simulation size N, the numerical cost induced by the estimators $Y^{N,q}_{h,\underline{n}}$, $N \geqslant 1$, reads

$$\operatorname{Cost}(\bar{Y}^{N,q}_{h,\underline{n}}) = \sum_{j=1}^{R} N_j \sum_{i=1}^{R} \frac{1}{h} n_i \mathbf{1}_{\{\mathbf{T}^j_i \neq 0\}} = N\,\kappa(\pi) \tag{19}$$

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

$$\kappa(\pi) = \frac{1}{h} \sum_{j=1}^{R} q_j \sum_{i=1}^{R} n_i \mathbf{1}_{\{\mathbf{T}^j_i \neq 0\}}. \tag{20}$$

However, it may happen, like for nested Monte Carlo (see Section 4.2), that the internal consistency of the family Y_h leads to reduce the computational cost since the computational complexity entirely results from the most refined "scheme". If so,

$$\kappa(\pi) = \frac{1}{h} \sum_{j=1}^{R} q_j \max_{1 \leqslant i \leqslant R} \big(n_i \mathbf{1}_{\{\mathbf{T}^j_i \neq 0\}}\big).$$

It follows that the effort of such a Multilevel estimator is given by

$$\phi(\pi) = \nu(\pi)\,\kappa(\pi) = \left(\sum_{j=1}^{R} \frac{1}{q_j} \operatorname{var}\left(\big\langle \mathbf{T}^j, Y^{(j)}_{h,\underline{n}} \big\rangle \right) \right) \kappa(\pi). \tag{21}$$

Bias error of a Multilevel estimator

We now establish the bias error in this general framework. The following bias error result follows straightforwardly from the weak error decomposition $(WE_{\alpha,\bar{R}})$ and the definition of an allocation matrix $\mathbf{T}$.

Proposition 3.4. *Assume $(WE_{\alpha,\bar{R}})$.*

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

$$\mu(\pi_0, q) = (-1)^{R-1} c_R \left(\frac{h^R}{\underline{n}!} \right)^{\alpha} \big(1 + \eta_{R,\underline{n}}(h)\big) \tag{22}$$

where $\eta_{R,\underline{n}}(h) = (-1)^{R-1} \underline{n}!^{\alpha} \sum_{r=1}^{R} \frac{\mathbf{w}_r}{n_r^{\alpha R}} \eta_R\left(\frac{h}{n_r}\right)$ (see (14)) with η_R defined in $(WE_{\alpha,\bar{R}})$.

(b) MLMC estimator: *Let $R \geqslant 2$ be the depth of an MLMC estimator. For any admissible stratification strategy $q = (q_1, \dots, q_R)$, the bias error reads*

$$\mu(\pi_0, q) = c_1 \left(\frac{h}{n_R}\right)^{\alpha} \left(1 + \eta_1\left(\frac{h}{n_R}\right)\right) \tag{23}$$

with η_1 defined in ($WE_{\alpha,\bar{R}}$).

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 stratification strategies $q = (q_j)_{1 \leqslant j \leqslant R}$ (as a function of a fixed bias parameter h). In practice, we will optimize an upper-bound $\bar{\phi}$ of the true problem

$$q^* = \underset{q \in \mathcal{S}_+(R)}{\operatorname{argmin}}\, \bar{\phi}(\pi_0, q), \quad \text{where} \quad \phi(\pi) \leqslant \bar{\phi}(\pi), \quad \text{and} \quad \phi^*(\pi_0) = \phi(\pi_0, q^*). \tag{24}$$

This phase is solved in Theorem 3.6 below (an explicit expression for $\bar{\phi}$ is provided in (27)). The quantity $\phi^*(\pi_0)$ is called the *optimal stratified 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):

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

This second phase is solved asymptotically when ε goes to 0 in Theorem 3.8 and Proposition 3.11, with some closed forms for some h^* and R^* as functions of ε and of the structural parameters coming from assumptions ($WE_{\alpha,\bar{R}}$) and (SE_β).

3.2 Optimally stratified effort (Step 1)

Throughout our investigations on these estimators, we will make extensive use in what follows of the following lemma which is a straightforward consequence of Schwarz's Inequality.

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

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

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

Theorem 3.6. *Assume (SE_β) holds and let θ be defined by (9). Then, the optimally stratified effort ϕ^* defined by (24) satisfies*

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

where $q^* = q^*(\pi_0)$ is an optimal strategy given by

$$\begin{cases} q_1^*(\pi_0) = \mu^*(1 + \theta h^{\frac{\beta}{2}}) \\ q_j^*(\pi_0) = \mu^* \theta h^{\frac{\beta}{2}} \left(\sum_{i=1}^{R} |\mathbf{T}_i^j| n_i^{-\frac{\beta}{2}}\right)\left(\sum_{i=1}^{R} n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}\right)^{-\frac{1}{2}}, \; j = 2, \dots, R, \end{cases} \tag{25}$$

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

Proof. Under assumption (17), we have $\langle \mathbf{T}^1, Y^{(1)}_{h,\underline{n}}\rangle = Y^{(1)}_h$ and, for every $j \in \{2,\dots,R\}$, $\langle \mathbf{T}^j, Y^{(j)}_{h,\underline{n}}\rangle = \langle \mathbf{T}^j, Y^{(j)}_{h,\underline{n}} - Y^{(j)}_0 \mathbf{1}\rangle$ since $\langle \mathbf{T}^j, \mathbf{1}\rangle = 0$. Hence, using Minkowski's inequality and the strong error assumption, we obtain

$$\forall j \geqslant 2, \quad \operatorname{var}\left(\langle \mathbf{T}^j, Y^{(j)}_{h,\underline{n}}\rangle\right) \leqslant \left\| \sum_{i=1}^R \mathbf{T}^j_i \left(Y^{(j)}_{\frac{h}{n_i}} - Y^{(j)}_0\right)\right\|_2^2$$
$$\leqslant V_1 h^\beta \left(\sum_{i=1}^R |\mathbf{T}^j_i| n_i^{-\frac{\beta}{2}}\right)^2.$$

The variance of the Multilevel estimator is then

$$\operatorname{var}\left(\bar{Y}^{N,q}_{h,\underline{n}}\right) \leqslant \frac{1}{N}\left(\frac{\operatorname{var}\left(Y^{(1)}_h\right)}{q_1} + V_1 h^\beta \sum_{j=2}^R \frac{1}{q_j}\left(\sum_{i=1}^R |\mathbf{T}^j_i| n_i^{-\frac{\beta}{2}}\right)^2\right). \tag{26}$$

On the other hand we have,

$$\begin{aligned}\operatorname{var}\left(Y^{(1)}_h\right) = \operatorname{var}(Y_h) &\leqslant \mathbf{E}\left[Y_h - \mathbf{E}[Y_0]\right]^2\\ &\leqslant \|Y_h - Y_0\|_2^2 + 2\mathbf{E}\left[(Y_h - Y_0)(Y_0 - \mathbf{E}[Y_0])\right] + \operatorname{var}(Y_0)\\ &\leqslant \operatorname{var}(Y_0) + V_1 h^\beta + 2\sqrt{V_1} h^{\beta/2}\sqrt{\operatorname{var} Y_0} = \operatorname{var}(Y_0)(1+\theta h^{\frac{\beta}{2}})^2.\end{aligned}$$

Combining (20), the above inequality (26) and the above upper-bound for $\operatorname{var}\left(Y^{(1)}_h\right)$, we derive the following upper bound for the effort $\phi(\pi) \leqslant \bar{\phi}(\pi)$ with

$$\bar{\phi}(\pi) = \frac{\operatorname{var}(Y_0)}{h}\left(\frac{(1+\theta h^{\frac{\beta}{2}})^2}{q_1} + \theta^2 h^\beta \sum_{j=2}^R \frac{1}{q_j}\left(\sum_{i=1}^R |\mathbf{T}^j_i| n_i^{-\frac{\beta}{2}}\right)^2\right)\left(\sum_{i,j=1}^R q_j n_i \mathbf{1}_{\{\mathbf{T}^j_i \neq 0\}}\right). \tag{27}$$

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

Remark 3.7 (About variance minimization)**.** *We proved in the above proof that for every stratification strategy $q = (q_1,\dots,q_R)$,*

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

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

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

with an optimal choice (to minimize the variance): $q_1^\dagger = \mu^\dagger(1+\theta h^{\frac{\beta}{2}})$, $q_j^\dagger = \mu^\dagger \theta h^{\frac{\beta}{2}}\left(\sum_{i=1}^R |\mathbf{T}^j_i| n_i^{-\frac{\beta}{2}}\right)$ ($\mu^\dagger$ normalizing constant such that $\sum_{j=1}^n q_j^\dagger = 1$). Note that this choice is non-optimal when dealing with the effort optimization approach.

3.3 Resulting cost optimization (Step 2)

3.3.1 Bias parameter optimization (first approach)

In this first approach, we fix the depth $R \geqslant 2$, the allocation matrix $\mathbf{T}$ and the refiners $n_1, \dots, n_R$ and we only optimize the bias parameter $h \in \mathcal{H}$ with respect to $\varepsilon > 0$, so that

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

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

$$\bar{\phi}^*(h) = \frac{\operatorname{var}(Y_0)}{h}\left(1 + \theta h^{\frac{\beta}{2}} \sum_{j=1}^{R}\left(\sum_{i=1}^{R} |\mathbf{T}_i^j| n_i^{-\frac{\beta}{2}}\right)\left(\sum_{i=1}^{R} n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}\right)^{\frac{1}{2}}\right)^2. \tag{28}$$

Theorem 3.8 (Bias parameter optimization). *Assume* ($WE_{\alpha,\bar{R}}$) *and* (SE_β). *Let* $R \geqslant 2$ *and let* n_i, $i = 1, \dots, R$ *,be fixed refiners.*

(*a*) ML2R estimator: *Assume* $R \in \{2, \dots, \bar{R}\}$ *is such that* $c_R \neq 0$*. An ML2R estimator of depth R satisfies*

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

with q^* *defined in* (25)*. This asymptotically optimal bound is achieved with a bias parameter given by*

$$h^*(\varepsilon, R) = (1+2\alpha R)^{-\frac{1}{2\alpha R}} \left(\frac{\varepsilon}{|c_R|}\right)^{\frac{1}{\alpha R}} \underline{n}!^{\frac{1}{R}}. \tag{29}$$

(*b*) MLMC estimator: *Assume* $c_1 \neq 0$*. An MLMC estimator of depth R satisfies*

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

with q^* *defined in* (25)*. This done asymptotically optimal bound is achieved with a bias parameter given by*

$$h^*(\varepsilon, R) = (1+2\alpha)^{-\frac{1}{2\alpha}} \left(\frac{\varepsilon}{|c_1|}\right)^{\frac{1}{\alpha}} n_R. \tag{30}$$

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

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

It follows from (28) 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 (22). First note that $\lim_{h \to 0} h\bar{\phi}(h, q^*) = \operatorname{var}(Y_0)$. 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

$$\xi(a, R') := \operatorname{argmax}_{\xi > 0} g_{a,R'}(\xi) = \left((2R'+1)^{\frac{1}{2}} a\right)^{-\frac{1}{R'}} \quad \text{and} \quad \max_{(0,+\infty)} g_{a,R'} = \frac{2R'}{(2R'+1)^{1+\frac{1}{2R'}}} a^{-\frac{1}{R'}}.$$

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

$$\inf_{h\in\mathcal{H},\,|\mu(h,q^*)|<\varepsilon} \frac{\phi^*(h)}{\varepsilon^2 - \mu^2(h,q^*)} \leqslant \Big(1+\frac{1}{2\alpha R}\Big)(2\alpha R+1)^{\frac{1}{2\alpha R}}|c_R|^{\frac{1}{\alpha R}} \frac{h(\varepsilon)\bar{\phi}^*(h(\varepsilon))}{\underline{n}!^{\frac{1}{R}}\varepsilon^{2+\frac{1}{\alpha R}}} \frac{1}{1-\frac{(\eta_{R,\underline{n}}(h(\varepsilon))+1)^2-1}{2\alpha R}}. \quad (31)$$

The "limsup" 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 (21) of the effort ϕ that

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

Then Schwarz's Inequality implies

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

since $n_i \geqslant n_1 = 1$, $i = 1,\ldots,R$. Denoting $g(h) = \max_{1\leqslant j\leqslant R}\mathrm{var}\left(\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)}\rangle\right)$ one clearly has $\lim_{h\to 0} g(h) = \mathrm{var}(Y_0)$ under the strong assumption (SE_β) and, as a consequence, $\lim_{h\to 0} h\,\phi(h) = \mathrm{var}(Y_0)$. 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 $\eta\in\big(0,1\big)$. There exists $\varepsilon_\eta > 0$ such that, for every $h\in(0, h(\varepsilon_\eta))$,

$$|g(h) - \mathrm{var}(Y_0)| \leqslant \eta\,\mathrm{var}(Y_0) \quad \text{and} \quad |\eta_{R,\underline{n}}(h)| \leqslant \eta.$$

We derive from Equation (22) that

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

Consequently, if $\varepsilon < \frac{\varepsilon_\eta(1-\eta)}{\sqrt{2\alpha R+1}}$, for every $h\in(0, h(\varepsilon_\eta))$, $\mu(h,q^*)^2 \leqslant \mu(h(\varepsilon_\eta),q^*)^2$ and, as soon as $\mu^2(h,q^*) < \varepsilon^2$, one has $|\,\mu(h,q^*)| \geqslant c_R\left(\frac{h^R}{\underline{n}!}\right)^{\alpha}(1-\eta)$, so that

$$\frac{g(h)}{h(\varepsilon^2 - \mu(h,q^*)^2)} \geqslant \frac{\mathrm{var}(Y_0)(1-\eta)}{h\big(\varepsilon^2 - (1-\eta)^2(c_R/\underline{n}!^{\alpha})^2 h^{2\alpha R}\big)}.$$

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

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

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

(ii) Owing to (23), the bias $\mu(h,q)$ is now given by

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

Following the lines of the proof of (i) with $R' = \alpha$ completes the proof. □

Remark 3.9. • *The fact that the function* $\lim_{h\to 0} h\phi^*(h) = \mathrm{var}(Y_0)$ *follows from the* L^2*-strong convergence of* Y_h *toward* Y_0*. Its rate of convergence plays no explicit rôle in this asymptotic rate of the cost as* $\varepsilon \to 0$*. However, this strong rate is important to design a practical stratification among the* R *independent Brownian motions, 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 3.10. *The asymptotic number* N *of simulations given by (8) satisfies*

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

for an ML2R estimator and

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

for an MLMC estimator.

3.3.2 Templates for R-level allocation matrix $\mathbf{T}$

We now specify the allocation matrix $\mathbf{T}$.

MLMC estimator The standard Multilevel Monte Carlo allocation matrix used by [Hei01, Gil08] is derived from the telescopic summation

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

This corresponds in our general framework to the allocation matrix $\mathbf{T}$ defined by $\mathbf{T}^j = e_j - e_{j-1}$ for $j \in \{2, \dots, R\}$ *i.e.*

$$\mathbf{T} = \begin{pmatrix} 1 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -1 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}. \qquad \text{(MLMC)}$$

In this particular case, the resulting upper-bound $\bar{\phi}^*$ of ϕ^* writes

$$\bar{\phi}^*(\pi_0) = \frac{\mathrm{var}(Y_0)}{h} \left(1 + \theta h^{\frac{\beta}{2}} \sum_{j=1}^R \left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right) \sqrt{n_{j-1} + n_j}\right)^2 \qquad (32)$$

with the convention $n_0 = (n_0)^{-1} = 0$.

ML2R estimator The corresponding allocation matrix $\mathbf{T}$ for the ML2R estimator is defined by $\mathbf{T}^j = -\mathbf{W}_j e_{j-1} + \mathbf{W}_j e_j$ for $j \in \{2, \dots, R\}$ with $\mathbf{W}_j = \sum_{k=j}^{R} \mathbf{w}_k$, $\mathbf{w}$ is given by (10) *i.e.*

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

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

$$\bar{\phi}^*(\pi_0) = \frac{\operatorname{var}(Y_0)}{h} \left(1 + \theta h^{\frac{\beta}{2}} \sum_{j=1}^{R} |\mathbf{W}_j| \left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right) \sqrt{n_{j-1} + n_j}\right)^2, \tag{33}$$

In the sequel, we will focus on the above choice (ML2R) for the allocation matrix $\mathbf{T}$ which leads to the ML2R estimator (3) proposed in the introduction. With this allocation matrix (ML2R) the ML2R estimator writes as a weighted version of MLMC

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

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

3.3.3 Bias parameter and depth R optimization (second approach) for geometric refiners

In this second approach, we consider geometric refiners with *root* $M \geqslant 2$ of the form

$$n_i = M^{i-1}, \ i = 1, \dots, R.$$

These are the refiners already considered in regular multilevel Monte Carlo framework.

$$\pi_0(\varepsilon) = \big(h(\varepsilon, M, R(\varepsilon), \mathbf{T}), R(\varepsilon, M, \mathbf{T})\big).$$

Theorem 3.11. *Assume* (SE_β) *holds for* $\beta > 0$.

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

$$\limsup_{\varepsilon\to 0} v(\beta,\varepsilon) \times \inf_{\substack{h\in\mathcal{H}, R\geqslant 2 \\ |\mu(h,R,q^*)|<\varepsilon}} \operatorname{Cost}\left(\bar{Y}_{h,\underline{n}}^{N,q}\right) \leqslant K(\alpha,\beta,M) \tag{34}$$

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

These bounds are achieved with an order

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

satisfying $\lim_{\varepsilon\to 0} R^*(\varepsilon) = +\infty$ *and a bias parameter* $h^* = h^*(\varepsilon, R(\varepsilon))$ *given by* (29). *The finite real constant* $K(\alpha,\beta,M)$ *depends on* M *and on the structural parameters* $\alpha, \beta, V_1, \mathrm{var}(Y_0), \mathbf{h}$, *namely*

$$K(\alpha,\beta,M) = \begin{cases} \frac{2V_1}{\alpha}\left(\frac{\mathbf{W}_\alpha(M)M(1+M)(1+M^{-\frac{1}{2}})^2}{\log(M)}\right) & \text{if } \beta = 1,\\ \frac{\mathrm{var}(Y_0)M}{\mathbf{h}}\left(1+\theta\,\mathbf{h}^{\frac{\beta}{2}}\frac{\mathbf{W}_\alpha(M)M^{\frac{\beta-1}{2}}\sqrt{1+M}(1+M^{-\frac{\beta}{2}})}{1-M^{\frac{1-\beta}{2}}}\right)^2 & \text{if } \beta > 1,\\ V_1\mathbf{h}^{1-\beta}\,\widetilde{c}^{\frac{(1-\beta)}{\alpha}}\left(\frac{\mathbf{W}^2_\alpha(M)M(1+M)(1+M^{-\frac{\beta}{2}})^2}{(M^{\frac{1-\beta}{2}}-1)^2}\right) & \text{if } \beta < 1.\end{cases} \tag{35}$$

(*b*) MLMC estimator: *Assume* $(WE_{\alpha,1})$ *and* $c_1 \neq 0$. *The MLMC estimator (with allocation matrix* $\mathbf{T}$ *defined in* (MLMC)*) satisfies*

$$\limsup_{\varepsilon\to 0} v(\beta,\varepsilon) \times \inf_{\substack{h\in\mathcal{H}, R\geqslant 2\\ |\mu(h,R,q^*)|<\varepsilon}} \mathrm{Cost}\left(\bar{Y}^{N,q}_{h,\underline{n}}\right) \leqslant K(\alpha,\beta,M) \tag{36}$$

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

These bounds are achieved with an order

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

satisfying $\lim_{\varepsilon\to 0} R^*(\varepsilon) = +\infty$ *and a bias parameter* $h^* = h^*(\varepsilon, R(\varepsilon))$ *given by* (30). *The finite real constant* $K(\alpha,\beta,M)$ *depends on* M *and the structural parameters* $\alpha, \beta, V_1, \mathrm{var}(Y_0), \mathbf{h}$, *namely*

$$K(\alpha,\beta,M) = \begin{cases}\left(1+\frac{1}{2\alpha}\right)\frac{V_1}{\alpha^2}\left(\frac{M(1+M)(1+M^{-\frac{1}{2}})^2}{\log(M)^2}\right) & \text{if } \beta = 1,\\ \left(1+\frac{1}{2\alpha}\right)\frac{\mathrm{var}(Y_0)M}{\mathbf{h}}\left(1+\theta\,\mathbf{h}^{\frac{\beta}{2}}\frac{M^{\frac{\beta-1}{2}}\sqrt{1+M}(1+M^{-\frac{\beta}{2}})}{1-M^{\frac{1-\beta}{2}}}\right)^2 & \text{if } \beta > 1,\\ \frac{(1+2\alpha)^{1+\frac{1-\beta}{2\alpha}}}{2\alpha}V_1\mathbf{h}^{1-\beta}|c_1|^{\frac{(1-\beta)}{\alpha}}\left(\frac{M(1+M)(1+M^{-\frac{\beta}{2}})^2}{(M^{\frac{1-\beta}{2}}-1)^2}\right) & \text{if } \beta < 1.\end{cases}$$

Remark 3.12.

- *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* $(c_R)_{R\geqslant 2}$ *in (a) of the above proposition are reasonable though probably impossible to check in practice. In particular, note that as soon as the sequence* $(c_R)_{R\geqslant 2}$ *has at most a polynomial growth as a function of* R, *it satisfies the assumption since* $\tilde{c} = 1$.
- *When* $\tilde{c} = 0$, *one can replace* c_R *in the proof below by* ϵ_0^R *(see also Remark 3.9) and carry on the computations (with* $\widetilde{c} = \epsilon_0$*). This constant has an impact when* $\beta < 1$: *when* $\epsilon_0 \to 0$, $K(\alpha,\beta,M)$ *goes to* 0 *which emphasizes that we are not in the right asymptotics.*
- *If* $\beta = 1$, *ML2R is asymptotically more efficient than MLMC by a factor* $\log(1/\varepsilon) \to +\infty$ *as* $\varepsilon \to 0$. *When* $\beta < 1$, *ML2R (with* $M = 2$*) is asymptotically infinitely 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*

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

It is clear that it is for this setting that ML2R is the most powerful compared to regular MLMC. When $\beta > 1$, *both estimators achieve the same rate* ε^{-2} *as a virtual unbiased Monte Carlo method based on the direct simulation of* Y_0.

Proof. We provide a detailed poof of claim (a), that of (b) following the same lines.
STEP 1: We start from Equation (31) in the proof of Theorem 3.8 which reads

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

with

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

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

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

Note that, by assumption we have $\lim_{R\to+\infty}|c_R|^{-\frac{1}{\alpha R}}=\widetilde{c}^{-\frac{1}{\alpha}}$ with $\widetilde{c}\in(0,+\infty)$. We choose to try to saturate the constraint $h^*\leqslant\mathbf{h}$ which leads to impose formally (for $R\geqslant 2$)

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

(where we temporarily forget that R should be an integer). Let $A=\widetilde{c}^{\frac{1}{\alpha}}\mathbf{h}$. As a consequence, we search for the positive zero $R_+(\varepsilon)$ of the polynomial

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

that is $R_+(\varepsilon)=\frac{1}{2}+\frac{\log(A)}{\log(M)}+\sqrt{\left(\frac{1}{2}+\frac{\log(A)}{\log(M)}\right)^2+2\frac{\log\left(\sqrt{1+4\alpha}/\varepsilon\right)}{\alpha\log M}}$ and denoting $R^*(\varepsilon)=\lfloor R_+(\varepsilon)\rfloor$ we obtain $P(R^*(\varepsilon))\leqslant 0$. Hence, $h^*(\varepsilon,R^*(\varepsilon))=\mathbf{h}e^{\frac{P(R^*)}{R^*}-\frac{P(R_+)}{R_+}}\leqslant\mathbf{h}$.

Let us show that our choice $h^*(\varepsilon,R^*(\varepsilon))$ for the bias parameter (see (29)) is admissible – *i.e.* $\mu(h^*(\varepsilon,R^*(\varepsilon)),R^*(\varepsilon),q^*)^2<\varepsilon^2$ – at least for small enough ε. Elementary computations show that

$$\begin{aligned}\mu\left(h^*(\varepsilon,R^*(\varepsilon))\right)^2 &= (c_{R^*(\varepsilon)}\widetilde{\mathbf{w}}_{R^*(\varepsilon)+1})^2(h^*(\varepsilon,R^*(\varepsilon)))^{2\alpha R^*(\varepsilon)}\\ &= \varepsilon^2 e^{-\alpha R^*(\varepsilon)\log(1+2\alpha R^*(\varepsilon))}\left(1+\eta_{R^*(\varepsilon),\underline{n}}(h^*(\varepsilon,R^*(\varepsilon)))\right)^2.\end{aligned}$$

Our choice for $R^*(\varepsilon)$ implies that $h^*(\varepsilon,R^*(\varepsilon))$ is upper-bounded by $\mathbf{h}$. Claim 6 of Proposition A.2 in Appendix A and the assumption on η_R imply that,

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

As a consequence of the assumption made on the functions η_R, it is clear that $\mu\left(h^*(\varepsilon,R^*(\varepsilon)),R^*(\varepsilon),q^*\right)^2=o(\varepsilon^2)$ 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 that $\lim_{\varepsilon\to 0}\frac{\left(\eta_{R^*(\varepsilon),\underline{n}}(h(\varepsilon,R^*(\varepsilon)))+1\right)^2-1}{2\alpha R^*(\varepsilon)}=0$. We have then proved that

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

with

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

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

$$\left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1}+n_j} = M^{\beta-1} M^{j\frac{1-\beta}{2}}\big(1+M^{-\frac{\beta}{2}}\big)\big(1+M\big)^{\frac{1}{2}}. \tag{37}$$

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

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

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,

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

and, as $R_+^2(\varepsilon) \sim \frac{2}{\alpha \log(M)} \log(1/\varepsilon)$ as $\varepsilon \to 0$, we get (34) with $K(\alpha, 1, M)$ given by (35) keeping in mind that $V_1 = \mathrm{var}(Y_0)\theta^2$.

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

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

which yields (34) with $K(\alpha, \beta, M)$ given by (35).

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

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

As $R_+(\varepsilon)$ satisfies $h^*(\varepsilon, R_+(\varepsilon)) = \mathbf{h}$, we obtain $M^{\frac{R_+(\varepsilon)-1}{2}} = \mathbf{h}\,\tilde{c}^{\frac{1}{\alpha}} \varepsilon^{-\frac{1}{\alpha R_+(\varepsilon)}}$. We have $\varepsilon^{-\frac{1}{\alpha R_+(\varepsilon)}} \sim e^{\sqrt{\frac{\log(M)}{2\alpha}\log(1/\varepsilon)}}$ as $\varepsilon \to 0$. Elementary, although tedious computations yield (34) with $K(\alpha, \beta, M)$ given by (35).

(b) The choice for $R^*(\varepsilon)$ follows from the formal constraint

$$\limsup_{\varepsilon \to 0} \left[h^*(\varepsilon, R^*(\varepsilon)) = (1+2\alpha)^{-\frac{1}{2\alpha}} |c_1|^{-\frac{1}{\alpha}} \varepsilon^{\frac{1}{\alpha}} M^{R^*(\varepsilon)-1}\right] = \mathbf{h}.$$

Then, the proof follows the same lines as that of (a). □

Remark 3.13 (On the constraint $\mathbf{h}$). *In the proof we choose to saturate the constraint $h^* \leqslant \mathbf{h}$. 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 3.11 depends on χ and one verifies the following facts:*

- *When $\beta < 1$, one can write $K(\alpha, \beta, M, \chi) = \chi^{1-\beta} K(\alpha, \beta, M, 1)$ which this time suggests to start the simulation with a small upper bias parameter $\chi < \mathbf{h}$.*

- *When $\beta > 1$, the asymptotic cost of the simulation increases in ε^2 like a (virtual) unbiased one. In that very case, it appears that the asymptotic constant $K(\alpha, \beta, M, \chi)$ can itself be optimized as a function of χ. Namely, if we set*

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

then

$$\chi_{opt} = \beta^{-\frac{2}{\beta+1}} \kappa_2^{-\frac{1}{\beta+1}} \qquad \textit{and} \qquad K(\alpha, \beta, M, \chi_{opt}) = (\beta+1)^2 \beta^{-\frac{2}{\beta+1}} \kappa_1 \kappa_2^{\frac{1}{\beta+1}}.$$

- *When $\beta = 1$, the asymptotic constant $K(\alpha, \beta, M, \chi)$* 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.*

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 [TT90] on weak error expansion, followed by the case of non smooth functions in [BT96] 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 $(W_t)_{t \geqslant 0}$ 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)*

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t,\ t \in [0, T],\ X_0 = x_0 \in \mathbf{R}^d, \tag{38}$$

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

$$\bar{X}^h_t = X_0 + \int_0^t b\big(\bar{X}^h_{\underline{s}}\big)\mathrm{d}s + \int_0^t \sigma\big(\bar{X}^h_{\underline{s}}\big)\mathrm{d}W_s, \qquad \textit{where} \qquad \underline{s} = kh \ \textit{on}\ [kh, (k+1)h).$$

(*a*) *Smooth setting (Talay-Tubaro [TT90]): 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 $R \in \mathbf{N}^*$*

$$\mathbf{E}\left[f(\bar{X}^h_T)\right] - \mathbf{E}\left[f(X_T)\right] = \sum_{k=1}^{R} c_k h^k + O\left(h^{R+1}\right), \tag{39}$$

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

(*b*) *(Hypo-)Elliptic setting (Bally-Talay [BT96]): If b and σ are infinitely differentiable with bounded partial derivatives and if σ is uniformly elliptic in the sense that*

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

or, more generally, if (b, σ) satisfies the strong Hörmander hypo-ellipticity assumption, then (39) 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 [KM02], in a uniformly elliptic setting, see also [Guy06] 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 more technical than a key assumption. For a recent review on weak error, we refer to [JKH11].

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 $h \in \mathcal{H} = \{\frac{T}{n},\, n \geqslant 1\}$. We set $Y_h = f(\bar{X}^h_T)$ and $Y_0 = f(X_T)$ for a smooth enough function f with polynomial growth. The above theorem says that condition ($WE_{\alpha,\bar{R}}$) 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 background 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 ($WE_{\alpha,\bar{R}}$) 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 ($q \geqslant 2$), it requires the simulation of Lévy areas, for which there is no known efficient method (except in dimension 2). In a recent work [GS12], Giles and Szpruch introduce a suitable *antithetic multilevel correction estimator* which avoids the simulation of these Lévy areas. This approach could be easily combined with our weighted version of MLMC.

Note that in the $\beta > 1$ case, Rhee and Glynn introduce in [RG12] 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), $\beta = 1$, with $\mathcal{H} = \{\frac{T}{n},\, n \geqslant 1\}$, (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 $F(\mathbf{w}) = f\Big(\int_0^T \mathbf{w}(s)ds\Big)$ 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 $\mathbf{R}^d$ which show that they satisfy ($WE_{\alpha,\bar{R}}$) with $\alpha = 1$ and $\bar{R} = 1$ (see *e.g.* [LT01, Gob00]). More recently, new results on first order weak error expansions have been obtained for functionals of the form $F(\mathbf{w}) = f\big(\mathbf{w}(T), \sup_{t\in[0,T]} \mathbf{w}(t)\big)$ (see [GHM09] and [AJKH13]). Thus, for the weak error expansion, it is shown in [AJKH13] that, for every $\eta > 0$, there exists a real constant $C_\eta > 0$ such that

$$\big|\mathbf{E}\big[f\big(X_T, \sup_{t\in[0,T]} X_t\big)\big] - \mathbf{E}\big[f\big(\bar{X}^n_T, \sup_{t\in[0,T]} \bar{X}^n_t\big)\big]\big| \leqslant \frac{C_\eta}{N^{\frac{2}{3}-\eta}}.$$

For a review of recent results on approximation of solutions of SDEs, we again refer to [JKH11].

Remark 4.2. *Note that, as concerned 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 [BK12]. A similar approach applied to the ML2R estimator with allocation matrix* (ML2R) *(weighted version of MLMC), should yield a similar Central Limit Theorem.*

4.2 Nested Monte Carlo

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

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

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 $[f]_{\text{Lip}}$.

Such quantities often appear in financial application, like compound option pricing or risk estimation (see [BDM11]) and in actuarial sciences (see [DL09]) where nested Monte Carlo is widely implemented. The idea of replacing conditional expectations by Monte Carlo estimates also appears in [BSD13] where authors devise 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

$$\mathbf{E}\left[X \mid Y\right](\omega) = \Big(\mathbf{E}\left[F(Z,y)\right]\Big)_{|y=Y(\omega)} = \int_{\mathbf{R}^{q_Z}} F(z, Y(\omega))\mathbf{P}_Z(dz).$$

To comply with the multilevel framework, we set

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

where $(Z_k)_{k\geqslant 1}$ 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 ($WE_{\alpha,\bar{R}}$) 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 [JJ09, GJ10] 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.*

The proof follows form the two lemmas below.

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

$$\left\|Y_0 - Y_h\right\|_2^2 \leqslant [f]_{\mathrm{Lip}}^2 \left(\left\|X\right\|_2^2 - \left\|\mathbf{E}\left[X \mid Y\right]\right\|_2^2\right) h \tag{40}$$

so that $(Y_h)_{h\in\mathcal{H}}$ satisfies (SE_β) with $\beta = 1$.

Proof. Let $h = 1/K$ and set $\mathbf{E}_Y[X] = \mathbf{E}\left[X \mid Y\right]$ for convenience. By Lipschitz continuity,

$$\begin{aligned}\left\|Y_0 - Y_h\right\|_2^2 &= \left\|f\Big(\mathbf{E}_Y[X]\Big) - f\left(\frac{1}{K}\sum_{k=1}^{K} F(Z_k, Y)\right)\right\|_2^2 \\ &\leqslant [f]_{\mathrm{Lip}}^2 \left\|\mathbf{E}_Y[X] - \frac{1}{K}\sum_{k=1}^{K} F(Z_k, Y)\right\|_2^2\end{aligned}$$

and

$$\begin{aligned}\left\|\mathbf{E}_Y[X] - \frac{1}{K}\sum_{k=1}^{K} F(Z_k, Y)\right\|_2^2 = \left\|\mathbf{E}_Y[X]\right\|_2^2 + \frac{1}{K^2}\sum_{k=1}^{K}\left\|F(Z_k, Y)\right\|_2^2 - \frac{2}{K}\sum_{k=1}^{K}\mathbf{E}\big[\mathbf{E}_Y[X]\, F(Z_k, Y)\big] \\ + \frac{2}{K^2}\sum_{1\leqslant j<k\leqslant K}\mathbf{E}\big[F(Z_j, Y)F(Z_k, Y)\big].\end{aligned}$$

Conditioning on Y gives

$$\mathbf{E}\big[\mathbf{E}_Y[X]\,F(Z_k,Y)\big] = \mathbf{E}\big[\mathbf{E}_Y[X]\,\mathbf{E}_Y[F(Z_k,Y)]\big] = \big\|\mathbf{E}_Y[X]\big\|_2^2$$

and $\mathbf{E}\big[F(Z_j,Y)F(Z_k,Y)\big] = \big\|\mathbf{E}_Y[X]\big\|_2^2$ for $j\neq k$. Plugging these two identities in the first expansion finally yields

$$\big\|Y_0 - Y_h\big\|_2^2 \leqslant \frac{[f]_{\rm Lip}^2}{K}\Big(\big\|F(Z,Y)\big\|_2^2 - \big\|\mathbf{E}_Y[X]\big\|_2^2\Big).$$

□

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

$$\forall h\in\mathcal{H},\quad \mathbf{E}\,[Y_h] = \mathbf{E}\,[Y_0] + \sum_{r=1}^{R-1} c_r h^r + O\big(h^R\big). \tag{41}$$

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

Proof. Let $K\geqslant 1$ and $\widetilde X_k = F(Z_k,Y) - \mathbf{E}_Y[F(Z_k,Y)] = X_k - \mathbf{E}_Y(X_k)$, $k=1,\dots,K$. By the multinomial formula we get

$$(\widetilde X_1+\cdots+\widetilde X_K)^k = \sum_{k_1+\cdots+k_K=k}\frac{k!}{k_1!\cdots k_K!}\widetilde X_1^{k_1}\cdots\widetilde X_K^{k_K}$$

so that, taking conditional expectation given Y, yields

$$\mathbf{E}_Y\Big[(\widetilde X_1+\cdots+\widetilde X_K)^k\Big] = k!\sum_{k_1+\cdots+k_K=k}\prod_{i=1}^K\frac{\mathbf{E}_Y\big[\widetilde X^{k_i}\big]}{k_i!}$$

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

$$\mathbf{E}_Y\Big[(\widetilde X_1+\cdots+\widetilde X_K)^k\Big] = k!\sum_{k_1+\cdots+k_K=k,\,k_i\neq 1}\prod_{i=1}^K\frac{\mathbf{E}_Y\big[\widetilde X^{k_i}\big]}{k_i!}.$$

Let $I=I(k_1,\dots,k_K)=\{i\,|\,k_i\neq 0,1\}$. It is clear that $1\leqslant |I|\leqslant k/2$. By symmetry, we have now that

$$\begin{aligned}\sum_{k_1+\cdots+k_K=k,\,k_i\neq 1}\prod_{i=1}^K\frac{\mathbf{E}_Y\big[\widetilde X^{k_i}\big]}{k_i!} &= \sum_{1\leqslant\ell\leqslant(k/2)\wedge K}\;\sum_{I\subset\{1,\dots,K\},|I|=\ell,\sum_{i\in I}k_i=k,k_i\geqslant 2}\;\prod_{i=1}^K\frac{\mathbf{E}_Y\big[\widetilde X^{k_i}\big]}{k_i!}\\ &= \sum_{1\leqslant\ell\leqslant k/2}\binom{K}{\ell}\sum_{\sum_{1\leqslant i\leqslant\ell}k_i=k-2\ell}\;\prod_{i=1}^{\ell}\frac{\mathbf{E}_Y\big[\widetilde X^{2+k_i}\big]}{(2+k_i)!}.\end{aligned}$$

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

$$\begin{aligned}\mathbf{E}_Y\big[Y_h\big] &= \mathbf{E}_Y\big[Y_0\big] + \sum_{k=1}^{2R-1}\frac{f^{(k)}\big(\mathbf{E}_Y[X]\big)}{k!K^k}\mathbf{E}_Y(\widetilde X_1+\cdots+\widetilde X_K)^k + \mathbf{R}_{2R-1}(Y)\\ &= \mathbf{E}_Y[Y_0] + \sum_{k=1}^{2R-1}\frac{f^{(k)}\big(\mathbf{E}_Y[X]\big)}{k!K^k}\sum_{1\leqslant\ell\leqslant(k/2)\wedge K}\binom{K}{\ell}a_{k,\ell} + \mathbf{R}_{2R-1}(Y)\end{aligned}$$

where

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

and

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

By the Marcinkiewicz-Zygmund Inequality we get

$$\begin{aligned}|\mathbf{R}_{2R-1}(Y)| &\leqslant (B^{MZ}_{2R})^{2R}\frac{\|f^{(2R)}\|_{\sup}}{(2R)!}\frac{1}{K^{2R}}\mathbf{E}_Y\Big[\big|\widetilde{X}_1^2+\cdots+\widetilde{X}_K^2\big|^{R}\Big]\\ &\leqslant \|f^{(2R)}\|_{\sup}\frac{(B^{MZ}_{2R})^{2R}}{(2R)!}\frac{1}{K^{R}}\mathbf{E}_Y\big[\widetilde{X}^{2R}\big]\end{aligned}$$

where $B^{MZ}_p = 18\frac{p^{\frac{3}{2}}}{(p-1)^{\frac{1}{2}}}$, $p>1$ (see [Shi96] p.499). Now we write the polynomial $x(x-1)\cdots(x-\ell+1)$ 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,

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

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

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

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

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

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 $\left(\frac{(B^{MZ}_{2R})^{2R}}{(2R)!}\right)^{\frac{1}{R}} \sim (36R)^2\left(\frac{2R}{e}\right)^{-2} \sim 18\,e^2$ as $R\to+\infty$ owing to Stirling's formula. This suggests that, e.g. 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 Theorems 3.6, 3.8 and 3.11 that the asymptotic optimal parameters (as ε goes to 0) R, h, q and N depend on structural parameters α, β, V_1, c_1, $\mathrm{var}(Y_0)$ and $\mathbf{h}$ (recall that $\theta = \sqrt{V_1/\mathrm{var}(Y_0)}$). Note we do not have optimized the design of the multilevel estimators, namely the allocation matrix $\mathbf{T}$ and the refiners n_i, $i=2,\ldots,R$. We propose in this Section a numerical procedure to choose a good value of M in the case $n_i = M^{i-1}$.

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

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

Since Y_0 cannot be simulated at a reasonable 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 for not too small step h. Then, one derives from Minkowski's Inequality that, for every integer $M \geqslant 1$,

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

so that

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

As a consequence, if we choose $M = M_{\max}$ large enough (see (46) below), we are led to consider the following estimator

$$\widehat{V}_1(h) = \big(1 + M_{\max}^{-\frac{\beta}{2}}\big)^{-2} h^{-\beta}\|Y_h - Y_{\frac{h}{M_{\max}}}\|_2^2. \tag{42}$$

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

Note that, even in a crude Monte Carlo method, these structural parameters are useful (and sometimes necessary) to deal with the bias error (see Proposition 2.2).

Design of the Multilevel

The allocation 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 allocation matrix $\mathbf{T}$ satisfying (17) like $\mathbf{T}^j = -\,\mathbf{w}_j\, e_1 + \mathbf{w}_j\, e_j$ for $j \in \{2, \dots, R\}$ which reads

$$\mathbf{T} = \begin{pmatrix} 1 & -\,\mathbf{w}_2 & -\,\mathbf{w}_3 & \cdots & -\,\mathbf{w}_R \\ 0 & \mathbf{w}_2 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{w}_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \mathbf{w}_R \end{pmatrix}. \tag{43}$$

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

$$\mathbf{T} = \begin{pmatrix} \widetilde{\mathbf{W}}_1 & 0 & \cdots & \cdots & \cdots & 0 \\ -\widetilde{\mathbf{W}}_1 & \widetilde{\mathbf{W}}_2 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & -\widetilde{\mathbf{W}}_{R-2} & \widetilde{\mathbf{W}}_{R-1} & 0 \\ 0 & \cdots & \cdots & \cdots & -\widetilde{\mathbf{W}}_{R-1} & 1 \end{pmatrix} \quad \text{where} \quad \widetilde{\mathbf{W}}_j = \sum_{k=1}^{j} \mathbf{w}_k\,. \tag{44}$$

The refiners can be specified by users but it turns out that the parametrized family $n_i = M^{i-1}$, $i = 1, \dots, R$ ($M \in \mathbf{N}$, $M \geqslant 2$) seems the best compromise between variance control and implementability. The parameter α being settled, all the related quantities like $(\mathbf{W}_i(R, M))_{1\leqslant i\leqslant M}$ 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 *i.e.* c_r, $r = 2,\ldots,R-1$ (instead of R). One easily shows that, if $\big(\mathbf{w}_r^{(R-1)}\big)_{r=1,\ldots,R-1}$ 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

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

Asymptotic optimal parameters

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 the table 5.1 for the (ML2R) estimator and in the table 5.2 for the (MLMC) estimator.

R	$\left\lfloor \frac{1}{2} + \frac{\log(\widetilde{c}^{\frac{1}{\alpha}}\mathbf{h})}{\log(M)} + \sqrt{\left(\frac{1}{2} + \frac{\log(\widetilde{c}^{\frac{1}{\alpha}}\mathbf{h})}{\log(M)}\right)^2 + 2\frac{\log(A/\varepsilon)}{\alpha\log(M)}}\right\rfloor,\quad A = \sqrt{1+4\alpha}$
h^{-1}	$\left\lceil (1+2\alpha R)^{\frac{1}{2\alpha R}} \varepsilon^{-\frac{1}{\alpha R}} M^{-\frac{R-1}{2}} \right\rceil_{\mathbf{h}}$
q	$q_1 = \mu^*(1+\theta h^{\frac{\beta}{2}})$ $q_j = \mu^*\theta h^{\frac{\beta}{2}} \left(\lvert \mathbf{W}_j(R,M)\rvert \frac{n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}}{\sqrt{n_{j-1}+n_j}} \right),\ j = 2,\ldots,R;\ \sum_{1\leqslant j\leqslant R} q_j = 1$
N	$\left(1+\frac{1}{2\alpha R}\right) \frac{\operatorname{var}(Y_0)\left(1+\theta h^{\frac{\beta}{2}}\sum_{j=1}^{R}\lvert\mathbf{W}_j(R,M)\rvert \left(n_{j-1}^{-\frac{\beta}{2}}+n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1}+n_j}\right)^2}{\varepsilon^2 \sum_{j=1}^{R} q_j(n_{j-1}+n_j)}$

Table 5.1: Optimal parameters for the ML2R estimator.

R	$\left\lfloor 1 + \frac{\log(\lvert c_1\rvert^{\frac{1}{\alpha}}\mathbf{h})}{\log(M)} + \frac{\log(A/\varepsilon)}{\alpha\log(M)} \right\rfloor,\quad A = \sqrt{1+2\alpha}$
h^{-1}	$\left\lceil (1+2\alpha)^{\frac{1}{2\alpha}} \varepsilon^{-\frac{1}{\alpha}} \lvert c_1\rvert^{\frac{1}{\alpha}} M^{-(R-1)} \right\rceil_{\mathbf{h}}$
q	$q_1 = \mu^*(1+\theta h^{\frac{\beta}{2}})$ $q_j = \mu^*\theta h^{\frac{\beta}{2}} \left(\frac{n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}}{\sqrt{n_{j-1}+n_j}} \right),\ j = 2,\ldots,R;\ \sum_{1\leqslant j\leqslant R} q_j = 1$
N	$\left(1+\frac{1}{2\alpha}\right) \frac{\operatorname{var}(Y_0)\left(1+\theta h^{\frac{\beta}{2}}\sum_{j=1}^{R}\left(n_{j-1}^{-\frac{\beta}{2}}+n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1}+n_j}\right)^2}{\varepsilon^2 \sum_{j=1}^{R} q_j(n_{j-1}+n_j)}$

Table 5.2: Optimal parameters for the MLMC estimator.

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

$$C_\varepsilon(M) = \mathrm{Cost}(\bar{Y}^{N(M),q(M)}_{h(M),\underline{n}}) = N(M)\,\kappa(h(M), R(M), q(M)), \tag{45}$$

where $\kappa(h, R, q) = \frac{1}{h}\sum_{j=1}^{R} q_j \sum_{i=1}^{R} n_i \mathbf{1}_{\{\mathbf{T}^j_i \neq 0\}}$ (in the framework of Section 4.1) and $\kappa(h, R, q) = \frac{1}{h}\sum_{j=1}^{R} q_j \max_{1\leqslant i\leqslant R} n_i \mathbf{1}_{\{\mathbf{T}^j_i \neq 0\}}$ (in the framework of Section 4.2). This function can be optimized for likely values of M. In numerical experiments we consider

$$M = \underset{M\in\{2,\dots,M_{\max}\}}{\operatorname{argmin}} \quad C_\varepsilon(M) \quad \text{with} \quad M_{\max} = 10. \tag{46}$$

Simulating consistent Brownian increments 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 $[\frac{(k-1)T}{n}, \frac{kT}{n}]$, $k = 2, \dots, n$. 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 $m = gcd(n_1, \dots, n_R)$. One can also produce once and for all an *abacus* of coefficients to compute by induction the needed increments from small subintervals up to the root interval of length $\frac{T}{n}$. This is done *e.g.* in [Pag07] up to $R = 5$ for $\alpha = 1$ and up to $R = 3$ for $\alpha = \frac{1}{2}$.

5.2 Methodology

We compare the two MLMC and ML2R estimators for different biased problems. In the sequel, we consider the allocation matrix (ML2R) for the ML2R estimator. After a crude evaluation of $\mathrm{var}(Y_0)$ and V_1 (using (42)) we compute the "optimal" parameter M solution to (46). The others parameters are specified according to Tables (5.1) and (5.2) with $\tilde{c} = c_1 = 1$. In the numerical simulations we do not round down R, we round to the nearest integer.

The empirical bias error of the estimator $\bar{Y}^{N,q}_{h,\underline{n}}$ is estimated by

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

using $L = 256$ independent replications, where $I_0 = \mathbf{E}[Y_0]$ is the true value. To compute this bias error, it is necessary to use many replications of the estimator.

Let $\bar{\nu}_N$ be the empirical (unitary) variance of the estimator $\bar{Y}^{N,q}_{h,\underline{n}}$ defined by

$$\bar{\nu}_N = \sum_{j=1}^{R} \frac{1}{N_j(N_j - 1)} \sum_{k=1}^{N_j} \left(\langle \mathbf{T}^j, Y^{(j),k}_{h,\underline{n}} \rangle - m_{N_j} \right)^2,$$

here $m_{N_j} = \dfrac{1}{N_j}\displaystyle\sum_{k=1}^{N_j} \langle \mathbf{T}^j, Y^{(j),k}_{h,\underline{n}} \rangle$. Note that $\bar{\nu}_N$ is an estimator of $\mathrm{var}(\bar{Y}^{N,q}_{h,\underline{n}})$ that we compute online to obtain a (biased) confidence interval for one run of $\bar{Y}^{N,q}_{h,\underline{n}}$. Since we use L independent replications of $\bar{Y}^{N,q}_{h,\underline{n}}$ (to obtain the bias error), then we consider the empirical means over the $L = 256$ runs

$$\tilde{\nu}_L = \frac{1}{L}\sum_{\ell=1}^{L} \bar{\nu}^{(\ell)}_N,$$

as an estimator of $\operatorname{var}(\bar{Y}^{N,q}_{h,\underline{n}})$. 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 then given by

$$\tilde{\varepsilon}_{L} = \sqrt{(\tilde{\mu}_{L})^2 + \tilde{\nu}_{L}}. \tag{47}$$

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).

5.3 Euler scheme of a geometric Brownian motion

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 $t = T$:

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

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 $e^{-rT}\mathbf{E}\left[\varphi(S_T)\right]$ and the price of a *path dependent* option with functional payoff φ is given by $e^{-rT}\mathbf{E}\left[\varphi((S_t)_{t\in[0,T]})\right]$. Since $(S_t)_{t\in[0,T]}$ is solution to the diffusion SDE

$$dS_t = S_t(r\mathrm{d}t + \sigma dW_t), \quad S_0 = s_0 > 0,$$

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 $(\bar{S}_{kh})_{0\leqslant k\leqslant n}$, $h = \frac{T}{n}$ (even if we are aware that S_T is simulatable). The bias parameter set $\mathcal{H}$ is then defined by $\mathcal{H} = \left\{T/n,\ n \geqslant 1\right\}$ and $\mathbf{h} = T$.

Although nobody would adopt any kind of Monte Carlo simulation to compute option price in this model in practice 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 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).

5.3.1 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 $\mathcal{C}^\infty_b$ and the payoff function is Lipschitz continuous), one has $\alpha = \beta = 1$. The parameters $\theta = \sqrt{V_1/\operatorname{var}(Y_0)}$ and $\operatorname{var}(Y_0)$ have been roughly estimated following the procedure (42) on a sample of size 100 000 described in Section 5.1 leading to the values $V_1 \simeq 56$ and $\operatorname{var}(Y_0) \simeq 876$ (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 $I_0 = 29.4987$.

The results are summarized in Table 5.3 for the ML2R estimator and in Table 5.4 for the MLMC estimator.

As an example, the third line of the Table 5.3 reads as follows: for a prescribed RMSE error $\varepsilon = 2^{-3} = 0.125$, the ML2R estimator $\bar{Y}^{N,q}_{h,\underline{n}}$ (with allocation 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 stratification weights q_i (not reported in this Table) are such that the numerical cost $\operatorname{Cost}(\bar{Y}^{N,q}_{h,\underline{n}}) \simeq 709\,800$. For such parameters, the empirical RMSE $\tilde{\varepsilon}_{L} \simeq 0.0928$ and the computational time of $Y^{N,q}_{h,\underline{n}} \simeq 0.559$ seconds. The empirical bias error $\tilde{\mu}_{L}$ is reported in the 5th column (*bias*) and the empirical unitary variance $\tilde{\nu}_{L}$ is reported in the 6th column (*var*). Recall that $\tilde{\varepsilon}_{L} = \sqrt{(\tilde{\mu}_{L})^2 + \tilde{\nu}_{L}}$.

Note first that, as expected, the depth parameter $R \geqslant 2$ and the numerical cost $\operatorname{Cost}(\bar{Y}^{N,q}_{h,\underline{n}})$ grow slower for ML2R than for MLMC as ε goes to 0. Consequently, regarding the CPU–time for a

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$–error	time (s)	bias	var	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 5.3: Call option ($\alpha = 1, \beta = 1$): Parameters and results of ML2R estimator.

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$–error	time (s)	bias	var	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	$5.17{\cdot}10^{+08}$	$1.62{\cdot}10^{+09}$

Table 5.4: Call option ($\alpha = 1, \beta = 1$): Parameters and results of MLMC estimator.

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.* $\tilde{\varepsilon}_L \leqslant \varepsilon$ (see also Figure 5(a) in Appendix C). We can conclude that the automatic tuning of the algorithm parameters is satisfactory for both estimators.

In Figure 1 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 $\tilde{\varepsilon}_L$ shows that ML2R goes from 1.28 up to 2.8 faster, within the range of our simulations. In Appendix C, Figure 5(b) represents the product (CPU–time)$\times\varepsilon^2$ as a function of ε. Note that this function is almost constant (as expected) for both estimators.

5.3.2 Lookback option $\alpha = 0.5$, $\beta = 1$

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

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

where $\lambda \geqslant 1$. 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 functional Lipschitz continuous payoff, (SE_β) 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 others structural parameters yields $\text{var}(Y_0) \simeq 41$ and $V_1 \simeq 3.58$ (and then $\theta \simeq 0.29$). Both estimators are implemented using the automatic tuning previously exposed.

The results are summarized in Table 5.5 for the ML2R and in Table 5.6 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 ($k = 9$), idem for 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 (see 2(a) in Appendix C).

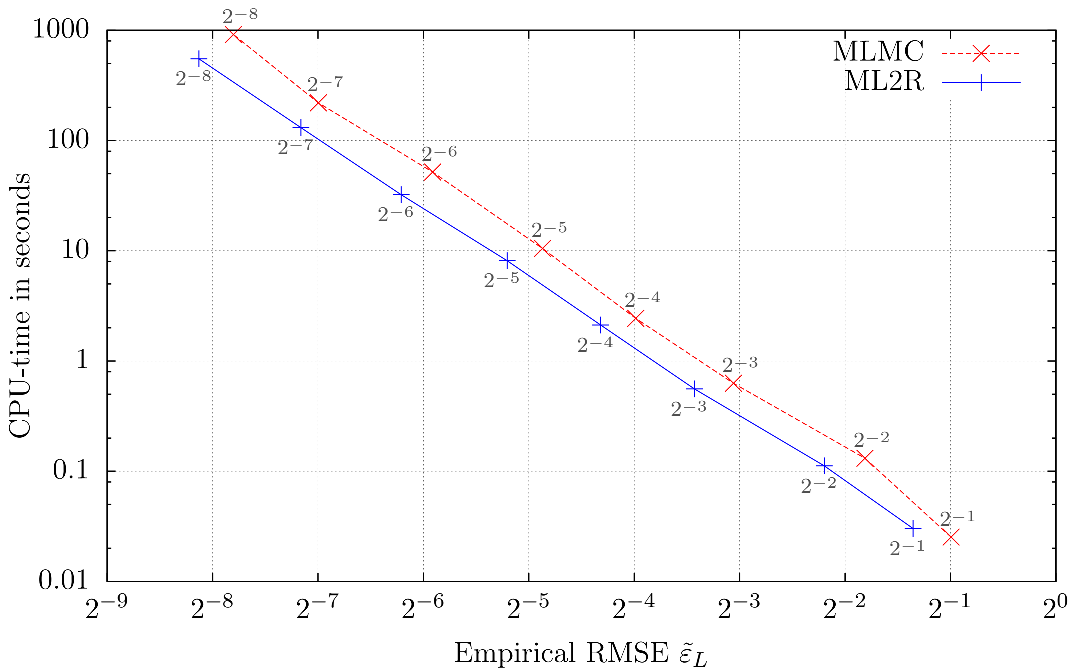


Figure 1: Call option in a Black-Scholes model. CPU–time (y–axis, log scale) as a function of $\tilde{\varepsilon}_L$ (x–axis, $\log_2$ scale).

One observes that the $\mathbf{L}^2$–error of ML2R has a very small bias $\tilde{\mu}_L$ (5th column) due to the particular choice of the weights $(\mathbf{W}_i)_{1\leqslant i\leqslant R}$.

Figure 2 provides a graphical representation of the performance of both estimators, now as a function of the empirical RMSE $\widetilde{\varepsilon}$. It shows that ML2R is faster then MLMC by a factor that goes from 18 up to 48 within the range of our simulations. Additional Figures 6(a) and 6(b) are given in Appendix C.

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$–error	time (s)	bias	var	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 5.5: Lookback option ($\alpha = 0.5, \beta = 1$): Parameters and results of ML2R estimator.

5.3.3 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

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

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 obtained 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 $\max_{k\in\{1,\dots,n\}} \bar{S}_{kh}$. This allows us to compare both estimators in the case $\beta = 0.5$. As in the Lookback option we assume

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$–error	time (s)	bias	var	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	$2.71{\cdot}10^{+08}$	$1.66{\cdot}10^{+09}$

Table 5.6: Lookback option ($\alpha = 0.5, \beta = 1$): Parameters and results of MLMC estimator.

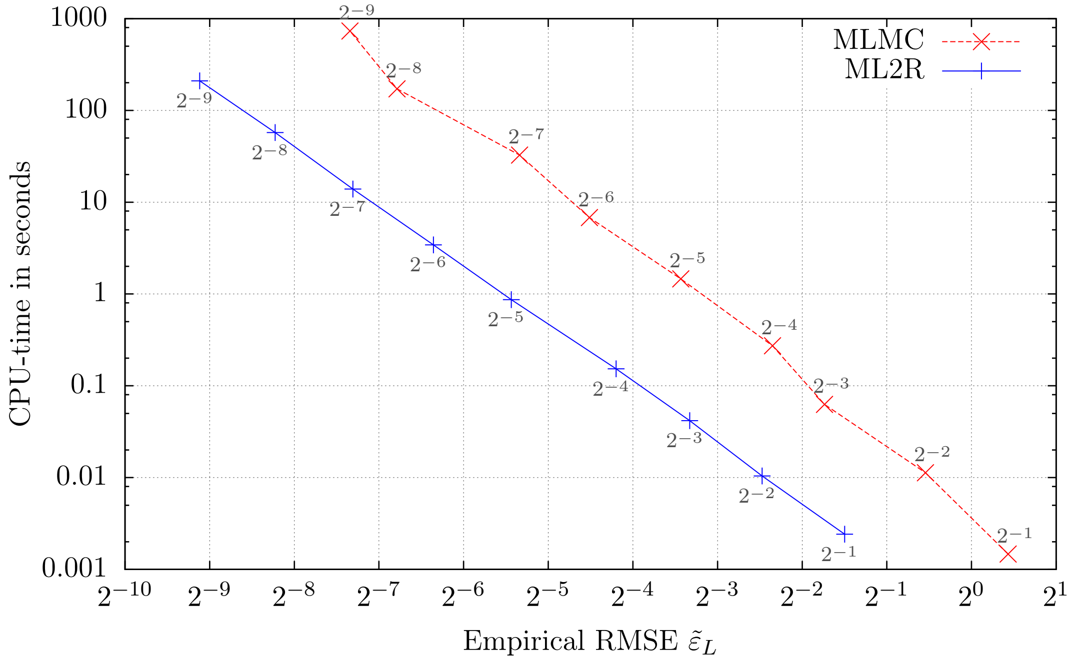


Figure 2: Lookback option in a Black-Scholes model. CPU–time (y–axis, log scale) as a function of $\tilde{\varepsilon}_L$ (x–axis, $\log_2$ scale).

that ($WE_{\alpha,\bar{R}}$) 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 5.7 for ML2R and in Table 5.8 for MLMC.

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$–error	time (s)	bias	var	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 5.7: Barrier option ($\alpha = 0.5, \beta = 0.5$): Parameters and results of ML2R estimator.

See Figure 3 for a graphical representation. First, 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 (see Figure 7 in

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$–error	time (s)	bias	var	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	$4.37{\cdot}10^{+08}$	$1.67{\cdot}10^{+10}$

Table 5.8: Barrier option ($\alpha = 0.5, \beta = 0.5$): Parameters and results of MLMC estimator.

Appendix C) which agrees with the theoretical asymptotic rates from Theorem 3.11.

In fact, in this highly biased example with slow strong convergence, the ratio Cost(MLMC)/ Cost(ML2R) as a function of the prescribed $\varepsilon = 2^{-k}$ goes from 1.1 ($k = 2$) up to 22 ($k = 8$), idem for the ratio between CPU–times. 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.

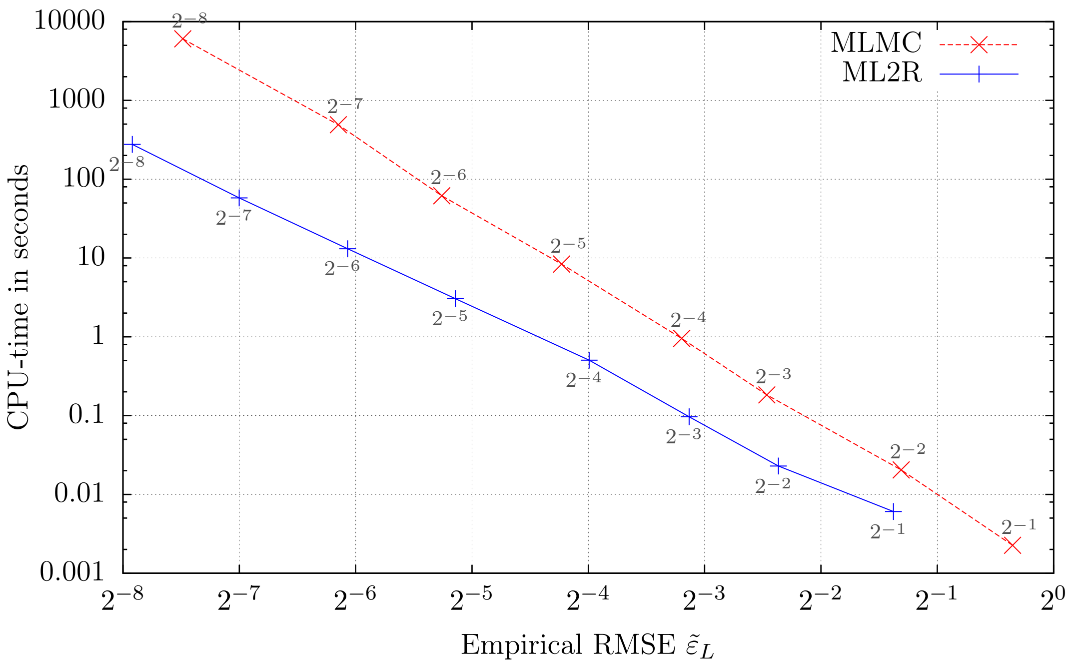


Figure 3: Barrier option in a Black-Scholes model. CPU–time (y–axis, log scale) as a function of $\tilde{\varepsilon}_L$ (x–axis, $\log_2$ scale).

5.4 Nested Monte Carlo for compound option pricing

A compound option is simply an option on an option. The exercise 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). On the first expiration date T_1, the holder has the right to sell a new *Call* option using 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

$$(K_1 - \mathbf{E}\left[(S_{T_2} - K_2)_+ \mid S_{T_1}\right])_+$$

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

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

where $(Z^k)_{k\geqslant 1}$ 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 the underlying process $(S_t)_{t\in[0,T_2]}$ is not discretized in time. The bias error is then due to the inner Monte Carlo estimator of the conditional expectation.

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 suggest that (SE_β) and $(WE_{\alpha,\bar{R}})$ are satisfied with $\beta = \alpha = 1$. A crude computation of others 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 5.9 for ML2R and in Table 5.10 for MLMC.

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$–error	time (s)	bias	var	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 5.9: Nested compound option ($\alpha = 1$, $\beta = 1$): Parameters and results of ML2R estimator.

k	$\varepsilon = 2^{-k}$	$\mathbf{L}^2$–error	time (s)	bias	var	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	$1.07{\cdot}10^{+08}$	$6.06{\cdot}10^{+08}$

Table 5.10: Nested compound option ($\alpha = 1$, $\beta = 1$): Parameters and results of MLMC estimator.

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

A Appendix

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

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

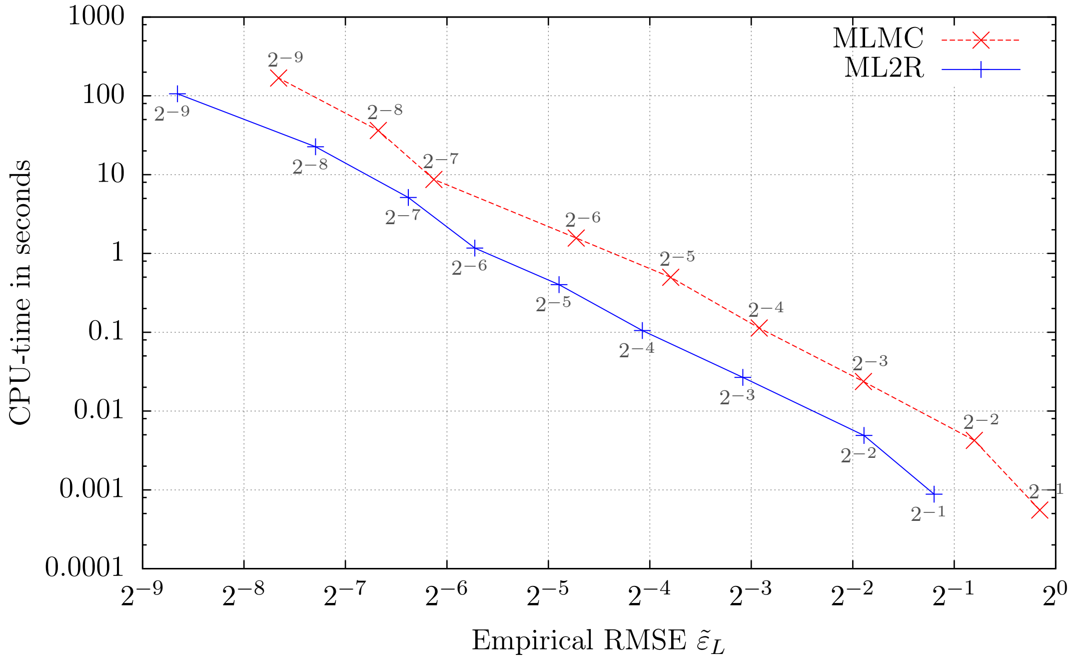


Figure 4: Nested compound option in a Black-Scholes model. CPU–time (y–axis, log scale) as a function of $\tilde{\varepsilon}_L$ (x–axis, $\log_2$ scale).

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

(*b*) *Furthermore*

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

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 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 $\det(V) = \prod_{1\leqslant j<k\leqslant n}(a_k - a_j)$. So we have for every $i \in \{1, \ldots, R\}$

$$\mathbf{w}_i = \frac{\displaystyle\prod_{1\leqslant j<k\leqslant R; j,k\neq i}(a_k - a_j)\prod_{1\leqslant j<i}(-a_j)\prod_{i<k\leqslant R} a_k}{\displaystyle\prod_{1\leqslant j<k\leqslant R}(a_k - a_j)} = \frac{\displaystyle\prod_{1\leqslant j<i}(-a_j)\prod_{i<k\leqslant R} a_k}{\displaystyle\prod_{1\leqslant j<i}(a_i - a_j)\prod_{i<k\leqslant R}(a_k - a_i)}$$

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

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

and

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

which completes the proof.

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

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

□

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 $\mathbf{w}_i$, $i = 1, \ldots, R$:*

$$\mathbf{w}_i = \mathbf{w}_i(R, M) = (-1)^{R-i} \frac{M^{-\frac{\alpha}{2}(R-i)(R-i+1)}}{\prod_{1\leqslant j\leqslant i-1}(1 - M^{-j\alpha}) \prod_{1\leqslant j\leqslant R-i}(1 - M^{-j\alpha})}, \; i = 1, \ldots, R.$$

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:*

$$\sup_{R\in\mathbf{N}^*} \sum_{i=1}^{R-1} |\mathbf{w}_i(R, M)| \leqslant \frac{M^{-\alpha}}{\pi^2_{\alpha,M}} \sum_{k\geqslant 0} M^{-\alpha\frac{k(k+3)}{2}} \quad \text{and} \quad 1 \leqslant \mathbf{w}_R(R, M) \leqslant \frac{1}{\pi_{\alpha,M}}$$

where $\pi_{\alpha,M} = \prod_{k\geqslant 1}(1 - M^{-\alpha k})$.

4. *Asymptotics of the coefficients $\mathbf{w}_i$ when $M \to +\infty$:*

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

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 $\mathbf{W}_1 = 1$,*

$$\max_{1\leqslant i\leqslant R} |\mathbf{W}_i(R, M)| \leqslant \mathbf{W}_\alpha(M) := \frac{M^{-\alpha}}{\pi^2_{\alpha,M}} \sum_{k\geqslant 0} M^{-\alpha\frac{k(k+3)}{2}} + \frac{1}{\pi_{\alpha,M}} \tag{48}$$

and

$$\max_{1\leqslant i\leqslant R} |\mathbf{W}_i(R, M) - 1| \leqslant \mathbf{W}_\alpha(M) - 1 \sim M^{-\alpha} \to 0 \quad \text{as } M \to +\infty.$$

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*

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

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

$$\frac{|\mathbf{w}_r(R, M)|}{n_r^{\alpha R}} \leqslant \frac{M^{-\frac{\alpha}{2}((R-r)(R-r+1)+2(r-1)R)}}{\pi^2_{\alpha,M}}$$

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

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

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

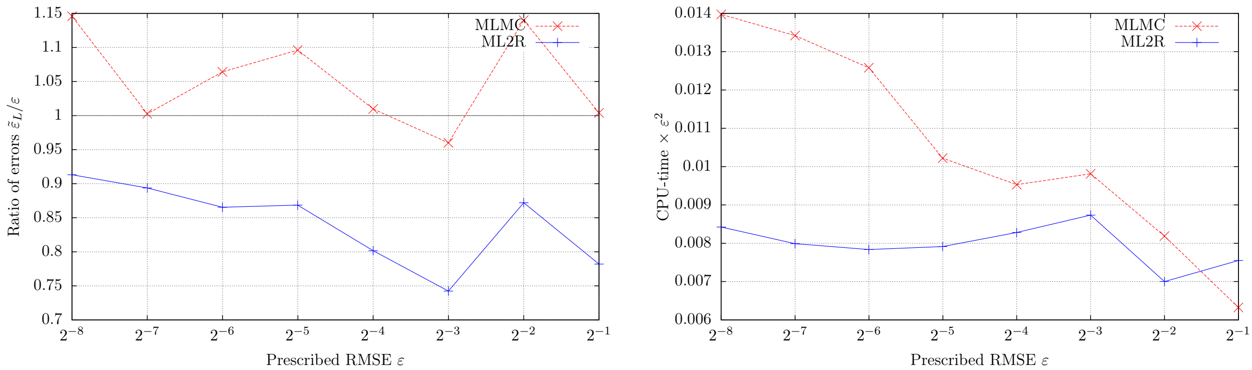
B Appendix: sketch of proof of Propositions 2.2 and 2.4

The multistep Richardson-Romberg estimator with the formal framework of Section 3, is characterized by the allocation 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 $\mathbf{E}\big[\bar{Y}^N_{h,\underline{n}}\big]$ follows from Proposition 2.4. No stratification is needed here since only one Brownian motion is involved. The proof of Theorem 3.8 applies here with $q = (1, 0, \ldots, 0)$. Furthermore

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

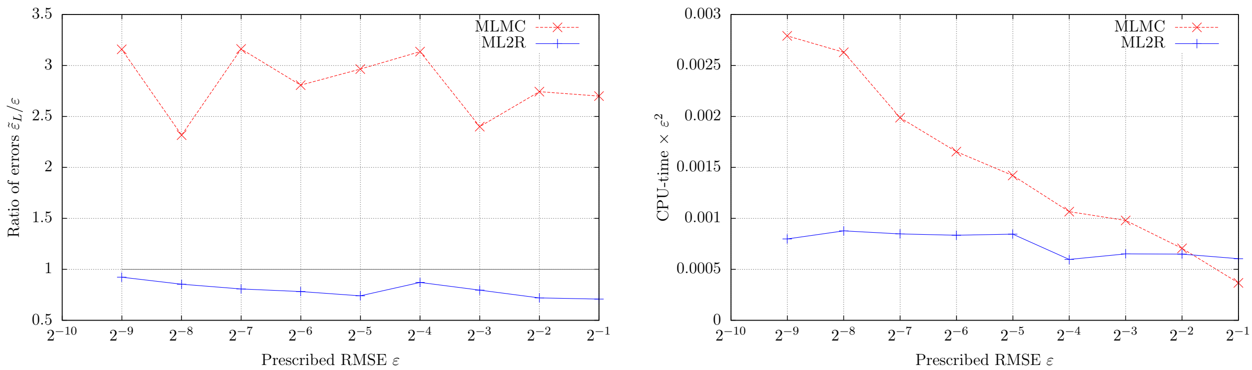
since $Y^1_{h,\underline{n}} \to Y_0 \mathbf{1}$ in $\mathbf{L}^2$ and $\sum_{i=1}^R \mathbf{w}_i = 1$. □

C Appendix: Additional Figures



(a) Ratio $\varepsilon_L/\varepsilon$ as a function of ε (b) CPU–time $\times\ \varepsilon^2$ as a function of ε

Figure 5: Call option in a Black-Scholes model ($\log_2$ scale for the x–axis)



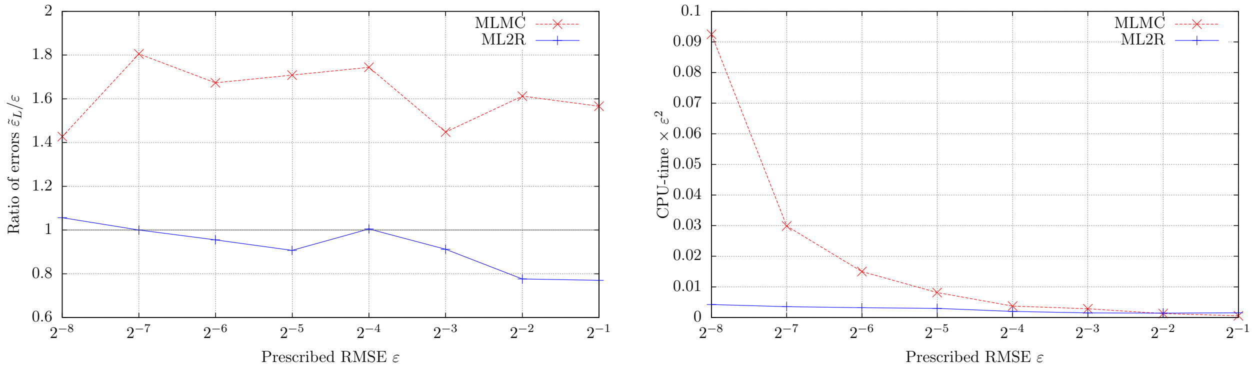
(a) Ratio $\varepsilon_L/\varepsilon$ as a function of ε (b) CPU–time $\times\ \varepsilon^2$ as a function of ε

Figure 6: Lookback option in a Black-Scholes model ($\log_2$ scale for the x–axis)

References

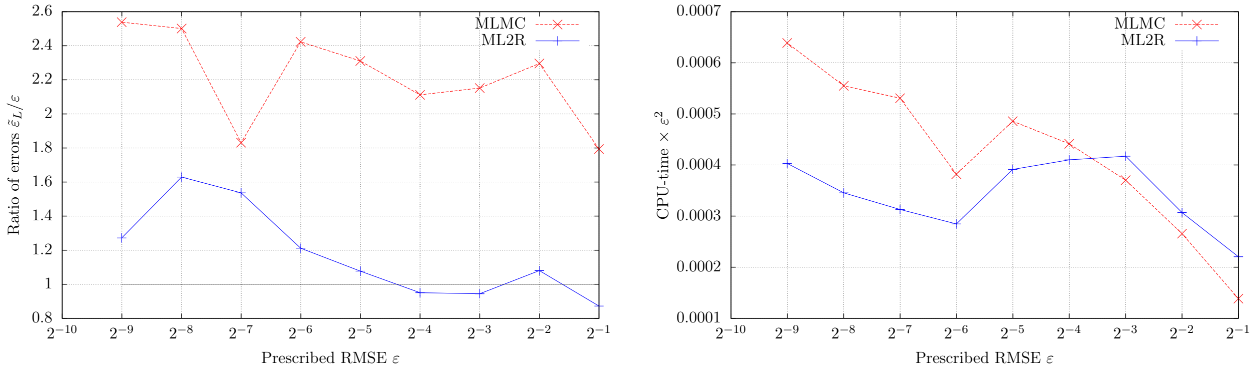
[AJKH13] A. Alfonsi, B. Jourdain, and A. Kohatsu-Higa. Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme. *The Annals of Applied Probability*, (To appear in), 2013.

[BDM11] M. Broadie, Y. Du, and C.C. Moallemi. Efficient risk estimation via nested sequential simulation. *Management Science*, 57(6):1172–1194, 2011.



(a) Ratio $\varepsilon_L/\varepsilon$ as a function of ε (b) CPU–time × ε^2 as a function of ε

Figure 7: Barrier option in a Black-Scholes model ($\log_2$ scale for the x–axis)



(a) Ratio $\varepsilon_L/\varepsilon$ as a function of ε (b) CPU–time × ε^2 as a function of ε

Figure 8: Nested compound option in a Black-Scholes model ($\log_2$ scale for the x–axis)

[BK12] M. Ben Alaya and A. Kebaier. Central Limit Theorem for the Multilevel Monte Carlo Euler Method and Applications to Asian Options. Technical report, LAGA, Univ. Paris 13 (France), May 2012.

[BLS13] A. Barth, A. Lang, and C. Schwab. Multilevel Monte Carlo method for parabolic stochastic partial differential equations. *BIT*, 53(1):3–27, 2013.

[BSD13] D. Belomestny, J. Schoenmakers, and F. Dickmann. Multilevel dual approach for pricing American style derivatives. *Finance and Stochastics*, 17(4):717–742, 2013.

[BT96] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probab. Theory Related Fields*, 104(1):43–60, 1996.

[Der11] S. Dereich. Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *The Annals of Applied Probability*, 21(1):283–311, 2011.

[DG95] D. Duffie and P.W. Glynn. Efficient Monte Carlo simulation of security prices. *The Annals of Applied Probability*, pages 897–905, 1995.

[DH11] S. Dereich and F. Heidenreich. A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stochastic Processes and their Applications*, 121(7):1565–1587, 2011.

[DL09] L. Devineau and S. Loisel. 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(17):188–221, 2009.

[GHM09] M.B. Giles, D.J. Higham, and X. Mao. Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff. *Finance and Stochastics*, 13(3):403–413, 2009.

[Gil08] M.B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[GJ10] M.B. Gordy and S. Juneja. Nested simulation in portfolio risk measurement. *Management Science*, 56(10):1833–1848, 2010.

[Gob00] E. Gobet. Weak approximation of killed diffusion using Euler schemes. *Stochastic Process. Appl.*, 87(2):167–197, 2000.

[GS12] M.B. Giles and L. Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. 2012.

[Guy06] J. Guyon. Euler scheme and tempered distributions. *Stochastic Processes and their Applications*, pages 877–904, 2006.

[Hei01] S. Heinrich. Multilevel Monte Carlo methods. In *Large-scale scientific computing*, pages 58–67. Springer, 2001.

[JJ09] L. Jeff Hong and S. Juneja. Estimating the mean of a non-linear function of conditional expectation. In M.D. Rossetti, R.R. Hill, B. Johansson, A. Dunkin, and R.G. Ingalls, editors, *Proceedings of the 2009 Winter Simulation Conference* , pages 1223–1236. IEEE, 2009.

[JKH11] B. Jourdain and A. Kohatsu-Higa. A review of recent results on approximation of solutions of stochastic differential equations. In *Stochastic analysis with financial applications*, volume 65 of *Progr. Probab.*, pages 121–144. Birkhäuser/Springer Basel AG, Basel, 2011.

[Keb05] A. Kebaier. Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. *Ann. Appl. Probab.*, 15(4):2681–2705, 2005.

[KM02] V. Konakov and E. Mammen. Edgeworth type expansions for Euler schemes for stochastic differential equations. *Monte Carlo Methods and Applications*, 8(3):271–286, 2002.

[LT01] B. Lapeyre and E. Temam. Competitive Monte Carlo methods for the pricing of Asian Options. *J. of Comput. Fin.*, 5(1):39–59, 2001.

[Pag07] G. Pagès. Multi-step Richardson-Romberg extrapolation: remarks on variance control and complexity. *Monte Carlo Methods Appl.*, 13(1):37–70, 2007.

[RG12] C. Rhee and P.W. Glynn. A new approach to unbiased estimation for SDE's. In *Proceedings of the Winter Simulation Conference*, page 17. Winter Simulation Conference, 2012.

[Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.

[TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.