Weighted Multilevel Langevin Simulation of Invariant Measures

Gilles Pagès *Fabien Panloup †

April 30, 2016

Abstract

We investigate a weighted Multilevel Richardson-Romberg extrapolation for the ergodic approximation of invariant distributions of diffusions adapted from the one introduced in [LP13] for regular Monte Carlo simulation. In a first result, we prove under weak confluence assumptions on the diffusion, that for any integer $R \geq 2$, the procedure allows us to attain a rate $n^{\frac{R}{2R+1}}$ whereas the original algorithm convergence is at a weak rate $n^{\frac{1}{3}}$. Furthermore, this is achieved without any explosion of the asymptotic variance. In a second part, under stronger confluence assumptions and with the help of some second order expansions of the asymptotic error, we go deeper in the study by optimizing the choice of the parameters involved by the method. In particular, for a given $\varepsilon > 0$, we exhibit some semi-explicit parameters for which the number of iterations of the Euler scheme required to attain a Mean-Squared Error lower than ε^2 is about $\varepsilon^{-2}\log(\varepsilon^{-1})$.

Finally, we numerically test our method on several examples including the simple one-dimensional Ornstein-Uhlenbeck process but also an high dimensional diffusion motivated by a statistical problem. These examples seem to confirm the theoretical efficiency of the method.

Keywords: Ergodic diffusion, Invariant measure, Multilevel, Richardson-Romberg, Monte Carlo.

AMS Classification: 60J60, 37M25, 65C05.

1 Introduction

Let $(X_t)_{t\in[0,T]}$ be the unique strong solution to the stochastic differential equation (SDE)

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t$$

starting at X_0 where W is a standard $\mathbb{R}^q$-valued standard Brownian motion, independent of X_0, both defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathcal{M}(d, q, \mathbb{R})$ are locally Lipschitz continuous functions at most linear growth. The process $(X_t)_{t\geq 0}$ is a Markov process and we denote by $\mathbb{P}_\mu$ its distribution starting from $X_0 \sim \mu$. Let $\mathcal{L}$ denote its infinitesimal generator, defined on twice differentiable functions $g : \mathbb{R}^d \to \mathbb{R}$ by

$$\mathcal{L}g = (b|\nabla g) + \frac{1}{2}\mathrm{Tr}\big(\sigma^* D^2 g\,\sigma\big).$$

As soon as there exists a continuously twice differentiable *Lyapunov* function $V : \mathbb{R}^d \to \mathbb{R}_+$ such that

$$\sup_{x\in\mathbb{R}^d} \mathcal{L}V(x) < +\infty \quad \text{and} \quad \overline{\lim_{|x|\to+\infty}} \mathcal{L}V(x) < 0, \tag{1.1}$$

*UPMC, Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 5, France. E-mail: gilles.pages@upmc.fr

†Institut de Mathématiques de Toulouse, Université Paul Sabatier & INSA Toulouse, 135, av. de Rangueil, F-31077 Toulouse Cedex 4, France. E-mail: fabien.panloup@math.univ-toulouse.fr

then there exists an invariant probability measure ν for the diffusion in the sense that X is a stationary process under $\mathbb{P}_\nu$, so that $X_t \sim \nu$ for every $t \in \mathbb{R}_+$. Under appropriate (hypo-)ellipticity assumptions on σ or global confluence assumptions (on this topic, see *e.g.* [LPP15]), this invariant measure ν is unique, hence ergodic. In particular,

$$\mathbb{P}_\nu(d\omega)\text{-}a.s. \quad \nu_t(\omega, d\xi) = \frac{1}{t}\int_0^t \delta_{X_s(\omega)} ds \overset{(\mathbb{R}^d)}{\Longrightarrow} \nu$$

where $\overset{(\mathbb{R}^d)}{\Longrightarrow}$ denotes weak convergence of distributions on $\mathbb{R}^d$ (see *e.g.* [Bil78] or [Kre85] for background). We will assume that this uniqueness holds throughout the paper. Under additional assumptions, one shows that the diffusion is *stable* in the sense that

$$\forall\, x \in \mathbb{R}^d,\ \mathbb{P}_x(d\omega)\text{-}a.s. \quad \nu_t(\omega, d\xi) \overset{(\mathbb{R}^d)}{\Longrightarrow} \nu.$$

This $\mathbb{P}_x$-$a.s.$ convergence is ruled by Bhattacharya's CLT (see [Bha82] for detailed assumptions), namely, if $f : \mathbb{R}^d \to \mathbb{R}$ is such that the Poisson equation $f - \nu(f) = -\mathcal{L}g$ admits a solution, then

$$\sqrt{t}\big(\nu_t(\omega, f) - \nu(f)\big) \overset{(\mathbb{R}^d)}{\Longrightarrow} \mathcal{N}\big(0, \sigma^2(f)\big) \tag{1.2}$$

with $\sigma^2(f) = \int_{\mathbb{R}^d} |\sigma^* \nabla g|^2 d\nu$ where σ^* denotes the transpose matrix of σ.

In a series of papers (see *e.g.* [LP02, LP03, Lem07, PP09, PP14, Pan08]), the above properties have been exploited in order to compute by ergodic simulation integrals $\int f d\nu = \mathbb{E}_\nu f(X_t)$ or, more generally, $\mathbb{E}_\nu F\Big((X_t)_{t\in[0,T]}\Big)$ where F is a (path-dependent) functional defined on the space $\mathcal{C}([0,T], \mathbb{R}^d)$ (see also [Tal90] or [PS94] for other references on the topic or more recently [GT15]).

The starting idea is to mimic (1.2). First we replace the diffusion X by a (simulable) discretization scheme with *decreasing step*. To be more precise, we consider, for given a non-increasing sequence of steps $\gamma_n > 0$, $n \geq 1$, the associated Euler scheme with decreasing step defined by

$$\bar{X}_{n+1} = \bar{X}_n + \gamma_{n+1} b(\bar{X}_n) + \sigma(\bar{X}_n)\big(W_{\Gamma_n} - W_{\Gamma_{n-1}}\big),\ n \geq 0,\ \bar{X}_0 = X_0, \tag{1.3}$$

where $\Gamma_n = \gamma_1 + \ldots + \gamma_n$, $n \geq 1$. Then we introduce (for technical matter to be explained further on) a *weight* sequence $(\eta_n)_{n\geq 1}$ and the related η-weighted empirical (or occupation) measures of the above Euler scheme, namely

$$\nu_n^{\eta,\gamma}(\omega, dx) = \frac{1}{H_n}\sum_{k=1}^n \eta_k \delta_{\bar{X}_{k-1}(\omega)}.$$

The computation of $\nu_n^{\eta,\gamma}(f)$ can be performed recursively, once noted that that

$$\nu_n^{\eta,\gamma}(f) = \frac{\eta_n}{H_n} f(\bar{X}_n) + \Big(1 - \frac{\eta_n}{H_n}\Big)\nu_{n-1}^{\eta,\gamma}(f),\ \nu_0^{\eta,\gamma}(f) = 0. \tag{1.4}$$

It is clear that, in order to let the scheme explore the whole state space $\mathbb{R}^d$ and to let the empirical measures take into account new values as n grows, we must require the pair $(\eta_n, \gamma_n)_{n\geq 1}$ satisfies

$$H_n := \eta_1 + \cdots + \eta_n \to +\infty \quad \text{and} \quad \Gamma_n := \gamma_1 + \cdots + \gamma_n \to +\infty \quad \text{as } n \to +\infty. \tag{1.5}$$

When $\eta = \gamma$, the γ-empirical measure $\nu^{\gamma,\gamma}$ is the natural counterpart of ν_t and one expects that, under natural *mean-reverting* assumptions similar to (1.1) (or slightly more stringent), $\mathbb{P}_x(d\omega)$-$a.s.$ $\nu_n^{\eta,\gamma}(\omega, dx) \overset{(\mathbb{R}^d)}{\Longrightarrow} \nu$ taking advantage of the fact that the step $\gamma_n \downarrow 0$. The major difference with the above continuous time pointwise Birkhoff's ergodic theorem is that, provided b, σ can be computed easily, these random measures taken against a function f (computable as well) can in turn be simulated. This opens the way to simulation based ergodic methods to compute $\nu(f)$. Note that, though we will not go deeper in that direction, when $\nu = h.\lambda_d$ is absolutely continuous such a method appears as a probabilistic numerical scheme for the resolution of the stationary Fokker-Planck equation $\mathcal{L}^* h = 0$ by providing the values of as many integrals $\int f h \lambda_d$ as required.

Let us first recall one simple convergence result for the $a.s.$ weak convergence of the weighted empirical measures $(\nu_n^{\eta,\gamma})_{n\geq 1}$ (see Theorem V.2 borrowed and slightly adapted from [Lem05]).

Proposition 1.1. *Assume b and σ satisfy the mean-reverting assumption*

($\mathbf{S}$): *There exists a positive $\mathcal{C}^2$-function $V:\mathbb{R}^d\to\mathbb{R}_+$ and $\rho\in(0,+\infty)$ such that*

$$\lim_{|x|\to+\infty}\frac{V(x)}{|x|^\rho}=+\infty,\quad |\nabla V|^2\le CV\quad\text{and}\quad \sup_{x\in\mathbb{R}^d}\|D^2V(x)\|<+\infty$$

and there exist some real constants $C_b>0$, $\alpha>0$ and $\beta\ge0$ such that:

$$\textit{(i)}\quad |b|^2\le C_bV,\quad \mathrm{Tr}(\sigma\sigma^*)(x)=o\big(V(x)\big)\quad\text{as } |x|\to+\infty\qquad \textit{(ii)}\quad (\nabla V|b)\le\beta-\alpha V.$$

Then (SDE) admits at least one invariant distribution ν and for every $x\in\mathbb{R}^d$ and $p>0$, $\sup_n\mathbb{E}_xV^p(\bar X_n)<+\infty$.

Assume ν is the unique invariant measure of (SDE). If the pair $(\eta_n,\gamma_n)_{n\ge1}$ satisfies (1.5)

$$\sum_{n\ge2}\frac{1}{H_n}\left(\frac{\eta_n}{\gamma_n}-\frac{\eta_{n-1}}{\gamma_{n-1}}\right)_+<+\infty\quad\textit{and}\quad\sum_{n\ge1}\left(\frac{\eta_n}{H_n\sqrt{\gamma_n}}\right)^2<+\infty \tag{1.6}$$

then, $\mathbb{P}_x(d\omega)$-a.s. $\nu_n^{\eta,\gamma}(\omega,dx)\overset{(\mathbb{R}^d)}{\Longrightarrow}\nu$.

Moreover, $\mathbb{P}_x$-a.s., for every ν-a.s. continuous functions $\mathbb{R}^d\to\mathbb{R}$ with V-polynomial growth,

$$\nu^{\eta,n}(\omega,f)\to\nu(f)\qquad\textit{as}\qquad n\to+\infty. \tag{1.7}$$

Remark 1.1. ▷ By V-polynomial growth we mean that $f=O(V^p)$ at infinity for some $p>0$.

▷ The condition ($\mathbf{S}$) is stronger than (1.1). It implies that there exists $\alpha'\in(0,+\infty)$ and $\beta\in\mathbb{R}$ such that $\mathcal{L}V\le\beta'-\alpha'V$. In fact the conclusions of the above proposition are also true for the continuous time occupation measure $\nu_t(\omega)=\frac{1}{t}\int_0^t\delta_{X_s(\omega)}ds$ of the diffusion itself.

▷ The above result remains true under weaker Lyapunov assumptions of the following type: $\mathcal{L}V\le\beta'-\alpha'V^a$ with $a\in(0,1]$. For the sake of simplicity, we choose in this paper to state the results under ($\mathbf{S}$) only but all what follows can be extended to the weaker setting owing to additional technicalities (involving the control of the moments of the diffusion or of the Euler scheme (1.3)).

▷ In the above proposition, the condition $\lim_{|x|\to+\infty}\frac{V(x)}{|x|^\rho}=+\infty$ can be relaxed into $\lim_{|x|\to+\infty}V(x)=+\infty$. For the sequel, the interest of this slightly reinforced assumption is to ensure that every function f with polynomial growth has a V-polynomial growth.

Definition 1.1. A pair $(\eta_n,\gamma_n)_{n\ge1}$ (with decreasing γ_n) satisfying (1.5) and (1.6) is called an *averaging system.*

Examples. If $\gamma_n=\gamma_1n^{-a}$ and $\eta_n=\eta_1n^{-c}$, then the pair $(\eta_n,\gamma_n)_{n\ge1}$ is averaging as soon as $0<a<1$ and $0<c<1$. In practice, we will extensively use that, furthermore, the pairs of the form $(\gamma_n^\ell,\gamma_n)_{n\ge1}$ are averaging for $\ell\in\{1,\dots,\lceil\frac{1}{a}\rceil-1\}$ so that $a\ell<1$.

The rate of convergence of $\nu_n^{\eta,\gamma}(f)$ toward $\nu(f)$ has also been elucidated and reads as follows (when $d=1$ and $\eta_n=\gamma_n$ for the sake of simplicity, keeping in mind that even in that setting various averaging systems are needed). Set $\Gamma_n^{(2)}=\sum_{k=1}^n\gamma_k^2$, $n\ge1$.

Assume the Poisson Equation $f-\nu(f)=-\mathcal{L}g$ has a smooth enough solution and that $\frac{\Gamma_n^{(2)}}{\sqrt{\Gamma_n}}\to\widetilde\beta$, then

$$\sqrt{\Gamma_n}\big(\nu_n^{\gamma,\gamma}(f)-\nu(f)\big)\quad\overset{(\mathbb{R})}{\longrightarrow}\quad\mathcal{N}\Big(\widetilde\beta\nu(\Psi_2);\sigma_1^2(f)\Big)\qquad\text{if }\widetilde\beta\in[0,+\infty), \tag{1.8}$$

$$\frac{\Gamma_n}{\Gamma_n^{(2)}}\big(\nu_n^{\gamma,\gamma}(f)-\nu(f)\big)\quad\overset{a.s.}{\longrightarrow}\quad\nu(\Psi_2)\qquad\text{if }\widetilde\beta=+\infty \tag{1.9}$$

with $\sigma_1^2(f) = \nu\big(|\sigma^*\nabla g|^2\big) = -2\nu(g.Lg)$ and

$$\Psi_2(x) := \frac{1}{2}D^2 g(x)b(x)^{\otimes 2} + \frac{1}{24}\mathbb{E}[D^{(4)}g(x)(\sigma(x)U)^{\otimes 4}], \quad U \sim \mathcal{N}(0, I_q).$$

When $\gamma_n = n^{-a}$ the unbiased CLT ($\widetilde{\beta} = 0$) holds for $a \in \big(\frac{1}{3}, 1\big]$, the biased CLT for $a = \frac{1}{3}$ and the biased convergence in probability for $a \in (0, \frac{1}{3})$.

On can interpret this result as follows: if (γ_n) decreases to 0 fast enough ($\widetilde{\beta} = 0$), the empirical measures $\nu_n^{\gamma,\gamma}$ behaves like the empirical measures ν_t of the diffusion. When (γ_n) goes to 0 too slowly, there is a discretization effect which slows down the convergence of the empirical measure at rate $\frac{\Gamma_n}{\Gamma_n^{(2)}}$. The convergence then holds $a.s.$ (or at least in probability) which confirms that what slows down the convergence is a bias term whose rate of decay is lower than $1/\sqrt{\Gamma_n}$. The top rate of convergence is obtained with a biased CLT.

We will see in Theorem 2.1 further on that, in fact, there are many of these bias terms which go to 0 slower than the CLT rate for slowly decreasing steps. So killing these terms is a major issue to speed up such ergodic simulations (or Langevin Monte Carlo method) compared to the regular Monte Carlo method.

The Multilevel paradigm has been introduced by M. Giles in the late 2000's (2008, see [Gil08]). Ever since, it has been extensively adapted to various types of simulations (nested Monte Carlo, see [LP13], stochastic approximation [Fri16]) and dynamics (Lévy driven diffusion, random maps, etc) as a bias killer. The principle is the following: assume that a quantity of interest to be computed does have a representation as an expectation, say $\mathbb{E}\, Y_0$, but that the random variable Y_0 cannot be simulated at a reasonable computational cost. Then one usually approximates Y_0 by a family $(Y_h)_{h>0}$ of random vectors that can be simulated with a low complexity, relying on discretization schemes of the underlying dynamics. The typical situation is the $Y_0 = f(X_T)$ or $F\big((X_t)_{t\in 0,T]}\big)$ where (X_t) is a Brownian diffusion as above and $Y_h = f(\bar{X}_T^n)$ or $F\big((\bar{X}_t^n)_{t\in 0,T]}\big)$ where $(\bar{X}_t)_{t\in[0,T]}$ is a discretization scheme, say an Euler or a Milstein scheme with step $h = \frac{T}{n} \in \mathbb{H} = \big\{\frac{T}{m},\, m \in \mathbb{N}^*\big\}$. A multilevel estimator with depth $L \in \mathbb{N}^*$ of $\mathbb{E}Y_0$ is designed by implementing a non-homogeneous Multilevel Monte Carlo ($MLMC$) estimator of size $N \in \mathbb{N}^*$ of the form

$$\frac{1}{N_1}\sum_{k=1}^{N_1} Y_{\mathbf{h}}^{(1),k} + \sum_{\ell=2}^{L}\frac{1}{N_\ell}\sum_{k=1}^{N_\ell} Y^{(\ell),k}_{\frac{\mathbf{h}}{M^{\ell-1}}} - Y^{(\ell),k}_{\frac{\mathbf{h}}{M^{\ell-2}}}$$

where $\mathbf{h} \in \mathbb{H}$ is a fixed *coarse* step, $\big((Y_h^{(\ell),k})_{h\in\mathbb{H}}\big)_{\ell=1,\dots,L,k\geq 0}$ are independent copies of $(Y_h)_{h\in\mathbb{H}}$, $M \geq 2$, is a fixed integer and $N_1, \dots, N_R$ is an appropriate (optimized) allocation *policy* of the simulated paths across the levels ℓ such that $N_1 + \cdots + N_R = N$ (in practice, at a given level ℓ, only $Y^{(\ell)}_{\frac{\mathbf{h}}{M^{\ell-1}}}$ and $Y^{(\ell)}_{\frac{\mathbf{h}}{M^{\ell-2}}}$ have to be simulated). The level $\ell = 1$ is the *coarse* level whereas the levels $\ell \geq 2$ are the *refined* levels. Within a refined given level ℓ, $Y^{(\ell),k}_{\frac{h}{M^{\ell-2}}}$ denotes the coarse scheme and $Y^{(\ell),k}_{\frac{h}{M^{\ell-1}}}$ the refined scheme. For some fixed k and ℓ, the random variables are "consistent" in the sense that they have been simulated from the same underlying Brownian motion $W^{(\ell)}$. A quantitative translation of this consistency is that Y_h converges in (squared) quadratic norm to Y_0 at a h^β rate, namely $\|Y_h - Y_0\|_2^2 \leq V_1|h|^\beta$, $h \in \mathbb{H}$. The parameter β depends on f or F in a diffusion framework. If f or F are locally Lipschitz continuous with polynomial growth (with respect to the sup norm as for F), $\beta = 1$. This parameter β and the constant V_1 are key parameters to optimize the allocations of the paths to the various levels (see [Gil08, LP13]).

Among other results, M. Giles proved that if $\alpha = 1$ and $\beta = 1$ – which is the standard situation in a diffusion discretized by its Euler scheme – when $Y_0 = f(X_T)$, $Y_h = f(\bar{X}_T^n)$ (Euler scheme with step $h = \frac{T}{n}$), f, b, σ smooth enough (or σ uniformly elliptic if f is simply Borel and bounded), the resulting complexity of the optimized Multilevel Monte Carlo estimator to attain a prescribed Mean Squared Error ε^2 behaves like $O\Big(\Big(\log(1/\varepsilon)/\varepsilon\Big)^2\Big)$ as $\varepsilon \to 0$. When $\beta > 1$ (fast strong approximation like with the Milstein scheme), this rates attains $O(\varepsilon^{-2})$ *i.e.* the rate of a (virtual) unbiased simulation. The case $\beta < 1$ provides even better improvements compared to a crude Monte Carlo simulation.

In a recent paper (see [LP13]) a weighted version of the above multilevel estimator has been devised to take advantage of a higher order expansion of the weak error (bias expansion) up to an order $R \in \mathbb{N}^*$, namely

$$\mathbb{E}\, Y_h = \sum_{r=1}^{R} c_r h^{\alpha r} + O\big(h^{\alpha(R+1)}\big),$$

still under the above quadratic convergence rate assumption. Then, the so-called *Multilevel Richardson-Romberg estimator* ($ML2R$ in short) is still based on the simulation of independent copies of $(Y_h)_{h\in\mathbb{H}}$ and reads

$$\frac{\mathbf{W}_1^{(R)}}{N_1}\sum_{k=1}^{N_1} Y_{\mathbf{h}}^{(1),k} + \sum_{r=2}^{R}\frac{\mathbf{W}_r^{(R)}}{N_r}\sum_{k=1}^{N_r} Y^{(r),k}_{\frac{\mathbf{h}}{M^{r-1}}} - Y^{(r),k}_{\frac{h}{M^{r-2}}}$$

where the R-tuple $(\mathbf{W}_r^{(R)})_{1\le r\le R}$ of weights has a closed form *entirely determined by* α*, and* M and not on $(Y_h)_{h\ge 0}$ (that means on the specific form of f, b, σ in a diffusion framework). For this weighted estimator, the complexity is reduced *mutatis mutandis* to $O\Big(\log(1/\varepsilon)/\varepsilon^2\Big)$ in the setting $\beta = 1$. When $\beta < 1$ this estimator dramatically outperforms the above "regular" multilevel method since it only differs from a (virtual) unbiased simulation by a factor $\exp^{-\frac{1-\beta}{\alpha}\sqrt{\log(2)\log(1/\varepsilon)/2}}$ (when $M = 2$) instead of $\varepsilon^{\frac{1-\beta}{\alpha}}$ with $MLMC$. The underlying idea for this weighted Multilevel method is to combine the multilevel paradigm with a multistep Richardson-Romberg extrapolation introduced in [Pag07], hence its name. We refer to [LP13] for more precise results and proofs.

The aim of this paper is to transpose the weighted multilevel paradigm to the Langevin Monte Carlo simulation with decreasing step described above, with the issue that, in contrast with regular Monte Carlo simulation, canceling the bias terms directly impacts the rate of convergence of the method by enlarging the range of step parameters for which a CLT holds at rate $\sqrt{\Gamma_n}$ to coarser steps (so that Γ_n go faster to infinity where the stationary regime takes places). So we will adapt the ML2R estimator to the occupation measure $\nu_n^{\gamma} = \nu_n^{\gamma,\gamma}$ introduced before. Like in the regular Monte Carlo setting, we introduce, for a function f, a weighted estimator involving $\nu_n(f)$ and some correcting terms denoted by $\mu_n^{(r,M)}(f)$, $r = 1,\dots,R$ based on some pairs of coupled refined schemes (see (2.12) for details). Since the ergodic estimation of the invariant measure is based on only one path, the idea here is to replace the allocation policy of realizations $N_1,\dots,N_R$ of the ML2R method by a *sizing* policy $q_1,\dots q_R$ of the length of the coarse path (involved in $\nu_n(f)$) and those of the correcting sequences $\mu_n^{(r,M)}(f)$.

In order to asymptotically kill the successive terms of the bias induced by the estimator, we will need some asymptotic expansions of the ν_n and $\mu_n^{(r,M)}$ such as (2.13) and (2.14) below. These expansions, which require the invertibility of the infinitesimal generator (or equivalently the existence of solutions to the Poisson equation) can be viewed as the counterpart of the classical weak error/bias expansion $\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)]$ in finite horizon. Concerning the strong convergence rate property which leads to the control of the variance of the corrective terms in the standard Multilevel method, its counterpart in our ergodic setting is the following mean confluence result which says that

$$\frac{1}{\Gamma_n^{(2)}}\sum_{k=1}^{n}\gamma_k|X_{\Gamma_k} - \bar{X}_k|^2 \longrightarrow 0 \quad a.s. \quad \text{as } n\to+\infty.$$

It says that the (γ^2,γ)-empirical measure of the couple $(X,\bar{X})$ concentrates on the diagonal of $\mathbb{R}^d$ at rate $o\Big(\frac{\Gamma_n}{\Gamma_n^{(2)}}\Big)$. Such a property holds *e.g.* when the diffusion itself is exponentially confluent (typically a mean-reverting Ornstein-Uhlenbeck process) under an exponential confluence property which holds under $(\mathbf{S})$.

Throughout the proofs, we will work in one dimension for notational convenience. The extension to the multidimensional case would only generate technicalities.

Outline. The paper is organized as follows. We begin by introducing precisely the weighted empirical sequence built for the estimation of the invariant measure, called **ML2Rgodic** and denoted by $\widetilde{\nu}_n^{R,\mathbf{W}}$. Then, our main results are divided in three parts. In Theorem 2.1, we obtain some CLTs for $\widetilde{\nu}_n^{R,\mathbf{W}}$: we show that the **ML2Rgodic**-Algorithm with $R-1$ levels of corrections and an appropriate sequence (γ_n) has an optimal rate of order $n^{\frac{R}{2R+1}}$ with an asymptotic variance which is the same as the one of the original procedure. Then, in view of the optimization of the choices of the parameters, we exhibit in Theorem 2.2 some first and second order asymptotic expansions of the Mean-Squared Error. Based on this result, we proceed to the optimization in Theorem 2.3 and provide some choices of the parameters involved by the algorithm which lead to a complexity of order $\varepsilon^{-2}\log(\frac{1}{\varepsilon})$ (instead of ε^{-3} for the original procedure). The main tools for the establishment of Theorems 2.1 and 2.2 appear in Sections 3 and 4. Then, the proofs of Theorems 2.1, 2.2 and 2.3 are achieved in Section 5. Finally, we end this paper by some numerical computations in Section 6.

2 The Multilevel-Romberg Ergodic (ML2Rgodic) procedure

2.1 Design of the ML2Rgodic Langevin estimator

We aim at adapting the multilevel paradigm to devise an ergodic estimator for the approximation of the invariant distribution. For a given integer $R \geq 2$, the idea is to modify the original procedure with the aim to kill the R first terms of the expansion of the discretization error without impacting too much the simulation cost of simulation.

Let $\gamma = (\gamma_n)_{n\geq 1}$ be a sequence of steps, and M and R be two integers such that $R \geq 2$ and $M \geq 2$. First we consider an Euler scheme $\bar{X}^{(1)} = \bar{X}$ with decreasing step γ associated to a standard Brownian motion $W^{(0)} = W$. We associate to this scheme $R-1$ *independent* coupled schemes $(\bar{X}^{(r)}, \bar{Y}^{(r,M)})$, $r = 2, \ldots, R$, independent of $\bar{X}^{(0)}$ where

- $\bar{X}^{(r)}$ is an Euler scheme with decreasing step $\gamma^{(r,M)} = \frac{\gamma}{M^{r-2}}$ (so that $\gamma^{(2,M)} = \gamma$) associated to a Brownian motion $W^{(r)}$.
- $\bar{Y}^{(r,M)}$ is a refined Euler scheme with decreasing step $\widetilde{\gamma}^{(r,M)}$ associated *to the same* Brownian motion $W^{(r)}$ where

$$\forall\, m \in \{1, \ldots, M\}, \quad \widetilde{\gamma}^{(r,M)}_{M(n-1)+m} = \frac{\gamma_n^{(r,M)}}{M} = \frac{\gamma_n}{M^{r-2}},\ n \geq 1. \tag{2.10}$$

Set, for every integers $\ell \geq 1$ and $r \geq 2$,

$$\Gamma_n^{(\ell,r)} = \sum_{k=1}^{n} (\gamma_k^{(r,M)})^{\ell} = M^{-(r-2)\ell} \sum_{k=1}^{n} \gamma_k^{\ell} = M^{-(r-2)\ell}\Gamma_n^{(\ell)} \tag{2.11}$$

where $\Gamma_n^{(\ell)} = \Gamma_n^{(\ell,2)} = \sum_{k=1}^n \gamma_k^{\ell}$. Note that $\Gamma_n^{(\ell)} = \Gamma_n^{(\ell,2)}$.

Then, we define for every $r = 2, \ldots, R$ the sequence of difference of the empirical measures of the two schemes by

$$\begin{aligned} \mu_n^{(r,M)}(dx) &= \frac{1}{\Gamma_n^{(1,r)}} \sum_{k=1}^{n} \left(\left(\sum_{m=0}^{M-1} \widetilde{\gamma}^{(r)}_{M(k-1)+m} \delta_{\bar{Y}^{(r)}_{M(k-1)+m}} \right) - \gamma_k^{(r)} \delta_{\bar{X}^{(r)}_{k-1}} \right),\ n \geq 1, \\ &= \frac{1}{\Gamma_n^{(1,r)}} \sum_{k=1}^{n} \frac{\gamma_n}{M^{r-2}} \left(\frac{1}{M} \sum_{m=0}^{M-1} \delta_{\bar{Y}^{(r)}_{M(k-1)+m}} - \delta_{\bar{X}^{(r)}_{k-1}} \right),\ n \geq 1. \end{aligned} \tag{2.12}$$

The expected weak limit of $\mu_n^{(r,M)}(f)$ is 0 as a difference of occupation measures of two Euler schemes with decreasing step. Thus, this empirical measure plays the role of a correcting term.

Now, let $q_1, \ldots, q_R$ denote some positive real numbers, called *re-sizers* from now on, satisfying

$$\forall r \in \{1, \ldots, R\}, \quad 0 < q_r < 1, \quad q_1 + \ldots + q_R = 1,$$

and, for a given integer $n \geq 1$, set

$$n_r = \lfloor q_r n \rfloor, \; r = 1, \ldots, R.$$

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a smooth function, coboundary for the infinitesimal generator $\mathcal{L}$ (existence of solutions to the Poisson equation $f - \nu(f) = \mathcal{L}(g)$). Under some appropriate assumptions (including weak confluence) we can prove in a sense made precise later on (see Propositions 3.2(b) and 3.3(b) that the sequences $(\nu_{n_1}(f))_{n\geq 1}$ and $(\mu_{n_r}^{(r,M)}(f))_{n\geq 1}$ satisfy the following asymptotic generic type-expansions:

$$\nu_{n_1}(f) \;=\; \nu(f) + \sum_{\ell=2}^{R+1} \frac{\Gamma_{n_1}^{(\ell)}}{\Gamma_{n_1}} \nu(\Psi_\ell) + \frac{\mathbf{M}_n}{\Gamma_n} + o\Big(\frac{1}{\sqrt{\Gamma_{n_1}}} \wedge \frac{\Gamma_{n_1}^{(R+1)}}{\Gamma_{n_1}}\Big) \tag{2.13}$$

$$\mu_{n_r}^{(r,M)}(f) \;=\; \sum_{\ell=2}^{R+1} M^{(r-2)(1-\ell)}(M^{1-\ell} - 1) \frac{\Gamma_{n_r}^{(\ell)}}{\Gamma_{n_r}} \nu(\Psi_\ell) + o\Big(\frac{1}{\sqrt{\Gamma_{n_r}}} \wedge \frac{\Gamma_{n_r}^{(R+1)}}{\Gamma_{n_r}}\Big), \tag{2.14}$$

where $(\mathbf{M}_n)_{n\geq 1}$ is a martingale and $(\Psi_\ell)_{\ell\geq 1}$ is a sequence of functions made precise further on. At this stage, the reader can remark that there is no martingale term in the main part of the second expansion. This point, which is strongly linked with the weak confluence assumption $(\mathbf{C_w})$ introduced below, can be understood as follows: the martingale term induced by $\mu_n^{(r,M)}$ is asymptotically negligible against the one of $\nu_{n_1}(f)$. In a rough sense, this means that if we build an appropriate combination of $\nu_{n_1}(f)$ and $\mu_{n_r}^{(r,M)}(f)$, $r = 1, \ldots, R$, we will be able to kill the bias error without growing the asymptotic variance. But a numerical computation holds in a finite (non-asymptotic) setting so that this heuristic needs to be refined in practice. One of the objectives of the paper is thus to go deeper in the study of the expansion in order to be able to propose an efficient and potentially optimized method of approximation of the invariant distribution.

The ML2Rgodic-algorithm: As mentioned before, the first step toward our **ML2Rgodic** estimator is to design an appropriate combination of the formerly defined empirical measures in order to “kill” the bias. Furthermore, we require that this combination does not depend upon the size n of the estimator. We thus define a sequence of empirical measures denoted by $(\widetilde{\nu}_n^{(R,W)})_{n\geq 1}$ by:

$$\widetilde{\nu}_n^{(R,\mathbf{W})} = \mathbf{W}_1 \nu_{n_1} + \sum_{r=2}^{R} \mathbf{W}_r \mu_{n_r}^{(r,M)}, \quad n \geq 1, \tag{2.15}$$

where $\mathbf{W} = (\mathbf{W}_r)_{r=1}^R$ is a sequence of real numbers. For the sake of simplicity, we do not mention the dependency of $\widetilde{\nu}_n^{(R,\mathbf{W})}$in M and γ. Let us now specify $\mathbf{W}$. First, by (2.13) and (2.14), one remarks that it is necessary to assume that $\mathbf{W}_1 = 1$ in order to ensure the convergence towards ν.

Let us now consider the construction of $\mathbf{W}_2, \ldots, \mathbf{W}_R$. To this end, we consider from now on step sequences with polynomial decay

$$\gamma_k = \gamma_1 k^{-a} \;\text{ with }\; \gamma_1 > 0, \; a \in (0, 1). \tag{2.16}$$

Then by plugging the expansions of the bias resulting from (2.13) and (2.14) in the definition (2.15) of the **ML2Rgodic** estimator we derive that

$$\begin{aligned}
\mathbb{E}\big(\widetilde{\nu}_n^{(R,\mathbf{W})}\big) &= \mathbf{W}_1 \mathbb{E}\nu_{n_1}(f) + \sum_{r=2}^{R} \mathbf{W}_r \mathbb{E}\,\mu_{n_r}^{(r,M)}(f)\\
&= \mathbf{W}_1\nu(f) + \sum_{\ell=2}^{R+1}\left[\mathbf{W}_1\frac{\Gamma^{(\ell)}_{n_1}}{\Gamma_{n_1}} + \sum_{r=2}^{R}\mathbf{W}_r M^{(r-2)(1-\ell)}\big(M^{1-\ell}-1\big)\frac{\Gamma^{(\ell)}_{n_r}}{\Gamma_{n_r}}\right]\nu(\Psi_\ell) + o\Big(\frac{\Gamma_n^{(R+1)}}{\Gamma_n}\Big)\\
&\approx \mathbf{W}_1\nu(f) + \sum_{\ell=2}^{R+1}\frac{\Gamma^{(\ell)}_n}{\Gamma_n}\nu(\Psi_\ell)\left[\mathbf{W}_1 q_1^{-a(\ell-1)} + \sum_{r=2}^{R}\mathbf{W}_r M^{(r-2)(1-\ell)}\big(M^{1-\ell}-1\big)q_r^{-a(\ell-1)}\right]\\
&\quad + o\Big(\frac{\Gamma_n^{(R+1)}}{\Gamma_n}\Big)
\end{aligned}$$

where the notation $\approx$ is used to keep in mind that one implicitly assumes that $\frac{\Gamma^{(\ell)}_{n_r}}{\Gamma_{n_r}} - q_r^{-a(\ell-1)}\frac{\Gamma^{(\ell)}_n}{\Gamma_n}$ is negligible (see further on the proof of Theorems 2.1 and 2.2). Then as soon as the weights $(\mathbf{W}_r)_{1\le r\le R}$ are solutions to the linear system

$$\mathbf{W}_1 = 1, \quad \mathbf{W}_1 q_1^{-a(\ell-1)} + (M^{1-\ell}-1)\sum_{r=2}^{R}\mathbf{W}_r M^{-(r-2)(\ell-1)}q_r^{-a(\ell-1)} = 0, \qquad \ell = 2,\ldots,R, \tag{2.17}$$

the bias is "killed" up to order R and reads

$$\mathbb{E}\big(\widetilde{\nu}_n^{(R,\mathbf{W})}\big) \approx \frac{1-a}{1-a(R+1)}\,\gamma_1^R\,\nu(\Psi_{R+1})\widetilde{\mathbf{W}}_{R+1}n^{-aR} + o\Big(\frac{\Gamma_n^{(R+1)}}{\Gamma_n}\Big)$$

where we set, more generally,

$$\widetilde{\mathbf{W}}_{R+i} = \mathbf{W}_1 q_1^{-a(R+i)} + (M^{-R-i+1}-1)\sum_{r=2}^{R}\mathbf{W}_r M^{-(r-2)(R+i-1)}q_r^{-a(R+i)},\; i\ge 0. \tag{2.18}$$

The main difference at this stage with regular weighted multilevel estimator is that these weights *depend on* the re-sizers q_r which will make a complete optimization of these allocation parameters out of reach.

In the following lemma the linear system (2.17) is solved. In short, it shows that the weights are uniquely defined provided the re-sizers q_r satisfy $\frac{q_r}{M^{r/a}} \neq \frac{q_s}{M^{s/a}}$, $s\neq r$. Note that these weights depend on the exponent a (and the (q_r)) but not on γ_1.

Another important point is that, by contrast with the regular weighted Multilevel Monte Carlo setting, this system in its general form *is not a regular Vandermonde system* though it shows some similarities. In fact it can be related to a sequence of $(R-1)\times(R-1)$-Vandermonde systems with closed solutions. A notable exception to this situation occurs in the very special of *uniform* re-sizers $q_r = \frac{1}{R}$, $r = 1,\ldots,R$ where we retrieve exactly the weights of the regular Monte Carlo *ML2R* introduced in [LP13].

Lemma 2.1. (a) General re-sizers*: If* $\mathbf{q} := (q_1,\ldots,q_R)\in\mathcal{S}_R := \big\{(x_1,\ldots,x_R)\in(0,+\infty)^R, \sum_{i=1}^R x_i = 1\big\}$ *and satisfies* $\frac{q_r}{M^{r/a}} \neq \frac{q_s}{M^{s/a}}$, $s\neq r$, *then the above system* (2.17) *has a unique solution given by*

$$\mathbf{W}_r^{(R)} = M^{r-2}\left(\frac{q_r}{q_1}\right)^a \sum_{k\ge 0}\frac{1}{M^k}\prod_{s=2,s\neq r}^{R}\frac{1-M^{s-2-k}(q_s/q_1)^a}{1-M^{s-r}(q_s/q_r)^a},\; r = 2,\ldots,R. \tag{2.19}$$

Moreover, the coefficients $\widetilde{\mathbf{W}}^{(R)}_{R+i}$, $i = 1,2$, *as defined in* (2.18) *read*

$$\widetilde{\mathbf{W}}^{(R)}_{R+1} = \frac{(1-M^{-R})}{q_1^{aR}}\sum_{k\ge 0}\frac{1}{M^{kR}}\prod_{r=0}^{R-2}\Big(1-M^{k-r}\Big(\frac{q_1}{q_{r+2}}\Big)^a\Big). \tag{2.20}$$

and

$$\widetilde{\mathbf{W}}^{(R)}_{R+2} = \frac{(1-M^{-(R+1)})}{q_1^{a(R+1)}} \sum_{k\geq 0} \frac{1}{M^{k(R+1)}} \Big(1+\sum_{r=0}^{R-2} M^{k-r}\Big(\frac{q_1}{q_{r+2}}\Big)^a\Big) \prod_{r=0}^{R-2} \Big(1-M^{k-r}\Big(\frac{q_1}{q_{r+2}}\Big)^a\Big). \tag{2.21}$$

(b) Uniform re-sizers*: If* $q_r = \frac{1}{R}$*,* $r = 1,\ldots,R$*, the following simpler closed form holds for the weights* $\mathbf{W}^{(R)}_r$*:*

$$\mathbf{W}^{(R)}_r = \mathbf{w}^{(R)}_r + \cdots + \mathbf{w}^{(R)}_R,\ r = 1,\ldots,R \tag{2.22}$$

with

$$\mathbf{w}^{(R)}_r = \prod_{s=1,s\neq r}^{R} \frac{M^{-(s-1)}}{M^{-(s-1)}-M^{-(r-1)}} = \prod_{s=1,s\neq r}^{R} \frac{1}{1-M^{s-r}},\ r = 1,\ldots,R. \tag{2.23}$$

These weights $(\mathbf{W}^R_r)_{r=1,\ldots,R}$*,* $R\geq 1$*, are bounded i.e.* $\sup_{r=1,\ldots,R,R\geq 1}|\mathbf{W}^{(R)}_r| < +\infty$*. Furthermore*

$$\widetilde{\mathbf{W}}^{(R)}_{R+1} = (-1)^{R-1}R^{aR}M^{-\frac{R(R-1)}{2}} \quad \text{and} \quad \widetilde{\mathbf{W}}^{(R)}_{R+2} = (-1)^R R^{a(R+1)} M^{-\frac{R(R-1)}{2}} \frac{1-M^R}{1-M^{-1}}. \tag{2.24}$$

The proof is postponed to Appendix A.

Examples. • $R=2$: $\mathbf{W}^{(2)}_1 = 1$, $\mathbf{W}^{(2)}_2 = \frac{M}{M-1}\left(\frac{q_2}{q_1}\right)^a$

• $R=3$:

$$(\mathbf{W}^{(3)}_2, \mathbf{W}^{(3)}_3) = \frac{M}{M-1}\left(\left(\frac{q_2}{q_1}\right)^a \frac{1-\frac{M^2}{M+1}\left(\frac{q_3}{q_1}\right)^a}{1-M\left(\frac{q_3}{q_2}\right)^a}, \left(\frac{q_3}{q_1}\right)^a \frac{1-\frac{M^2}{M+1}\left(\frac{q_2}{q_1}\right)^a}{1-M^{-1}\left(\frac{q_2}{q_3}\right)^a}\right).$$

When there is no ambiguity the superscript $^{(R)}$ will be dropped in the notations $\mathbf{W}^{(R)}$, $\mathbf{w}^{(R)}_r$ and $\mathbf{W}^{(R)}_{R+1}$. In the sequel, $\widetilde{\nu}^{(R,\mathbf{W})}_n$ will be always defined with $\mathbf{W}$ satisfying (2.17) or (2.19).

Assumptions. We introduce below the assumptions for the first theorem. As recalled in the introduction, the study of the rate of convergence brings into play the Poisson equation related to the SDE. In this paper where we are going deeper in the expansion of the error, we will need to use it successively. For the sake of simplicity, we thus assume the following (strong) assumption:

$(\mathbf{P})$: For every $\mathcal{C}^\infty$ function f, there exists a unique (up to an additive constant) $\mathcal{C}^\infty$-function g, such that $f-\nu(f) = -\mathcal{L}g$. Furthermore, if f is a function with polynomial growth, then g also is.

For instance, it can be shown that, when σ is bounded and uniformly elliptic (in the sense that $(\sigma\sigma^*(x)x|x) \geq \lambda_0|x|^2$ for some $\lambda_0 > 0$), Assumption $(\mathbf{S})$ is in force and f, b and σ are smooth have polynomial growth as well as their derivatives, then $(\mathbf{P})$ holds true. Actually, we first recall that under the ellipticity and Lyapunov assumptions, the semi-group converges exponentially fast towards ν (in total variation) so that $g(x) = \int_0^\infty P_s f(x) - \nu(f)ds$ is well-defined and it is classical background that g is the unique (up to a constant) solution to the Poisson equation $f-\nu(f) = -\mathcal{L}g$ (see *e.g.* [PV01]). Then, by [GT83, Theorem 6.17], under uniform ellipticity, g is in fact $\mathcal{C}^\infty$ as soon as f, b and σ are. The polynomial growth of g and ∇g has been proved in [PV01, Theorem 1]. The property is obtained through the a priori estimate, see Equation (9.40) in [GT83], which in fact also holds for $D^2 g$. Then, we can establish by induction that all the partial derivatives of g have a polynomial growth. Assume it is true up to order k. First note that $u = \partial_{i_1,\ldots,i_{k-1}} g$ is a solution to $\mathcal{L}u = -f_g$ where f_g is a function which depends on f, b and σ and their first order partial derivatives and some derivatives of g up to order k. Hence, f_g has polynomial growth and the *a priori* error bound (9.40) in [GT83] for the second order partial derivatives of u yields the polynomial growth of the partial derivatives $\partial_{i_1,\ldots,i_{k+1}} g$.

The second additional assumption has been introduced in [PP14] and deeply studied [LPP15]: it requires the diffusion to be *weakly confluent*, *i.e.* that two paths of the diffusion, with different initial values, but driven by the same Brownian motion, asymptotically cluster in a weak (or statistical) sense as follows: let $(X_t, Y_t)_{t\geq 0}$ be the *duplicated diffusion* (or *two-point motion*) associated with the diffusion (SDE) by

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t)dW_t \\ dY_t = b(Y_t)dt + \sigma(Y_t)dW_t, \end{cases} \tag{2.25}$$

where X_0, Y_0 are two starting values independent of W. If ν is an invariant distribution for (SDE), $\nu_\Delta := \nu \circ (x \mapsto (x,x))^{-1}$ is trivially invariant for the couple (X,Y). The (SDE) is said *weakly confluent* if ν_Δ is *the only invariant distribution* for (X,Y) (which implies implicitly that ν itself is the unique invariant distribution of (SDE)). In the sequel, this assumption referred to as

$(\mathbf{C_w})$: (SDE) is weakly confluent.

Remark 2.2. ▷ Under slight additional assumptions on the stability of (SDE), it can be shown (see [LPP15]) that, if $(\mathbf{C_w})$ holds, the diffusion is statistically confluent in the sense that

$$\frac{1}{t}\int_0^t \delta_{(X_s,Y_s)}ds \overset{(\mathbb{R}^{2d})}{\Longrightarrow} \nu_\Delta \quad a.s. \quad \text{as } t \to +\infty.$$

▷ For the empirical measure $\widetilde{\nu}_n^{(R,\mathbf{W})}$, the impact of $(\mathbf{C_w})$ is to ensure that the empirical measures $\mu_n^{r,L}$, built with some differences of schemes $\bar{X}_n^{(r)}$ and $\bar{Y}_n^{(r)}$ have a negligible asymptotic variance (with respect to that of ν_n). This property will be made precise in Section 4.

We are now in position to state the first main theorem.

Theorem 2.1 (CLT)**.** *Assume* $(\mathbf{S})$, $(\mathbf{P})$ *and* $(\mathbf{C_w})$. *Let* $(R,M) \in (\mathbb{N}^* \setminus \{1\})^2$ *and let* $(\mathbf{W}_r)_{1\leq r\leq R}$ *denote the* R*-tuple of weights defined by* (2.19). *Let* $q = (q_r)_{1\leq r\leq R} \in \mathcal{S}_\mathcal{R}$ *be an* R*-tuple of re-sizers satisfying* $\frac{q_r}{M^r} \neq \frac{q_s}{M^s}$, $s \neq r$. *Let* $\gamma_n = \gamma_1 n^{-a}$, $n \in \mathbb{N}^*$, $a \in (0, 1/R)$, *be a discretization step sequence. Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be a* $\mathcal{C}^\infty$*-function and denote by* g *the solution to* $f - \nu(f) = -\mathcal{L}g$. *Let* $\mathbf{W} = (\mathbf{W}_r)_{r=1,\ldots,R}$ *be defined by* (2.19).

(a) If $a \in \left(\frac{1}{2R+1}, \frac{1}{R}\right)$, *then*

$$n^{\frac{1-a}{2}}\left(\widetilde{\nu}_n^{(R,\mathbf{W})}(f) - \int_{\mathbb{R}} f d\nu\right) \overset{(\mathbb{R})}{\Longrightarrow} \mathcal{N}\Big(0; \sigma_f^2(a,q,R)\Big) \qquad as \qquad n \to +\infty$$

with

$$\sigma_f^2(a,q,R) = \frac{1-a}{\gamma_1}\frac{\sigma_1^2(f)}{q_1^{1-a}} \qquad with \qquad \sigma_1^2(f) = \nu(|\sigma^*\nabla g|^2). \tag{2.26}$$

(b) If $a = \frac{1}{2R+1}$, *the CLT holds at an optimal rate towards a biased Gaussian distribution, namely*

$$n^{\frac{R}{2R+1}}\left(\widetilde{\nu}_n^{(R,\mathbf{W})}(f) - \int_{\mathbb{R}^d} f d\nu\right) \overset{(\mathbb{R})}{\Longrightarrow} \mathcal{N}\Big(m_f(q,R); \sigma_f^2(q,R)\Big) \qquad as \qquad n \to +\infty$$

with $\sigma_f^2(q,R) := \sigma_f^2(\frac{1}{2R+1}, q, R)$ *and* $m_f(q,R) := 2\gamma_1^R \widetilde{\mathbf{W}}_{R+1} c_{R+1}$ *where* $\widetilde{\mathbf{W}}_{R+1}$ *is given by* (2.18) *and* $c_{R+1} = \nu(\Psi_{R+1})$, Ψ_{R+1} *being a* $\mathcal{C}^\infty$*-function with polynomial growth (whose explicit expression in the one-dimensional case is given by* (3.38)*).*

(c) If $a \in \left(0, \frac{1}{2R+1}\right)$, *then*

$$n^{aR}\left(\widetilde{\nu}_n^{(R,\mathbf{W})}(f) - \int_{\mathbb{R}} f d\nu\right) \overset{\mathbb{P}}{\to} m_f(a,q,R) \quad \text{as} \qquad n \to +\infty$$

with

$$m_f(a,q,R) := \frac{1-a}{1-a(R+1)}\gamma_1^R \widetilde{\mathbf{W}}_{R+1} c_{R+1}. \tag{2.27}$$

Remark 2.3. Note that the definitions of $m_f(a,q,R)$ and $m_f(q,R)$ in the above claims (b) and (c) are consistent since $m_f(q,R) = m_f(a,q,R)$ when $a = \frac{1}{2R+1}$.

From an asymptotic point of view, the above result says in particular that when R grows, the optimal rate of convergence tends to $n^{\frac{1}{2}}$ without increasing the (asymptotic) variance. However, from a non-asymptotic point of view, one has certainly to go deeper in the result to try to optimize the choice of the parameters. This implies to take into account the effect of the choice of $\mathbf{q}$, M and R on the residual bias term, the variance and on the computational cost. This is the purpose of the next paragraph.

L^2-expansions of the error. The aim of this part is to study the quadratic error to prepare the optimization of the parameter of the multilevel estimator (a, q, R, n) algorithm subject to a prescribed quadratic error $\varepsilon > 0$. To this end, we will not only provide a re-formulation of Theorem 2.1 in quadratic norm, we will also go deeper in the study of the asymptotic error. In particular, in the previous result, the variance induced by the correcting terms $\mu_n^{R,M}$ does not appear and we would like to quantify it. We will also need to control the residual error terms not only in n but also with respect to the depth R, since this parameter is intended to go to $+\infty$ in the optimization phase. This will lead us to carry out the expansion to the order $R+2$ and not R or $R+1$ like in the above theorem and to introduce a second and more constraining confluence assumption denoted by $(\mathbf{C_s})$:

$(\mathbf{C_s})$: There exists $\alpha > 0$ and a positive matrix S such that for every $x, y \in \mathbb{R}^d$,

$$(b(x) - b(y)|x - y)_{_S} + \frac{1}{2}\|\sigma(x) - \sigma(y)\|_{_S}^2 \le -\alpha\|x - y\|_{_S}^2$$

where $(\,.\,|\,.\,)_{_S}$ and by $|\,.\,|_{_S}$ stand for the inner product and norm on $\mathbb{R}^d$ defined by $(x|y)_{_S} = (x|Sy)$ and $|x|_{_S}^2 = (x|x)_{_S}$, and for $A \in \mathcal{M}(d, d, \mathbb{R})$, $\|A\|_{_S}^2 = \mathrm{Tr}(A^*SA)$.

Furthermore to get closer to practical aspects, we only consider the optimal case $a = \bar{a} = 1/(2R+1)$ which clearly provides the highest possible rate of convergence for a given complexity. Finally we will focus on the uniform re-sizing vector $q_r = \frac{1}{R}$, $r = 1, \dots, R$. They turn out to be most likely rate optimal and, as emphasized in Remark B.11, in that case the first term of the bias of the **ML2Rgodic** estimator *does vanish* whereas for other choices of vectors q a residual bias (at rate $O(n^{-1-\bar{a}})$) still remains. Though theoretically negligible, it turns out to have a strong numerical impact on simulations.

Theorem 2.2 (Mean Squared Error for $a = \bar{a} = \frac{1}{2R+1}$)**.** *(a) Suppose that the assumptions of the previous theorem hold and let $a = \frac{1}{2R+1}$. Then,*

$$\left\|\widetilde{\nu}_n^{(R,\mathbf{W})}(f) - \nu(f)\right\|_2^2 = n^{-\frac{2R}{2R+1}}\left(\sigma_f^2(q, R) + m_f^2(q, R) + o(1)\right) \quad \text{as} \quad n \to +\infty.$$

(b) If, furthermore, $(\mathbf{C_s})$ holds

$$\left\|\widetilde{\nu}_n^{(R,\mathbf{W})}(f) - \nu(f)\right\|_2^2 = n^{-\frac{2R}{2R+1}}\left(\sigma_f^2(q, R) + m_f^2(q, R)\right) + \frac{1}{n}\left(\widetilde{\sigma}_f^2(q, R) + \widetilde{m}_f(q, R) + o(1)\right) \quad \text{as} \quad n \to +\infty$$

where, on the one hand

$$\widetilde{\sigma}_f^2(q, R) = \frac{1}{q_1}\sigma_{2,1}^2(f) + \left(1 - \frac{1}{M}\right)\Psi(R, M)\sigma_{2,2}^2(f) \tag{2.28}$$

with

$$\Psi(R, M) = \frac{4R^2}{4R^2 - 1}\sum_{r=2}^{R}(\mathbf{W}_r^{(R)})^2 \tag{2.29}$$

and $\sigma_{2,1}^2(f)$ and $\sigma_{2,2}^2(f)$ some variance terms explicitly defined further on by (4.42) and (4.45) in Propositions 4.4 and 4.5 respectively. On the other hand $\widetilde{m}_f(q, R)$ is given by

$$\widetilde{m}_f(q, R) = \frac{8R}{R-1}c_{_{R+1}}c_{_{R+2}}\gamma_1^{2R+1}\widetilde{\mathbf{W}}_{_{R+1}}^{(R)}\widetilde{\mathbf{W}}_{_{R+2}}^{(R)}.$$

(c) If furthermore the re-sizers are uniform, namely $q_r = \bar{q}_r = \frac{1}{R}$, $r = 1, \dots, R$, then the weights $\mathbf{W}_r^{(R)}$ are given by (2.23) and $\widetilde{\mathbf{W}}_{_{R+1}}^{(R)}$ and $\widetilde{\mathbf{W}}_{_{R+2}}^{(R)}$ by (2.24) so that

$$\widetilde{m}_f(\bar{q}, R) = -\frac{4R}{R-1}c_{_{R+1}}c_{_{R+2}}\gamma_1^{2R+1}R\,M^{-R(R-1)}\frac{1 - M^{-R}}{1 - M^{-1}}. \tag{2.30}$$

2.2 Optimization procedure

It remains to optimize the parameters to minimize the complexity of the estimator for a given prescribed *mean square error* (MSE). In view of the above Theorem 2.1, it is clear that the parameter a should be settled at

$$a = \bar{a} = \frac{1}{2R+1}.$$

We start from Theorem 2.2(b) with

$$a = \bar{a} = \frac{1}{2R+1} \quad \text{and} \quad q_r = \bar{q}_r := \frac{1}{R},\ r = 1, \dots, R.$$

Then the weights $\mathbf{W}_r$, $r = 1\dots, R$ and $\widetilde{\mathbf{W}}_{R+1}$ are given by (2.23) and (2.24) (those coming out in standard multilevel Monte Carlo *e.g.* in the case of the approximation of a diffusion by its Euler scheme).

We denote by $\varpi = (R, \gamma_1, n, M) \in \Pi = \mathbb{N}^* \times (0, +\infty) \times \mathbb{N}^* \times \mathbb{N}^*$ the remaining set of free simulation parameters that we wish to optimize. With this specification for a and the allocation vector $\bar{q}$, the $MSE(\varpi)$ reads

$$\left\|\nu_n^{R,\mathbf{W}} - \nu(f)\right\|_2^2 = \frac{1}{n^{\frac{2R}{2R+1}}}\Big(\sigma_f^2(\bar{a}, \bar{q}, R) + m_f^2(\bar{a}, \bar{q}, R) + \Big) + \frac{1}{n}\Big(\widetilde{\sigma}_f^2(\bar{a}, \bar{q}, R) + \widetilde{m}_f(q, R) + o(1)\Big) \tag{2.31}$$

as n goes to ∞ where, owing to (2.27), (2.30), (2.26) and (2.28),

$$\begin{aligned}
m_f(\bar{a}, \bar{q}, R) &= 2\gamma_1^R(-1)^{R-1}R^{\frac{R}{2R+1}}M^{-\frac{R(R-1)}{2}}c_{R+1}\\
\widetilde{m}_f(\bar{q}, R) &= -\frac{8R}{R-1}c_{R+1}c_{R+2}\gamma_1^{2R+1}R\,M^{-R(R-1)}\frac{1-M^R}{1-M^{-1}}.\\
\sigma_f^2(\bar{a}, \bar{q}, R) &= = \frac{2R}{2R+1}R^{\frac{2R}{2R+1}}\sigma_1^2(f)\gamma_1^{-1}\\
\widetilde{\sigma}_f^2(\bar{a}, \bar{q}, R) &= = R\left[\sigma_{2,1}(f)^2 + \left(1 - \frac{1}{M}\right)\Psi(R, M)\sigma_{2,2}^2(f)\right].
\end{aligned}$$

On the other hand, the complexity $K(\varpi, n, M)$ of the multilevel Langevin estimator devised in (2.15) reads

$$\begin{aligned}
K(\varpi, n, M) &= n\big(q_1 + (M+1)(q_2 + \cdots + q_R)\big)\\
&= n\big(1 + M(1 - q_1)\big)\kappa_0 = n\Big(1 + M\Big(1 - \frac{1}{R}\Big)\Big)\kappa_0
\end{aligned}$$

where κ_0 denotes the unitary computational cost of one iteration of an Euler scheme.

To calibrate the above parameter ϖ, we want to minimize the complexity subject to a prescribed $RMSE$ $\varepsilon > 0$, that is solving the constrained optimization problem:

$$\inf_{MSE(\varpi) \le \varepsilon^2} K(\varpi).$$

To state the main result of this section, whose proof is postponed to Section 5, we need to introduce a function related to the weights $\mathbf{W}_r^{(R)}$ and on the depth of the simulation. We know from Lemma 2.1(c) that $\sup_{1\le r\le R, R\ge 2}|\mathbf{W}_r^{(R)}| < +\infty$. Consequently, M being fixed, $\Psi(R, M) = O(R)$ as $R \to +\infty$. This leads us to define

$$\mathbf{\Psi}(M) = \sup_{R\ge 1}\frac{\Psi(R, M)}{R}. \tag{2.32}$$

where Ψ is defined by (2.29) (see Table 2 for some numerical values of ψ and $\mathbf{\Psi}$).

Theorem 2.3. *Under the assumptions of Theorem 2.2 and if, furthermore,* $\lim_{R\to+\infty} \frac{1}{R}\left|\frac{c_{R+1}}{c_R}\right| = 0$ *and* $|c_R|^{\frac{1}{R}} \to \widetilde{c} \in (0,+\infty)$*, then*

$$\inf_{MSE(\varpi)\leq\varepsilon^2,\varpi\in\Pi} K(\varpi) \precsim K(f,M).\varepsilon^{-2}\Big(\log\left(\frac{1}{\varepsilon}\right)\Big) \quad as \quad \varepsilon\to 0,$$

where

$$K(f,M) = \frac{2\kappa_0(M+1)}{\log M}\left(\frac{(M-1)\mathbf{\Psi}(M)}{\widetilde{c}\,\theta_1(f)}+1\right)\widetilde{c}\,\sigma_1^2(f) \tag{2.33}$$

with $\theta_1(f) = \frac{\sigma_1^2(f)}{\sigma_{2,2}^2(f)}$.

(b) The above bound can be achieved by the (sub-)optimal ϖ^* *given by* $q^* = \frac{\mathbf{1}}{R}$, $R^* = R(\varepsilon,M) = \lceil x(\varepsilon,M)\rceil$ *where* $x(\varepsilon,M)$ *is the unique solution to the equation* $\frac{\log(M)}{2}x(x-1)+x\log x+\log(\varepsilon)=0$ *and*

$$\gamma^*(\varepsilon,M) = \left(\frac{2R}{2R+1}\right)^{\frac{1}{2R+1}}(8R)^{-\frac{1}{2R+1}}|c_{R+1}|^{-\frac{2}{2R+1}}\sigma_1^2(f)^{\frac{1}{2R+1}}M^{\frac{R(R-1)}{2R+1}}.$$

Furthermore, as $\varepsilon\to 0$,

$$x(\varepsilon,M) = \sqrt{\frac{2\log\left(\frac{1}{\varepsilon}\right)}{\log M}} - \frac{\log_{(2)}\left(\frac{1}{\varepsilon}\right)}{2\log M}+\frac{1}{2}+\frac{\log(\log M)-\log 2}{2\log M}+O\left(\frac{\log_{(2)}(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}}\right) \qquad as\ \varepsilon\to 0$$

and the (minimal) number of iterations $n(\varepsilon,M)$ *necessary to attain an MSE lower than* ε^2 *satisfies*

$$n(\varepsilon,M) \precsim \frac{2}{\log M}\left(\frac{(M-1)\mathbf{\Psi}(M)}{\widetilde{c}\,\theta_1(f)}+1\right)\sigma_1^2(f)\varepsilon^{-2}\log\left(\frac{1}{\varepsilon}\right) \qquad as\ \varepsilon\to 0. \tag{2.34}$$

Remark 2.4. Though difficult to check in practice, note that the assumptions on the sequence $(c_r)_{r\geq 1}$ are satisfied as soon as

$$\lim_{R\to+\infty}\left|\frac{c_{R+1}}{c_R}\right| = \widetilde{c}\in(0,+\infty).$$

Remark 2.5. Note that the choice of $R(\varepsilon,M)$ does not depend on the parameters. In Table 1, we give the values of $x(\varepsilon,M)$ for several choices of M and ε. As expected, one can check that $R(\varepsilon,M)$ increases very slowly when ε decreases.

	$\varepsilon=10^{-1}$	$\varepsilon=10^{-2}$	$\varepsilon=10^{-3}$	$\varepsilon=10^{-4}$
$M=2$	2.08	2.79	3.38	3.89
$M=3$	1.94	2.56	3.06	3.50
$M=4$	1.87	2.44	2.90	3.30

Table 1: Values of $x(\varepsilon,M)$

Remark 2.6. A remarkable point to be noted is that we retrieve the same asymptotic rate as that obtained with the original ML2R Monte Carlo simulation at finite horizon, that is for the computation of expectations $\mathbb{E}\,f(X_T)$ where $X=(X_t)_{t\in[0,T]}$ is a standard diffusion discretized by its Euler scheme.

Practical aspects are investigated in the practitioners' corner (see Section 6.1) especially how to calibrate the parameters which are involved in the definition of ϖ^*.

3 Expansion of the error

For the sake of simplicity, the proofs are detailed in dimension 1. In the following subsections, we begin by decomposing the quantity $\nu_n^{\gamma,\eta}(f)-\nu(f)$ for a given smooth coboundary function f (*i.e.* such that the Possion equation $f-\nu(f)=-\mathcal{L}g$ has a smooth enough solution) and for a general weight sequence (η_n). Then, in the next subsections, we successively propose some expansions of the error, $\nu_n^{\gamma}(f)-\nu(f)$ for the original sequence $(\nu_n^{\gamma}(f))_{n\geq 1}$ (implemented on the coarse level) and for the sequences of correcting empirical measures $(\mu_n^{(r,M)}(f))$ for $r=2,\dots,R$ defined in (2.12) and corresponding to the successive refined levels of our estimator.

Note that by expansion, we mean an expansion of the bias of our estimators (level by level then globally) until we reach an order at which we reach a martingale term involved in the weak rate of convergence.

3.1 Higher order expansion of $\nu_n^{\gamma}(f)-\nu(f)$ (coarse level)

For every integer $n\geq 1$, for every sequence $(v_n)_{n\geq 1}$, we set

$$\Delta v_n = v_n - v_{n-1},\; U_n=\gamma_n^{-\frac12}(W_{\Gamma_n}-\Gamma_{n-1})\overset{d}{=}\mathcal{N}\big(0;I_q\big) \text{ and } \rho_m=\mathbb{E}[U_1^m],\; m\in\mathbb{N}.$$

Lemma 3.2. *Let $L\in\mathbb{N}$. Assume that $f-\nu(f)=-\mathcal{L}g$ where g is a $\mathcal{C}^{2L+3}$-function. Then, for every integer $n\geq 1$,*

$$\Delta g(\bar X_n)=-\gamma_n(f(\bar X_{n-1})-\nu(f))+\left[\sum_{\ell=2}^{L+1}\gamma_n^{\ell}\varphi_\ell(f)(\bar X_{n-1})\right]+\sum_{i=1}^{3}\Delta M_n^{(i,g)}+\Delta R_{n,L}^{(1,g)}+\Delta R_{n,L}^{(2,g)}+\Delta R_{n,L}^{(3,g)} \tag{3.35}$$

where

$$\varphi_\ell(f)(x)=\sum_{(m_1,m_2),m_1+\frac{m_2}{2}=\ell} g^{(m_1+m_2)}(x)\frac{\rho_{m_2}}{m_1!\,m_2!}b^{m_1}(x)\sigma^{m_2}(x)$$

$$\Delta M_n^{(1,g)}=\sqrt{\gamma_n}(g'\sigma)(\bar X_{n-1})U_n,\quad \Delta M_n^{(2,g)}=\frac12\gamma_n g''(\bar X_{n-1})\sigma^2(\bar X_{n-1})\left[U_n^2-1\right],$$

$$\Delta M_n^{(3,g)}=\gamma_n^{\frac32}\left(\frac12 g''(\bar X_{n-1})b(\bar X_{n-1})\sigma(\bar X_{n-1})U_n+\frac16 g^{(3)}(\bar X_{n-1})\sigma^3(\bar X_{n-1})U_n^3\right),$$

$$\begin{aligned}\Delta R_{n,L}^{(1,g)}=&\sum_{\ell=2}^{2L+1}\gamma_n^{\ell+\frac12}\sum_{(m_1,m_2),m_1+\frac{m_2}{2}=\ell+\frac12} g^{(m_1+m_2)}(\bar X_{n-1})\frac{1}{m_1!\,m_2!}b^{m_1}(\bar X_{n-1})\sigma^{m_2}(\bar X_{n-1})U_n^{m_2}\\&+\sum_{\ell=2}^{2L+1}\gamma_n^{\ell}\sum_{(m_1,m_2),m_1+\frac{m_2}{2}=\ell} g^{(m_1+m_2)}(\bar X_{n-1})\frac{1}{m_1!\,m_2!}b^{m_1}(\bar X_{n-1})\sigma^{m_2}(\bar X_{n-1})[U_n^{m_2}-\rho_{m_2}],\end{aligned}$$

$$\Delta R_{n,L}^{(2,g)}=\sum_{\ell=L+2}^{2L+2}\gamma_n^{\ell}\sum_{(m_1,m_2),m_1+\frac{m_2}{2}=\ell} g^{(m_1+m_2)}(\bar X_{n-1})\frac{\rho_{m_2}}{m_1!\,m_2!}b^{m_1}(\bar X_{n-1})\sigma^{m_2}(\bar X_{n-1})$$

$$\Delta R_{n,L}^{(3,g)}=g^{(2L+3)}(\xi_n)(\gamma_n b(\bar X_{n-1})+\sqrt{\gamma_n}\sigma(\bar X_{n-1})U_n)^{2L+3},\; \xi_n\in[\bar X_{n-1},\bar X_n].$$

As a consequence,

$$\begin{aligned}\nu_n^{\eta,\gamma}(f)-\nu(f)=&-\frac{1}{H_n}\sum_{k=1}^{n}\frac{\eta_k}{\gamma_k}\Delta g(\bar X_k)+\sum_{\ell=2}^{L+1}\frac{\sum_{k=1}^n\eta_k\gamma_k^{\ell-1}}{H_n}\nu_n^{\eta\gamma^{\ell-1},\gamma}(\varphi_\ell(f))\\&+\frac{1}{H_n}\sum_{k=1}^{n}\frac{\eta_k}{\gamma_k}\left(\sum_{i=1}^{3}\Delta M_k^{(i,g)}+\Delta R_{k,L}^{(1,g)}+\Delta R_{k,L}^{(2,g)}+\Delta R_{k,L}^{(3,g)}\right).\end{aligned} \tag{3.36}$$

Proof. By the Taylor formula with order $2L+2$, we have for every x and y in $\mathbb{R}^d$,

$$g(x+y)-g(x)=\sum_{\ell=1}^{2L+2}\frac{1}{k!}g^{(k)}(x)y^k+g^{(2L+3)}(\xi)y^{2L+3}$$

where $\xi\in[x,x+y]$. Then, if $y=\gamma b(x)+\sqrt{\gamma}\sigma(x)u$ with $u\in\mathbb{R}^d$,

$$\frac{1}{k!}y^k=\sum_{m_1+m_2=k}\frac{1}{m_1!m_2!}\gamma^{m_1+\frac{m_2}{2}}b^{m_1}(x)\sigma^{m_2}(x)u^{m_2}.$$

The decomposition of $\Delta g(x)$ easily follows by separating odd and even m_2 and by remarking that

$$g'(x)y+\frac{1}{2}g''(x)y^2=-\gamma\mathcal{L}g(x)+\sqrt{\gamma}\sigma(x)u+\frac{1}{2}\gamma\sigma^2(x)(u^2-1)+\frac{1}{2}g''(x)\left(\gamma^2b^2(x)+2\gamma^{\frac{3}{2}}\sigma(x)u\right).$$

Since

$$\nu_n^{\eta,\gamma}(f)-\nu(f)=\frac{1}{H_n}\sum_{k=1}^{n}\frac{\eta_k}{\gamma_k}\left(\gamma_k(f(\bar{X}_{k-1})-\nu(f)\right),$$

the second part of the lemma is a direct consequence. □

For notational convenience, we will denote by $\mathcal{Q}f$ in what follows the solution of the Poisson equation $f-\nu(f)=-\mathcal{L}\big(\mathcal{Q}f\big)$ satisfying $\nu(\mathcal{Q}f)=0$. (Under Assumption (**P**), $\mathcal{Q}f$ is well-defined).

Definition 3.2. (a) Under Assumption (**P**), one may define a mapping $\varphi_\ell^{[1]}(.)$ from $\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ into itself defined for every $f\in\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ by

$$\varphi_\ell^{[1]}(f)(.)=\sum_{(m_1,m_2),m_1+\frac{m_2}{2}=\ell}\frac{\rho_{m_2}}{m_1!\,m_2!}b^{m_1}(.)\sigma^{m_2}(.)(\mathcal{Q}f)^{(m_1+m_2)}(.)\tag{3.37}$$

where $h^{(k)}$ denotes the k^{th} derivative of a function h. Then, for every $\ell\in\mathbb{N}$, one sets $\varphi_\ell^{[m]}=\varphi_\ell^{[m-1]}\circ\varphi_\ell^{[1]}$. To alleviate notations, we will often write $\varphi_m(f)$ instead of $\varphi_m^{[1]}(f)$ in what follows.
(b) Still under Assumption (**P**), we define the mappings Ψ_ℓ, $\ell\in\mathbb{N}^*$,

$$\Psi_\ell=\sum_{k=1}^{\ell-1}\sum_{\substack{(m_1,\ldots,m_k)\in[\![2,\ell]\!]^k,\\ m_1+\ldots+m_k=\ell+k-1}}\varphi_{m_1}\circ\ldots\circ\varphi_{m_k}.\tag{3.38}$$

For example, note that

$$\Psi_2=\varphi_2,\quad\Psi_3=\varphi_3+\varphi_2^{[2]}\quad\text{and}\quad\Psi_4=\varphi_4+\varphi_3\circ\varphi_2+\varphi_2\circ\varphi_3+\varphi_2^{[3]}.$$

We have the following expansions of the error, depending on the averaging properties of the step sequence γ.

Proposition 3.2 (Bias error expansion for the coarse level). *Assume* (**S**), (**P**) *(and uniqueness of the invariant distribution ν). Let $R\in\mathbb{N}$, $R\geq 2$ and let $f\in\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ with polynomial growth and $g=\mathcal{Q}f$.*
(a) *If $(\gamma_n^\ell,\gamma_n)_{n\geq1}$ is averaging for every $\ell\in\{1,\ldots,R\}$,*

$$\nu_n^\gamma(\omega,f)-\nu(f)-\sum_{\ell=2}^{R}\frac{\Gamma_n^{(\ell)}}{\Gamma_n}\nu\big(\Psi_\ell(f)\big)=\frac{M_n^{(1,g)}}{\Gamma_n}+o_{L^2}\left(\frac{\sqrt{\Gamma_n}\vee\Gamma_n^{(R)}}{\Gamma_n}\right).$$

(b) *If, furthermore, the pair $(\gamma_n^{R+1},\gamma_n)_{n\geq1}$ is averaging,*

$$\nu_n^\gamma(\omega,f)-\nu(f)-\sum_{\ell=2}^{R}\frac{\Gamma_n^{(\ell)}}{\Gamma_n}\nu(\Psi_\ell(f))=\frac{M_n^{(1,g)}}{\Gamma_n}+\frac{\Gamma_n^{(R+1)}}{\Gamma_n}\nu\big(\Psi_{R+1}(f)\big)+o_{L^2}\left(\frac{\sqrt{\Gamma_n}\vee\Gamma_n^{(R+1)}}{\Gamma_n}\right).$$

(c) The following sharper expansion also holds when $(\gamma_n^{R+2}, \gamma_n)_{n\geq 1}$ is averaging

$$\nu_n^\gamma(\omega, f) - \nu(f) - \sum_{\ell=2}^{R} \frac{\Gamma_n^{(\ell)}}{\Gamma_n} \nu(\Psi_\ell(f)) = \frac{M_n^{(1,g)} + N_n}{\Gamma_n}$$
$$+ \frac{\Gamma_n^{(R+1)}}{\Gamma_n} \nu\big(\Psi_{R+1}(f)\big) + \frac{\Gamma_n^{(R+2)}}{\Gamma_n} \nu\big(\Psi_{R+2}(f)\big) + o_{L^2}\left(\frac{\sqrt{\Gamma_n^{(3)}} \vee \Gamma_n^{(R+2)}}{\Gamma_n}\right),$$

where $N_0 = 0$ and

$$\Delta N_n = \Delta M_k^{2,g} + \Delta M_k^{3,g} + \gamma_k^{\frac{3}{2}} (\sigma g_2')(\bar{X}_{k-1}) U_k,$$

with $g_2 = \mathcal{Q}(\varphi_2(f))$, i.e. the solution to $\varphi_2(f) - \nu(\varphi_2(f)) = -\mathcal{L} g_2$.

Remark 3.7. The first expansion is adapted to the proof of Theorem 2.1(a), the second one for Theorem 2.1(b) and (c) and Theorem 2.2(a). Statement (c) is written in view of Theorem 2.2(b) where one needs to handle the second order term of the asymptotic expansion of the MSE. Note that the bias term of order $R+2$ in (c) will contribute to $\widetilde{m}_f(\bar{q}, R)$ in Theorem 2.2(b). At this stage, it can be justified by the following remark: when $a = 1/(2R+1)$,

$$\frac{\Gamma_n^{(R+1)}}{\Gamma_n} \frac{\Gamma_n^{(R+2)}}{\Gamma_n} \overset{n\to+\infty}{\sim} \left(\frac{2R}{2R+1}\right)^2 \frac{1}{n}.$$

As concerns the contribution of the martingale correction ΔN_n, we refer to Proposition 4.4 for details. Finally, remark that all the negligible terms are given with the L^2-norm. For Theorem 2.1, "$o_{\mathbb{P}}$" is enough.

Proof. (a) and (b): Let $R \geq 2$ be an integer. Let us consider the decomposition given by (3.35) in Lemma 3.2. When $(\gamma_n)_{n\geq 1} = \eta = (\gamma_n)_{n\geq 1}$, $L = R$ and $g = \mathcal{Q}f$, we get

$$\nu_n^\gamma(f) - \nu(f) - \sum_{\ell=1}^{R} \frac{\Gamma_n^{(\ell)}}{\Gamma_n} \nu(\varphi_\ell(f)) = \frac{g(\bar{X}_0) - g(\bar{X}_n)}{\Gamma_n} + \sum_{\ell=2}^{R} \frac{\Gamma_n^{(\ell)}}{\Gamma_n} \left(\nu_n^{\gamma^\ell,\gamma}(\varphi_\ell(f)) - \nu(\varphi_\ell(f))\right) \tag{3.39}$$
$$+ \frac{\Gamma_n^{(R+1)}}{\Gamma_n} \nu_n^{\gamma^{R+1},\gamma}(\varphi_{R+1}(f)) + \frac{M_n^{1,g}}{\Gamma_n}$$
$$+ \frac{1}{\Gamma_n} \sum_{k=1}^{n} \left(\sum_{i=2}^{3} \Delta M_k^{(i,g)} + \Delta R_{k,R}^{(1,g)} + \Delta R_{k,R}^{(2,g)} + \Delta R_{k,R}^{(3,g)}\right).$$

By Lemma 3.3(i) applied with $(\eta_n) = (\gamma_n)$,

$$\left\|\frac{g(\bar{X}_0) - g(\bar{X}_n)}{\Gamma_n}\right\|_2 \leq \frac{C}{\Gamma_n}$$

As well, by Lemma 3.3(ii)applied for different choices of (θ_n), h and $(Z_n)_{n\geq 1}$, we have

$$\left\|\frac{1}{\Gamma_n} \sum_{k=1}^{n} \left(\Delta M_k^{(2,g)} + \Delta M_k^{(3,g)} + \Delta R_{k,R}^{(1,g)}\right)\right\|_2 \leq C \frac{\sqrt{\Gamma_n^{(2)}}}{\Gamma_n}.$$

Finally, Lemma 3.3(iii) and (iv) are adapted to manage $\Delta R_{k,R}^{(2,g)}$ and $\Delta R_{k,R}^{(3,g)}$ respectively. This yields

$$\left\|\frac{1}{\Gamma_n} \sum_{k=1}^{n} \left(\Delta R_{k,R}^{(2,g)} + \Delta R_{k,R}^{(3,g)}\right)\right\|_2 \leq C \left(\frac{\Gamma_n^{(R+2)}}{\Gamma_n} + \frac{\Gamma_n^{(R+\frac{3}{2})}}{\Gamma_n}\right) \leq C \frac{\Gamma_n^{(R+\frac{3}{2})}}{\Gamma_n}.$$

The above terms are thus negligible in expansions (a) and (b). As concerns $\nu_n^{\gamma^{R+1},\gamma}(\varphi_{R+1}(f))$, one can deduce from the polynomial growth of $\varphi_{R+1}(f)$ and from (3.41) that there exists $C > 0$ such that

$$\forall n \geq 1, \quad \left\|\nu_n^{\gamma^{R+1},\gamma}(\varphi_{R+1}(f))\right\|_2 \leq C.$$

This means that this term is negligible in the expansion (a). In (b), $(\gamma_n^{R+1}, \gamma_n)$ is averaging so that by Proposition 1.1,

$$\nu_n^{\gamma^{R+1},\gamma}(\varphi_{R+1}(f)) \xrightarrow{n\to+\infty} \nu(\varphi_{R+1}(f)) \quad a.s.$$

But using again (3.41), one checks that there is a $\delta > 0$ such that $(\|\nu_n^{\gamma^{R+1},\gamma}(\varphi_{R+1}(f))\|_{2+\delta})_n$ is a bounded sequence. Thus, an uniform integrability argument yields that

$$\nu_n^{\gamma^{R+1},\gamma}(\varphi_{R+1}(f)) \xrightarrow{n\to+\infty} \nu(\varphi_{R+1}(f)) \quad \text{in } L^2.$$

But for any ℓ, φ_ℓ is the component corresponding to $k = 1$ in the definition (3.38) of Ψ_ℓ. In (b), $\nu(\varphi_{R+1}(f))$ will thus contribute to $\nu(\Psi_{R+1})$. As well, the terms $\nu(\varphi_\ell)$, $\ell = 2, \ldots, R$ exhibited in this first expansion will certainly contribute to $\nu(\Psi_\ell)$, $\ell = 2, \ldots, R$.

Now, we focus on the second bias term of the right-hand side of (3.39). More precisely, for each $\ell \in \{2, \ldots, R\}$, we have to repeat the previous procedure: we apply the expansion (3.35) of Lemma 3.2 with $\eta = (\gamma_n^\ell)_{n\geq 1}$, $L = R - \ell + 1$, $f_\ell = \varphi_\ell$ and $g_\ell = \mathcal{Q}\varphi_\ell$ (defined above). After several transformations, this yields

$$\begin{aligned}
\frac{\Gamma_n^{(\ell)}}{\Gamma_n}\left(\nu_n^{\gamma^\ell,\gamma}(f_\ell) - \nu(f_\ell)\right) - \sum_{m=2}^{R-\ell+1} \frac{\Gamma_n^{(\ell+m-1)}}{\Gamma_n}\nu(\varphi_m^{[1]} \circ \varphi_\ell^{[1]}(f)) &= -\frac{1}{\Gamma_n}\sum_{k=1}^n \gamma_k^{\ell-1}\Delta\mathcal{Q}\varphi_\ell(\bar{X}_k) \\
&+ \sum_{m=2}^{R-\ell+1} \frac{\Gamma_n^{(\ell+m-1)}}{\Gamma_n}\left(\nu_n^{\gamma^{\ell+m-1},\gamma} - \nu\right)(\varphi_m \circ \varphi_\ell(f)) \\
&+ \frac{\Gamma_n^{(R+1)}}{\Gamma_n}\nu_n^{\gamma^{R+1},\gamma}(\varphi_{R-\ell+2} \circ \varphi_\ell(f)) \\
&+ \frac{1}{\Gamma_n}\sum_{k=1}^n \gamma_k^{\ell-1}\left(\sum_{i=1}^3 \Delta M_k^{(i,g_\ell)} + \Delta R_{k,R-\ell+1}^{(i,g_\ell)}\right).
\end{aligned} \tag{3.40}$$

Applying again Lemma 3.3 allows us to control the L^2-norm of the negligible terms:

$$\Big\|\frac{1}{\Gamma_n}\sum_{k=1}^n \gamma_k^{\ell-1}\Delta\mathcal{Q}\varphi_\ell(\bar{X}_k)\Big\|_2 \leq \frac{C\gamma_1^{\ell-1}}{\Gamma_n}$$

and

$$\Big\|\frac{1}{\Gamma_n}\sum_{k=1}^n \gamma_k^{\ell-1}\left(\sum_{i=1}^3 \Delta M_k^{(i,g_\ell)} + \Delta R_{k,R-\ell+1}^{(i,g_\ell)}\right)\Big\|_2 \leq C\frac{\sqrt{\Gamma_n^{(2\ell-1)}} \vee \Gamma_n^{(R+\frac{3}{2}+\ell-1)}}{\Gamma_n}.$$

Again, the penultimate term of the previous decomposition is negligible for expansion (a) and satisfies the following convergence property when (γ^{R+1}, γ) is averaging:

$$\frac{\Gamma_n}{\Gamma_n^{(R+1)}}\left(\frac{\Gamma_n^{(R+1)}}{\Gamma_n}\nu_n^{\gamma,\gamma^{L+1}}\big(\varphi_{R-\ell+2}^{[1]} \circ \varphi_\ell^{[1]}(f)\big)\right) \xrightarrow{n\to+\infty} \nu\big(\varphi_{L-\ell+2}^{[1]} \circ \varphi_\ell^{[1]}(f)\big) \quad a.s. \text{ and in } L^2.$$

This brings a second "contribution" to $\nu(\Psi_{R+1})$.

Finally, it remains to consider for every $\ell \in \{2, \ldots, R\}$ each term of (3.40). Setting $\ell = m_1$, $m = m_2$, the sequel of the proof consists in repeating the procedure until $k := \inf\{i : m_1 + \ldots m_i = R + i\}$. The result follows.

(c) The proof is based on the same principle but is slightly more involved since we aim at keeping all the terms which are going to play a role in the second order expansion of Theorem 2.2(b). This implies to start the previous proof with $L = R + 1$ (and in the second step with $L = R - \ell + 2$). Furthermore, the main other difference comes from the martingale component. As a complement of $M_n^{(1,g)}$, one also keeps whole the martingale terms whose L^2-norm is not negligible with respect to $\sqrt{\frac{\Gamma_n^{(3)}}{\Gamma_n}}$. In short, this corresponds to the martingale increments with a factor γ_k or $\gamma_k^{\frac{3}{2}}$. This yields the two martingale increments $\Delta M_k^{(2,g)}$ and $\Delta M_k^{(3,g)}$ of the first expansion but also the dominating martingale increment of the second expansion above : $\gamma_k \Delta M_k^{(1,g_\ell)}$. The result follows. □

Lemma 3.3. *Assume* (**S**). *Let h be a smooth function with polynomial growth. We know from Proposition 1.1 that, for every $p \in (0,+\infty)$,*

$$C_{h,p} = \sup_{n\geq 1} \|h(X_n)\|_p < +\infty. \tag{3.41}$$

Then,

(i) If $(\eta_n/\gamma_n)_{n\geq 1}$ is a non-increasing sequence of real numbers,

$$\Big\| \sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} \Delta h(\bar{X}_k) \Big\|_2 \leq C_{h,2} \frac{\eta_1}{\gamma_1}.$$

(ii) If $(Z_k)_{k\geq 1}$ is a sequence of i.i.d centered random variables with finite variance, then for any deterministic sequence $(\theta_k)_{k\geq 0}$,

$$\Big\| \sum_{k=1}^{n} \theta_k h(\bar{X}_{k-1}) Z_k \Big\|_2 \leq C_{h,2} \|Z_1\|_2 \sqrt{\sum_{k=1}^{n} \theta_k^2}.$$

(iii) For any sequence $(\theta_k)_{k\geq 1}$ of real numbers,

$$\Big\| \sum_{k=1}^{n} \theta_k h(\bar{X}_{k-1}) \Big\|_2 \leq C_{h,2} \sum_{k=1}^{n} |\theta_k|$$

(iv) For any sequence $(\theta_k)_{k\geq 1}$ of real numbers and any $r > 0$, there exists a real constant $C = C_{r,b,\sigma,h,\gamma}$ such that

$$\Big\| \sum_{k=1}^{n} \theta_k \sup_{u\in[0,1]} |h(\bar{X}_{k-1} + u\Delta \bar{X}_k)| |\Delta \bar{X}_k|^r \Big\|_2 \leq C \sum_{k=1}^{n} |\theta_k| \gamma_k^{\frac{r}{2}}.$$

Proof. Using that $(\eta_n/\gamma_n)_{n\geq 1}$ is a non-increasing sequence, we have

$$\left| \sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} \Delta h(\bar{X}_k) \right| = \frac{\eta_1}{\gamma_1} |h(\bar{X}_0)| + \sum_{k=1}^{n-1} \left(\frac{\eta_k}{\gamma_k} - \frac{\eta_{k+1}}{\gamma_{k+1}} \right) |h(\bar{X}_k)| + \frac{\eta_n}{\gamma_n} |h(\bar{X}_n)|$$

so that

$$\left\| \sum_{k=1}^{n} \frac{\eta_k}{\gamma_k} \Delta h(\bar{X}_k) \right\|_2 \leq C_{h,2} \left(\frac{\eta_1}{\gamma_1} + \sum_{k=1}^{n-1} \left(\frac{\eta_k}{\gamma_k} - \frac{\eta_{k+1}}{\gamma_{k+1}} \right) + \frac{\eta_n}{\gamma_n} \right) = C_{h,2} \frac{\eta_1}{\gamma_1}.$$

This concludes the proof of (i). Items (ii) and (iii) are straightforward consequences of the fact that $\sup_{n\geq 1} \mathbb{E}[|h(X_n)|^2] < +\infty$. For (iv), the polynomial growth of h implies that there exists $p > 0$ and a constant $C > 0$ such that for any $x, y \in \mathbb{R}^d$,

$$\sup_{u\in[0,1]} |h(x + uy)| \leq C(1 + |x|^p + |y|^p).$$

Using that b and σ are sublinear functions and Minkowskii's Inequality

$$\Big\| \sup_{u\in[0,1]} |h(\bar{X}_{k-1} + u\Delta \bar{X}_k)| |\Delta \bar{X}_k|^r \Big\|_2 \leq C\big(1 + \| |X_{k-1}|^p \|_4 + \| |\Delta X_k|^p \|_4 \big) \| |\Delta X_k|^r \|_4 \leq \tilde{C} \gamma_k^{\frac{r}{2}}$$

The last statement follows using again Minkowskii's Inequality. □

3.2 Error expansion of the correcting levels

For a given sequence $\gamma := (\gamma_n)$, let us denote by $(\bar{X}_k)_{k\geq 0}$ and $(\bar{Y}_k)_{k\geq 0}$ the two Euler schemes of the diffusion $(X_t)_{t\geq 0}$ driven by the same Brownian motion W and with the step sequences (γ_n) and (γ_n/M) respectively. We then define a sequence of empirical measures $(\mu_n^{M,\gamma})$ by

$$\mu_n^{M,\gamma}(dx) = \frac{1}{\Gamma_n}\sum_{k=1}^{n}\left(\left(\sum_{m=0}^{M-1}\frac{\gamma_k}{M}\delta_{\bar{Y}_{M(k-1)+m}}\right) - \gamma_k\delta_{\bar{X}_{k-1}}\right),\ n\geq 1.$$

By the definition (2.12), one first notes that for $r = 2,\ldots,R$, $\mu_n^{(r,M)} = \mu_n^{M,\gamma^{(r)}}$ built with the Euler schemes $\bar{X}^{(r)}$ and $\bar{Y}^{(r)}$ (keep in mind that $\gamma_k^{(r)} = \frac{\gamma_k}{M^{r-2}}$). As a consequence, expanding $(\mu_n^{M,\gamma}(f))_{n\geq 1}$ will elucidate the behavior of the refined levels in the **ML2Rgodic** procedure.

In the proposition below, we thus state a result similar to Proposition 3.2 but for the sequence $(\mu_n^{M,\gamma}(f))_{n\geq 1}$.

Proposition 3.3 (Bias error expansion for the refined levels)**.** *Assume* (**S**), (**P**) *and uniqueness of the invariant distribution ν of the diffusion is unique. Let $R\in\mathbb{N}^*$, $R\geq 2$ and let $f\in\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ with polynomial growth and let $g = \mathcal{Q}f$.*

(a) *Assume that for every $\ell\in\{1,\ldots,R\}$, the pair $(\gamma_n^\ell,\gamma_n)_{n\geq 1}$ is averaging. Then,*

$$\mu_n^{M,\gamma}(f) - \sum_{\ell=2}^{R}(M^{1-\ell}-1)\frac{\Gamma_n^{(\ell)}}{\Gamma_n}\nu(\Psi_\ell(f)) = -\frac{\mathcal{M}_n(\sigma g')}{\Gamma_n} + o_{L^2}\left(\frac{\sqrt{\Gamma_n}\vee\Gamma_n^{(R)}}{\Gamma_n}\right)$$

where for a Borel function $\varphi:\mathbb{R}^d\to\mathbb{R}$

$$\mathcal{M}_n(\varphi) = \sum_{k=1}^{n}\varphi(\bar{X}_{k-1})\big(W_{\Gamma_k}-W_{\Gamma_{k-1}}\big) - \sum_{m=0}^{M-1}\varphi(\bar{Y}_{M(k-1)+m})\big(W_{\Gamma_{k-1+\frac{m+1}{M}}} - W_{\Gamma_{k-1+\frac{m}{M}}}\big).$$

(b) *If furthermore, the pair $(\gamma_n^{R+1},\gamma_n)_{n\geq 1}$ is averaging, then the following sharper expansion also holds:*

$$\begin{aligned}\mu_n^{M,\gamma}(\omega,f) - \sum_{\ell=2}^{R}(M^{1-\ell}-1)\frac{\Gamma_n^{(\ell)}}{\Gamma_n}\nu(\Psi_\ell(f)) &= -\frac{\mathcal{M}_n(\sigma g')}{\Gamma_n} + \big(M^{-R}-1\big)\frac{\Gamma_n^{(R+1)}}{\Gamma_n}\nu\big(\Psi_{R+1}(f)\big)\\ &\quad + o_{L^2}\left(\frac{\sqrt{\Gamma_n^{(2)}}\vee\Gamma_n^{(R+1)}}{\Gamma_n}\right).\end{aligned}$$

(c) *The following sharper expansion also holds when $(\gamma_n^{R+2},\gamma_n)_{n\geq 1}$ is averaging :*

$$\mu_n^{M,\gamma}(f) - \sum_{\ell=2}^{R}(M^{1-\ell}-1)\frac{\Gamma_n^{(\ell)}}{\Gamma_n}\nu\big(\Psi_\ell(f)\big) = -\frac{\mathcal{M}_n(\sigma g') + \mathcal{N}_n(\frac{1}{2}\sigma^2 g'')}{\Gamma_n}$$

$$+\big(M^{-R}-1\big)\frac{\Gamma_n^{(R+1)}}{\Gamma_n}\nu\big(\Psi_{R+1}(f)\big) + \big(M^{-R-1}-1\big)\frac{\Gamma_n^{(R+2)}}{\Gamma_n}\nu\big(\Psi_{R+2}(f)\big) + o_{L^2}\left(\frac{\sqrt{\Gamma_n^{(2)}}\vee\Gamma_n^{(R+2)}}{\Gamma_n}\right),$$

where, for a Borel function $\varphi:\mathbb{R}^d\to\mathbb{R}$,

$$\mathcal{N}_n(\varphi) = \sum_{k=1}^{n}\varphi(\bar{X}_{k-1})\Big(\big(W_{\Gamma_k}-W_{\Gamma_{k-1}}\big)^2-\gamma_k\Big) - \sum_{m=0}^{M-1}\varphi(\bar{Y}_{M(k-1)+m})\left(\big(W_{\Gamma_{k-1+\frac{m+1}{M}}} - W_{\Gamma_{k-1+\frac{m}{M}}}\big)^2 - \frac{\gamma_k}{M}\right).$$

Proof. With the notation introduced in (2.10). Set

$$\nu_n^{\tilde{\gamma}^{2,M}}(\bar{Y},f) = \left(\sum_{k=1}^{n}\tilde{\gamma}_k^{2,M}\right)^{-1}\sum_{k=1}^{n}\tilde{\gamma}_k^{2,M}\delta_{\bar{Y}_{k-1}}.$$

One can check that for every $n \geq 1$,

$$\mu_n^{M,\gamma}(\omega, f) = \Big(\nu_{nM}^{\tilde{\gamma}^{2,M}}(\bar{Y}, f) - \nu(f)\Big) - (\nu_n^{\gamma}(f) - \nu(f)).$$

For (a) and (b), it remains now to apply Proposition 3.2(a) and (b) to both terms in the right-hand side of the above equation (with step $\tilde{\gamma}^{2,M}$ for $\nu_{nM}^{\tilde{\gamma}^{2,M}}(\bar{Y}, f)$). The result follows by concatenating martingale components and by noting that for any integer $\ell \geq 2$,

$$\frac{\sum_{k=1}^{nM} (\tilde{\gamma}_k^{2,M})^{\ell}}{\sum_{k=1}^{nM} \tilde{\gamma}_k^{2,M}} = \frac{M^{1-\ell}\Gamma_n^{(\ell)}}{\Gamma_n}.$$

For the proof of (c), the only difference with Proposition 3.2(c) is that one only keeps the martingale increment $M_n^{(2,g)}$ of the corrective term N_n. More precisely, the terms of N_n appearing with a factor $\gamma_n^{\frac{3}{2}}$ are here viewed as negligible terms. Using Lemma 3.3(ii), one easily check that these martingale corrections are bounded in L_2 by $\sqrt{\Gamma_n^{(3)}}/\Gamma_n$ (which is $o(\sqrt{\Gamma_n^{(2)}}/\Gamma_n)$). □

Remark 3.8. The fact that we keep less martingale terms in Expansion (c) can be understood as follows: in section (4.5), we will show that the apparently dominating martingale component $\mathcal{M}_n(\sigma g')$ is in fact negligible at the first order of the expansion under confluence assumptions. This implies that the covariance terms induced by the product of this martingale and the martingale corrections appearing with a factor $\gamma_k^{\frac{3}{2}}$ in N_n (see Proposition 3.2) will be also negligible at a second order.

4 Rate of convergence for the dominating martingales

In the continuity of Propositions 3.2 and 3.3, we now propose to elucidate the weak or L^2 rate of convergence of the dominating martingales , that is the martingales coming out in the above error expansions established in the former section.

4.1 The dominating martingale term involved in $\nu_n^{\gamma}(f) - \nu(f)$

We begin by stating some asymptotic results for the first and second order martingales $(M_n^{(1,g)})_{n\geq 1}$ and $(N_n)_{n\geq 1}$ which appear in the expansions of Proposition 3.2. The associated statements describe the asymptotic martingale contributions of the first (dominating) term of the **ML2Rgodic** procedure. With the view to Theorem 2.1, the first statement concerns the convergence in distribution of the dominating martingale $(M_n^{(1,g)})_{n\geq 1}$ whereas the second and third ones are crucial steps in the proof of Theorem 2.2 (a) and (b) respectively.

Proposition 4.4. *Assume* **(S)** *and* **(P)**. *Let* $g = \mathcal{Q}f$*.Then,*

(a)

$$\frac{1}{\sqrt{\Gamma_n}} M_n^{(1,g)} \overset{(\mathbb{R})}{\Longrightarrow} \mathcal{N}\Big(0; \int_{\mathbb{R}} (\sigma g')^2 d\nu\Big).$$

(b)

$$\mathbb{E}\left[\frac{(M_n^{(1,g)})^2}{\Gamma_n}\right] = \int_{\mathbb{R}} (\sigma g')^2 d\nu + o(1) \quad \text{as } n \to +\infty.$$

(c) *If* (γ_n, γ_n^2) *is averaging,*

$$\mathbb{E}\left[\frac{(M_n^{(1,g)} + N_n)^2}{\Gamma_n}\right] = \int_{\mathbb{R}} (\sigma g')^2 d\nu + \frac{\Gamma_n^{(2)}}{\Gamma_n}\Big(\sigma_{2,1}^2(f) + o(1)\Big) \quad \text{as } n \to +\infty,$$

where

$$\sigma_{2,1}^2(f) = \int_{\mathbb{R}} \left[\varphi_2((\sigma g')^2) + \frac{1}{2}(\sigma^2 g'')^2 + (\sigma g')\big(g^{(3)}\sigma^3 + 2(\sigma g_2')\big)\right] d\nu, \tag{4.42}$$

where $g_2 = \mathcal{Q}\varphi_2(f)$*, i.e. the solution to* $\varphi_2(f) - \nu(\varphi_2(f)) = -\mathcal{L}g_2$*.*

Remark 4.9. If $\gamma_n = \gamma_1 n^{-\frac{1}{2R+1}}$,

$$\frac{1}{\Gamma_n} \overset{n\to+\infty}{\sim} \frac{2R}{(2R+1)\gamma_1} n^{-\frac{2R}{2R+1}} \quad \text{and} \quad \frac{\Gamma_n^{(2)}}{\Gamma_n} \overset{n\to+\infty}{\sim} \frac{2R}{(2R+1)n}.$$

One thus retrieves the orders of the expansions established in Theorem 2.2.

Proof. (a) Using Proposition 1.1,

$$\frac{\langle M^{(1,g)}\rangle_n}{\Gamma_n} = \nu_n^\gamma\big((\sigma g')^2\big) \xrightarrow{n\to+\infty} \nu\big((\sigma g')^2\big) \quad a.s. \tag{4.43}$$

Furthermore, by Cauchy-Schwarz inequality and (3.41), we have for every $\varepsilon > 0$,

$$\sum_{k=1}^n \mathbb{E}\left[(\Delta M_k^{(1,g)})^2 1_{(\Delta M_k^{(1,g)})^2>\varepsilon}\right] \le \frac{1}{\varepsilon^2}\sum_{k=1}^n \mathbb{E}\big[(\Delta M_k^{(1,g)})^4\big] \le C\frac{\Gamma_n^{(2)}}{\Gamma_n^2} \xrightarrow{n\to+\infty} 0.$$

This second convergence implies that the so-called *Lindeberg condition* is fulfilled. Then, (a) is a consequence fo the CLT for martingale arrays (see [HH80, Corollary 3.1]).

(b) By Jensen inequality, for a given function f,

$$\mathbb{E}\big[(\nu_n^\gamma(f))^2\big] \le \mathbb{E}\big[\nu_n^\gamma(f^2)\big]$$

and it follows again from Proposition 1.1 and the fact that $\sigma g'$ has (at most) polynomial growth that

$$\sup_n \mathbb{E}\big[(\nu_n^\gamma((\sigma g')^2))^2\big] \le \sup_n \mathbb{E}\big[1 + |\bar{X}_n|^r\big] < +\infty. \tag{4.44}$$

owing to $(\mathbf{S})$ and (3.41). As a consequence, $\big(\nu_n^\gamma((\sigma g')^2)\big)_{n\ge1}$ is a uniformly integrable sequence so that $\nu_n^\gamma((\sigma g')^2)$ toward $\nu((\sigma g')^2)$ also holds in L^1. The second statement then follows from (4.43).

(c) First, using that $\mathbb{E}[U_n(U_n^2-1)] = 0$ and that $\mathbb{E}[U_n^4] = 1$, one can check that

$$\frac{1}{\Gamma_n}\mathbb{E}\big[(M_n^{(1,g)} + N_n)^2\big] = \mathbb{E}\big[\nu_n^{\gamma,\gamma}((\sigma g')^2)\big] + \frac{\Gamma_n^{(2)}}{\Gamma_n}\mathbb{E}\big[\nu_n^{\gamma^2,\gamma}(F)\big],$$

where

$$F(x) = \Big[\frac{1}{2}(\sigma^2 g'')^2 + (\sigma g')(g^{(3)}\sigma^3 + 2(\sigma g_2'))\Big](x).$$

On the one hand, since $(\gamma_n^2, \gamma_n)_{n\ge1}$ is averaging, we deduce from Proposition 1.1 that

$$\nu_n^{\gamma^2,\gamma}(F) \xrightarrow{n\to+\infty} \nu(F) \quad a.s.$$

But using uniform integrability arguments similar to (4.44), the convergence also holds in L^1. On the other hand, let us focus on $\mathbb{E}[\nu_n^{\gamma,\gamma}((\sigma g')^2)]$. We set $h = (\sigma g')^2$. Using Proposition 3.2(a) (and the fact that $\Psi_2 = \varphi_2$) with $R = 2$, we have

$$\nu_n^\gamma(h) - \nu(h) = \frac{M_n^{(1,\mathcal{Q}h)}}{\Gamma_n} + \frac{\Gamma_n^{(2)}}{\Gamma_n}\nu\big(\varphi_2(h)\big) + o_{L^2}\left(\frac{\sqrt{\Gamma_n}\vee\Gamma_n^{(2)}}{\Gamma_n}\right).$$

By (4.43), we deduce that

$$\mathbb{E}\left[\frac{(M_n^{(1,g)} + N_n)^2}{\Gamma_n}\right] = \int_{\mathbb{R}} (\sigma g')^2 d\nu + \frac{\Gamma_n^{(2)}}{\Gamma_n}\Big(\nu(\varphi_2(h) + F) + o(1)\Big).$$

The last statement follows. $\square$

4.2 The dominating martingale in the error expansion of $(\mu_n^{M,\gamma}(f))_{n\geq 1}$

In this section, we focus on the behavior of the martingale terms involved by the refined levels of the **ML2Rgodic** procedure. Thus, this corresponds to the variance induced by this procedure. On a finite horizon, Euler schemes are pathwise close (in an L^2-sense for instance) and this property implies one of the important features of multilevel procedures: reducing the bias without increasing significantly the variance. As mentioned before, on a long run scale, such a property is not true in general. More precisely, without additional assumptions, the martingale $(\mathcal{M}_n)_{n\geq 1}$ defined in Proposition 3.3 is *a priori* not negligible compared to the one induced by the first term of the **ML2Rgodic** procedure. However, this turns out to be true in presence of an asymptotic confluence assumption. This is the first statement of the next proposition. In the second one, we go deeper in the analysis of the martingale contribution of $(\mu_n^{M,\gamma}(f))_{n\geq 1}$ under a stronger confluence assumption. The second property will contribute only to Theorem 2.2(b).

Proposition 4.5. *Assume* $(\mathbf{S})$ *and* $(\mathbf{P})$. *Let h_1 and h_2 be locally Lipschitz functions with polynomial growth.*

(a) *If* $(\mathbf{C_w})$ *holds, then* $\left(\frac{\mathcal{M}_n(h_1)}{\sqrt{\Gamma_n}}\right)_{n\geq 1}$ *converges to* 0 *in* L^2.

(b) *Assume* $(\mathbf{C_s})$ *holds and that* $(\gamma_n, \gamma_n^2)_n$ *is averaging. Assume that h_1 is $\mathcal{C}^2$ and that h_1 and its derivatives have polynomial growth. Then, the martingales $(\mathcal{M}_n(h_1))$ and $(\mathcal{N}_n(h_2))$ are orthogonal and*

$$\frac{1}{\Gamma_n^{(2)}}\mathbb{E}\left[(\mathcal{M}_n(h_1)+\mathcal{N}_n(h_2))^2\right] \xrightarrow{n\to+\infty} \left(1-\frac{1}{M}\right)\left[\frac{1}{2}\int_{\mathbb{R}}(h_1'\sigma)^2 d\nu + 2\int h_2^2 d\nu\right].$$

In particular, when $h_1=\sigma g'$ and $h_2=\frac{1}{2}\sigma^2 g''$ (with $g=\mathcal{Q}f$), this variance is denoted by $\sigma^2_{2,2}(f)$ which subsequently reads

$$\sigma^2_{2,2}(f) = \left[\frac{1}{2}\int_{\mathbb{R}}(h_1'\sigma)^2 d\nu + 2\int h_2^2 d\nu\right] = \int \sigma^2\left((\sigma g'')^2 + \sigma\sigma' g' g'' + \frac{1}{2}(\sigma' g')^2\right)d\nu. \tag{4.45}$$

Proof. (a) Set $\varphi = h_1$. First, using that $\bar{X}$ and $\bar{Y}$ are built with the same Wiener increments,

$$\langle \mathcal{M}(\varphi)\rangle_n = \sum_{k=1}^n \frac{\gamma_k}{M}\sum_{m=0}^{M-1}\left(\varphi(\bar{X}_{k-1}) - \varphi(\bar{Y}_{M(k-1)+m})\right)^2$$

so that

$$\frac{\langle \mathcal{M}(\varphi)\rangle_n}{\Gamma_n} = M\sum_{m=0}^{M-1}\hat{\nu}_n^{\gamma,m}(\hat{\varphi}^2)$$

where $\hat{\nu}_n^{\gamma,m}(f) = \frac{1}{\Gamma_n}\sum_{k=1}^n \gamma_k f(\bar{X}_{k-1}, \bar{Y}_{M(k-1)+m})$ and $\hat{\varphi}(x,y) = \varphi(x)-\varphi(y)$. With similar arguments as for the proof of Proposition 1.1, for every $m\in\{0,\ldots,M-1\}$, $(\hat{\nu}_n^{\gamma,m})_n$ converges $a.s.$ to the unique invariant distribution of the duplicated diffusion ν_Δ (since Assumption $(\mathbf{C_w})$ holds). By uniform integrability arguments, one can check that the convergence holds along continuous functions with polynomial growth so that

$$\hat{\nu}_n^{\gamma,m}(\hat{\varphi}^2) \xrightarrow{n\to+\infty} \int (\varphi(x)-\varphi(y))^2\,\nu_\Delta(dx,dy) = 0 \quad a.s.$$

Again with uniform integrability arguments (using that $\sup_n \mathbb{E}[|\bar{X}_n|^r] < +\infty$ for every positive r), one can check that $\mathbb{E}[\hat{\nu}_n^{\gamma,m}(\hat{\varphi}^2)] \xrightarrow{n\to+\infty} 0$. It follows that $\mathbb{E}\left[\frac{\langle\mathcal{M}(\varphi)\rangle_n}{\Gamma_n}\right] \xrightarrow{n\to+\infty} 0$.

(b) The proof of this statement is the purpose of the end of the section. First, remark that the orthogonality of $\mathcal{M}(h_1)$ and $\mathcal{N}(h_2)$ follows from independency of the increments of the Brownian motion and from the fact that for every $s<t$, $\mathbb{E}[(W_t-W_s)((W_t-W_s)^2-(t-s))]=0$. Then, it remains to study these two martingales separately. In Lemma 4.4, we go deeper in the study of the long run behavior of the martingale $\mathcal{M}(h_1)$ under Assumption $(\mathbf{C_s})$ and in Lemma 4.5, we investigate the one of the martingale $\mathcal{N}(h_2)$. □

4.2.1 Long run behavior of $\mathcal{M}(\varphi)$ under strong confluence.

Lemma 4.4. *Under the assumptions of Proposition 4.5(b),*

$$\frac{1}{\Gamma_n^{(2)}}\mathbb{E}\left[\mathcal{M}_n(h_1)^2\right]\xrightarrow{n\to+\infty}\frac{1}{2}\Big(1-\frac{1}{M}\Big)\int_{\mathbb{R}}(h_1'\sigma)^2d\nu.$$

Proof. We temporarily write φ instead of h_1.

Step 1: We decompose $\mathcal{M}(\varphi)$ as the sum of terms involving the limiting diffusion process X:

$$\mathcal{M}(\varphi) = \mathcal{M}^{(1)}-\sum_{m=0}^{M-1}\mathcal{M}^{(2,m)}+\sum_{m=1}^{m-1}\mathcal{M}^{(3,m)}$$

where

$$\mathcal{M}_n^{(1)} = \sum_{k=1}^{n}\Big(\varphi(\bar{X}_{k-1})-\varphi(X_{\Gamma_{k-1}})\Big)\Delta W_{\Gamma_k},$$

$$\mathcal{M}_n^{(2,m)} = \sum_{k=1}^{n}\Big(\varphi(\bar{Y}_{M(k-1)+m})-\varphi(X_{\Gamma_{k-1+\frac{m}{M}}})\Big)(W_{\Gamma_{k-1+\frac{m+1}{M}}}-W_{\Gamma_{k-1+\frac{m}{M}}}),$$

$$\mathcal{M}_n^{(3,m)} = \sum_{k=1}^{n}\Big(\varphi(X_{\Gamma_{k-1+\frac{m}{M}}})-\varphi(X_{\Gamma_{k-1}})\Big)(W_{\Gamma_{k-1+\frac{m+1}{M}}}-W_{\Gamma_{k-1+\frac{m}{M}}}).$$

We first deal with $\mathcal{M}^{(1)}$ whose predictable bracket given by

$$\begin{aligned}\langle\mathcal{M}^{(1)}\rangle_n &\le \sum_{k=1}^{n}\gamma_k\big(\varphi(\bar{X}_{k-1})-\varphi(X_{\Gamma_{k-1}})\big)^2\\ &\le [\varphi]_{\rm Lip}\sum_{k=1}^{n}\gamma_k\big|\bar{X}_{k-1}-X_{\Gamma_{k-1}}\big|^2.\end{aligned}$$

Let $A^{(2)}$ be the infinitesimal generator of the duplicated diffusion $(X_t^x,X_t^{x'})_{t\ge0}$ and let us denote by $\widetilde{b}:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}^d\times\mathbb{R}^d$ and $\widetilde{\sigma}:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{M}_{2d,2q}$ the associated drift and diffusion coefficients. If we temporarily set $S(x,y)=(x-y)^2$, then

$$A^{(2)}S(x,y)=(b(x)-b(y))(x-y)+\frac{1}{2}\big(\sigma(x)-\sigma(y)\big)^2.$$

and $(\mathbf{C_s})$ reads, $A^{(2)}S\le-\alpha\,S$ or equivalently $0\le S\le-\frac{1}{\alpha}A^{(2)}S$.

Now, by mimicking the proof of (1.9) (where the result has been established for functions of the Euler scheme alone), we get that, as soon as $\frac{\sqrt{\Gamma_n}}{\Gamma_n^{(2)}}\to0$, for every smooth function $f:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$

$$\frac{1}{\Gamma_n^{(2)}}\sum_{k=1}^{n}\gamma_kA^{(2)}f(X_{\Gamma_{k-1}},\bar{X}_{k-1})\xrightarrow{a.s.}m(f)=\nu_\Delta\Big(\frac{1}{2}D^2f(.).\widetilde{b(.)}^{\otimes2}\Big)+\frac{1}{24}\mathbb{E}\big[D^{(4)}f(.)(\sigma(.)U)^{\otimes4}\big]$$

where $U\sim\mathcal{N}(0,I_q)$ and ν_Δ is the image of ν on the diagonal of $\mathbb{R}^2$ (which is the unique invariant distribution of the duplicated diffusion). Straightforward computations show that $m(S)=0$ since $\nabla S(x,y)=2\begin{pmatrix}x-y\\y-x\end{pmatrix}$, $D^{(2)}S(x,y)=2\begin{bmatrix}1&-1\\-1&1\end{bmatrix}$ and $D^{(\ell)}S\equiv0$, $\ell\ge3$. Thus, taking advantage of the strong confluence, we derive that

$$\lim_n\frac{1}{\Gamma_n^{(2)}}\sum_{k=1}^{n}\gamma_k\big(X_{\Gamma_{k-1}}-\bar{X}_{k-1}\big)^2\le-\frac{1}{\alpha}m(S)=0\quad a.s.$$

Uniform integrability arguments imply that the above convergence also holds in L^1. Thus,

$$\mathbb{E}\left[\frac{\langle\mathcal{M}^{(1)}\rangle_n}{\Gamma_n^{(2)}}\right]\longrightarrow0\quad\text{as}\quad n\to+\infty.$$

The same method of proof shows a similar result for $\mathcal{M}^{(2,m)}$, $m=0,\ldots,M-1$ (by considering the scheme $(\bar{Y}_{Mk+m})_{k\geq 0}$ and the filtration $\mathcal{G}^m_k=\mathcal{F}^W_{\Gamma_{k-1+\frac{m}{M}}}$). It follows that $\mathbb{E}\left[\frac{\langle\mathcal{M}^{(2,m)}\rangle_n}{\Gamma^{(2)}_n}\right]\longrightarrow 0$ as $n\to+\infty$.

Step 2: Now we deal with $\mathcal{M}^{(3,m)}$, $m=1,\ldots,M-1$. First we compute the predictable bracket

$$\langle\mathcal{M}^{(3,m)}\rangle_n=\frac{1}{M}\sum_{k=1}^n\gamma_k\big(\varphi(X_{\Gamma_{k-1+\frac{m}{M}}})-\varphi(X_{\Gamma_{k-1}})\big)^2.$$

Then, we decompose

$$\begin{aligned}\varphi(X_{\Gamma_{k-1+\frac{m}{M}}})-\varphi(X_{\Gamma_{k-1}})&=\underbrace{\varphi'(X_{\Gamma_{k-1}})(X_{\Gamma_{k-1+\frac{m}{M}}}-X_{\Gamma_{k-1}})}_{(a)_k}\\&\quad+\underbrace{\big(\varphi'(\Xi_{k-1})-\varphi'(X_{\Gamma_{k-1}})\big)(X_{\Gamma_{k-1+\frac{m}{M}}}-X_{\Gamma_{k-1}})}_{(b)_k},\ \Xi_{k-1}\in(X_{\Gamma_{k-1}},X_{\Gamma_{k-1+\frac{m}{M}}}).\end{aligned}$$

Let us deal first with $(b)_k$. The function φ'' being with polynomial growth, there exists some positive C and p such that for every x and y in $\mathbb{R}^d$,

$$|\varphi'(x+y)-\varphi'(x)|\leq C(1+|x|^p+|y|^p)|y|.$$

Thus,

$$\frac{1}{\Gamma^{(2)}_n}\sum_{k=1}^n\gamma_k(b)^2_k\leq\frac{C}{\Gamma^{(2)}_n}\sum_{k=1}^n\gamma_k(X_{\Gamma_{k-1+\frac{m}{M}}}-X_{\Gamma_{k-1}})^4(1+|X_{\Gamma_{k-1}})|^{2p})(1+|U_k|^{2p}).$$

Using that $\sup_t\mathbb{E}[|X^x_t|^r]<+\infty$, one easily checks that for every $r\geq 2$,

$$\sup_k\mathbb{E}[|X_{\Gamma_{k-1+\frac{m}{M}}}-X_{\Gamma_{k-1}}|^r]\leq C\gamma_k^{\frac{r}{2}}$$

so that with the help of the Cauchy-Schwarz inequality,

$$\lim_n\frac{1}{\Gamma^{(2)}_n}\sum_{k=1}^n\gamma_k\mathbb{E}(b)^2_k=0.$$

For $(a)_k$ we write

$$(a)_k=(\varphi'\sigma)\big(X_{\Gamma_{k-1}}\big)\big(W_{\Gamma_{k-1+\frac{m}{M}}}-W_{\Gamma_{k-1}}\big)+(\widetilde{a})_k$$

where

$$(\widetilde{a})_k=\varphi'\big(X_{\Gamma_{k-1}}\big)\left(\int_{\Gamma_{k-1}}^{\Gamma_{k-1+\frac{m}{M}}}b(X_s)ds+\int_{\Gamma_{k-1}}^{\Gamma_{k-1+\frac{m}{M}}}\big(\sigma(X_s)-\sigma(X_{\Gamma_{k-1}})\big)dW_s\right).$$

It is clear, owing to Doob's Inequality, that

$$\mathbb{E}(\widetilde{a})^2_k\leq\|\varphi'\|^2_{\sup}\left(\gamma_k^2\sup_{t\geq 0}\mathbb{E}|b(X_t)|^2+\gamma_k[\sigma]^2_{\rm Lip}\mathbb{E}\Big(\sup_{t\in[X_{\Gamma_{k-1}},\Gamma_{k-\frac{1}{2}})}|X_s-X_{\Gamma_{k-1}}|^2\Big)\right)\leq C_{b,\sigma,\varphi}\gamma_k^2.$$

Then $\displaystyle\frac{1}{\Gamma^{(2)}_n}\sum_{k=1}^n\gamma_k(\widetilde{a})^2_k\xrightarrow{L^1}0$ as above.

The last term of interest is again a martingale increment. We note that

$$\mathbb{E}\left((\varphi'\sigma)^2\big(X_{\Gamma_{k-1}}\big)\big(W_{\Gamma_{k-1+\frac{m}{M}}}-W_{\Gamma_{k-1}}\big)^2\,|\,\mathcal{F}^W_{\Gamma_{k-1}}\right)=\frac{m\gamma_k}{M}(\varphi'\sigma)(X_{\Gamma_{k-1}})^2.$$

The sequence $(\gamma_n, \gamma_n^2)_{n\geq 1}$ being averaging,

$$\frac{1}{\Gamma_n^{(2)}}\sum_{k=1}^{n}\gamma_k^2(\varphi'\sigma)(X_{\Gamma_{k-1}})^2 \xrightarrow{a.s.} \int_{\mathbb{R}}(\varphi'\sigma)^2 d\nu \quad \text{as} \quad n\to+\infty.$$

Uniform integrability arguments imply that $\frac{1}{\Gamma_n^{(2)}}\sum_{k=1}^n \gamma_k^2\,\mathbb{E}[(\varphi'\sigma)(X_{\Gamma_{k-1}})^2]\longrightarrow\int_{\mathbb{R}}(\varphi'\sigma)^2d\nu$ and one deduces that

$$\mathbb{E}\left[\frac{\langle\mathcal{M}^{(3,m)}\rangle_n}{\Gamma_n^{(2)}}\right]\longrightarrow \frac{m}{M^2}\int_{\mathbb{R}}(\varphi'\sigma)^2 d\nu.$$

The result then follows from the orthogonality of the martingales $\mathcal{M}^{(3,m)}$, $m=1,\ldots,M-1$ (the fact that the martingales $\mathcal{M}^1$ and $\mathcal{M}^{2,m}$ are negligible also implies by Schwarz's Inequality that so is their cross product).

□

4.2.2 Long run behavior of $\mathcal{N}(h_2)$.

We consider now the martingale

$$\mathcal{N}_n(h_2) = \mathcal{N}_n^1 - \sum_{m=0}^{M-1}\mathcal{N}^{2,m}$$

where

$$\mathcal{N}_n^1 = \sum_{k=1}^n h_2(\bar{X}_{k-1})\big((W_{\Gamma_k}-W_{\Gamma_{k-1}})^2-\gamma_k\big)$$

$$\mathcal{N}_n^{2,m} = \sum_{k=1}^n h_2(\bar{Y}_{M(k-1)+m})\big((W_{\Gamma_{k-1+\frac{m+1}{M}}}-W_{\Gamma_{k-1+\frac{m}{M}}})^2-\frac{\gamma_k}{M}\big).$$

Lemma 4.5. *Under Assumptions of Proposition 4.5(b),*

$$\frac{1}{\Gamma_n^{(2)}}\mathbb{E}\left[\mathcal{N}_n(h_2)^2\right]\xrightarrow{n\to+\infty} 2\left(1-\frac{1}{M}\right)\int_{\mathbb{R}}h_2^2 d\nu.$$

Proof. Like in the previous proof, we write φ instead of h_2. We focus on the asymptotic behavior of $\langle\mathcal{N}\rangle_n$.

First, noting that for a random variable $Z\sim\mathcal{N}(0;1)$, $\mathbb{E}\big((Z^2-1)^2\big)=2$, we get since $(\gamma_n,\gamma_n^2)_{n\geq1}$ is averaging,

$$\frac{\langle\mathcal{N}^1\rangle_n}{\Gamma_n^{(2)}} = \frac{2}{\Gamma_n^{(2)}}\sum_{k=1}^n\gamma_k^2\varphi^2(\bar{X}_{k-1})\longrightarrow 2\int_{\mathbb{R}}\varphi^2 d\nu \quad a.s. \quad \text{as} \quad n\to+\infty. \tag{4.46}$$

likewise one shows that for $m=0,\ldots,M-1$ that

$$\frac{\langle\mathcal{N}^{2,m}\rangle_n}{\Gamma_n^{(2)}}\longrightarrow\frac{2}{M^2}\int_{\mathbb{R}}\varphi^2 d\nu \quad a.s. \quad \text{as} \quad n\to+\infty.$$

By uniform integrability arguments, the above convergence extends to the expectations. Second, we focus on the "slanted" brackets. Let us set $\Delta_{m,k}=(W_{\Gamma_{k+\frac{m+1}{M}}}-W_{\Gamma_{k+\frac{m}{M}}})^2-\gamma_k/M$. Using the chaining rule for conditional expectations, we note that, for every $m\neq m'$,

$$\mathbb{E}_{k-1}\left(\varphi(\bar{Y}_{M(k-1)}+m)\varphi(\bar{Y}_{M(k-1)+m'})\Delta_{m,k-1}\Delta_{m',k-1}\right)=0$$

so that $\langle\mathcal{N}^{2,m},\mathcal{N}^{2,m'}\rangle_n\equiv 0$.

Now, let us compute $\langle\mathcal{N}^1,\mathcal{N}^{2,m'}\rangle_n$ where $m\in\{0,\ldots,M-1\}$ and $(\mathcal{N}^1,\mathcal{N}^{2,m'})$ is viewed as a couple of $(\mathcal{F}_k)$-martingales. Writing the increment $W_{\Gamma_k}-W_{\Gamma_{k-1}}$ as follows:

$$W_{\Gamma_k}-W_{\Gamma_{k-1}}=\big(W_{\Gamma_k}-W_{\Gamma_{k-1+\frac{m+1}{M}}}\big)+\big(W_{\Gamma_{k-1+\frac{m+1}{M}}}-W_{\Gamma_{k-1+\frac{m}{M}}}\big)+\big(W_{\Gamma_{k-1+\frac{m}{M}}}-W_{\Gamma_{k-1}}\big)$$

and using some standard properties of the increments of the Brownian Motion, one can check that

$$\langle \mathcal{N}^1, \mathcal{N}^{2,m'} \rangle_n = \frac{2}{M^2} \sum_{k=1}^{n} \gamma_k^2 \varphi(\bar{X}_{k-1}) \mathbb{E}\left(\varphi(\bar{Y}_{M(k-1)+m}) | \mathcal{F}_{k-1}\right).$$

Using second order Taylor expansions of φ between $\varphi(\bar{Y}_{M(k-1)+\ell-1})$ and $\varphi(\bar{Y}_{M(k-1)+\ell})$ for $\ell = 1, \ldots, m$, combined with the fact that $\sup_j \mathbb{E}[|\bar{Y}_j|^r] < +\infty$ for every $r > 0$, one derives

$$\langle \mathcal{N}^1, \mathcal{N}^{2,m'} \rangle_n = \frac{2}{M^2} \sum_{k=1}^{n} \gamma_k^2 \left(\varphi(\bar{X}_{k-1})\varphi(\bar{Y}_{M(k-1)}) + O_{L^1}(\gamma_k)\right) = \frac{\Gamma_n^{(2)}}{2M^2} \hat{\nu}_n^{\gamma,\gamma^2}(\varphi \otimes \varphi) + O_{L^1}(\Gamma_n^{(3)})$$

where $\hat{\nu}_n^{\gamma,\gamma^2}(f) = \frac{1}{\Gamma_n^{(2)}} \sum_{k=1}^{n} f(\bar{X}_{k-1}, \bar{Y}_{M(k-1)})$. Thus, the sequence $(\hat{\nu}_n^{\gamma,\gamma^2})_n$ of empirical measures associated to the duplicated diffusion (2.25) has a unique invariant distribution ν_Δ. By an adaptation of the proof of Proposition 1.1, it can thus be proved that

$$\hat{\nu}_n^{\gamma,\gamma^2}(\varphi \otimes \varphi) \xrightarrow{n \to +\infty} \nu_\Delta(\varphi \otimes \varphi) = \int \varphi^2 d\nu.$$

Once again, by a uniform integrability argument (and using what precedes), one obtains

$$\frac{1}{\Gamma_n^{(2)}} \mathbb{E}[\langle \mathcal{N}^1, \mathcal{N}^{2,m'} \rangle_n] \xrightarrow{n \to +\infty} \frac{2}{M^2} \int \varphi^2 d\nu.$$

As a conclusion of the previous convergences, one deduces that

$$\frac{1}{\Gamma_n^{(2)}} \mathbb{E}\left[\left\langle \mathcal{N}^1 - \sum_{m=0}^{M-1} \mathcal{N}^{2,m'} \right\rangle_n\right] \xrightarrow{n \to +\infty} \left(2 + \frac{2}{M^2}(M - 2M)\right) \int \varphi^2 d\nu = 2\left(1 - \frac{1}{M}\right) \int \varphi^2 d\nu.$$

□

5 Proofs of the main theorems (CLT and optimization)

Owing to the results established in the previous sections, we are now in position to prove the three main results: Theorems 2.1, 2.2 and 2.3. First keep in mind that in these theorems the step sequence reads $\gamma_n = \gamma_1 n^{-a}$ for some $\gamma_1 > 0$ and $a \in (0, 1)$.

5.1 Proof of Theorem 2.1.

We mainly detail the proof of Theorem 2.1(b) and we will only give some elements of the ones of (a) and (c) (which are based on the same principle) at the end of this section.

First, by (2.15), one reminds that $\widetilde{\nu}_n^{(R,\mathbf{W})}$ is a linear combination of ν_{n_1} and of $\mu_{n_r}^{(r,M)}$ with $n_r = \lfloor q_r n \rfloor$, $r = 2, \ldots, R$, . For $\nu_{n_1}(f)$ and $\mu_{n_2}^{(2,M)}(f)$, we will make use of the expansions given in Propositions 3.2(b) and 3.3(b) respectively. For $\mu_{n_r}^{(r,M)}(f)$, $r = 3, \ldots, R$, as defined by (2.14), we apply Proposition 3.3(b) with step sequence $(\gamma_n / M^{r-2})_{n \geq 1}$. More precisely, by (2.11),

$$(M^{1-\ell} - 1) \frac{\Gamma_{n_r}^{(\ell,r)}}{\Gamma_{n_r}^{(1,r)}} = m_{r,\ell} \frac{\Gamma_{n_r}^{(\ell)}}{\Gamma_{n_r}} \quad \text{with} \;\; m_{r,\ell} = (M^{1-\ell} - 1) M^{-(r-2)(\ell-1)},$$

so that by Proposition 3.3(b), we have for every $r \in \{2, \ldots, R\}$,

$$\mu_{n_r}^{(r,M)}(f) - \sum_{\ell=2}^{R} m_{r,\ell} \frac{\Gamma_{n_r}^{(\ell)}}{\Gamma_{n_r}} \nu(\Psi_\ell(f)) = \quad c_{R+1} m_{r,R+1} \frac{\Gamma_{n_r}^{(R+1)}}{\Gamma_{n_r}} - \frac{M^{r-2} \mathcal{M}_{n_r}^{(r)}(\sigma g')}{\Gamma_{n_r}} + o_{L^2}\left(\frac{\sqrt{\Gamma_{n_r}} \vee \Gamma_{n_r}^{(R+1)}}{\Gamma_{n_r}}\right)$$

where $(\mathcal{M}^{(r)}_{n_r})_{n\geq 1}$ is defined similarly to $\mathcal{M}_n$ but with the step sequence $(\gamma_n/M^{r-2})_{n\geq 1}$ In particular, $(\bar{X}_n, \bar{Y}_{Mn+m})$ is now a couple of Euler schemes with step sequences $(\gamma_n/M^{r-2})_{n\geq 1}$ and $(\gamma_n/M^{r-1})_{n\geq 1}$ respectively.

It follows from the expansions of *order* $R+1$ of each term of $\widetilde{\nu}_n^{(R,\mathbf{W})}$ established in Propositions 3.2(b) and 3.3(b) respectively that

$$
\begin{aligned}
\widetilde{\nu}_n^{(R,\mathbf{W})}(f) - \nu(f) &= \nu_{n_1}(f) - \nu(f) + \sum_{r=2}^{R} \mathbf{W}_r \mu_{n_r}^{(r,M)}(f) \\
&= c_{R+1} \widetilde{\mathbf{W}}_{R+1} \frac{\Gamma_n^{(R+1)}}{\Gamma_n} + \frac{M_{n_1}^{(1,g)}}{\Gamma_{n_1}} - \sum_{r=2}^{R} \mathbf{W}_r \frac{M^{r-2}\mathcal{M}^{(r)}_{n_r}(\sigma g')}{\Gamma_{n_r}} \\
&\quad + \mathrm{Bias}^{(1)}(a,R,q,n) + \mathrm{Bias}^{(2)}(a,R,q,n) + o_{L^2}\left(\frac{\sqrt{\Gamma_n} \vee \Gamma_n^{(R+1)}}{\Gamma_n}\right)
\end{aligned}
\tag{5.47}
$$

where $\mathrm{Bias}^{(1)}(a,R,q,n)$ is defined in Lemma B.9(b) and

$$
\mathrm{Bias}^{(2)}(a,R,q,n) = c_{R+1}\Big[\mathbf{W}_1\left(\frac{\Gamma_{n_1}^{(R+1)}}{\Gamma_{n_1}} - q_1^{-aR}\frac{\Gamma_n^{(R+1)}}{\Gamma_n}\right) + \sum_{r=2}^{R} \mathbf{W}_r m_{r,R+1}\left(\frac{\Gamma_{n_r}^{(R+1)}}{\Gamma_{n_r}} - q_r^{-aR}\frac{\Gamma_n^{(R+1)}}{\Gamma_n}\right)\Big].
$$

By Lemma B.9

$$
\left|\mathrm{Bias}^{(1)}(a,R,q,n)\right| + \left|\mathrm{Bias}^{(2)}(a,R,q,n)\right| \leq \frac{C}{n^{1-a}} = o\left(\frac{1}{\sqrt{\Gamma_n}}\right). \tag{5.48}
$$

As concerns the martingale components, one deduces from Propositions 4.4(a) and 4.5(a) that

$$
\sqrt{\Gamma_{n_1}}\left(\frac{M_{n_1}^{(g)}}{\Gamma_{n_1}} - \sum_{r=2}^{R} \mathbf{W}_r \frac{M^{r-2}\mathcal{M}^{(r)}_{n_r}(\sigma g')}{\Gamma_{n_r}}\right) \xrightarrow{(\mathbb{R})} \mathcal{N}\Big(0; \int_{\mathbb{R}} (\sigma g')^2 d\nu\Big).
$$

Theorem 2.1(b) then follows by Slutsky Theorem and the following remarks:

$$
\Gamma_{n_1} \overset{n\to+\infty}{\sim} \frac{\gamma_1 q_1^{1-a}}{1-a} n^{1-a}, \quad \Gamma_n^{(R+1)}\Gamma_n = \frac{1-a}{1-a(R+1)}\gamma_1^R n^{aR}
$$

and that when $a = \frac{1}{2R+1}$,

$$
1-a = 2aR = \frac{2R}{2R+1} \quad \text{and} \quad \frac{1-a}{1-a(R+1)} = 2.
$$

For the proof of Theorem 2.1(c), the only difference comes from the fact that the martingale component becomes negligible since $1-a > 2aR$ when $a \in (0, (2R+1)^{-1})$ so that $(\widetilde{\nu}_n^{(R,\mathbf{W})})_{n\geq 1}$ converges in probability towards $m_f(a,q,R)$. Finally, the proof of Theorem 2.1(a) follows the same lines but with the help of the expansions of Propositions 3.2(a) and 3.3(a).

5.2 Proof of Theorem 2.2.

(a) is an L^2-version of Theorem 2.1(b) so that it relies on the same decomposition. More precisely, it is a direct consequence of (5.47) and (5.48) combined with Propositions 4.4(b) and 4.5(a).

Claim (b) is based on the (sharper) second expansions of Propositions 3.2(c) and 3.3(c) up to order $R+2$. More precisely, using the same strategy as in (5.47), one obtains

$$(\widetilde{\nu}_n^{(R,\mathbf{W})}(f)-\nu(f)) = c_{_{R+1}}\widetilde{\mathbf{W}}_{_{R+1}}\frac{\Gamma_n^{(R+1)}}{\Gamma_n} + c_{_{R+2}}\widetilde{\mathbf{W}}_{_{R+2}}\frac{\Gamma_n^{(R+2)}}{\Gamma_n}$$
$$+ \frac{M_{n_1}^{(1,g)} + N_{n_1}}{\Gamma_{n_1}} - \sum_{r=2}^{R}\mathbf{W}_r\frac{M^{r-2}\left(\mathcal{M}_{n_r}^{(r)}(\sigma g') + \mathcal{N}_{n_r}^{(r)}(\frac{1}{2}\sigma^2 g'')\right)}{\Gamma_{n_r}}$$
$$+ \sum_{i=1}^{3}\text{Bias}^{(i)}(a,R,q,n) + \eta_n^{(1)} + \eta_n^{(2)}$$

where $\widetilde{\mathbf{W}}_{_{R+2}}$ is defined by (2.18) (and explicitly given by (2.21)), $\text{Bias}^{(3)}$ is given by

$$\text{Bias}^{(3)}(a,R,q,n) = c_{_{R+2}}\Big[\mathbf{W}_1\left(\frac{\Gamma_{n_1}^{(R+2)}}{\Gamma_{n_1}} - q_1^{-aR}\frac{\Gamma_n^{(R+2)}}{\Gamma_n}\right) + \sum_{r=2}^{R}\mathbf{W}_r m_{r,R+2}\left(\frac{\Gamma_{n_r}^{(R+2)}}{\Gamma_{n_r}} - q_r^{-aR}\frac{\Gamma_n^{(R+2)}}{\Gamma_n}\right)\Big],$$

and $\eta_n^{(1)}$ (resp. $\eta_n^{(2)}$) denotes a remainder term induced by the coarse level (resp. by the levels $r = 2,\ldots,R$). By Propositions 3.2(c) and 3.3(c), one obtains when $a = 1/(2R+1)$,

$$\|\eta_n^{(1)}\|_2 = o(n^{-\frac{R+1}{2R+1}}) \quad\text{and}\quad \eta_n^{(2)} = \mathcal{S}_n + o(n^{-\frac{R+1}{2R+1}}),$$

where $\mathcal{S}_n$ is a centered random variable independent of $M_{n_1}^{(1,g)}$ and N_{n_1} and such that $\mathbb{E}[\mathcal{S}_n^2] = o(\frac{\Gamma_n^{(2)}}{\Gamma_n^2}) = o(\frac{1}{n})$ (In fact, for $\eta_n^{(2)}$, one is slightly more precise than in Proposition 3.3(c) by separating the martingale component and the bias component in the o_{L_2}).
On the other hand, we obtain similarly to (5.48):

$$\sum_{i=1}^{3}|\text{Bias}^{(i)}(a,R,q,n)| \le \frac{C}{n^{1-a}} = \frac{C}{n^{\frac{2R}{2R+1}}}.$$

With the help of these properties (and from the independence of the stratas), one deduces that

$$\|(\widetilde{\nu}_n^{(R,\mathbf{W})}(f)-\nu(f))\|_2^2 = \left(c_{_{R+1}}\widetilde{\mathbf{W}}_{_{R+1}}\frac{\Gamma_n^{(R+1)}}{\Gamma_n} + c_{_{R+2}}\widetilde{\mathbf{W}}_{_{R+2}}\frac{\Gamma_n^{(R+2)}}{\Gamma_n}\right)^2$$
$$+\mathbb{E}\left[\left(\frac{M_{n_1}^{(1,g)} + N_{n_1}}{\Gamma_{n_1}}\right)^2\right] + \sum_{r=2}^{R}\mathbf{W}_r^2 M^{2(r-2)}\mathbb{E}\left[\left(\frac{\mathcal{M}_{n_r}^{(r)}(\sigma g') + \mathcal{N}_{n_r}^{(r)}(\frac{1}{2}\sigma^2 g'')}{\Gamma_{n_r}}\right)^2\right] + o\left(\frac{1}{n}\right).$$

The result is then a consequence of Propositions 4.4(c) and 4.5(c) combined with the following expansion available for any $\rho \in (0,1)$: $\sum_{k=1}^n k^{-\rho} = (1-\rho)^{-1}n^{1-\rho} + O(1)$ (see (B.66)). In particular, it is worth noting that when $a = 1/(2R+1)$,

$$\frac{\Gamma_n^{(R+1)}\Gamma_n^{(R+2)}}{\Gamma_n^2} \overset{n\to+\infty}{\sim} \frac{4R}{R-1}\frac{\gamma_1^{2R+1}}{n},$$

which induces the rectangular term $\widetilde{m}_f(q,R)$.

5.3 Proof of Theorem 2.3

Step 1(*Optimization of the step parameter* γ_1): This step is devoted to the optimization of the starting step γ_1, in order to *equalize the impact of the bias and of the variance* in the first term of the expansion of the MSE in (2.31). It amounts to solving the elementary minimization problem

$$\min_{\gamma_1>0}\left[\sigma_f^2(\varpi) + m_f^2(\varpi) = R^{\frac{2R}{2R+1}}\Big(\frac{2R}{2R+1}\sigma_1^2(f)\gamma_1^{-1} + 4\gamma_1^{2R}M^{-R(R-1)}c_{_{R+1}}^2\Big)\right].$$

We rely on the following elementary lemma (whose proof is left to the reader).

Lemma 5.6. *Let $A, B, R > 0$. Then,*

$$u^* := \text{argmin}_{u>0}\left[Au^{-1} + Bu^{2R}\right] = \left(\frac{A}{2RB}\right)^{\frac{1}{2R+1}}$$

and

$$\min_{u>0}\left[Au^{-1} + Bu^{2R}\right] = (2R+1)B(u^*)^{2R} = A^{\frac{2R}{2R+1}}B^{\frac{1}{2R+1}}(2R)^{\frac{1}{2R+1}}\left(1+\frac{1}{2R}\right).$$

Consequently,

$$\min_{\gamma_1>0}\left[\sigma_f^2(q,R) + m_f^2(q,R)\right] = \left(2^{\frac{1}{R}}R(2R+1)^{\frac{1}{2R}}M^{-\frac{R-1}{2}}\sigma_1^2(f)|c_{R+1}|^{\frac{1}{R}}\right)^{\frac{2R}{2R+1}}$$

attained at $\gamma_1^* = \gamma_1^*(R, M)$ given by

$$\gamma_1^* = \left(\frac{2R}{2R+1}\right)^{\frac{1}{2R+1}}(8R)^{-\frac{1}{2R+1}}|c_{R+1}|^{-\frac{2}{2R+1}}\sigma_1^2(f)^{\frac{1}{2R+1}}M^{\frac{R(R-1)}{2R+1}}. \tag{5.49}$$

Step 2 (*Optimization of the size of the coarse level*): We introduce an auxiliary allocation parameter $\rho \in (0,1)$ to dispatch the target global MSE ε^2 so that the contribution of the first and the second term in the right hand side of (2.31) are $\rho\varepsilon^2$ to $(1-\rho)\varepsilon^2$ respectively. The first of these two equalities reads

$$n^{-\frac{2R}{2R+1}}\left[\sigma_f^2(\varpi) + m_f^2(\varpi)\right] \leq \rho\varepsilon^2$$

where the step parameter $\gamma_1 = \gamma_0^*(R, M)$ is given by (5.49). One straightforwardly derives that

$$n = n(\varepsilon, R, M, \rho) = \left\lceil \rho^{-(1+\frac{1}{2R})}\mu(R)R\,\sigma_1^2(f)M^{-\frac{R-1}{2}}\varepsilon^{-2-\frac{1}{R}}\right\rceil \tag{5.50}$$

where

$$\mu(R) = 2^{\frac{1}{R}}(2R+1)^{\frac{1}{2R}}|c_{R+1}|^{\frac{1}{R}} \longrightarrow \widetilde{c} \quad \text{as} \quad R \to +\infty.$$

Step 3 (*Calibrating the depth R*): To calibrate $R = R(\varepsilon)$, we will now deal with the second term $\frac{\widetilde{\sigma}_f^2 + \widetilde{m}_f}{n}$ of the MSE expansion (2.31). Since we have no clue on the sign of the residual bias term $\widetilde{m}_f(\bar{q}, R, \gamma_1)$, we will replace it by its absolute value. Moreover, we can plug in its formula the above expression (5.49) of the optimal stepsize $\gamma_1^*(R, M)$ which yields

$$|\widetilde{m}_f(\bar{q}, R, \gamma_1^*)| = \mathbf{1}_{\{c_{R+1}\neq 0\}}\frac{|c_{R+2}|}{|c_{R+1}|}\frac{R}{R-1}\frac{1-M^{-R}}{1-M^{-1}}\sigma_1^2(f).$$

Consequently, using the function $\mathbf{\Psi}$ introduced in (2.32) and the obvious fact that $1 - M^{-R} \leq 1$, this second term will be upper-bounded by $(1-\rho)\varepsilon^2$ as soon as

$$\frac{R}{n(\varepsilon)}\left(\eta(f, R, M)\sigma_1^2(f) + \sigma_{2,1}^2(f) + \mathbf{\Psi}(M)R\left(1 - \frac{1}{M}\right)\sigma_{2,2}^2(f)\right) \leq (1-\rho)\varepsilon^2. \tag{5.51}$$

where $\eta(f, R, M) = \mathbf{1}_{\{c_{R+1}\neq 0\}}\frac{|c_{R+2}|}{|c_{R+1}|}\frac{1}{(R-1)(1-M^{-1})} \to 0$ as $R \to +\infty$ owing to the assumption made on the sequence $(c_r)_{r\geq 1}$.

Given the expression obtained for $n(\varepsilon, R, M, \rho)$, this inequality is satisfied in turn as soon as

$$\sigma_{2,1}^2(f) + \mathbf{\Psi}(M)R\left(1 - \frac{1}{M}\right)\sigma_{2,2}^2(f) \leq (1-\rho)\rho^{-(1+\frac{1}{2R})}\mu(R)\,\sigma_1^2(f)M^{-\frac{R-1}{2}}\varepsilon^{-\frac{1}{R}},$$

or equivalently

$$\varepsilon^{\frac{1}{R}}M^{\frac{R-1}{2}}R \leq \frac{1-\rho}{\rho}\rho^{-\frac{1}{2R}}\frac{\mu(R)\theta_1(f)}{\left(1-\frac{1}{M}\right)\mathbf{\Psi}(M) + R^{-1}\big(\theta_2(f) + \eta(f, R, M)\big)}, \tag{5.52}$$

where

$$\theta_1(f) = \frac{\sigma_1^2(f)}{\sigma_{2,2}^2(f)} \quad \text{and} \quad \theta_2(f) = \frac{\sigma_{2,1}^2(f)}{\sigma_{2,2}^2(f)}$$

Note that under the assumptions made on the sequence $(c_r)_{r\geq 1}$, $\theta_3(f)$, In order to ensure the above condition, we begin by rewriting the left-hand side as follows:

$$\varepsilon^{\frac{1}{R}} M^{\frac{R-1}{2}} R = \exp\left(\frac{1}{R}\left(\frac{\log M}{2} R(R-1) + R\log R + \log\varepsilon\right)\right) \tag{5.53}$$

and will apply the next lemma with $\delta = (\log M)/2$ and $R = \lceil x(\varepsilon)\rceil$:

Lemma 5.7. *Let $\delta \in (0, +\infty)$. Then, for every $\varepsilon \in (0, 1]$, there exists a unique $x(\varepsilon) \in (1, +\infty)$ solution to*

$$\delta x(x-1) + x\log x + \log(\varepsilon) = 0.$$

The function $\varepsilon \mapsto x(\varepsilon)$ is increasing and satisfies

$$\lim_{\varepsilon\to 0} x(\varepsilon) = +\infty, \quad x(\varepsilon) \leq \frac{1}{2} + \sqrt{\frac{\log\left(\frac{1}{\varepsilon}\right)}{\delta} + \frac{1}{4}} \tag{5.54}$$

and

$$x(\varepsilon) = \sqrt{\frac{\log\left(\frac{1}{\varepsilon}\right)}{\delta}} - \frac{\log_{(2)}\left(\frac{1}{\varepsilon}\right)}{4\delta} + \frac{1}{2} + \frac{\log\delta}{4\delta} + O\left(\frac{\log_{(2)}(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}}\right) \quad \text{as} \quad \varepsilon \to 0 \tag{5.55}$$

where $\log_{(2)} x = \log\log x$, $x > 1$.

Proof. The function $h : (\varepsilon, x) \mapsto \delta x(x-1) + x\log x + \log\varepsilon$ defined on $(0,1)\times[1,+\infty)$ is continuous, increasing in both ε and x, $h(\varepsilon, 1) = \log\varepsilon \leq 0$ and $\lim_{x\to+\infty} h(\varepsilon, x) = +\infty$ which ensures the existence of a unique solution $x(\varepsilon) \in [1, +\infty)$ to the equation $h(\varepsilon, x) = 0$. The monotony of $x(\varepsilon)$ follows from that of h. Its limit at infinity follows from the fact that $\lim_{\varepsilon\to 0} h(\varepsilon, x) = +\infty$ and the inequality in (5.54) is a consequence of the fact that $\delta x(\varepsilon)^2 - \delta x(\varepsilon) - \log\left(\frac{1}{\varepsilon}\right) \leq 0$ as $x(\varepsilon) \geq 1$. For the expansion, we first note that $x(\varepsilon)$ satisfies the second order equation

$$\delta x(\varepsilon)^2 + b x(\varepsilon) - \log\left(\frac{1}{\varepsilon}\right) = 0$$

with $b = \log\big(x(\varepsilon)/\alpha\big)$ where $\alpha = \exp(\delta)$ so that

$$x(\varepsilon) = \sqrt{\frac{\log\left(\frac{1}{\varepsilon}\right)}{\delta}}\left(\sqrt{1 + \frac{(\log\big(x(\varepsilon)/\alpha\big))^2}{4\delta\log\left(\frac{1}{\varepsilon}\right)}} - \frac{\log\big(x(\varepsilon)/\alpha\big)}{2\sqrt{\delta\log\left(\frac{1}{\varepsilon}\right)}}\right). \tag{5.56}$$

We derive from the inequality in Equation (5.54) that, for small enough ε,

$$0 \leq \frac{\log\big(x(\varepsilon)/\alpha\big)}{\sqrt{\log\left(\frac{1}{\varepsilon}\right)}} = O\left(\frac{\log_{(2)}(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}}\right) = o(1) \quad \text{as} \quad \varepsilon \to 0.$$

Consequently, we derive from (5.56) that

$$x(\varepsilon) = \sqrt{\frac{\log\left(\frac{1}{\varepsilon}\right)}{\delta}}\left(1 + O\left(\frac{\log_{(2)}(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}}\right)\right) \tag{5.57}$$

so that $\log x(\varepsilon) = \frac{1}{2}\left(\log_{(2)}(1/\varepsilon) - \log\delta\right) + O\left(\frac{\log_{(2)}(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}}\right)$. Plugging this back into (5.57) yields

$$\begin{aligned} x(\varepsilon) &= \sqrt{\frac{\log\left(\frac{1}{\varepsilon}\right)}{\delta}}\left(1 - \frac{\log_{(2)}\left(\frac{1}{\varepsilon}\right) - \log\delta - 2\,\delta}{4\sqrt{\delta\log\left(\frac{1}{\varepsilon}\right)}} + O\left(\frac{\log_{(2)}(1/\varepsilon)}{\log(1/\varepsilon)}\right)\right) \\ &= \sqrt{\frac{\log\left(\frac{1}{\varepsilon}\right)}{\delta}} - \frac{\log_{(2)}\left(\frac{1}{\varepsilon}\right)}{4\delta} + \frac{1}{2} + \frac{\log\delta}{4\delta} + O\left(\frac{\log_{(2)}(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}}\right). \quad \square \end{aligned}$$

Now let $x(\varepsilon, M)$ be the solution of the above equation where $\delta = \delta(M) = \frac{\log M}{2}$. We have

$$x(\varepsilon, M) = \sqrt{\frac{2\log\left(\frac{1}{\varepsilon}\right)}{\log M}} - \frac{\log_{(2)}\left(\frac{1}{\varepsilon}\right)}{2\log M} + \frac{1}{2} + \frac{\log(\log M) - \log 2}{2\log M} + O\left(\frac{\log_{(2)}(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}}\right).$$

Now, we set

$$R(\varepsilon) = R(\varepsilon, M) = \lceil x(\varepsilon, M)\rceil.$$

We derive from the above lemma the following useful estimates for $R(\varepsilon)$:

$$R(\varepsilon) \sim \sqrt{\frac{2\log\left(\frac{1}{\varepsilon}\right)}{\log M}} \xrightarrow{\varepsilon\to 0} +\infty \quad \text{and} \quad R(\varepsilon) \le \frac{3}{2} + \sqrt{\frac{2\log\left(\frac{1}{\varepsilon}\right)}{\log M} + \frac{1}{4}}.$$

Now, it follows from the very definitions of $x(\varepsilon, M)$ and $R(\varepsilon)$ that

$$h(\varepsilon, R(\varepsilon)) \ge h(\varepsilon, x(\varepsilon, M)) = 0 \ge h(\varepsilon, R(\varepsilon) - 1).$$

where h is defined in the proof of the previous lemma. Plugging these inequalities into (5.53) yields

$$1 \le \varepsilon^{\frac{1}{R(\varepsilon)}} M^{\frac{R(\varepsilon)-1}{2}} R(\varepsilon) \le M\left(1 - \frac{1}{R(\varepsilon)}\right)^{-1+\frac{1}{R(\varepsilon)}}\left(\frac{R(\varepsilon)}{M}\right)^{\frac{1}{R(\varepsilon)}}. \tag{5.58}$$

The above inequality on the right implies that (6.59) will be true as soon as $\rho = \rho(\varepsilon, M) \in (0, 1)$ satisfies

$$\frac{1-\rho}{\rho}\rho^{-\frac{1}{2R(\varepsilon)}} \ge M\left(1 - \frac{1}{R(\varepsilon)}\right)^{-1+\frac{1}{R(\varepsilon)}}\left(\frac{R(\varepsilon)}{M}\right)^{\frac{1}{R(\varepsilon)}}\left(\frac{R^{-1}(\varepsilon)\theta_2(f) + \left(1 - \frac{1}{M}\right)\mathbf{\Psi}(M)}{\mu(R(\varepsilon))\theta_1(f)}\right).$$

In fact, one will try to saturate the above condition, *i.e.* to choose $\rho(\varepsilon, M)$ such that

$$\frac{1-\rho(\varepsilon, M)}{\rho(\varepsilon, M)}\rho(\varepsilon, M)^{-\frac{1}{2R(\varepsilon)}} = M\left(1 - \frac{1}{R(\varepsilon)}\right)^{-1+\frac{1}{R(\varepsilon)}}\left(\frac{R(\varepsilon)}{M}\right)^{\frac{1}{R(\varepsilon)}}\left(\frac{R(\varepsilon)^{-1}\theta_2(f) + \left(1 - \frac{1}{M}\right)\mathbf{\Psi}(M)}{\mu(R(\varepsilon))\theta_1(f)}\right).$$

As the function $\rho \mapsto \frac{1-\rho}{\rho}\rho^{-\frac{1}{2R(\varepsilon)}}$ is a decreasing homeomorphism from $(0, 1)$ onto $(0, +\infty)$ this equation always has a solution $\rho = \rho(\varepsilon, M)$. Unfortunately it turns out to be of little interest in its present form for practical implementation since both $\theta_i(f)$ are unknown.

However, as $R(\varepsilon) \to +\infty$ as $\varepsilon \to 0$, and $\mu(R(\varepsilon)) \to \widetilde{c}$ as $\varepsilon \to 0$, we derive that

$$\frac{1-\rho(\varepsilon, M)}{\rho(\varepsilon, M)} \sim \frac{(M-1)\mathbf{\Psi}(M)}{\widetilde{c}\,\theta_1(f)} \quad \textit{i.e.} \quad \rho(\varepsilon, M) \sim \frac{1}{1 + \frac{(M-1)\mathbf{\Psi}(M)}{\widetilde{c}\,\theta_1(f)}} \quad \text{as } \varepsilon \to 0.$$

STEP 4 (*MSE, number of iterations and resulting complexity*):

▷ Resulting MSE: From what precedes, we deduce that after $n(\varepsilon, R(\varepsilon, M, \rho(\varepsilon))$ iterations, the MSE is lower than ε^{-2}.

▷ Size: it follows from Equation (5.50) in Step 1 combined with the left inequality in Equation (5.58) that

$$\begin{aligned} n(\varepsilon, R(\varepsilon), M, \rho(\varepsilon)) &\sim \left(1 + \frac{(M-1)\mathbf{\Psi}(M)}{\widetilde{c}\,\theta_1(f)}\right) \sigma_1^2(f)\, \widetilde{c}\, R(\varepsilon) \big(M^{\frac{R(\varepsilon)-1}{2}} \varepsilon^{\frac{1}{R(\varepsilon)}}\big)^{-1} \varepsilon^{-2} \\ &\precsim \left(1 + \frac{(M-1)\mathbf{\Psi}(M)}{\widetilde{c}\,\theta_1(f)}\right) \sigma_1^2(f) \widetilde{c}\, R(\varepsilon)^2 \varepsilon^{-2} \\ &\sim \frac{2}{\log M} \left(\widetilde{c} + \frac{(M-1)\mathbf{\Psi}(M)}{\theta_1(f)}\right) \sigma_1^2(f) \varepsilon^{-2} \log\left(\frac{1}{\varepsilon}\right) \quad \text{as } \varepsilon \to 0. \end{aligned}$$

▷ Complexity: Set $n(\varepsilon, M) = n(\varepsilon, R(\varepsilon), M, \rho(\varepsilon))$. The asymptotic resulting complexity satisfies

$$K(n(\varepsilon, M), M) = n(\varepsilon, M)\Big(1 + (M+1)\Big(R(\varepsilon) - 1\Big)\Big)\kappa_0 \sim (M+1)R(\varepsilon)n(\varepsilon, M)\kappa_0 \quad \text{as } \varepsilon \to 0$$

so that

$$K(n(\varepsilon, M), M) \precsim \frac{2\kappa_0(M+1)}{\log M} \left(\widetilde{c} + \frac{(M-1)\mathbf{\Psi}(M)}{\theta_1(f)}\right) \sigma_1^2(f) \varepsilon^{-2} \log\left(\frac{1}{\varepsilon}\right) \quad \text{as } \varepsilon \to 0.$$

▷ *Initialization of the step:* it follows from (5.49), the assumption on c_{R+1} and the convergence of $R(\varepsilon) \to +\infty$ that

$$\gamma_1^*(\varepsilon) \quad \sim \quad \widetilde{c}^{\,-1} M^{\frac{R(\varepsilon)}{2}} M^{-\frac{3}{4}} \quad \text{as} \quad \varepsilon \to 0$$

where we used that $\frac{R(R-1)}{2R+1} = \frac{R}{2} - \frac{3}{4} + \frac{3}{4}\frac{1}{2R+1}$. Finally using the expression of $x(\varepsilon)$, we get

$$\gamma_0^*(\varepsilon) \sim \widetilde{c}^{\,-1} \underbrace{M^{-\frac{3}{4} + \frac{\lceil x(\varepsilon, M)\rceil - x(\varepsilon, M)}{2}}}_{\in (M^{-\frac{1}{4}}, M^{-\frac{3}{4}}]} \left(\frac{\log M}{2}\right)^{\frac{1}{4}} \exp\left(\sqrt{\frac{\log M \log\left(\frac{1}{\varepsilon}\right)}{2}}\right) \left(\log\left(\frac{1}{\varepsilon}\right)\right)^{-\frac{1}{4}} \quad \text{as} \quad \varepsilon \to 0.$$

6 Numerical experiments

6.1 Practitioner's corner

In this section, we want to provide some helpful informations for some practical use of the optimized algorithm given in Theorem 2.3. Let $\varepsilon > 0$ denote the prescribed RMSE and let M be an integer greater than 2. In what follows we assume we aim at computing $\nu(f)$ for a given function f.

▷ **The weights $\mathbf{W}_r^{(R)})_{r=1,\ldots,R}$** . When the re-sizers are uniform they are computed by an instant closed form(2.23). Otherwise, they are given in full generality by the R-tuple of series (2.19) whose computation is also (almost) instantaneous. When $R = 2, 3$ one has again an instant closed form (see Examples below Lemma 2.1).

▷ **Computation of $R(\varepsilon, M)$.** We recall that $R(\varepsilon, M) = \lceil x(\varepsilon, M)\rceil$ where $x(\varepsilon, M)$ is the unique solution to $\frac{\log(M)}{2}x(x-1) + x\log x + \log(\varepsilon) = 0$. For the computation of $x(\varepsilon, M)$, we use the classical (one-dimensional) zero search Newton algorithm. For standard values of R and M, the reader may use Table 1. Finally, note that, "though" $R(\varepsilon) \sim \sqrt{\frac{2\log\left(\frac{1}{\varepsilon}\right)}{\log M}}$, one has $\lim_{\varepsilon\to 0} R(\varepsilon) - \sqrt{\frac{2\log\left(\frac{1}{\varepsilon}\right)}{\log M}} = -\infty$.

▷ **Values for $\mathbf{\Psi}(M)$ and choice of M.** The quantity $\mathbf{\Psi}(M)$ appears in the size parameter $n(\varepsilon, M)$ (and in the complexity parameter $K(f, M)$ given by (2.33)). Going back to the optimization procedure of the previous section, one remarks that for some fixed R and M, one can replace $\mathbf{\Psi}(M)$ by $\frac{\Psi(R,M)}{R}$. This strategy leads to sharper bounds on the size parameter $n(\varepsilon, M)$ for a given RMSE ε. We refer to the first paragraph of Section 6.2 for further investigations on this topic (see (6.60) below and what precedes). Consequently, in Table 2, we give some values of $\mathbf{\Psi}(M)$ but also for $\frac{\Psi(R,M)}{R}$ corresponding to some standard specifications encountered in practical simulations. This also allows to check how $\frac{\Psi(R,M)}{R}$ varies for such low values of R compared to $\mathbf{\Psi}(M)$. The conclusion is that $\mathbf{\Psi}(M)$ is an acceptable proxy of $\frac{\Psi(R,M)}{R}$.

$\frac{\Psi(R,M)}{R}$	$R = 2$	$R = 3$	$R = 4$	$\mathbf{\Psi}(M)$
$M = 2$	2.133	2.591	2.674	2.674
$M = 3$	1.200	1.278	1.245	1.278
$M = 4$	0.948	1.021	1.024	1.024

Table 2: Values of $\Psi(R, M)$ and $\mathbf{\Psi}(M)$

▷ **Computation of $n(\varepsilon, M)$.** The specification of the size of the coarse level $n(\varepsilon, M)$ and, which is less important, the *a priori* estimation of the global complexity, denoted $K(f, \varepsilon, M)$, both require to estimate, at least theoretically, the parameters $\tilde{c}$, $\theta_1(f)$ and $\sigma_1^2(f)$. We will focus on their calibration in the next paragraph. To some extent, the estimation of $\theta_2(f)$ is less important and any way out of reach at a reasonable cost.

But even at this stage, it is inserting to analyze their impact on $n(\varepsilon, M)$ in order to optimize the choice of the root M. To this end, we assume for a moment that $C = \tilde{c}\,\theta_1(f)$ is known. Going back to the sharper upper-bound of at our disposal, namely (2.33), it suggests to minimize, for fixed C, the function

$$g_C : M \longmapsto \frac{M+1}{\log M}\left(\frac{(M-1)\mathbf{\Psi}(M)}{C} + 1\right).$$

Without going further, let us just note that $2\mathbf{\Psi}(3) \leq \mathbf{\Psi}(2)$ so that $g_C(3) \leq g_C(2)$ for any C since $3/\log 2 > 4/\log 3$ so that it seems that $M = 3$ is always a better choice than $M = 2$. But as emphasized in the next section 6.2 (first paragraph devoted to a "toy" Ornstein-Uhlenbeck setting), a sharper study of the complexity involving $\frac{\Psi(R,M)}{R}$ leads to temper the answer.

▷ **Calibration of the parameters** This calibration can be performed as a pre-processing phase based on a preliminary short Monte Carlo simulation, having in mind that only rough estimates are needed.

– *Estimation of $\sigma_1^2(f)$ and $\theta_1(f)$.* First, let us consider $\sigma_1^2(f)$. Through an L^2-version of (1.8), one deduces that for a family of independent random empirical measures $(\nu_n^{(\ell)})_{\ell=1}^L$, namely

$$\frac{1}{\Gamma_n}\sum_{\ell=1}^{L}\mathbb{E}[(\nu_n^{(\ell)}(f) - \bar{\nu}_n^{(L)}(f))^2)] \xrightarrow{n\to+\infty} \sigma_1^2(f) \qquad \text{as} \quad L,\, n \to +\infty$$

where $\gamma_n = \gamma_1 n^{-a}$ with $a > 1/3$ (say $a = \frac{1}{2}$ in practice to get rid of the bias effect even for small values of n) and $\bar{\nu}_n^{(L)}(f)) = \frac{1}{L}\sum_{\ell=1}^{L}\nu_n^{(\ell)}(f)$.

As $\theta_1(f) = \frac{\sigma_1^2(f)}{\sigma_{2,2}^2(f)}$, it remains to provide an estimator of $\sigma_{2,2}^2(f)$. To do so we take advantage of the fact that $\sigma_{2,2}^2(f)$ is the (normalized) asymptotic variance of $(\mu_n^{M,\gamma})_{n\geq 1}$. We thus may use the same strategy as above. More precisely, under Assumption $(\mathbf{C_s})$, we deduce from Propositions 3.3 and 4.5 that

$$\frac{1}{\Gamma_n^{(2)}}\sum_{\ell=1}^{L}\mathbb{E}[(\mu_n^{(\ell)}(f) - \bar{\mu}_n^{(L)}(f))^2)] \xrightarrow{n\to+\infty} \sigma_{2,2}^2(f)$$

if $\gamma_n = \gamma_1 n^{-a}$ with $a > 1/5$ (say $a = \frac{1}{4}$ in practice to get rid of the bias effect even for small values of n) with $\bar{\mu}_n^{(L)}(f) = \frac{1}{L}\sum_{\ell=1}^{L} \mu_n^{(\ell)}(f)$.

– *About* $\widetilde{c}$ *and* $\theta_2(f)$. The coefficient $\widetilde{c}$ will probably always remain mysterious. On the other hand in practice what we really need is rather $|c_{R(\varepsilon)}|^{\frac{1}{R(\varepsilon}}$. However, under the assumption $\lim_{R\to+\infty} |c_R|^{\frac{1}{R}} = \widetilde{c} \in (0,+\infty)$ made on c_R in Theorem 2.3, one can make the guess from its very definition that its value is not too far from 1 or is at least of order a few units. In particular, if the coefficients c_R have a polynomial growth or even $c_R = O(\exp|R|^{\vartheta_0})$, $\vartheta_0 \in [0,1)$, $\widetilde{c} = 1$. If they have an exponential growth it remains finite (but possibly large). The point of interest is that, anyway, this value is much more stable than the first coefficient itself c_1 which would come out in a standard $MLMC$ Langevin simulation framework.

The parameter $\theta_2(f)$ seems to be unaccessible as well, but for another reason: it is the variance induced by a second order martingale. However as noticed in Section 6.2 (first paragraph), $\theta_2(f)$ is the *ratio of two variance terms* so that it seems not so dependent too much on the magnitude of the diffusion coefficient (in fact it can be noted that the same property holds for $\theta_1(f)$).

Remark 6.10. The numerical investigations of the next section show that the algorithm is very robust to the choice of the parameters. For simple practice, we thus recommend to get a rough estimation of $\sigma_1^2(f)$ and possibly of $\theta_1(f)$ and to fix $\theta_2(f) = \widetilde{c} = 1$.

6.2 Numerical tests

We propose in this section to provide some numerical tests of our algorithm.

Orstein-Uhlenbeck process: oracle and blind simulation We begin with the Ornstein-Uhlenbeck process in dimension 1 solution to

$$dX_t = -\frac{1}{2}X_t dt + \sigma dW_t.$$

with $f(x) = x^2$. We recall that this case is a toy example since whole the computations can be made explicit. In particular, $\nu \sim \mathcal{N}(0,\sigma^2)$ so that $\nu(f) = \sigma^2$. Furthermore, $g(x) = x^2$ is the unique solution (up to a constant) to the Poisson equation $f - \nu(f) = -\mathcal{L}g$ and it follows that

$$\sigma_1^2(f) = \sigma_{2,2}^2(f) = 4\sigma^4, \quad \text{and} \quad \sigma_{2,1}^2(f) = 5\sigma^4.$$

The reader can remark that in this case, the ratios $\theta_1(f)$ and $\theta_2(f)$ do not depend on σ. Even though this property can not be really generalized, it however emphasizes a stability of these parameters with respect to the variance of the model. The bias terms can also be computed: using that $\varphi_2(f) = \frac{1}{4}f$ and that $\varphi_\ell = 0$ for $\ell \geq 3$, we get $c_{R+1} = \sigma^2/4^R$ (so that $\tilde{c} = 1/4$).

We want in this part to get a sharp estimate of the complexity for several choices of couples (R, M). Following the optimization procedure, we go back to the definition of $n(\varepsilon, R, M, \rho)$ given in (5.50):

$$n = n(\varepsilon, R, M, \rho) = \left\lceil \rho^{-(1+\frac{1}{2R})}\mu(R) R\,\sigma_1^2(f) M^{-\frac{R-1}{2}}\varepsilon^{-2-\frac{1}{R}} \right\rceil$$

and for each value of R and M, we solve by a Newton method the following equation for $\rho \in [0,1]$:

$$\varepsilon^{\frac{1}{R}} M^{\frac{R-1}{2}} R = \frac{1-\rho}{\rho}\rho^{-\frac{1}{2R}} \frac{\mu(R)\theta_1(f)}{R^{-1}\theta_2(f) + \left(1-\frac{1}{M}\right)R^{-1}\Psi(R,M)} \tag{6.59}$$

where the values of $\Psi(R, M)$ for $R, M = 2, 3, 4$ are given in Table 2.

We denote by $\rho^\star$ the solution of this equation. Then, the complexity $K(\varepsilon, M)$ (where we assume that $\kappa_0 = 1$) is given by

$$K(\varepsilon, R, M) = \left(1 + M\left(1 - \frac{1}{R}\right)\right) n(\varepsilon, R, M, \rho^\star). \tag{6.60}$$

$\sigma = 1$	$R = 2$	R=3	R=4
$M = 2$	$\mathbf{1.09 * 10^6}$	$1.58 * 10^6$	$2.55 * 10^6$
$M = 3$	$1.11 * 10^6$	$1.43 * 10^6$	$2.05 * 10^6$
$M = 4$	$1.18 * 10^6$	$1.57 * 10^6$	$2.27 * 10^6$

$\sigma = 4$	$R = 2$	R=3	R=4
$M = 2$	$7.02 * 10^8$	$5.23 * 10^8$	$7.34 * 10^8$
$M = 3$	$7.17 * 10^8$	$\mathbf{4.76 * 10^8}$	$6.10 * 10^8$
$M = 4$	$7.56 * 10^8$	$4.99 * 10^8$	$6.55 * 10^8$

Table 3: $K(\varepsilon, R, M)$ for $\varepsilon = 10^{-2}$

This yields the following results for $\varepsilon = 10^{-2}$:

On this example, we retrieve the property which says that $M = 2$ is a good choice when $\tilde{c}\theta_1$ is small whereas $M = 3$ can be greater when this quantity increases. However, as expected, the main parameter is the level R of the method which increases when $\varepsilon \to 0$.

Taking only the first term of the expansion of the MSE for the crude procedure, the optimized complexity (with $\kappa_0 = 1$) for a MSE lower than $\varepsilon = 10^{-2}$ is equal to $K(\varepsilon) = 6.93 * 10^6$ and $K(\varepsilon) = 1.77 * 10^9$ if $\sigma = 1$ or $\sigma = 4$ respectively.

In Figure 1, we compare numerically the evolution of **ML2Rgodic** with the crude algorithm for $\sigma = 1$ and $\sigma = 4$. Note that to obtain a rigorous comparison, the graphs are drawn in terms of the complexity, that once again with a slight abuse of language, is the number of iterations of the Euler scheme involved by procedure. One remarks that the effect of the Multilevel-RR procedure

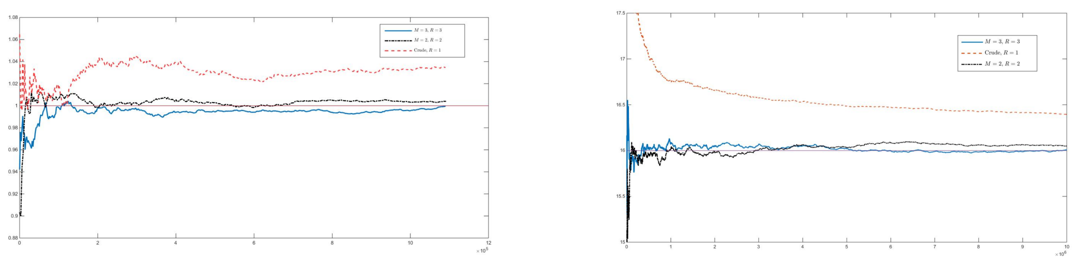

Figure 1: Comparison of the evolution in terms of the complexity of the ML2Rgodic with the crude algorithm

is increased in the case $\sigma = 4$ where the bias is larger. One also remarks in this case that, even though the algorithm is robust to the choice M and R, the best choice seems to be the one given in Table 3.

Of course, in practice, one can not make use of the exact parameters. As explained in Section 6.1, it is possible to get a rough estimation of $\sigma_1^2(f)$ and $\theta_1(f)$ using the CLTS induced by the procedure. The coefficient c_{R+1} can also be estimated but for this coefficient, this requires to use a Multistep method or the procedure ML2Rgodic itself with one more stratum than in the algorithm that we will implement after. Finally, the coefficient $\theta_2(f)$ seems to be impossible to estimate. This implies that the natural question that the practitioner may ask is: is it possible to get rid of the estimation of the above parameters ?

To answer to this question, we propose in the case $\sigma = 4$ to look at the dynamics of the procedure when we choose to fix

- $c_{R+1} = \theta_2(f) = 1$ and to estimate $\sigma_1^2(f)$ and $\theta_1(f)$,
- $c_{R+1} = \theta_2(f) = \sigma_1^2(f) = \theta_1(f)$.

With these two choices of parameters and with $\varepsilon = 10^{-2}$, we follow the procedure described in the previous section to estimate $\gamma_1^\star$, R, ρ and M. Note that we again obtain $R = 3$ and $M = 3$ as an optimal choice. In Figure 2, we thus compare the evolution of the previous method (with semi-estimated or not estimated) parameters and we can remark on this example that the algorithm seems to be very robust to the choice of the parameters.

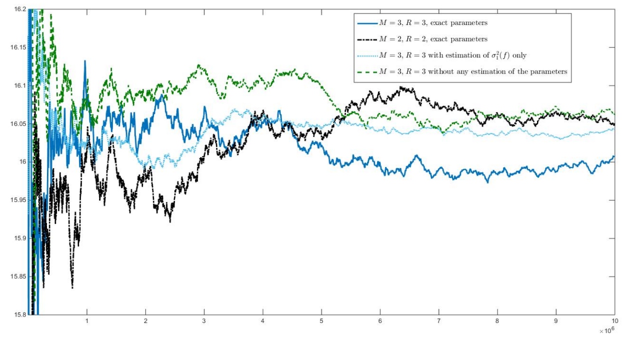

Figure 2: Evolution of the algorithm in terms of the estimation of the parameters, Exact Value : $\nu(f) = 16$.

Double-well potential We consider a second example in dimension 1

$$dX_t = -V_1'(X_t)dt + \sigma dW_t$$

where $V_1(x) = x^2 - \log(1 + x^2)$ which is a non-convex potential (with two local minima in -1 and 1) so that Assumption $(\mathbf{C_s})$ is not fulfilled. However, Assumption $(\mathbf{C_w})$ is true (see [LPP15], Theorem ??). Let us also recall that the invariant distribution ν satisfies

$$\nu(dx) = \frac{1}{Z_{V_1}} \exp\left(-\frac{V_1(x)}{2\sigma^2}\right) \lambda(dx)$$

where $Z_{V_1} = \int_{\mathbb{R}} \exp\left(-\frac{V_1(x)}{2\sigma^2}\right) \lambda(dx)$.

We test the algorithm in this setting with $f(x) = x^2$ and $\sigma = 2$. Figure 3 shows that ML2Rgodic is still efficient in this setting. The results are obtained using a rough estimation of $\sigma_1^2(f)$ and $\theta_1(f)$ and the other parameters are fixed to 1. Once again, the evolution is compared with the crude algorithm with an optimized choice of γ_1^* and the evolution is drawn as a function of the complexity.

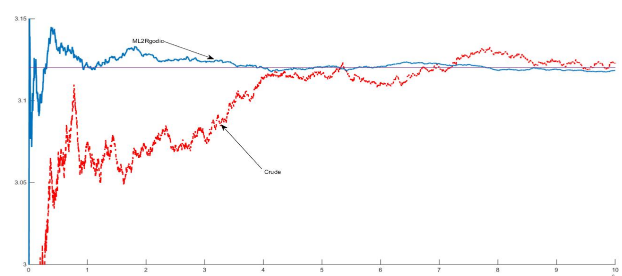

Figure 3: Approximation of $\nu(f)$ with $f(x) = x^2$, $\sigma = 2$, Exact value : 3.1207.

Statistical example In [DT12], the authors consider the problem of Sparse Regression Learning by Aggregation. For the sake of simplicity, we only recall the case of linear regression: let p denote the number of variables and n the number of observations and suppose we are given n couples of observations $(\mathbf{X}_1, \mathbf{Y}_1), \ldots (\mathbf{X}_n, \mathbf{Y}_n)$ where $\mathbf{X}_i = (X_i^1, \ldots, X_i^p)$ (called the *predictor*) and $\mathbf{Y}_i$ is a scalar (called the *response*). Suppose that there exists $\theta_0 \in \mathbb{R}^p$ such that

$$\forall i \in \{1, \ldots, n\}, \quad \mathbf{Y}_i = \mathbf{X}_i \theta_0 + \xi_i$$

where $(\xi_i)_{i=1}^n$ denotes a sequence of *i.i.d.* random variables with distribution $\mathcal{N}(0, \sigma^2)$ for a given (generally unknown) $\sigma > 0$. Then, the classical question is: how to estimate θ_0 ? When $p >> n$, the classical methods (such as the least-square method) do not work and it is necessary to introduce some alternative procedures. Dalalyan and Tsybakov propose an estimator of θ_0 called *EWA* (Exponentially Weighted Aggregate) which is built as follows:

$$\hat{\theta} = \int \theta \pi_{V_2}(d\theta)$$

where Π_{V_2} is the Gibbs probability measure defined by:

$$\pi_{V_2}(d\theta) = \frac{1}{Z_{V_2}} \exp\left(-V_2(\theta)\right) \lambda(d\theta)$$

where Z_{V_2} is a normalizing coefficient and V_2 is defined by

$$\forall \theta \in \mathbb{R}^p, \quad V_2(\theta) = \frac{|\mathbf{Y} - \mathbf{X}\theta|^2}{\beta} + \sum_{j=1}^{d} \log(\tau^2 + \theta_j^2) + \omega(\alpha\theta_j)$$

with $\omega(\theta) = \theta^2 \vee (2|\theta| - 1)$.

As mentioned (and already numerically tested) in [DT12], $\hat{\theta}$ is thus the expectation related to the invariant distribution of the following SDE

$$d\theta_t = -\nabla V_2(\theta_t)dt + \sqrt{2}dW_t,$$

it can be estimated through a Langevin Monte-Carlo procedure. The difficulty in this context is the fact that p is potentially large so that the numerical computation needs some adaptations.

Below, we test our algorithm with $p = 200$, $n = 100$ and $S = 20$ where S denotes the sparsity parameter. More precisely, we assume that θ_0 has only S coordinates different from 0 (but one does not know which ones). Then, the matrices $\mathbf{X}$ and $\mathbf{Y}$ are computed numerically (taking $\sigma = 1$). Denoting by $\hat{\theta}_n$ the approximation of $\hat{\theta}$ obtained after n iterations of the scheme, we draw in Figure 4, the evolution of $(\|\hat{\theta}_n - \theta_0\|_2)$ with n. Note that $(\|\hat{\theta}_n - \theta_0\|_2)$ converges to $.\|\hat{\theta} - \theta_0\|_2$ (which in particular is not equal to 0).

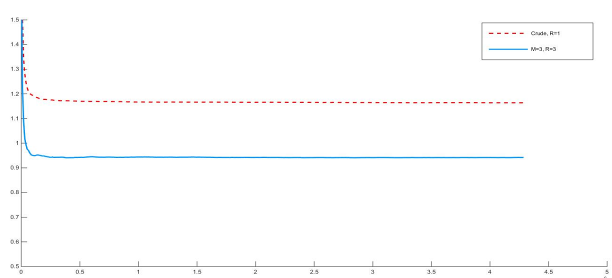

Figure 4: $n \mapsto \|\hat{\theta}_n - \theta_0\|_2$ for the Crude and **ML2Rgodic** procedures.

Double-well potential We consider a second example in dimension 1

$$dX_t = -V_1'(X_t)dt + \sigma dW_t$$

where $V_1(x) = \sqrt{1 + (x^2 - 1)^2}$ which is a non-convex potential.

References

[Bha82] R. N. Bhattacharya. On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete*, 60(2):185–201, 1982.

[Bil78] Patrick Billingsley. *Ergodic theory and information.* Robert E. Krieger Publishing Co., Huntington, N.Y., 1978. Reprint of the 1965 original.

[DT12] A. S. Dalalyan and A. B. Tsybakov. Sparse regression learning by aggregation and Langevin Monte-Carlo. *J. Comput. System Sci.*, 78(5):1423–1443, 2012.

[Fri16] Noufel Frikha. Multi-level stochastic approximation algorithms. *Ann. Appl. Probab.*, 26(2):933–985, 2016.

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

[GT83] David Gilbarg and Neil S. Trudinger. *Elliptic partial differential equations of second order*, volume 224 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, second edition, 1983.

[GT15] Camilo Andrés García Trillos. A decreasing step method for strongly oscillating stochastic models. *Ann. Appl. Probab.*, 25(2):986–1029, 2015.

[HH80] P. Hall and C. C. Heyde. *Martingale limit theory and its application.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980. Probability and Mathematical Statistics.

[Kre85] Ulrich Krengel. *Ergodic theorems*, volume 6 of *de Gruyter Studies in Mathematics.* Walter de Gruyter & Co., Berlin, 1985. With a supplement by Antoine Brunel.

[Lem05] V. Lemaire. *Estimation récursive de la mesure invariante d'un processus de diffusion.* Thèse de doctorat, Université de Marne-la-Vallée (France), 2005.

[Lem07] V. Lemaire. Behavior of the Euler scheme with decreasing step in a degenerate situation. *ESAIM Probab. Stat.*, 11:236–247, 2007.

[LP02] D. Lamberton and G. Pagès. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3):367–405, 2002.

[LP03] D. Lamberton and G. Pagès. Recursive computation of the invariant distribution of a diffusion: the case of a weakly mean reverting drift. *Stoch. Dyn.*, 3(4):435–451, 2003.

[LP13] V. Lemaire and G. Pagès. Multilevel Richardson-Romberg extrapolation. *Bernoulli (to appear in)*, 2013.

[LPP15] V. Lemaire, G. Pagès, and F. Panloup. Invariant measure of duplicated diffusions and application to Richardson-Romberg extrapolation. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(4):1562–1596, 2015.

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

[Pan08] F. Panloup. Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. *Annals of Applied Probability*, 18(2):379–426, 2008.

[PP09] G. Pagès and F. Panloup. Approximation of the distribution of a stationary Markov process with application to option pricing. *Bernoulli*, 15(1):146–177, 2009.

[PP14] G. Pagès and F. Panloup. A mixed-step algorithm for the approximation of the stationary regime of a diffusion. *Stochastic Process. Appl.*, 124(1):522–565, 2014.

[PS94] Mauro Piccioni and Sergio Scarlatti. An iterative Monte Carlo scheme for generating Lie group-valued random variables. *Adv. in Appl. Probab.*, 26(3):616–628, 1994.

[PV01] E. Pardoux and A. Yu. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3):1061–1085, 2001.

[Tal90] Denis Talay. Second order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stoch. Stoch. Rep.*, 29(1):13–36, 1990.

A Proof of Lemma 2.1

Prior to the proof of Lemma 2.1, we need to prove this first technical lemma which will be used to estimate in a precise way the coefficients $\widetilde{\mathbf{W}}_{R+1}$ and $\widetilde{\mathbf{W}}_{R+2}$ involved in the asymptotic mean square error of the **ML2Rgodic** estimator in Theorems 2.1 and 2.2.

Lemma A.8. *Let $R \geq 2$ be an integer. If $(y_r)_{r=1,\ldots,R-1}$ is solution to the $(R-1)\times(R-1)$-Vandermonde system*

$$\sum_{r=1}^{R-1} x_r^{\ell-1} y_r = c^{\ell-1},\ \ell = 1,\ldots,R-1,$$

where the x_r are pairwise distinct, then

$$y_r = \frac{\prod_{s=1,s\neq r}^{R-1}(x_r - c)}{\prod_{s=1,s\neq r}^{R-1}(x_r - x_s)} \tag{A.61}$$

and

$$\sum_{r=1}^{R-1} y_r x_r^{R-1} = c^{R-1} - \prod_{r=1}^{R-1}(c - x_r) \tag{A.62}$$

$$\sum_{r=1}^{R-1} y_r x_r^{R} = c^{R} + \left(\sum_{r=1}^{R-1} x_r + c\right)\prod_{r=1}^{R-1}(c - x_r). \tag{A.63}$$

Proof. The above Vandermonde system $\mathrm{Vand}(x_r, r = 1 : R)\mathbf{w} = [0^{\ell-1}]_{\ell=1,R}$ can be explicitly solved by the Cramer formulas since its right hand side is of the form $[c^{\ell-1}]_{1\leq \ell\leq R}$ for some $c\in\mathbb{R}$. Namely

$$y_r = \frac{\det(\mathrm{Vand}(x_1,\ldots,x_{r-1}0, x_{r+1},\ldots,x_{R-1}))}{\det(\mathrm{Vand}(x_s, s = 1 : R-1))},\ r = 1,\ldots,R-1$$

where the column vector $[c^{\ell-1}]$ replaces the r^{th} column of the original Vanderlmonde matrix. Then, elementary computations show that it yields the announced solutions.

To compute the next two sums, we start from the following canonical decomposition of the rational fraction

$$\frac{1}{\prod_{r=1}^{R-1}(X - \frac{1}{x_r - c})} = \sum_{r=1}^{R-1} \frac{1}{(X - \frac{1}{x_r-c})\prod_{s\neq r}(\frac{1}{x_r-c} - \frac{1}{x_s-c})}.$$

Setting $X = 0$ yields after elementary computations

$$\sum_{r=1}^{R-1} y_r (x_r - c)^{R-1} = (-1)^R \prod_{r=1}^{R-1}(x_r - c).$$

Now, using that $(y_r)_{r=1,\ldots,R-1}$ solves the above Vandermonde system, we get

$$\begin{aligned}\sum_{r=1}^{R-1} y_r(x_r-c)^{R-1} &= \sum_{r=1}^{R-1} y_r \sum_{k=0}^{R-1}\binom{R-1}{k}(-1)^{R-1-k}x_r^k c^{R-1-k}\\ &= \sum_{k=0}^{R-1}\binom{R-1}{k}(-1)^{R-1-k}c^{R-1-k}\underbrace{\sum_{r=1}^{R-1} y_r x_r^k}_{=c^k \text{if } k<R-1}\\ &= \sum_{r=1}^{R-1} y_r x_r^{R-1} + c^{R-1}\big((1-1)^{R-1}-1\big)\end{aligned}$$

so that

$$\sum_{r=1}^{R-1} y_r x_r^{R-1} = c^{R-1} - (-1)^{R-1}\prod_{r=1}^{R-1}(x_r-c) = c^{R-1} - \prod_{r=1}^{R-1}(c-x_r).$$

The second identity follows likewise by differentiating the above rational fraction with respect to X and then setting $X=0$ again. □

Proof of Lemma 2.1. (a) We introduce the auxiliary variables and parameters

$$\overline{\mathbf{W}}_r = \left(\frac{q_1}{q_{r+1}}\right)^a \frac{W_{r+1}}{M^{r-1}}, \qquad x_r = M^{-(r-1)}\left(\frac{q_1}{q_{r+1}}\right)^a, \ r=1,\ldots,R-1. \tag{A.64}$$

Then $(\mathbf{W}_r)_{1\le r\le R-1}$ is solution to the sytem (2.17) if and only if $(\overline{\mathbf{W}}_r)_{1\le r\le R-1}$ is solution to

$$\sum_{r=1}^{R-1}\overline{\mathbf{W}}_r x_r^{\ell-1} = \frac{1}{1-M^{-\ell}},\ \ell=1,\ldots,R-1.$$

Expanding $\dfrac{1}{1-M^{-\ell}} = \sum_{k\ge 0}\dfrac{1}{M^k}\dfrac{1}{M^{k(\ell-1)}}$ yields by linearity of the above system that it suffices to solve the sequence of Vandermonde systems.

$$(\mathcal{V}_k)\equiv \sum_{r=1}^{R-1}\overline{\mathbf{W}}_{k,r}x_r^{\ell-1} = M^{-k(\ell-1)},\ \ell=1,\ldots,R-1,\ \ k\ge 0.$$

As the x_r are pairwise distinct, $(\mathcal{V}_k)$ has a unique solutions given by

$$\overline{\mathbf{W}}_{k,r} = \prod_{s=1,s\neq r}^{R-1}\frac{x_s-M^{-k}}{x_s-x_r},\ r=1,\ldots,R-1.$$

with thz usual convention $\prod_{\emptyset}=1$ Consequently, for every $r=2,\ldots,R$,

$$\overline{\mathbf{W}}_r = \sum_{k\ge 0}\frac{1}{M^k}\overline{\mathbf{W}}_{k,r} = \sum_{k\ge 0}\frac{1}{M^k}\prod_{s=1,s\neq r}^{R-1}\frac{x_s-M^{-k}}{x_s-x_r},\ r=1,\ldots,R-1.$$

Coming back to the weights of interest finally yields the expected formula.

One derives from the definition (2.18) of $\widetilde{\mathbf{W}}_{R+1}$, using the auxiliary variables, that

$$\widetilde{\mathbf{W}}_{R+1} = q_1^{-aR}\big(1+(M^{-R}-1)\widetilde{\overline{\mathbf{W}}}_R\big)$$

with

$$\widetilde{\overline{\mathbf{W}}}_R = \sum_{r=1}^{R-1}\overline{\mathbf{W}}_r x_r^{R-1}$$

where x_r is given by (A.64). Following the lines of (a), we derive that

$$\widetilde{\widetilde{\mathbf{W}}}_{R} = \sum_{k\geq 0} \frac{1}{M^k}\, \widetilde{\widetilde{\mathbf{W}}}_{R,k}$$

where the identity (**??**) established in lemma A.8 below yields

$$\widetilde{\widetilde{\mathbf{W}}}_{R,k} = M^{-k(R-1)} - \prod_{r=1}^{R-1}(M^{-k} - x_r).$$

Finally

$$\widetilde{\mathbf{W}}_{R+1} = q_1^{-aR}\left(1 + (M^{-R}-1)\sum_{k\geq 0}\frac{1}{M^{kR}}\Big(1 - \prod_{r=0}^{R-2}\Big(1 - M^{k-r}\Big(\frac{q_1}{q_{r+2}}\Big)^a\Big)\right).$$

Noting that $\sum_{k\geq 0}\frac{1}{M^{kR}} = \frac{1}{1-M^{-R}}$ completes the proof this claim.

(b) In the the starting system (2.17) for the weights $q_r^{a(\ell-1)}$ no longer depends on r and can be cancelled in each equation. This leads to the system

$$\mathbf{W}_1 = 1, \quad 1 + (M^{-(\ell-1)} - 1)\sum_{r=2}^{R} M^{-(r-2)(\ell-1)}\mathbf{W}_r = 0,\ \ell = 2:R.$$

After a standard Abel transform, we get that $\mathbf{W}_r = \mathbf{w}_r + \cdots + \mathbf{w}_R$ where the $\mathbf{w}_r$ are solution to the Vandermonde system

$$\sum_{r=1}^{R} M^{-(r-1)(\ell-1)}\mathbf{w}_r = 0^{\ell-1},\ \ell = 1:R.$$

Note that these weights corresponds to those coming out when dealing with $ML2R$ for regular Monte Carlo (see [LP13]) under a weak error expansion condition at rate $\alpha = 1$.

As for the boundedness, first note that the "small" weights $\mathbf{w}_r$ read $\mathbf{w}_r = b_{R-r}/a_r$, $r = 1,\ldots,R$, with

$$a_r = \prod_{k=1}^{r}(1 - M^{-k}) \quad \text{and} \quad b_r = (-1)^r M^{-\frac{r(r-1)}{2}} a_r^{-1}.$$

One straightforwardly checks that $a_r \downarrow a_\infty = \prod_{k\geq 1}(1-M^{-k}) > 0$ and $B_\infty = \sum_{r\geq 1}|b_r| < +\infty$. As a consequence

$$\forall\, R\in,\ \forall\, r\in\{1,\ldots,R\}, \quad |\mathbf{W}_r^{(R)}| \leq \frac{B_\infty}{a_\infty} < +\infty.$$

Finally, the same Abel transform shows that

$$\widetilde{\mathbf{W}}_{R+i} = R^{a(R+i)}\sum_{r=1}^{R} M^{-(r-1)(R+i-1)}\mathbf{w}_r$$

and one concludes by formula (A.62) and (A.63) from Lemma A.8. □

B An additional bias term

In this part of the appendix, we focus the bias induced by the approximation

$$\frac{\Gamma_{n_r}^{(\ell)}}{\Gamma_{n_r}} \approx q_r^{-a\ell}\frac{\Gamma_n^{(\ell)}}{\Gamma_n} \quad (\text{with } \gamma_n = \gamma_1 n^{-a},\ a\in(0,1)),$$

that we use to build some universal weights $(\mathbf{W}_r^{(R)})_{r=1,\ldots,R}$ (by universal, we mean that they do not depend on n). We have the following lemma:

Lemma B.9. *Assume that* $\gamma_n = \gamma_1 n^{-a}$ *with* $a \in (0,1)$.

(a) Let $\chi \in (0,1)$ *and* $L \in \mathbb{N}$ *such that* $La < 1$. *Then, for every* $n \geq n_0 = \lceil \frac{6^{\frac{1}{1-a}}}{\chi} \rceil$,

$$\left| \frac{\Gamma^{(\ell)}_{\lfloor \chi n \rfloor}}{\Gamma_{\lfloor \chi n \rfloor}} - \chi^{-a(\ell-1)} \frac{\Gamma^{(\ell)}_n}{\Gamma_n} \right| \leq 3\Big(1 + \frac{1-a}{1-a(R+1)}\Big) \frac{\gamma_1^{\ell-1}}{n^{1-a}} \frac{\chi^{-a\ell}}{\chi^{1-a} - 3n^{a-1}}$$
$$\leq \left(6\, \frac{2-aL}{1-aL} \gamma_1^{\ell-1} \chi^{-1-a(\ell-1)}\right) \frac{1}{n^{1-a}}. \tag{B.65}$$

(b) Set

$$\text{Bias}^{(1)}(a,R,q,n) = \sum_{\ell=2}^{R} \left[\Big[\frac{\Gamma^{(\ell)}_{n_1}}{\Gamma_{n_1}} - q_1^{-a(\ell-1)} \frac{\Gamma^{(\ell)}_n}{\Gamma_n} \Big] \mathbf{W}_1 + \sum_{r=2}^{R} m_{r,\ell} \mathbf{W}_r \Big[\frac{\Gamma^{(\ell)}_{n_r}}{\Gamma_{n_r}} - q_r^{-a(\ell-1)} \frac{\Gamma^{(\ell)}_n}{\Gamma_n} \Big] \right] c_\ell,$$

where $m_{r,\ell} = (M^{-(\ell-1)} - 1) M^{-(r-2)(\ell-1)}$. *We have:*

$$|\text{Bias}^{(1)}(a,R,q,n)| \leq \frac{C_{a,\mathbf{q},r}}{n^{1-a}},$$

where,

$$C_{a,q,r} = 6\, \frac{2-a(R+1)}{1-a(R+1)} \|\mathbf{W}\|_\infty\, q_*^{-1} \sum_{\ell=2}^{R} (\gamma_1 q_*^{-a})^{\ell-1} \left[1 + \sum_{r=2}^{R} m_{r,\ell} \right] |c_\ell|$$

with $q_* = \min_{1\leq r\leq R} q_r$ *and* $\|\mathbf{W}\|_\infty = \sup_{r\in\{1\ldots,R\},R\geq 2} \mathbf{W}^{(R)}_r$. *Furthermore, if* $q_1 = \ldots = q_R = \frac{1}{R}$, *then* $\text{Bias}^{(1)}(a,R,q,n) = 0$.

Remark B.11. Note that since $a < 1/2$, $n^{1-a} = o(n^{-\frac{1}{2}})$ so that this term is negligible at the first and second orders of the expansions obtained in this paper. Finally, it is worth noting that this term is equal to 0 when the q_i are equal to $\frac{1}{R}$, case where, in addition, the $\mathbf{W}_r$, $r = 1\ldots,R$ have a simple closed form given by (2.22) and (2.23) in Lemma 2.1.

Proof. First, we derive by an obvious comparison argument with integrals $\int_0^n x^{-a}dx$ and $\int_1^{n+1} x^{-a}dx$ that

$$\frac{n^{1-a}-2}{1-a} \leq \sum_{k=1}^{n} k^{-a} \leq \frac{n^{1-a}}{1-a},\ n \geq 1, a \in (0,1). \tag{B.66}$$

Elementary computations then show that, for every $a \in \big(0, \frac{1}{R}\big)$, $\chi \in (0,1)$ and every $n \geq 1$, every integer $\ell \in \{1,\ldots,R+1\}$

$$\left| \frac{\Gamma^{(\ell)}_{\lfloor \chi n \rfloor}}{\Gamma_{\lfloor \chi n \rfloor}} - \chi^{-a(\ell-1)} \frac{\Gamma^{(\ell)}_n}{\Gamma_n} \right| \leq \frac{3\gamma_1^\ell}{\Gamma_{\lfloor \chi n \rfloor}} \left(\frac{1}{1-a\ell} + \frac{\chi^{-a\ell}}{1-a} \right)$$

Using that $u \mapsto u^{1-a}$ is $(1-a)$-Hölder, we derive from the left inequality in (B.66) that $\Gamma_{\lfloor \chi n \rfloor} \geq \gamma_1 \frac{(\chi n)^{1-a}-3}{1-a}$ so that, for every $n \geq \frac{6^{\frac{1}{1-a}}}{\chi}$,

$$\left| \frac{\Gamma^{(\ell)}_{\lfloor \chi n \rfloor}}{\Gamma_{\lfloor \chi n \rfloor}} - \chi^{-a(\ell-1)} \frac{\Gamma^{(\ell)}_n}{\Gamma_n} \right| \leq 3\Big(1 + \frac{1-a}{1-a(R+1)}\Big) \frac{\gamma_1^{\ell-1}}{n^{1-a}} \frac{\chi^{-a\ell}}{\chi^{1-a} - 3n^{a-1}}$$
$$\leq 6\, \frac{2-a(R+1)}{1-a(R+1)} \frac{\gamma_1^{\ell-1}}{n^{1-a}} \chi^{-1-a(\ell-1)}. \tag{B.67}$$

Now, since $\|\mathbf{W}\|_\infty < +\infty$ (see Lemma 2.1(b)), we deduce by plugging the above inequality in $\mathrm{Bias}^{(1)}(a, R, q, n)$ that, for every $n \geq \frac{6^{\frac{1}{1-a}}}{q_*}$

$$|\mathrm{Bias}^{(1)}(a, R, q, n)| \quad \leq \quad 6\,\frac{2-a(R+1)}{1-a(R+1)}\,\frac{1}{n^{1-a}}\,\|\mathbf{W}\|_\infty\, q_*^{-1} \sum_{\ell=2}^{R} (\gamma_1 q_*^{-a})^{\ell-1} \left[1 + \sum_{r=2}^{R} m_{r,\ell}\right] |c_\ell|.$$

When $q_r = \frac{1}{R}$, $r = 1, \ldots, R$,

$$\mathrm{Bias}^{(1)}(a, R, q, n) = \sum_{\ell=2}^{R} \left[\frac{\Gamma_{n_1}^{(\ell)}}{\Gamma_{n_1}} - \bar{q}_1^{-a(\ell-1)}\frac{\Gamma_n^{(\ell)}}{\Gamma_n}\right] \left(\mathbf{W}_1 + \sum_{r=2}^{R} \mathbf{W}_r m_{r,\ell}\right) c_\ell = 0$$

since $\mathbf{W}$ is solution to (2.17). □