Multilevel Richardson-Romberg extrapolation

Vincent Lemaire, Gilles Pagès

Multilevel Richardson-Romberg extrapolation

Vincent Lemaire*, Gilles Pagès†

December 19, 2013

Abstract

We propose and analyze a Multilevel Richardson-Romberg ($MLRR$) estimator which combines the higher order bias cancellation of the Multistep Richardson-Romberg ($MSRR$) method introduced in [Pag07] and the variance control resulting from the stratification in the Multilevel Monte Carlo ($MLMC$) method (see [Hei01]). Thus we show that, in standard frameworks like discretization schemes of diffusion processes, an assigned quadratic error ε can be obtained with our $MLRR$ estimator with a global complexity of $\log(1/\varepsilon)/\varepsilon^2$ instead of $(\log(1/\varepsilon))^2/\varepsilon^2$ with the standard $MLMC$ method, at least when the weak error $\mathbf{E}[Y_h] - \mathbf{E}[Y_0]$ of the biased implemented estimator of Y_h can be expanded at any order in h. We analyze and compare these estimators on two numerical problems: the classical vanilla option pricing by MC simulation and the less classical Nested Monte Carlo simulation.

Keywords: Multilevel Monte Carlo methods ; Richardson-Romberg Extrapolation ; Multi-Step ; 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 method introduced by [Hei01] and popularized for financial applications 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 which can be represented as an expectation of a non-simulatable random variable Y_0 (at least at a reasonable unitary cost). Both above methods can then take advantage of the existence of a family of (easily) simulatable random variables Y_h, $h > 0$, which strongly approximate Y_0 as $h \to 0$ with a bias $\mathbf{E}[Y_h] - \mathbf{E}[Y_0]$ that can be expanded as a polynomial function of h (or h^α, $\alpha > 0$).

However, the multilevel Monte Carlo method does not fully take advantage of the existence of such an expansion beyond the first order where the multistep Richardson-Rombeg 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 we have a family $(Y_h)_{h\geqslant 0}$ of real-valued random variables in $\mathbf{L}^2(\mathbf{P})$ associated to a non degenerate $Y_0 \in \mathbf{L}^2(\mathbf{P})$ and satisfying $\lim_{h\to 0} \mathbf{E}\left[(Y_h - Y_0)^2\right] = 0$ where h takes values in an admissible subset of parameters $\mathcal{H} \subset (0, \mathbf{h}]$ having 0 as a limiting value and such that $\frac{\mathcal{H}}{n} \subset \mathcal{H}$ for every integer $n \geqslant 1$. We also assume that $\mathbf{h} \in \mathcal{H}$. Usually the random variable Y_h results from a time discretization scheme of parameter h or from an inner approximation in a Nested Monte Carlo so that we will speak of h as 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 with in mind to make it inverse linear in h.

*Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, UPMC, Case 188, 4 pl. Jussieu, F-75252 Paris Cedex 5, France, E-mail: `vincent.lemaire@upmc.fr`

†Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, UPMC, 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 $h \in (0, \mathbf{h}]$ and to consider a standard Monte Carlo estimator 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 well-known control of the renormalized *statistical error* $\sqrt{N}(I_N^{(h)} - \mathbf{E}[Y_h])$ which behaves as a centered Gaussian variable with variance $\mathrm{var}(Y_h)$. On the other hand, there is a *bias error* due to the approximation of I_0 by $I_h = \mathbf{E}[Y_h]$. Note that the bias error is also known as *weak error* when Y_h is a function of a time discretization scheme of a stochastic differential equation solution with step h. In many applications the bias error can be expanded as follows

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

where α is a positive real parameter (usually $\alpha = \frac{1}{2}, 1$ or 2). In this paper we will take account of this error expansion and provide a very efficient estimator which can be viewed as a coupling between a Multilevel estimator [Hei01, Keb05, Gil08] and a Multistep Richardson-Romberg extrapolation [Pag07].

We first present a formal description (in our abstract framework) of the original Multilevel Monte Carlo as described *e.g.* in [Gil08]. The main idea is to use the following telescopic summation

$$\mathbf{E}[Y_{h_R}] = \mathbf{E}[Y_h] + \sum_{j=1}^R \mathbf{E}\left[Y_{h_j} - Y_{h_{j-1}}\right],$$

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

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

where for every j, $(Y^{(j),k})_{k\geqslant 1}$ is a sequence of independent copies of $Y^{(j)}$. The analysis of the computational complexity and the study of the bias and the variance of this estimator will appear as particular case of a generalized multilevel paradigm introduced and analyzed in section 3. This framework, like the original multilevel MC simulation highly relies on the combination of a strong rate of approximation of Y_0 by Y_h and a weak error $\mathbf{E}[Y_h] - \mathbf{E}[Y_0]$ expansion. This method 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 principle of Richardson-Romberg extrapolation is to take advantage of the first order expansion (1) of the bias error to reduce the order (1 in h if $\alpha = 1$ in the above *crude* Monte Carlo simulation). To this end we consider one biased family denoted $Y = (Y_h)_{h\in\mathcal{H}}$ attached to the one random variable Y_0. The Richardson-Romberg Monte Carlo estimator writes then

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

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 satisfy 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 (as h decreases to 0) satisfies $\lim_{h\to 0} \mathrm{var}(I^N_{h,\frac{h}{2}}) = \frac{1}{N}\mathrm{var}(Y_0)$ 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 is considered in [Pag07] in the framework of discretization of diffusion processes.

The aim of this paper is to show that an appropriate combination of the Multilevel (ML) Monte Carlo estimator and the Multistep Richardson-Romberg (MRR) estimator outperform the standard Multilevel MC.

The paper is organized as follows: in Section 2, we propose a general parametrized framework to formalize the optimization of a *biased* Monte Carlo (MC) simulation based on the mean squared (quadratic) error minimization (also known as *MSE*). This *parameter* is a vector π usually representative of the bias expansion order R, the strong approximation rate, the selected bias refiners, the stratification strategy, but also the "inner" simulation size (for nested MC), etc. Crude MC and multistep RR 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 general family extended multilevel estimators attached to a "free" "allocation" matrix $\mathbf{T}$ and a set of *refiners*. Among them, we describe in more details our proposal: the hybrid *Multilevel Richardson-Romberg extrapolation* estimators, but also the standard multilevel MC estimators. Then we develop our optimization strategy: an order R being fixed, minimizing the *effort* (complexity $\times$ variance) by an appropriate stratification strategy, then, when the refiners are in geometric progression, the order R of the procedure is in turn optimized as a function of the target quadratic error ε. In Section 4 are presented two typical fields of application: the time discretization of stochastic processes (Euler scheme) and the nested MC 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 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 $x \in \mathbf{R}$, $\lceil x \rceil$ 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 $n \leqslant x < n + 1$.
• 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 $\lim_n \varepsilon = 1$, $a_n = O(b_n)$ 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 biased *linear* estimator. For a fixed parameter $\pi \in \Pi \subset \mathbf{R}^d$, we consider a linear statistical estimator I^N_π of $I_0 \in \mathbf{R}$ as N tends to infinity. By linear we mean on the one hand that

$$\mathbf{E}\left[I^N_\pi\right] = \mathbf{E}\left[I^1_\pi\right], \ N \geqslant 1,$$

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

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

where $\kappa(\pi) = \mathrm{Cost}(I^1_\pi)$ 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:

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

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 $\{I_\pi^N, \pi \in \Pi, N \in \mathbf{N}^*\}$ namely the estimator minimizing the computational cost for a given error $\varepsilon > 0$. In the sequel we treat N as a continuous variable *i.e.* as if $N \in \mathbf{R}_+$. A natural choice for measuring the random error $I_\pi^N - I_0$ is to consider the mean squared error $\mathbf{E}\big[(I_\pi^N - I_0)^2\big] = \big\|I_\pi^N - I_0\big\|_2^2$ (the squared $\mathbf{L}^2$-norm error), often denoted *MSE* in Statistics. Our aim is to minimize the cost of the simulation for a given target error, say $\varepsilon > 0$. So we are looking for the following generic problem,

$$\big(\pi(\varepsilon), N(\varepsilon)\big) = \underset{\|I_\pi^N - I_0\|_2 \leqslant \varepsilon}{\operatorname{argmin}} \ \mathrm{Cost}(I_\pi^N). \tag{3}$$

To deal with this minimization problem we introduce the definition of the *effort* $\phi\big(\pi\big)$ 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*

$$\phi\big(\pi\big) = \nu(\pi)\,\kappa(\pi). \tag{4}$$

By definition of I_π^N we have that

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

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

$$\mathrm{Cost}(I_\pi^N) = N\frac{\phi\big(\pi\big)}{\nu(\pi)}. \tag{5}$$

- If the estimators I_π^N are *unbiased i.e.* $\mathbf{E}\big[I_\pi^N\big] = I_0$ for any $\pi \in \Pi$, 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 (3) then reads

$$\pi(\varepsilon) = \frac{1}{\varepsilon^2}\operatorname*{argmin}_{\pi\in\Pi}\phi\big(\pi\big), \quad N(\varepsilon) = \frac{\nu(\pi(\varepsilon))}{\varepsilon^2} = \frac{\phi(\pi(\varepsilon))}{\kappa(\pi(\varepsilon))\varepsilon^2}. \tag{6}$$

  Consequently, the most performing estimator I_π^N is characterized as a minimizer of the effort $\phi\big(\pi\big)$ as defined above.

- When the estimators I_π^N, $\pi \in \Pi$, 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

$$\mu^2(\pi) = \big(\mathbf{E}\big[I_\pi^N\big] - I_0\big)^2 = \big(\mathbf{E}\big[I_\pi^1\big] - I_0\big)^2$$

  denotes the bias error (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 (3) reads

$$\pi(\varepsilon) = \underset{\pi\in\Pi,\,\mu(\pi)<\varepsilon}{\operatorname{argmin}}\left(\frac{\phi\big(\pi\big)}{\varepsilon^2 - \mu^2(\pi)}\right), \quad N(\varepsilon) = \frac{\nu(\pi(\varepsilon))}{\varepsilon^2 - \mu^2(\pi(\varepsilon))} = \frac{\phi(\pi(\varepsilon))}{\kappa(\pi(\varepsilon))(\varepsilon^2 - \mu^2(\pi(\varepsilon)))}. \tag{7}$$

2.2 Assumptions on weak and strong approximation errors

We consider a family $(Y_h)_{h\in\mathcal{H}}$ of real-valued random variables associated to a random variable Y_0, indexed by $\mathcal{H}\subset(0,\mathbf{h}]$. The index set $\mathcal{H}$ is a *consistent* set of *step* parameters in the sense that $\mathbf{h}\in\mathcal{H}$ 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 string 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 [Hei01]):

Bias error expansion (weak error rate):

$$\exists\,\alpha>0,\bar{R}\geqslant 1,\quad \mathbf{E}\left[Y_h\right]=\mathbf{E}\left[Y_0\right]+c_1h^{\alpha}+c_2h^{2\alpha}+\cdots+c_{\bar{R}}h^{\alpha\bar{R}}(1+\eta_{\bar{R}}(h)),\quad \lim_{h\to 0}\eta_{\bar{R}}(h)=0.\ (WE_{\alpha,\bar{R}})$$

Strong approximation error assumption:

$$\exists\,\beta>0,\quad \left\|Y_h-Y_0\right\|_2^2=\mathbf{E}\left[\left|Y_h-Y_0\right|^2\right]\leqslant V_1h^{\beta}.\qquad (SE_{\beta})$$

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).

Consistency of strong and weak errors imply in what follows that, if $c_1\neq 0$, $\beta\leqslant 2\alpha$. In the sequel we will consider a free parameter $R\in\left\{1,\dots,\bar{R}\right\}$ fir which $(WE_{\alpha,R})$ is always satisfied (with the same coefficients c_r up to $r=R$).

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 supposed to be defined on the same probability space. Note that, since the above properties (SE_{β}) and $(WE_{\alpha,\bar{R}})$, $\bar{R}\in\mathbf{N}\cup\{\infty\}$, only depend on the distribution of $(Y_h)_{h\in\mathcal{H}}$, all these copies will also satisfy these two properties.

We associate to a family $(Y_h)_{h\in\mathcal{H}}$ and a given bias parameter $h\in\mathcal{H}$, an $\mathbf{R}^R$-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, one chooses the refiners n_i as $n_i=M^{i-1}$ where M is an integer greater than 1. 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* $(WE_{\alpha,1})$ *and* $c_1\neq 0$. *The Monte Carlo estimators of* $\mathbf{E}\left[Y_0\right]$ *defined by*

$$\bar{Y}_h^N=\frac{1}{N}\sum_{k=1}^{N}Y_h^k,\ h>0,$$

where $\left(Y_h^k\right)_{k=1,\dots,N}$, *is an* i.i.d. N*-sample of* Y_h, *satisfy*

$$\mu(h)=c_1h^{\alpha}(1+\eta_1(h)),\quad \kappa(h)=\frac{1}{h},\quad \phi(h)=\frac{\mathrm{var}(Y_h)}{h}$$

and, for a prescribed L^2-error $\varepsilon > 0$, the optimal parameters $h^*(\varepsilon)$ and $N^*(\varepsilon)$ solution to (3) are

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

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. □

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

2.4 Multistep Richardson-Romberg extrapolation (MRR)

In [Pag07] is proposed a so-called *Multistep Richardson-Romberg estimator* 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 $n_i = i$, $i = 1,\dots,R$, (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 of them 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 $Y_h = \bar{X}^{(h)}$, $h \in \{\frac{T}{n}, n \geqslant 1\}$, makes up an admissible family in the above sense as will be seen further on in more details.

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

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

This solution $\mathbf{w}$ has a closed form given by Cramer's rule (see Lemma A.1 in Appendix A for more details).

$$\forall i \in \{1,\dots,R\}, \quad \mathbf{w}_i = \frac{(-1)^{R-i} n_i^{\alpha(R-1)}}{\prod_{1\leqslant j<i} (n_i^\alpha - n_j^\alpha) \prod_{i<j\leqslant R} (n_j^\alpha - n_i^\alpha)}. \tag{8}$$

We also derive the following identity of interest

$$\widetilde{\mathbf{w}}_{R+1} := \sum_{i=1}^R \frac{\mathbf{w}_i}{n_i^{\alpha R}} = \frac{(-1)^{R-1}}{\underline{n}!^\alpha}. \tag{9}$$

Note that all coefficients $\mathbf{w}_i$ depend on the order R of the combined extrapolation. For the standard choices $n_i = i$ or $n_i = M^{i-1}$, $i = 1,\dots,R$, we obtain the following expressions

$$\mathbf{w}_i = \begin{cases} \dfrac{(-1)^{R-i} i^{\alpha R}}{\prod_{j=0}(i^\alpha - j^\alpha) \prod_{i+1}^R (j^\alpha - i^\alpha)} & \text{if } n_j = j,\ j \in \{1,\dots,R\} \\ \dfrac{(-1)^{R-i} M^{-\frac{\alpha}{2}(R-i)(R-i+1)}}{\prod_{j=1}^{i-1}(1-M^{-j\alpha}) \prod_{j=1}^{R-i}(1-M^{-j\alpha})} & \text{if } n_j = M^{j-1},\ j \in \{1,\dots,R\} \end{cases} \tag{10}$$

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

Assume now that $(WE_{\alpha,\bar{R}})$ holds. In order to design a estimator which kills the bias put to order R, we focus on the random variable resulting from the linear combination $\langle \mathbf{w}, Y_{h,\underline{n}} \rangle$.

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

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

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 get

$$\mathbf{E}\big[\langle \mathbf{w}, Y_{h,\underline{n}} \rangle\big] = \mathbf{E}\left[Y_0\right] + c_{_R} \widetilde{\mathbf{w}}_{_{R+1}} h^{\alpha R} \big(1 + \eta_{R,\underline{n}}(h)\big) \tag{11}$$

where

$$\eta_{R,\underline{n}}(h) = \frac{1}{\widetilde{\mathbf{w}}_{_{R+1}}} \sum_{r=1}^{R} \frac{\mathbf{w}_r}{n_r^{\alpha R}} \eta_R\Big(\frac{h}{n_r}\Big) \to 0 \quad \text{as} \quad h \to 0. \tag{12}$$

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

$$\bar{Y}^N_{h,\underline{n}} = \frac{1}{N} \sum_{k=1}^{N} \langle \mathbf{w}, Y^k_{h,\underline{n}} \rangle = \langle \mathbf{w}, \frac{1}{N} \sum_{k=1}^{N} Y^k_{h,\underline{n}} \rangle \tag{13}$$

where $\left(Y^k_{h,\underline{n}}\right)_{k=1,\dots,N}$, is an i.i.d. *$N$-sample 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 L^2-error $\varepsilon > 0$ and a fixed $R \geqslant 2$, the optimal parameters $h^*(\varepsilon)$ and $N^*(\varepsilon)$ solution of (3) are

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

Furthermore, we have

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

Proof. The proof is postponed to Annex 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. The choice of the refiners in [Pag07], namely $n_i = i$, $i = 1, \dots, R$, is justified by the control of the ratio $\dfrac{|\underline{n}|}{\underline{n}!^{\frac{1}{R}}}$: 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 the above asymptotics which shows that when R increases as a function of the cost/the quadratic error, the multistep method behaves like an unbiased Monte Carlo simulation, one observes in practice that the factor $|\underline{n}|$ in front of $\operatorname{var}(Y_0)$ 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 $\mathrm{var}(\frac{X+X'}{2}) = \frac{1}{2}\mathrm{var}(X)$, combined with an appropriate stratification strategy to control the complexity of the resulting estimator. So, let us consider now R *independant* copies $(Y^{(j)}_{h,\underline{n}})$, $j = 1, \ldots, R$ of the random vector $Y_{h,\underline{n}}$ and the linear combination

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

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

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

Under Assumption ($WE_{\alpha,\bar{R}}$),

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

The convergence also holds (without rate) as soon as Y_h strongly converges toward Y_0 (in L^2).

As emphasized further on, we will also need that all column vectors $\mathbf{T}^j$, $j \neq 1$ have zero sum. In turn, this leads to introduce the notion of Multilevel estimator(s) as a family of stratified estimators of $\mathbf{E}\, Y_0$ attached to the random vectors $\langle \mathbf{T}^j, Y^{(j)}_{h,\underline{n}} \rangle$, $j = 1, \ldots, R$. This leads to the following definitions.

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

$$\mathbf{T}^1 = e_1, \quad \sum_{i,j=1}^{d} \mathbf{T}^j_i = 1 \quad \textit{and} \quad \sum_{i=1}^{R} \mathbf{T}^j_i = 0, \ j = 2, \ldots, R. \tag{14}$$

Definition 3.2 (General Multilevel estimator)**.** *Let $R \geqslant 2$. A Multilevel estimator of order R attached to a stratification strategy $q = (q_1, \ldots, q_R)$ with $q_j > 0$, $j = 1, \ldots, R$ and $\sum_j q_j = 1$ and an allocation matrix $\mathbf{T}$ is defined for every integer $k \geqslant 1$ by*

$$\bar{Y}^{N,q}_{h,\underline{n}} = \sum_{j=1}^{R} \frac{1}{N_j} \sum_{k=1}^{N_j} \langle \mathbf{T}^j, Y^{(j),k}_{h,\underline{n}} \rangle \tag{15}$$

where $\left(Y^{(j),k}_{h,\underline{n}}\right)_{k=1,\ldots,N}$ is an i.i.d. *N-sample of $Y^{(j)}_{h,\underline{n}}$ and, for all $j \in \{1, \ldots, R\}$, $N_j = \lceil q_j N \rceil$ (allocated budget to compute $\mathbf{E}\left[\langle \mathbf{T}^j, Y^{(j)}_{h,\underline{n}} \rangle\right]$).*

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

$$\sum_{j=1}^{R} \mathbf{T}^j = e_R, \ \textit{where } e_R = (0, 0, \ldots, 1),$$

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

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

$$\sum_{j=1}^{R} \mathbf{T}^j = \mathbf{w}, \text{ where } \mathbf{w} \text{ is the unique solution to (8),}$$

the estimator is called a **Multilevel Richardson-Romberg 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.*

Within the abstract framework of a parametrized Monte Carlo simulation described in Section 2.1, the structure parameter π of the multilevel estimator $(\bar{Y}^{N,q}_{h,\underline{n}})_{N\geqslant 1}$ defined by (15) is

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

Cost, complexity and effort of a Multilevel estimator

In order to optimize (minimize) the effort $\phi(\pi)$ of the estimator (15), let us evaluate its unitary computational complexity. For a simulation size N, the cost satisfies

$$\mathrm{Cost}(\bar{Y}^{N,q}_{h,\underline{n}}) = \sum_{j=1}^{R} N_j \sum_{i=1}^{R} \frac{1}{h} n_i \mathbf{1}_{\{\mathbf{T}^j_i \neq 0\}} = N\,\kappa(\pi) \tag{16}$$

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

$$\kappa(\pi) = \frac{1}{h} \sum_{j=1}^{R} q_j \sum_{i=1}^{R} n_i \mathbf{1}_{\{\mathbf{T}^j_i \neq 0\}}. \tag{17}$$

However, it may happen, like for nested Monte Carlo (see Section 4.2), that the internal consistency of the family Y_h leads to spare the computational cost, the computational complexity being entirely due to the most refined "scheme". In such a case one has

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

The effort of such a Multilevel estimator is given by

$$\phi(\pi) = \nu(\pi)\,\kappa(\pi) = \left(\sum_{j=1}^{R} \frac{1}{q_j} \mathrm{var}\left(\langle \mathbf{T}^j, Y^{(j)}_{h,\underline{n}} \rangle \right) \right) \kappa(\pi). \tag{18}$$

Bias error of a Multilevel estimator

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

Proposition 3.4. *(i)* Multilevel Richardson-Romberg estimator: *Assume ($WE_{\alpha,\bar{R}}$). Let $R \in \{2, \ldots, \bar{R}\}$ be the order of a Multilevel Richardson-Romberg estimator. For any admissible stratification strategy $q = (q_1, \ldots, q_R)$, the bias error reads*

$$\mu(\pi_0, q) = (-1)^{R-1} c_R \left(\frac{h^R}{\underline{n}!} \right)^{\alpha} \left(1 + \eta_{R,\underline{n}}(h) \right) \tag{19}$$

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 (12)) and η_R is defined in $(WE)_{\alpha,R}$).

(*ii*) Multilevel Monte Carlo estimator: *Assume* $(WE_{\alpha,1})$. *For any admissible stratification strategy* $q=(q_1,\dots,q_{R})$ *of a Multilevel Monte Carlo estimator of order* $R \geqslant 2$, *the bias error reads*

$$\mu(\pi_0, q) = c_1 \left(\frac{h}{n_{R}}\right)^{\alpha} \left(1 + \eta_1 \left(\frac{h}{n_{R}}\right)\right) \tag{20}$$

where η_1 *is defined in* $(WE_{\alpha,1})$.

Toward to the optimal parameters

The optimization problem (7) is not attainable, so we decompose it in two successive steps:

Step 1: Minimize 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

$$q^* = \operatorname*{argmin}_{q\in\mathcal{S}_+(R)} \bar{\phi}(\pi_0, q), \quad \text{where} \quad \phi(\pi) \leqslant \bar{\phi}(\pi), \quad \text{and} \quad \phi^*(\pi_0) = \phi(\pi_0, q^*). \tag{21}$$

This phase is solved in Theorem 3.6 below (an explicit expression for $\bar{\phi}$ is provided in (24)). 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: Minimizing the resulting cost as a function of the remaining parameters π_0 for a given target error ε (and determine the resulting size of the simulation and its cost):

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

This second phase is solved asymptotically when ε goes to 0 in Theorem 3.8 and Proposition 3.11, with 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)

Through 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 the equality case occurs if and only if $q_j = \mu\sqrt{a_j b_j^{-1}}$ *with* $\mu = \left(\sum_{k=1}^{R} \sqrt{a_k b_k^{-1}}\right)^{-1}$.

In practice we will optimize an upper-bound of the true problem.

Theorem 3.6. *Assume* (SE_{β}) *holds, and let* $\theta = \sqrt{\frac{V_1}{\mathrm{var}(Y_0)}}$. *Then the optimally stratified effort* ϕ^* *defined by* (21) *satisfies*

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

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

$$\begin{cases} q_1^*(\pi_0) = \mu^*(1+\theta h^{\frac{\beta}{2}}) \\ q_j^*(\pi_0) = \mu^* \theta h^{\frac{\beta}{2}} \left(\sum_{i=1}^R |\mathbf{T}_i^j| n_i^{-\frac{\beta}{2}}\right)\left(\sum_{i=1}^R n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}\right)^{-\frac{1}{2}}, \ j = 2, \dots, R, \end{cases} \tag{22}$$

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

Proof. Under assumption (14), we have $\left\langle \mathbf{T}^1, Y_{h,\underline{n}}^{(1)} \right\rangle = Y_h^{(1)}$ and, for every $j \in \{2, \dots, R\}$, $\left\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} \right\rangle = \left\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} - Y_0^{(j)} \mathbf{1} \right\rangle$ since $\left\langle \mathbf{T}^j, \mathbf{1} \right\rangle = 0$. Hence, using Minkowski inequality and the strong error assumption, we get

$$\begin{aligned} \forall j \geqslant 2, \quad \mathrm{var}\left(\left\langle \mathbf{T}^j, Y_{h,\underline{n}}^{(j)} \right\rangle\right) &= \left\| \sum_{i=1}^R \mathbf{T}_i^j \left(Y_{\frac{h}{n_i}}^{(j)} - Y_0^{(j)} \right) \right\|_2^2 \\ &\leqslant V_1 h^\beta \left(\sum_{i=1}^R |\mathbf{T}_i^j| n_i^{-\frac{\beta}{2}} \right)^2. \end{aligned}$$

The variance of the Multilevel estimator is then

$$\mathrm{var}\left(\bar{Y}_{h,\underline{n}}^{N,q}\right) \leqslant \frac{1}{N} \left(\frac{\mathrm{var}\left(Y_h^{(1)}\right)}{q_1} + V_1 h^\beta \sum_{j=2}^R \frac{1}{q_j} \left(\sum_{i=1}^R |\mathbf{T}_i^j| n_i^{-\frac{\beta}{2}} \right)^2 \right). \tag{23}$$

On the other hand we have,

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

Combining (17), the above inequality (23) and the above upper-bound for $\mathrm{var}\left(Y_h^{(1)}\right)$, we derive the following upper bound for the effort $\phi(\pi) \leqslant \bar{\phi}(\pi)$ with

$$\bar{\phi}(\pi) = \frac{\mathrm{var}(Y_0)}{h} \left(\frac{(1+\theta h^{\frac{\beta}{2}})^2}{q_1} + \theta^2 h^\beta \sum_{j=2}^R \frac{1}{q_j} \left(\sum_{i=1}^R |\mathbf{T}_i^j| n_i^{-\frac{\beta}{2}} \right)^2 \right) \left(\sum_{i,j=1}^R q_j n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}} \right). \tag{24}$$

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

Remark 3.7 (About variance minimization)**.** *Note that we have shown the following in the above proof : for every stratification strategy $q = (q_1, \dots, q_R)$,*

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

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

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

with an optimal choice (for minimizing the variance) is $q_1^\dagger = \mu^\dagger(1+\theta h^{\frac{\beta}{2}})$ *and* $q_j^\dagger = \mu^\dagger \theta h^{\frac{\beta}{2}} \left(\sum_{i=1}^{R} |\mathbf{T}_i^j| n_i^{-\frac{\beta}{2}}\right)$ *with* $\mu^\dagger$ *the normalizing constant such that* $\sum_{j=1}^{n} q_j^\dagger = 1$. *It has to be noticed that this choice is not optimal or even asymptotically optimal when dealing with the global cost minimization of the simulation.*

3.3 Resulting cost optimization (Step 2)

3.3.1 Bias parameter optimization (first approach)

In this first approach we fix the order $R \geqslant 2$, the allocation matrix $\mathbf{T}$ and the refiners $n_1, \dots, n_R$ so that we optimize only the bias parameter h with respect to ε 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

$$\bar{\phi}^*(h) = \frac{\mathrm{var}(Y_0)}{h} \left(1 + \theta h^{\frac{\beta}{2}} \sum_{j=1}^{R} \left(\sum_{i=1}^{R} |\mathbf{T}_i^j| n_i^{-\frac{\beta}{2}}\right) \left(\sum_{i=1}^{R} n_i \mathbf{1}_{\{\mathbf{T}_i^j \neq 0\}}\right)^{\frac{1}{2}}\right)^2. \tag{25}$$

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

(*a*) Multilevel *RR: Assume* $(WE_{\alpha,\bar{R}})$. *Let* $R \in \{2, \dots, \bar{R}\}$ *be such that* $c_R \neq 0$. *A Multilevel Richardson-Romberg estimator of order* R *satisfies*

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

with q^* *defined in* (22). *This asymptotically optimal bound is achieved with a bias parameter given by*

$$h^*(\varepsilon, R) = (1+2\alpha R)^{-\frac{1}{2\alpha R}} \left(\frac{\varepsilon}{c_R}\right)^{\frac{1}{\alpha R}} \underline{n}!^{\frac{1}{R}}. \tag{26}$$

(*b*) Multilevel *MC: A Multilevel Monte Carlo estimator* $\bar{Y}_{h,\underline{n}}^{N,q^*}$ *of order* $R \geqslant 2$ *satisfies*

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

with q^* *defined in* (22). *This asymptotically optimal bound is achieved with a bias parameter given by*

$$h^*(\varepsilon, R) = (1+2\alpha)^{-\frac{1}{2\alpha}} \left(\frac{\varepsilon}{c_1}\right)^{\frac{1}{\alpha}} n_R. \tag{27}$$

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

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

It follows from that the cost mimimization 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 (19). First note that $\lim_{h\to 0} h\bar{\phi}(h,q^*) = \mathrm{var}(Y_0)$. We will consider now the denominator $h(\varepsilon^2 - \mu^2(h,q^*))$. Elementary computations show that, for fixed real numbers $a,\ R' > 0$, the function $g_{a,R'}$ defined by $g_{a,R'}(\xi) = \xi(1 - a^2\xi^{2R'}),\ \xi > 0$, satisfies

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

Then, set $R' = R\alpha$, $\tilde{a} = \frac{|\widetilde{\mathbf{w}}_{R+1}c_R|}{\varepsilon}$. Inspired by what precedes we make the suboptimal choice $h(\varepsilon) = h(\varepsilon,R,\alpha) = \xi\Big(\tilde{a},\alpha R\Big) = \left(\dfrac{\varepsilon}{(2\alpha R+1)^{\frac{1}{2}}c_R}\right)^{\frac{1}{\alpha R}} \underline{n}!^{\frac{1}{R}}$ 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

$$\inf_{h\in\mathcal{H},\,\mu(h,q^*)<\varepsilon} \frac{\phi^*(h)}{\varepsilon^2 - \mu^2(h,q^*)} \leqslant \Big(1 + \frac{1}{2\alpha R}\Big)(2\alpha R+1)^{\frac{1}{2\alpha R}} c_R^{\frac{1}{\alpha R}} \frac{h(\varepsilon)\bar{\phi}^*(h(\varepsilon))}{\underline{n}!^{\frac{1}{R}}\varepsilon^{2+\frac{1}{\alpha R}}} \frac{1}{1 - \frac{(\eta_{R,\underline{n}}(h(\varepsilon))+1)^2-1}{2\alpha R}}. \quad (28)$$

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 (18) of the effort ϕ

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

and Schwarz's Inequality that

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

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 approximation assumption and, as a consequence, $\lim_{h\to 0} h\,\phi(h) = \mathrm{var}(Y_0)$. Hence, the cost mimimization problem is lower bounded by the more explicit problem

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

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

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

Let $\varepsilon\in(0,\varepsilon_\eta)$. We derive from Equation (19) that

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

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

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

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

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

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

(ii) Owing to (20), the bias $\mu(h,q)$ is now given by

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

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 strong convergence of* Y_h *toward* Y_0 *and the rate of this 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 prevent 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 further on).*

Remark 3.10. *The asymptotic number of simulation* N *provided by* (5.2) *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 a Multilevel Richardson-Romberg 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 a Multilevel Monte Carlo estimator.

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

We now fix the structure of the allocation matrix $\mathbf{T}$. The standard Multilevel allocation matrix used by [Hei01, Gil08] comes from the telescopic summation

$$\mathbf{E}\Big[Y_{\frac{h}{n_{_R}}}\Big] = \mathbf{E}\big[Y_h\big] + \sum_{j=1}^R \mathbf{E}\Big[Y_{\frac{h}{n_j}} - Y_{\frac{h}{n_{j-1}}}\Big],$$

In our general framework, we consider $\mathbf{T}$ the allocation matrix of type (b) defined $\mathbf{T}^j = e_j - e_{j-1}$ for $j\in\{2,\dots,R\}$ *i.e.*

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

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

$$\bar{\phi}^*(\pi_0) = \frac{\mathrm{var}(Y_0)}{h}\left(1+\theta h^{\frac{\beta}{2}}\sum_{j=1}^{R}\left(n_{j-1}^{-\frac{\beta}{2}}+n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1}+n_j}\right)^2, \tag{29}$$

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

The corresponding allocation matrix for our Multilevel Richardson-Romberg estimator is $\mathbf{T}$ an allocation matrix of type (c) defined by $\mathbf{T}^j = -\mathbf{W}_j\, e_{j-1} + \mathbf{W}_j\, e_j$ for $j \in \{2,\dots,R\}$ with $\mathbf{W}_j = \sum_{k=j}^{R} \mathbf{w}_k$ and $\mathbf{w}$ is given by (8) *i.e.*

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

The majorant $\bar{\phi}^*$ now reads

$$\bar{\phi}^*(\pi_0) = \frac{\mathrm{var}(Y_0)}{h}\left(1+\theta h^{\frac{\beta}{2}}\sum_{j=1}^{R}\left|\mathbf{W}_j\right|\left(n_{j-1}^{-\frac{\beta}{2}}+n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1}+n_j}\right)^2, \tag{30}$$

with the convention $n_0 = (n_0)^{-1} = 0$. In the sequel we will mainly focus on the above choice (MLRR) for the allocation matrix $\mathbf{T}$. Alternative choices for the allocation matrix $\mathbf{T}$ are proposed in Section 5.1.

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

In this second approach we consider geometric refiners (with "*root*" M) of the form

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

These are the refiners 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) **(MLRR) estimator.** *Assume that* $(WE_{\alpha,\infty})$ *holds for an* $\alpha \geqslant \beta/2$, $\sup_{R\in\mathbf{N}}\sup_{h'\in(0,h)}|\eta_R(h')| < +\infty$ *for every* $h \in \mathcal{H}$ *and that* $\lim_{R\to+\infty} c_R^{\frac{1}{R}} = \widetilde{c} \in (0,+\infty)$. *The Multilevel Richardson-Romberg estimator with allocation matrix* $\mathbf{T}$ *in* (MLRR) *satisfies*

$$\limsup_{\varepsilon\to 0} v(\beta,\varepsilon) \times \inf_{\substack{h\in\mathcal{H}, R\geqslant 2 \\ \mu(h,R,q^*)<\varepsilon}} \mathrm{Cost}\left(Y_{h,\underline{n}}^{N,q}\right) \leqslant K(\alpha,\beta,M) \tag{31}$$

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

These bounds are achieved with an order

$$R^*(\varepsilon) = \left\lfloor \frac{1}{2} + \frac{\log(\mathbf{h}\widetilde{c}^{\frac{1}{\alpha}})}{\log(M)} + \sqrt{\left(\frac{1}{2} + \frac{\log(\mathbf{h}\widetilde{c}^{\frac{1}{\alpha}})}{\log(M)}\right)^2 + 2\frac{\log(1/\varepsilon)}{\alpha\log(M)}} \right\rfloor,$$

satisfying $\lim_{\varepsilon\to 0} R^*(\varepsilon) = +\infty$ *and a bias parameter* $h^* = h^*(\varepsilon, R(\varepsilon))$ *given by* (26)*. 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*

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

(*b*) **(MLMC) estimator.** *Assume that* $(WE_{\alpha,1})$ *holds for an* $\alpha \geqslant \beta$*. The Multilevel Monte Carlo estimator (with allocation matrix* $\mathbf{T}$ *defined in* (MLMC)*) satisfies*

$$\limsup_{\varepsilon\to 0} v(\beta,\varepsilon) \times \inf_{\substack{h\in\mathcal{H}, R\geqslant 2 \\ \mu(h,R,q^*)<\varepsilon}} \mathrm{Cost}\left(Y_{h,\underline{n}}^{N,q}\right) \leqslant K(\alpha,\beta,M) \tag{32}$$

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

$$R^*(\varepsilon) = \left\lfloor 1 + \frac{\log\left((1+2\alpha)^{\frac{1}{2\alpha}}c_1^{\frac{1}{\alpha}}\mathbf{h}\right)}{\log(M)} + \frac{\log(1/\varepsilon)}{\alpha\log(M)} \right\rfloor,$$

satisfying $\lim_{\varepsilon\to 0} R^*(\varepsilon) = +\infty$ *and a bias parameter* $h^* = h^*(\varepsilon, R(\varepsilon))$ *given by* (27)*. 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.

- *In Appendix B, it is proved 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* $(c_R)_{R\geqslant 1}$ *in* (ii) *of the above proposition are reasonable although probably impossible to check in practice. In particular, note that as soon as the sequence* $(c_R)_{R\geqslant 1}$ *has at most a polynomial growth as a function of* R*, it satisfies the assumption since* $\widetilde{c} = 1$*.*
- *When* $\widetilde{c} = 0$*, one can replace* c_R *in the proof below by* ϵ_0^R *(see also Remark) and carry one 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 on the right asymptotics.*

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

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

It seems clear that it is for this setting that Multilevel RR is the most powerful with respect to regular Multilevel MC.
When $\beta > 1$, both Multilevel methods achieve the same rate ε^{-2} as a virtual unbiased MC 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 (28) in the proof of Theorem 3.8 which reads

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

with

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

with the 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^* does not explode. The choice of refiners $n_i = M^{i-1}$ implies that $\underline{n}! = M^{\frac{R(R-1)}{2}}$ then

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

Note that, under the assumption made on the sequence $(c_R)_{R\geqslant 1}$, we have $\lim_{R\to+\infty}(1+2\alpha R)^{-\frac{1}{2\alpha R}} c_R^{-\frac{1}{\alpha R}} = \widetilde{c}^{-\frac{1}{\alpha}}$. We choose to saturate the constraint $h^* \leqslant \mathbf{h}$ so this leads to impose formally (for big enough R)

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

(where we temporarily forget that R be should an integer). As a consequence we are naturally led to search for the positive zero $R_+(\varepsilon)$ of the polynomial

$$P(R) = \frac{R(R-1)}{2}\log(M) - R\log(\widetilde{c}^{-\frac{1}{\alpha}}\mathbf{h}) - \frac{1}{\alpha}\log(1/\varepsilon),$$

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

$$\frac{1}{h^*(\varepsilon, R^*(\varepsilon))} \leqslant \frac{1}{\mathbf{h}} e^{(R_+ - R^*)\left(\frac{\log(M)}{2} + \frac{\log(1/\varepsilon)}{\alpha (R^*)^2}\right)},$$

so that $\limsup_{\varepsilon\to 0} \dfrac{1}{h^*(\varepsilon, R^*(\varepsilon))} \leqslant \dfrac{M}{\mathbf{h}}$.

Let us show that our choice $h^*(\varepsilon, R^*(\varepsilon))$ for the bias parameter (see (26)) 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

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

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 $\boldsymbol{\mu}\left(h^*(\varepsilon, R^*(\varepsilon)), R^*(\varepsilon), q^*\right)^2 = o(\varepsilon^2)$ since $R^*(\varepsilon) \to +\infty$ as $\varepsilon \to 0$. Hence our choice of 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

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

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 $\max_{j=1,\dots,R} \left|\mathbf{W}_i\right| \leqslant \mathbf{W}_\alpha(M)$. On the other hand, standard computations show that, for every $j = 2, \dots, R$,

$$\left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right) \sqrt{n_{j-1}+n_j} = M^{\beta-1} M^{j\frac{1-\beta}{2}} \left(1 + M^{-\frac{\beta}{2}}\right)\left(1+M\right)^{\frac{1}{2}} \tag{33}$$

and that, 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 three cases for strong rate convergence parameter β.

Case $\beta = 1$ In this case we have

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

and since $R_+^2(\varepsilon) \sim \frac{2}{\alpha \log(M)} \log(1/\varepsilon)$ as $\varepsilon \to 0$, we get (31) with (keep in mind that $V_1 = \mathrm{var}(Y_0)\theta^2$)

$$K(\alpha, 1, M) = \frac{2V_1}{\alpha} \left(\frac{\mathbf{W}_\alpha(M) M (1+M)(1+M^{-\frac{1}{2}})^2}{\log(M)}\right).$$

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 have

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

which yields (31) with

$$K(\alpha, \beta, M) = \frac{\mathrm{var}(Y_0) M}{\mathbf{h}} \left(1 + \theta \mathbf{h}^{\frac{\beta}{2}} \frac{\mathbf{W}_\alpha(M) M^{\frac{\beta-1}{2}} \sqrt{1+M}(1+M^{-\frac{\beta}{2}})}{1-M^{\frac{1-\beta}{2}}}\right)^{2}.$$

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,

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

Recall that $R_+(\varepsilon)$ is such that $h^*(\varepsilon, R_+(\varepsilon)) = \mathbf{h}$ so that we obtain $M^{\frac{R_+(\varepsilon)-1}{2}} = \mathbf{h}\, \widetilde{c}^{\frac{1}{\alpha}} \varepsilon^{-\frac{1}{\alpha R_+(\varepsilon)}}$. 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 (31) with

$$\beta < 1, \quad K(\alpha, \beta, M) = V_1 \mathbf{h}^{(1-\beta)} \widetilde{c}^{\frac{2(1-\beta)}{\alpha}} \left(\frac{\mathbf{W}_\alpha^2(M) M(1+M)(1+M^{-\frac{\beta}{2}})^2}{(M^{\frac{1-\beta}{2}}-1)^2}\right).$$

(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 the above proof for part (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 depend 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*

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

then

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

- *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 Richardson-Romberg method is 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] (see also [Guy06] for more recent developments on this topic). It relies on the following theorem.

Theorem 4.1. *Let $b : [0,T] \times \mathbf{R}^d \to \mathbf{R}^d$, $\sigma : [0,T] \times \mathbf{R}^d \to \mathcal{M}(d,q)$ and $(W_t)_{t \geqslant 0}$ 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 the Stochastic Differential Equation (SDE)*

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \ t \in [0,T], \ X_0 = x_0 \in \mathbf{R}^d, \tag{34}$$

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

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

(*a*) *Regular 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 maturity* $T > 0$ *and for every integer* $R \in \mathbf{N}^*$

$$\mathbf{E}\left[f(\bar{X}^h_T)\right] - \mathbf{E}\left[f(X_T)\right] = \sum_{k=1}^{R} c_k h^k + O\left(h^{R+1}\right), \tag{35}$$

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*

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

or more generally if (b, σ) *satisfies the strong Hörmander's hypo-ellipticity assumption, then* (35) *holds true for every bounded Borel function* $f : \mathbf{R}^d \to \mathbf{R}$.

Claim (*b*) can in fact be extended to hypo-elliptic diffusions (whose coefficients satisfy a strong hypoelliptic assumption "à la Hörmander").

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 $h \in \mathcal{H} = \{\frac{T}{n},\, n \geqslant 1\}$, and for a smooth enough function f with polynomial growth, $Y_h = f(\bar{X}^h_T)$ and $Y_0 = f(X_T)$. The above Theorem says that condition $(WE_{\alpha,\bar{R}})$ is satisfied with $\bar{R} = +\infty$ and $\alpha = 1$. Note that for a fixed $\bar{R}$, one may relax the differentiability assumption on b, σ and f 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, 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 L^2 (in fact in every L^p-space) as the step h goes to 0. In such a setting, we can apply multilevel estimator with $\alpha = \beta = 1$ (which corresponds to claim (*a*) in Proposition 3.11).

Milstein scheme The Milstein scheme (provided it can be implemented) satisfies (SE_2) as a second order scheme but still $(WE_{\infty,1})$ as concerns weak error expansion (like the Euler scheme). Consequently, the multilevel RR extrapolation can be applied to that scheme with $\alpha = 1$ and $\beta = 2$ (which corresponds to claim (*b*) in Proposition 3.11).

Path-dependent functionals When a functional $F : \mathcal{C}([0,T], \mathbf{R}^d) \to \mathbf{R}$ is Lipschitz for the sup-norm, it is straightforward that $F(\bar{X}^h)$ and $F(X)$ satisfy (SE_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 α-Hölder, $\alpha \in (0,1]$, then this family satisfies (SE_α). True 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 Assumption $(WE_{1,1})$ (see *e.g.* [Lap01, Gob00]). This strongly suggests to carry out numerical experiments by implementing the multilevel RR extrapolation method with parameters $\beta = \alpha = 1$ on such functionals. This has already been done successfully with the multistep RR extrapolation in [Pag07] taking advantage of the fact that, in severals situations, the continuous Euler scheme can be simulated (this is the purpose of the so-called Brownian diffusion bridge method). More recently, new results on first order weak error expansions have been obtained for functionals of the form

$F(\mathbf{w}) = f\Big(\sup_{t\in[0,T]} \mathbf{w}(t)\Big)$ (see [GHM09] and [AJKHin]). Thus, for the weak error expansion, it is shown in [AJKHin] that, for every $\eta > 0$, there exists a real constant $C_\eta > 0$ such that

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

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 [BAK12]. It seems clear that a similar approach applied to the multilevel Richardson-Romberg estimator* (MLRR)*, which is, computationally speaking, but a weighted version of the Multilevel estimator, also yields 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

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

where (X, Y) is couple of $\mathbf{R}\times\mathbf{R}^{q_Y}$-valued random variable defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ with $X \in L^1(\mathbf{P})$ and $f : \mathbf{R} \to \mathbf{R}$ is a Lipschitz continuous function with Lipschitz coefficient $[f]_{\mathrm{Lip}}$. Such quantities often appear in financial application like compound option pricing or risk estimation (see [BDM11]) and in actuarial sciences (see [DL09]) where nested MC is widely implemented.

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

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

Then, if $X \in L^1(\mathbf{P})$, 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 random vectors with the same distribution as Z and 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 $X \in L^{2R}(\mathbf{P})$. 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 $\bar{R} = R$.*

Remark 4.4. *When f is no longer smooth, typically because 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 [JHJ09, 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 in a different scale). However, in this work we focus on the regular case (see [LP14] for applications to non regular case and actuarial sciences).*

The proof straightforwardly follows form the two lemmas below.

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

$$\big\|Y_0-Y_h\big\|_2^2\leqslant[f]_{\mathrm{Lip}}^2\Big(\|X\|_2^2-\big\|\mathbf{E}(X\,|\,Y)\big\|_2^2\Big)h$$

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

Proof. Let $h=\frac{1}{K}$ and set $\mathbf{E}^Y(X)=\mathbf{E}(\,X\,|\,Y)$ for convenience.

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

and

$$\begin{aligned}\Big\|\mathbf{E}^Y(X)-\frac{1}{K}\sum_{k=1}^K F(Z_k,Y)\Big\|_2^2 &= \|\mathbf{E}^Y(X)\|_2^2+\frac{1}{K^2}\sum_{k=1}^K\big\|F(Z_k,Y)\big\|_2^2-\frac{2}{K}\sum_{k=1}^K\mathbf{E}\big(\mathbf{E}^Y(X)F(Z_k,Y)\big)\\ &\quad+\frac{2}{K^2}\sum_{1\leqslant k<k'\leqslant K}\mathbf{E}\big(F(Z_k,Y)F(Z_{k'},Y)\big).\end{aligned}$$

Now

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

and

$$\begin{aligned}\mathbf{E}\big(F(Z_k,Y)F(Z_{k'},Y)\big) &= \int_{\mathbf{R}^{q_Y}}\mathbf{E}(F(Z_k,y)F(Z_{k'},y))\mathbf{P}_{_Y}(dy)\\ &= \int_{\mathbf{R}^{q_Y}}\mathbf{E}(F(Z_k,y))\mathbf{E}(F(Z_{k'},y))\mathbf{P}_{_Y}(dy)\\ &= \int_{\mathbf{R}^{q_Y}}\big(\mathbf{E}\,F(Z,y)\big)^2\mathbf{P}_{_Y}(dy)\\ &= \big\|\mathbf{E}^Y(X)\big\|_2^2.\end{aligned}$$

Plugging these two identities in the first expansion finally yields

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

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

$$\mathbf{E}Y_h=\mathbf{E}Y_0+\sum_{r=1}^{R-1}c_rh^r+O\big(h^R\big),\quad h\in\mathcal{H}\,.$$

Consequently $(Y_h)_{h\in\mathcal{H}\cup\{0\}}$ satisfies $(WE_{\alpha,R-1})$ with with $\alpha=1$.

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

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

so that taking conditional expectation given Y yields

$$\mathbf{E}^Y(\widetilde{X}_1+\cdots+\widetilde{X}_{_K})^k = k!\sum_{k_1+\cdots+k_K=k}\prod_{i=1}^K\frac{\mathbf{E}^Y\widetilde{X}^{k_i}}{k_i!}.$$

since $\mathbf{E}^Y\widetilde{X}_i^{k_i}=\mathbf{E}^Y\widetilde{X}^{k_i}$. Noting that $\mathbf{E}^Y\widetilde{X}_i=0$, we get that

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

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

$$\begin{aligned}\sum_{k_1+\cdots+k_K=k,\,k_i\neq 1}\prod_{i=1}^K\frac{\mathbf{E}^Y\widetilde{X}^{k_i}}{k_i!} &= \sum_{1\leqslant\ell\leqslant(k/2)\wedge K}\ \sum_{I\subset\{1,\ldots,K\},|I|=\ell,\sum_{i\in I}k_i=k,k_i\geqslant 2}\prod_{i=1}^K\frac{\mathbf{E}^Y\widetilde{X}^{k_i}}{k_i!}\\ &= \sum_{1\leqslant\ell\leqslant k/2}\left(\begin{array}{c}K\\ \ell\end{array}\right)\sum_{\sum_{1\leqslant i\leqslant\ell}k_i=k-2\ell}\prod_{i=1}^{\ell}\frac{\mathbf{E}^Y\widetilde{X}^{2+k_i}}{(2+k_i)!}\end{aligned}$$

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

$$\begin{aligned}\mathbf{E}^Y Y_h &= \mathbf{E}^Y Y_0+\sum_{k=1}^{2R-1}\frac{f^{(k)}\big(\mathbf{E}^Y(X)\big)}{k!K^k}\mathbf{E}^Y(\widetilde{X}_1+\cdots+\widetilde{X}_{_K})^k+\mathbf{R}_{2R-1}(Y)\\ &= \mathbf{E}^Y Y_0+\sum_{k=1}^{2R-1}\frac{f^{(k)}\big(\mathbf{E}^Y(X)\big)}{k!K^k}\sum_{1\leqslant\ell\leqslant(k/2)\wedge K}\left(\begin{array}{c}K\\ \ell\end{array}\right)c_{k,\ell}+\mathbf{R}_{2R-1}(Y)\end{aligned}$$

where

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

and

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

By the Marcinkiewicz-Zygmund Inequality we get

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

where $B_p^{MZ}=18\frac{p^{\frac{3}{2}}}{(p-1)^{\frac{1}{2}}}$, $p>1$ (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

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

so that

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

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

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

(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. *Although it is note the only term included in the final $O(K^{-R})$, it is worth noticing that $\Big(\frac{(B_{2R}^{MZ})^{2R}}{(2R)!}\Big)^{\frac{1}{R}} \sim (36R)^2 \Big(\frac{2R}{e}\Big)^{-2} \sim 18\, e^2$ as $R \to +\infty$ owing to Stirling's formula. This suggests that, e.g. if all the derivatives of f are uniformly bounded, $\limsup_{R\to+\infty} c_R^{\frac{1}{R}} < +\infty$. For results in that direction, we refer to [LP14].*

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 depends 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, $i = 2, \dots, R$. 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 Multilevel methods (MC or RR) needs to know both the weak and strong (mean quadratic) 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

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

Since Y_0 cannot be simulated easily, one can consider the following estimator

$$\hat{V}_1(h) = (1 + 2^{-\beta}) h^{-\beta} \big\| Y_h - Y_{\frac{h}{2}} \big\|_2, \tag{36}$$

which satisfy $\hat{V}_1(h) \sim V_1$ when h is small enough. The approximation of $\theta = \sqrt{V_1/\mathrm{var}(Y_0)}$ is simply given by $\sqrt{\hat{V}_1(h)/\mathrm{var}(Y_h)}$. In practice this estimator is relatively efficient and stable.

The estimation of the c_i (c_1 for a crude MC and a MLMC estimator and $\widetilde{c} = \lim_{R\to\infty} c_R^{\frac{1}{R}}$ for our MLRR estimator) is much more challenging. So these methods are usually implemented in a blind way by considering the coefficient $c_{_R}$ 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 (MLRR) for the multilevel Richardson-Romberg estimator and by the template (MLMC) for the multilevel Monte Carlo estimator. Alternative choices could be considered for the multilevel Richardson-Romberg estimator like $\mathbf{T}$ satisfying (14) with $\mathbf{T}^j = -\mathbf{w}_j\, e_1 + \mathbf{w}_j\, e_j$ for $j \in \{2, \ldots, R\}$ *i.e.*

$$\mathbf{T} = \begin{pmatrix} 1 & -\mathbf{w}_2 & -\mathbf{w}_3 & \cdots & -\mathbf{w}_R \\ 0 & \mathbf{w}_2 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{w}_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \mathbf{w}_R \end{pmatrix}. \tag{MLRR2}$$

We can also consider a lower triangular allocation matrix (which does not satisfies the conventional assumption $T^1 = e_1$)

$$\mathbf{T} = \begin{pmatrix} \tilde{\mathbf{W}}_1 & 0 & \cdots & \cdots & \cdots & 0 \\ -\tilde{\mathbf{W}}_1 & \tilde{\mathbf{W}}_2 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & -\tilde{\mathbf{W}}_{R-2} & \tilde{\mathbf{W}}_{R-1} & 0 \\ 0 & \cdots & \cdots & \cdots & -\tilde{\mathbf{W}}_{R-1} & 1 \end{pmatrix} \quad \text{where} \quad \tilde{\mathbf{W}}_j = \sum_{k=1}^{j} \mathbf{w}_k\,. \tag{37}$$

The refiners, can be specified by the users but it turns out that the parametrized family $n_i = M^{i-1}$, $i = 1, \ldots, R$ ($M \in \mathbf{N}$, $M \geqslant 2$) seems the best compromise between variance control and implementability. 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 erase the (remaining) coefficients up to order R *i.e.* c_r, $r = 2, \ldots, R-1$ (instead of R). One easily shows that, if $(w_r^{(R-1)})_{r=1,\ldots,R-1}$ denotes the weight vector at order $R-1$ associated to refiners $n_1 = 1 < n_2, \ldots, n_{R_1}$ (for a given scale parameter α of the weak error expansion), then the weight vector $\widetilde{w}^{(R)}$ at order R (with size $R-1$) reads

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

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 (MLRR) estimator and in the table 5.2 for the (MLMC) estimator.

Note that these optimal parameters depend on the structural parameters and on the user's choice of the "root" $M \geqslant 2$ for the refiners. If we emphasize the dependance in M *i.e.* $R(M)$, $h(M)$, $q(M)$ and $N(M)$ we have for a fixed $\varepsilon > 0$

$$\mathrm{Cost}(Y_{h(M),\underline{n}}^{N(M),q(M)}) = N(M)\, \kappa(h(M), R(M), q(M)),$$

where $\kappa(h, R, q) = \frac{1}{h} \sum_{j=1}^{R} q_j \sum_{i=1}^{R} n_i \mathbf{1}_{\left\{\mathbf{T}_i^j \neq 0\right\}}$ (in the framework of Section 4.1) and $\kappa(h, R, q) = \frac{1}{h} \sum_{j=1}^{R} q_j \max_{1 \leqslant i \leqslant R} n_i \mathbf{1}_{\left\{\mathbf{T}_i^j \neq 0\right\}}$ (in the framework of Section 4.2). This function can be optimized for likely values of M, say $M \in \{2, \ldots, 30\}$.

R	$\left\lfloor \frac{1}{2} + \frac{\log(\widetilde{c}^{\frac{1}{\alpha}}\mathbf{h})}{\log(M)} + \sqrt{\left(\frac{1}{2} + \frac{\log(\widetilde{c}^{\frac{1}{\alpha}}\mathbf{h})}{\log(M)}\right)^2 + 2\frac{\log(1/\varepsilon)}{\alpha\log(M)}} \right\rfloor$
h^{-1}	$\left\lceil (1+2\alpha R)^{\frac{1}{2\alpha R}} \varepsilon^{-\frac{1}{\alpha R}} M^{-\frac{R-1}{2}} \right\rceil$
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	$\left(1+\frac{1}{2\alpha R}\right) \frac{\mathrm{var}(Y_0)\left(1+\theta h^{\frac{\beta}{2}} \sum_{j=1}^{R} \lvert \mathbf{W}_j(R,M) \rvert \left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1}+n_j}\right)^2}{\varepsilon^2 \sum_{j=1}^{R} q_j (n_{j-1}+n_j)}$

Table 5.1: Optimal parameters for the Multilevel Richardson-Romberg estimator (MLRR).

R	$\left\lfloor 1 + \frac{\log(c_1/\varepsilon)}{\alpha \log M} \right\rfloor$
h^{-1}	$\left\lceil (1+2\alpha)^{\frac{1}{2\alpha}} \left(\frac{\varepsilon}{c_1}\right)^{-\frac{1}{\alpha}} M^{-(R-1)} \right\rceil$
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	$\left(1+\frac{1}{2\alpha}\right) \frac{\mathrm{var}(Y_0)\left(1+\theta h^{\frac{\beta}{2}} \sum_{j=1}^{R} \left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1}+n_j}\right)^2}{\varepsilon^2 \sum_{j=1}^{R} q_j (n_{j-1}+n_j)}$

Table 5.2: Optimal parameters for the Multilevel Monte Carlo estimator (MLMC).

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 $[\frac{(k-1)T}{n}, \frac{kT}{n}]$, $k = 2, \ldots, n$. This can be performed by simulating recursively the Brownian increments over all successive sub-intervals of interest, having in mind that the "quantum" size for the simulation is given by $\frac{T}{nm}$ where $m = gcd(n_1, \ldots, n_R)$. 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 Euler scheme of a geometric Brownian motion

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

$$S_t = x_0 e^{(r-\frac{\sigma^2}{2})t+\sigma W_t},\ t\in[0,T],\ S_0 = s_0 > 0,$$

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 $e^{-rT}\mathbf{E}\,\varphi(S_T)$. For many payoff functions $\varphi:(0,+\infty)\to(0,+\infty)$, like $\varphi(x)=(x-K)_+$ or $(K-x)_+$ and their linear combinations, the premium admits a closed form starting from the formula

$$e^{-rT}\mathbf{E}\,\varphi(S_T) = e^{-rT}\int_{\mathbf{R}}\varphi\Big(s_0 e^{(r-\frac{\sigma^2}{2})T+\sigma\sqrt{T}z}\Big)e^{-\frac{z^2}{2}}\frac{dz}{\sqrt{2\pi}}.$$

On the other hand, $(X_t)_{t\in[0,T]}$ is solution to the diffusion SDE

$$dS_t = S_t(rdt+\sigma dW_t)$$

and one can also compute $e^{-rT}\mathbf{E}\,\varphi(S_T)$ by a Monte Carlo simulation in which the true process $(S_t)_{t\in[0,T]}$ is replaced by its Euler scheme $(\bar{S}^n_{t_k})_{0\leqslant k\leqslant n}$ (even if we are aware that S_T is simulatable). Although nobody would adopt this model any kind of MC simulation in practice since a simple difference method on the BS parabolic PDE is much more efficient for numerical purpose, it turns out that the Black-Scholes model and its Euler scheme is a very demanding *benchmark* to test and evaluate the performances of MC method(s). As a consequence, it is quite appropriate to carry out numerical tests on a new variant like $MLRR$.

In that regular diffusion setting (both drift and diffusion coefficients are $\mathcal{C}^\infty_b$), one has $\alpha=\beta=1$.

The Black-Scholes parameters considered in the following numerical experiments are $s_0=100$, $r=0.04$ and $\sigma=0.2$. The payoff is a European *Call* with maturity $T=1$ year and strike $K=100$. The parameters $\theta=\sqrt{V_1/\operatorname{var}(Y_0)}$ and $\operatorname{var}(Y_0)$ have been estimated following the procedure (36) described in Section 5.1 leading to the values $\theta=0.068$ and $\operatorname{var}(Y_0)=208$ for all the estimators. The L^2-error is estimated using 400 runs the algorithm and the bias is computed using the true value of the price 9.92504.

The target accuracy ε for the L^2-error has been set at $\varepsilon=2^{-\ell}$, $\ell=1,\ldots,9$.

The results ([1]) are summarized in Table 5.3 for the (MLRR) and in Tables 5.4 and 5.5 for the (MLMC).

ε	L^2-error	time (s)	bias	var	R	M	h^{-1}	N	$\kappa(\pi)$
$5.00\cdot10^{-01}$	$3.52\cdot10^{-01}$	$2.74\cdot10^{-03}$	$-5.56\cdot10^{-02}$	$1.21\cdot10^{-01}$	2	5	1	$1.76\cdot10^{+03}$	1.22
$2.50\cdot10^{-01}$	$1.77\cdot10^{-01}$	$3.13\cdot10^{-03}$	$-3.11\cdot10^{-02}$	$3.02\cdot10^{-02}$	2	9	1	$7.00\cdot10^{+03}$	1.26
$1.25\cdot10^{-01}$	$8.16\cdot10^{-02}$	$3.11\cdot10^{-03}$	$1.27\cdot10^{-03}$	$6.66\cdot10^{-03}$	3	3	1	$3.16\cdot10^{+04}$	1.42
$6.25\cdot10^{-02}$	$4.13\cdot10^{-02}$	$4.10\cdot10^{-03}$	$-2.58\cdot10^{-04}$	$1.71\cdot10^{-03}$	3	4	1	$1.23\cdot10^{+05}$	1.43
$3.12\cdot10^{-02}$	$2.07\cdot10^{-02}$	$8.44\cdot10^{-03}$	$-1.63\cdot10^{-04}$	$4.30\cdot10^{-04}$	3	5	1	$4.89\cdot10^{+05}$	1.45
$1.56\cdot10^{-02}$	$1.04\cdot10^{-02}$	$2.10\cdot10^{-02}$	$9.20\cdot10^{-05}$	$1.07\cdot10^{-04}$	3	6	1	$1.96\cdot10^{+06}$	1.47
$7.81\cdot10^{-03}$	$5.17\cdot10^{-03}$	$6.56\cdot10^{-02}$	$-7.52\cdot10^{-05}$	$2.68\cdot10^{-05}$	3	7	1	$7.86\cdot10^{+06}$	1.49
$3.91\cdot10^{-03}$	$2.57\cdot10^{-03}$	$2.03\cdot10^{-01}$	$-1.07\cdot10^{-04}$	$6.61\cdot10^{-06}$	3	9	1	$3.18\cdot10^{+07}$	1.53
$1.95\cdot10^{-03}$	$1.24\cdot10^{-03}$	$8.03\cdot10^{-01}$	$-2.13\cdot10^{-05}$	$1.53\cdot10^{-06}$	4	4	1	$1.37\cdot10^{+08}$	1.64

Table 5.3: Multilevel Richardson-Romberg estimator (MLRR) (*Call* in a Black-Scholes model discretized by a Euler scheme).

[1] The computations were performed on a computer with 4 muliithreaded(16) octo-core processors (Intel(R) Xeon(R) CPU E5-4620 0 @ 2.20GHz).

ε	L^2-error	time (s)	bias	var	R	M	h^{-1}	N	$\kappa(\pi)$
$5.00{\cdot}10^{-01}$	$6.88{\cdot}10^{-01}$	$2.79{\cdot}10^{-03}$	$6.07{\cdot}10^{-01}$	$1.05{\cdot}10^{-01}$	2	4	1	$1.99{\cdot}10^{+03}$	1.16
$2.50{\cdot}10^{-01}$	$3.79{\cdot}10^{-01}$	$3.09{\cdot}10^{-03}$	$3.43{\cdot}10^{-01}$	$2.60{\cdot}10^{-02}$	2	7	1	$8.05{\cdot}10^{+03}$	1.21
$1.25{\cdot}10^{-01}$	$1.66{\cdot}10^{-01}$	$3.32{\cdot}10^{-03}$	$1.48{\cdot}10^{-01}$	$5.63{\cdot}10^{-03}$	3	4	1	$3.72{\cdot}10^{+04}$	1.35
$6.25{\cdot}10^{-02}$	$8.12{\cdot}10^{-02}$	$4.79{\cdot}10^{-03}$	$7.22{\cdot}10^{-02}$	$1.39{\cdot}10^{-03}$	3	6	1	$1.51{\cdot}10^{+05}$	1.42
$3.12{\cdot}10^{-02}$	$4.32{\cdot}10^{-02}$	$9.39{\cdot}10^{-03}$	$3.90{\cdot}10^{-02}$	$3.41{\cdot}10^{-04}$	3	8	1	$6.17{\cdot}10^{+05}$	1.47
$1.56{\cdot}10^{-02}$	$2.14{\cdot}10^{-02}$	$2.61{\cdot}10^{-02}$	$1.96{\cdot}10^{-02}$	$7.65{\cdot}10^{-05}$	4	5	1	$2.74{\cdot}10^{+06}$	1.60
$7.81{\cdot}10^{-03}$	$1.00{\cdot}10^{-02}$	$8.87{\cdot}10^{-02}$	$8.45{\cdot}10^{-03}$	$2.88{\cdot}10^{-05}$	4	5	2	$7.22{\cdot}10^{+06}$	2.88
$3.91{\cdot}10^{-03}$	$4.91{\cdot}10^{-03}$	$3.52{\cdot}10^{-01}$	$4.42{\cdot}10^{-03}$	$4.61{\cdot}10^{-06}$	4	8	1	$4.57{\cdot}10^{+07}$	1.72
$1.95{\cdot}10^{-03}$	$2.57{\cdot}10^{-03}$	$1.34{\cdot}10^{+00}$	$2.34{\cdot}10^{-03}$	$1.12{\cdot}10^{-06}$	4	10	1	$1.87{\cdot}10^{+08}$	1.78

Table 5.4: Multilevel Monte Carlo estimator (MLMC) (*Call* in a Black-Scholes model discretized by a Euler scheme).

ε	L^2-error	time (s)	bias	var	R	M	h^{-1}	N	$\kappa(\pi)$
$5.00{\cdot}10^{-01}$	$4.02{\cdot}10^{-01}$	$2.84{\cdot}10^{-03}$	$2.42{\cdot}10^{-01}$	$1.03{\cdot}10^{-01}$	2	10	1	$2.05{\cdot}10^{+03}$	1.25
$2.50{\cdot}10^{-01}$	$1.74{\cdot}10^{-01}$	$3.82{\cdot}10^{-03}$	$8.91{\cdot}10^{-02}$	$2.24{\cdot}10^{-02}$	3	5	1	$9.37{\cdot}10^{+03}$	1.39
$1.25{\cdot}10^{-01}$	$8.67{\cdot}10^{-02}$	$4.06{\cdot}10^{-03}$	$4.49{\cdot}10^{-02}$	$5.50{\cdot}10^{-03}$	3	7	1	$3.82{\cdot}10^{+04}$	1.45
$6.25{\cdot}10^{-02}$	$4.71{\cdot}10^{-02}$	$6.01{\cdot}10^{-03}$	$2.95{\cdot}10^{-02}$	$1.35{\cdot}10^{-03}$	3	9	1	$1.56{\cdot}10^{+05}$	1.49
$3.12{\cdot}10^{-02}$	$2.08{\cdot}10^{-02}$	$3.00{\cdot}10^{-02}$	$1.14{\cdot}10^{-02}$	$3.02{\cdot}10^{-04}$	4	6	1	$6.95{\cdot}10^{+05}$	1.65
$1.56{\cdot}10^{-02}$	$1.05{\cdot}10^{-02}$	$6.32{\cdot}10^{-02}$	$6.03{\cdot}10^{-03}$	$7.46{\cdot}10^{-05}$	4	7	1	$2.82{\cdot}10^{+06}$	1.68
$7.81{\cdot}10^{-03}$	$5.40{\cdot}10^{-03}$	$2.10{\cdot}10^{-01}$	$3.32{\cdot}10^{-03}$	$1.82{\cdot}10^{-05}$	4	9	1	$1.16{\cdot}10^{+07}$	1.75
$3.91{\cdot}10^{-03}$	$2.85{\cdot}10^{-03}$	$5.21{\cdot}10^{-01}$	$1.05{\cdot}10^{-03}$	$7.04{\cdot}10^{-06}$	4	9	2	$2.95{\cdot}10^{+07}$	3.09
$1.95{\cdot}10^{-03}$	$1.48{\cdot}10^{-03}$	$3.09{\cdot}10^{+00}$	$1.08{\cdot}10^{-03}$	$1.03{\cdot}10^{-06}$	5	7	1	$2.04{\cdot}10^{+08}$	1.92

Table 5.5: "Optimal" Multilevel Monte Carlo estimator (MLMC) with $c_1 = 2.6$ (*Call* in a Black-Scholes model discretized by a Euler scheme).

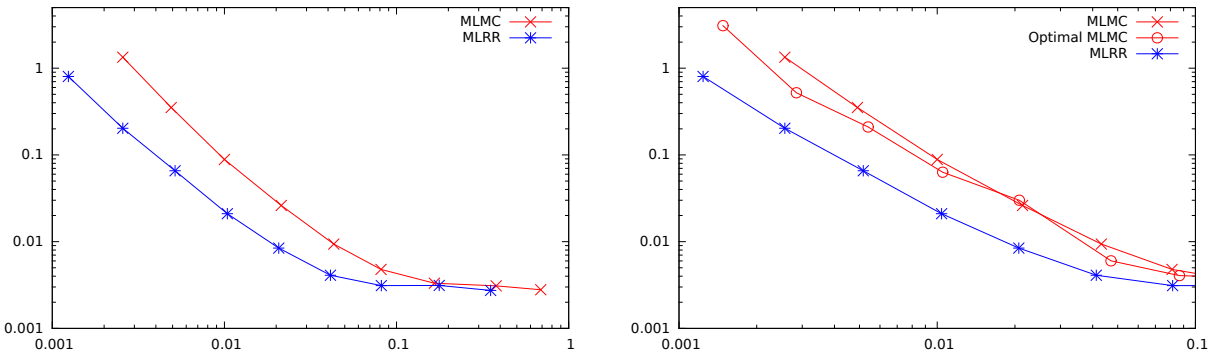

Figure 1: Time in seconds *vs* numerical L^2-error of the three estimators in the log-log scale

5.3 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 then has 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 $\mathcal{H} = \{1/K,\, K \geqslant 1\}$

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

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 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 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 $T_1 = 1$, $T_2 = 2$ and $K_1 = 20.6$, $K_2 = 100$.

ε	L^2-error	time (s)	bias	var	R	M	h^{-1}	N	$\kappa(\pi)$
$5.00\cdot10^{-01}$	$1.38\cdot10^{+00}$	$4.74\cdot10^{-03}$	$1.36\cdot10^{+00}$	$6.22\cdot10^{-02}$	2	2	2	$1.80\cdot10^{+03}$	2.57
$2.50\cdot10^{-01}$	$7.25\cdot10^{-01}$	$5.28\cdot10^{-03}$	$7.13\cdot10^{-01}$	$1.82\cdot10^{-02}$	2	4	2	$6.79\cdot10^{+03}$	3.19
$1.25\cdot10^{-01}$	$3.71\cdot10^{-01}$	$5.87\cdot10^{-03}$	$3.64\cdot10^{-01}$	$4.74\cdot10^{-03}$	2	8	2	$2.71\cdot10^{+04}$	3.91
$6.25\cdot10^{-02}$	$1.91\cdot10^{-01}$	$7.37\cdot10^{-03}$	$1.88\cdot10^{-01}$	$1.15\cdot10^{-03}$	2	16	2	$1.13\cdot10^{+05}$	4.81
$3.12\cdot10^{-02}$	$8.27\cdot10^{-02}$	$2.26\cdot10^{-02}$	$8.12\cdot10^{-02}$	$2.37\cdot10^{-04}$	3	6	2	$5.78\cdot10^{+05}$	5.37
$1.56\cdot10^{-02}$	$4.68\cdot10^{-02}$	$7.88\cdot10^{-02}$	$4.61\cdot10^{-02}$	$5.83\cdot10^{-05}$	3	8	2	$2.35\cdot10^{+06}$	6.00
$7.81\cdot10^{-03}$	$2.05\cdot10^{-02}$	$3.37\cdot10^{-01}$	$2.02\cdot10^{-02}$	$1.39\cdot10^{-05}$	3	12	2	$9.79\cdot10^{+06}$	6.98
$3.91\cdot10^{-03}$	$1.19\cdot10^{-02}$	$1.52\cdot10^{+00}$	$1.17\cdot10^{-02}$	$3.31\cdot10^{-06}$	3	16	2	$4.09\cdot10^{+07}$	7.77
$1.95\cdot10^{-03}$	$5.90\cdot10^{-03}$	$6.76\cdot10^{+00}$	$5.83\cdot10^{-03}$	$7.30\cdot10^{-07}$	4	8	2	$1.89\cdot10^{+08}$	8.14

Table 5.6: Nested MC: Multilevel Monte Carlo estimator (MLMC).

ε	L^2-error	time (s)	bias	var	R	M	h^{-1}	N	$\kappa(\pi)$
$5.00\cdot10^{-01}$	$3.92\cdot10^{-01}$	$5.54\cdot10^{-03}$	$2.80\cdot10^{-01}$	$7.55\cdot10^{-02}$	2	11	2	$1.72\cdot10^{+03}$	4.30
$2.50\cdot10^{-01}$	$1.81\cdot10^{-01}$	$5.58\cdot10^{-03}$	$1.32\cdot10^{-01}$	$1.52\cdot10^{-02}$	3	5	2	$9.01\cdot10^{+03}$	5.00
$1.25\cdot10^{-01}$	$8.17\cdot10^{-02}$	$7.00\cdot10^{-03}$	$5.39\cdot10^{-02}$	$3.77\cdot10^{-03}$	3	7	2	$3.63\cdot10^{+04}$	5.70
$6.25\cdot10^{-02}$	$4.38\cdot10^{-02}$	$1.00\cdot10^{-02}$	$3.18\cdot10^{-02}$	$9.11\cdot10^{-04}$	3	10	2	$1.50\cdot10^{+05}$	6.52
$3.12\cdot10^{-02}$	$2.01\cdot10^{-02}$	$2.98\cdot10^{-02}$	$1.37\cdot10^{-02}$	$2.17\cdot10^{-04}$	3	14	2	$6.25\cdot10^{+05}$	7.39
$1.56\cdot10^{-02}$	$1.09\cdot10^{-02}$	$1.11\cdot10^{-01}$	$8.25\cdot10^{-03}$	$5.12\cdot10^{-05}$	3	19	2	$2.63\cdot10^{+06}$	8.28
$7.81\cdot10^{-03}$	$4.84\cdot10^{-03}$	$4.61\cdot10^{-01}$	$3.46\cdot10^{-03}$	$1.15\cdot10^{-05}$	4	9	2	$1.20\cdot10^{+07}$	8.54
$3.91\cdot10^{-03}$	$2.10\cdot10^{-03}$	$2.01\cdot10^{+00}$	$1.30\cdot10^{-03}$	$2.73\cdot10^{-06}$	4	12	2	$5.00\cdot10^{+07}$	9.58
$1.95\cdot10^{-03}$	$1.20\cdot10^{-03}$	$8.85\cdot10^{+00}$	$8.90\cdot10^{-04}$	$6.51\cdot10^{-07}$	4	15	2	$2.08\cdot10^{+08}$	10.5

Table 5.7: Nested MC: "Optimal" Multilevel Monte Carlo estimator (MLMC) ($c_1 = 5.5$).

A Appendix

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

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

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

ε	L^2-error	time (s)	bias	var	R	M	h^{-1}	N	$\kappa(\pi)$
$5.00{\cdot}10^{-01}$	$3.21{\cdot}10^{-01}$	$5.36{\cdot}10^{-03}$	$1.01{\cdot}10^{-01}$	$9.30{\cdot}10^{-02}$	2	4	2	$1.65{\cdot}10^{+03}$	3.49
$2.50{\cdot}10^{-01}$	$1.90{\cdot}10^{-01}$	$5.07{\cdot}10^{-03}$	$1.13{\cdot}10^{-01}$	$2.33{\cdot}10^{-02}$	2	4	2	$6.60{\cdot}10^{+03}$	3.49
$1.25{\cdot}10^{-01}$	$9.81{\cdot}10^{-02}$	$6.08{\cdot}10^{-03}$	$6.14{\cdot}10^{-02}$	$5.86{\cdot}10^{-03}$	2	8	2	$2.42{\cdot}10^{+04}$	4.15
$6.25{\cdot}10^{-02}$	$4.80{\cdot}10^{-02}$	$6.91{\cdot}10^{-03}$	$3.01{\cdot}10^{-02}$	$1.40{\cdot}10^{-03}$	2	16	2	$9.81{\cdot}10^{+04}$	4.98
$3.12{\cdot}10^{-02}$	$1.74{\cdot}10^{-02}$	$1.97{\cdot}10^{-02}$	$2.58{\cdot}10^{-03}$	$2.97{\cdot}10^{-04}$	3	4	2	$5.02{\cdot}10^{+05}$	5.08
$1.56{\cdot}10^{-02}$	$9.51{\cdot}10^{-03}$	$6.32{\cdot}10^{-02}$	$4.03{\cdot}10^{-03}$	$7.42{\cdot}10^{-05}$	3	4	2	$2.01{\cdot}10^{+06}$	5.08
$7.81{\cdot}10^{-03}$	$4.51{\cdot}10^{-03}$	$2.42{\cdot}10^{-01}$	$1.35{\cdot}10^{-03}$	$1.85{\cdot}10^{-05}$	3	6	2	$7.66{\cdot}10^{+06}$	5.77
$3.91{\cdot}10^{-03}$	$2.15{\cdot}10^{-03}$	$9.98{\cdot}10^{-01}$	$8.30{\cdot}10^{-05}$	$4.60{\cdot}10^{-06}$	3	7	2	$3.05{\cdot}10^{+07}$	6.06
$1.95{\cdot}10^{-03}$	$1.07{\cdot}10^{-03}$	$4.00{\cdot}10^{+00}$	$8.30{\cdot}10^{-05}$	$1.14{\cdot}10^{-06}$	3	8	2	$1.22{\cdot}10^{+08}$	6.33

Table 5.8: Nested MC: Multilevel Richardson-Romberg estimator (MLRR).

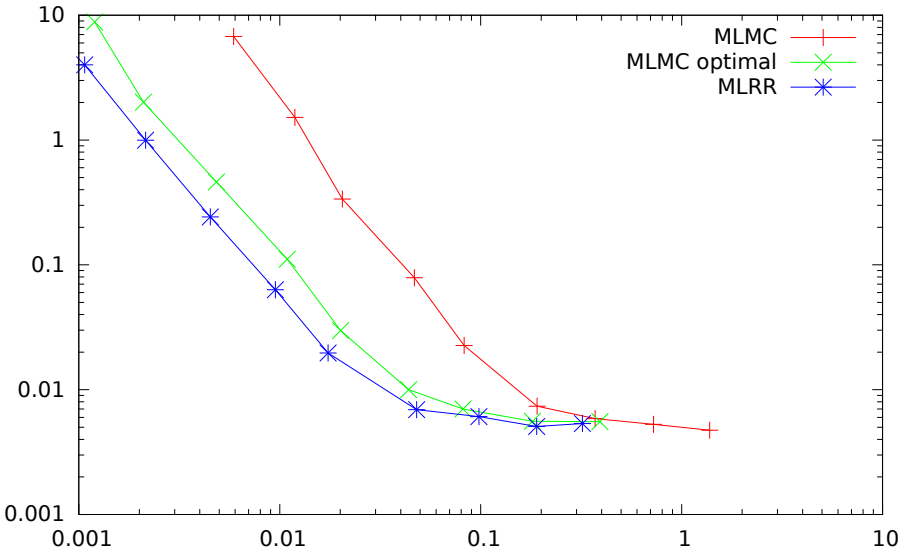

Figure 2: Nested MC: Rate of the L^2-error of the estimator (*log-log* scale).

(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) Following the proof 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 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 concludes 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. *On checks that*

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

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

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

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

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 (MLRR). It follows from what precedes that they satisfy $\mathbf{W}_1 = 1$,*

$$\max_{1\leqslant i\leqslant R} |\,\mathbf{W}_i(R, M)| \leqslant \mathbf{W}_\alpha(M) := \frac{M^{-\alpha}}{\pi^2_{\alpha,M}} \sum_{k\geqslant 0} M^{-\alpha\frac{k(k+3)}{2}} + \frac{1}{\pi_{\alpha,M}} \tag{38}$$

and

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

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

6. *One more useful inequality*

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

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

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

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

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

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

B Appendix: proof of Proposition 2.4

The multistep RR is characterized by the template matrix

$$\begin{pmatrix} \mathbf{w}_1 & 0 & \cdots & \cdots & 0 \\ \vdots & 0 & \cdots & \cdots & 0 \\ \mathbf{w}_i & 0 & \cdots & \cdots & 0 \\ \vdots & 0 & \cdots & \cdots & 0 \\ \mathbf{w}_R & 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

Note that the column is not e^1 but this has no influence on what follows. The expansion of $\mathbf{E}\big(\bar{Y}_{h,\underline{n}}^M\big)$ follows from Proposition 2.4. No stratification is needed here since only one Brownian motion is involved. Hence

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

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

References

[AJKHin] 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*, 2013 (To appear in).

[BAK12] 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.

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

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

[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.

[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.

[JHJ09] 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.

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

[Lap01] E. Lapeyre, B. Temam. Competitive Monte Carlo methods for the pricing of Asian Options. *J. of Comput. Fin.*, 5(1):39–59, 2001.

[LP14] V. Lemaire and G. Pagès. Multilevel Richardson-Romberg for nested Monte Carlo: the non regular case. In progress, 2014.

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

[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.