The Annals of Applied Probability
2018, Vol. 28, No. 6, 3358–3417
https://doi.org/10.1214/17-AAP1364

WEIGHTED MULTILEVEL LANGEVIN SIMULATION OF INVARIANT MEASURES

BY GILLES PAGÈS AND FABIEN PANLOUP

UPMC and Université d'Angers

We investigate a weighted multilevel Richardson–Romberg extrapolation for the ergodic approximation of invariant distributions of diffusions adapted from the one introduced in [*Bernoulli* **23** (2017) 2643–2692] 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 $n^{1/3}$. 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 this multilevel Langevin estimator on several examples including the simple one-dimensional Ornstein–Uhlenbeck process but also a high dimensional diffusion motivated by a statistical problem. These examples confirm the theoretical efficiency of the method.

CONTENTS

Received July 2016; revised June 2017.
MSC2010 subject classifications. 60J60, 37M25, 65C05.
Key words and phrases. Ergodic diffusion, invariant measure, multilevel, ergodicity, Richardson–Romberg, Monte Carlo, PAC-Bayesian.

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 $\sigma : \mathbb{R}^d \to \mathcal{M}(d, q, \mathbb{R})$ ($d \times q$-matrices with real entries) are locally Lipschitz continuous functions with 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

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

where $(\cdot|\cdot)$ denotes the canonical inner product on $\mathbb{R}^d$, $D^2 g$ denotes the Hessian matrix of g and Tr denotes the trace operator. As soon as there exists a continuously twice differentiable *Lyapunov* function $V : \mathbb{R}^d \to \mathbb{R}_+$ such that

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

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., [17]), this invariant measure ν is unique, hence ergodic. In particular,

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

where $\overset{(\mathbb{R}^d)}{\Longrightarrow}$ denotes weak convergence of distributions on $\mathbb{R}^d$ (see, e.g., [2] or [11] for background). We will assume that this uniqueness holds throughout the paper. Under slight additional assumptions, one shows that the diffusion is *stable* in the sense that

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

This $\mathbb{P}_x$-*a.s.* convergence is ruled by Bhattacharya's CLT (see [1] 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

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

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., [12, 13, 15, 19, 20, 22]), 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, $\mathbb{E}_\nu F((X_t)_{t\in[0,T]})$ where F is a (path-dependent) functional defined on the space $\mathcal{C}([0,T], \mathbb{R}^d)$ (see also [30] or [24] for other references on the topic or more recently [6]).

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

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

where $\Gamma_n = \gamma_1 + \cdots + \gamma_n$, $n \geq 1$. Then we introduce (for technical matter to be explained further on) a general nonnegative *weight* sequence $(\eta_n)_{n\geq 1}$ and the resulting 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

$$\nu_n^{\eta,\gamma}(f) := \int f \, d\nu_n^{\eta,\gamma}$$

can be performed recursively, once noted that

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

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 ask that the pair $(\eta_n, \gamma_n)_{n\geq 1}$ satisfies

$$\begin{aligned} &H_n := \eta_1 + \cdots + \eta_n \to +\infty, \qquad \gamma_n \downarrow 0 \quad \text{and} \\ &\Gamma_n := \gamma_1 + \cdots + \gamma_n \to +\infty, \end{aligned} \tag{1.5}$$

as $n \to +\infty$. When $\eta = \gamma$, the γ-empirical measure $\nu^{\gamma,\gamma}$ is the natural counterpart of μ_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.* $\nu_n^{\eta,\gamma}(\omega, dx) \stackrel{(\mathbb{R}^d)}{\Longrightarrow} \nu$ taking advantage of the fact that the step $\gamma_n \downarrow 0$. The major difference with the above continuous time Birkhoff's pointwise ergodic theorem is that, provided b and σ can be computed easily, these random measures taken against a (computable) function f can in turn be simulated. This suggests to look for ergodic simulation methods, also known as *Langevin Monte Carlo simulation* to compute $\nu(f)$. To be more precise, the term *Langevin Monte Carlo simulation* commonly refers to the case where the diffusion satisfies $b(x) = -\nabla V(x)$ and $\sigma(x) = \sigma > 0$ is constant whereas, most often but not always, the Euler discretization scheme also has a constant step (see, e.g., among others, [27, 28]). It can be seen as a general terminology for *ergodic simulation*. Note that, though we will not go deeper in that direction, when ν has a density h with respect to the Lebesgue measure λ_d on $\mathbb{R}^d$, such an approach appears as a probabilistic numerical scheme for solving the stationary Fokker–Planck equation $\mathcal{L}^* h = 0$ by providing the values of as many integrals $\int f h \lambda_d$ as requested.

Let us first recall one simple convergence result for the *a.s.* weak convergence of the weighted empirical measures $(\nu_n^{\eta,\gamma})_{n\ge1}$ [see Theorem V.2 borrowed and slightly adapted from [12] (see also [14])].

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

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

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

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

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

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

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

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

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

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

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

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

$\triangleright$ The condition (**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 $\mu_t(\omega) = \frac{1}{t}\int_0^t \delta_{X_s(\omega)}\, ds$ of the diffusion itself.

$\triangleright$ 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 (**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)].

$\triangleright$ In the above proposition, the condition

$$\lim_{|x|\to+\infty} \frac{V(x)}{|x|^\rho} = +\infty$$

can be relaxed into

$$\lim_{|x|\to+\infty} V(x) = +\infty.$$

The interest of this slightly strengthened assumption is to ensure that, in the sequel, every function f with regular polynomial growth has a V-polynomial growth as well.

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 $\gamma_n = \gamma_1 n^{-a}$ and $\eta_n = \eta_1 n^{-c}$, 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 involved):

Set $\Gamma_n^{(2)} = \sum_{k=1}^n \gamma_k^2$, $n \ge 1$. 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

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

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

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

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

When $\gamma_n = n^{-a}$, the unbiased *CLT* ($\widetilde{\beta} = 0$) holds for $a \in (\frac{1}{3}, 1]$, 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 μ_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 2000s (2008, see [8]). Ever since, it has been extensively adapted to various types of simulations (nested Monte Carlo, see [16], stochastic approximation [5]) 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 reasonable complexity, usually inversely proportional to h, relying on simulable time discretization schemes of the underlying dynamics. The typical situation is $Y_0 = f(X_T)$ or $F((X_t)_{t\in[0,T]})$ where (X_t) is a Brownian diffusion as above and $Y_h = f(\bar{X}_T^n)$ or $F((\bar{X}_t^n)_{t\in[0,T]})$ where $(\bar{X}_t)_{t\in[0,T]}$ is a discretization scheme, say an Euler or a Milstein scheme with step $h = \frac{T}{n} \in \mathbb{H} = \{\frac{T}{m}, m \in \mathbb{N}^*\}$. A multilevel estimator with depth $L \in \mathbb{N}^*$ of $\mathbb{E}Y_0$ is designed by implementing a nonhomogeneous multilevel Monte Carlo (MLMC) estimator of size $N \in \mathbb{N}^*$ of the form

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

where $\mathbf{h} \in \mathbb{H}$ is a fixed *coarse* step, $((Y_h^{(\ell),k})_{h\in\mathbb{H}})_{\ell=1,\dots,L,k\geq 0}$ 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 $Y_{\frac{\mathbf{h}}{M^{\ell-1}}}^{(\ell)}$ and $Y_{\frac{\mathbf{h}}{M^{\ell-2}}}^{(\ell)}$ 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 way to quantify this consistency is that Y_h converges in (squared) quadratic norm to Y_0 at an h^β rate, namely $\|Y_h - Y_0\|_2^2 \le V_1|h|^\beta$, $h \in \mathbb{H}$. The parameter β depends on f or F and on the selected discretization scheme in a diffusion framework. Thus, with an Euler scheme, 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 [8, 16]).

Among other results, it is proved in [8] that, if $\alpha = 1$ and $\beta = 1$—which is the standard situation in a diffusion discretized by an Euler scheme with step $h = \frac{T}{n}$ when $Y_0 = f(X_T)$, $Y_h = f(\bar{X}^n_T)$, 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 $O((\log(1/\varepsilon)/\varepsilon)^2)$ as $\varepsilon \to 0$. When $\beta > 1$ (fast strong approximation like with the Milstein scheme or its antithetic Giles–Szpruch simulable variant [9] in higher dimension), this rate attains $O(\varepsilon^{-2})$, that is, 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 [16]), 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

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

where the R-tuple $(\mathbf{W}_r^{(R)})_{1\le r\le R}$ of weights has a closed form *entirely determined by α, M and R* 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 $O(\log(1/\varepsilon)/\varepsilon^2)$ 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 (when $M = 2$) by a factor $\exp(\frac{1-\beta}{\alpha}\sqrt{\log(2)\log(1/\varepsilon)/2}) = o(\varepsilon^{-\eta})$, $\forall \eta > 0$, instead of $\varepsilon^{\frac{\beta-1}{\alpha}}$ with MLMC.

The underlying idea for this weighted multilevel method is to combine the multilevel paradigm with the multistep Richardson–Romberg extrapolation introduced in [18] in its modern form but which historical form goes back to [25, 26] and has been extensively used in the Monte Carlo literature.

We refer to [16] 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 goes 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 (2.3) 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 $\nu_n(f)$ and $\mu_n^{(r,M)}(f)$ such as (2.4) and (2.5) 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 weak error/bias expansion $\mathbb{E}[f(X_T)] - \mathbb{E}[f(\bar{X}_T)]$ in finite horizon.

As for the strong convergence rate used to control the variance of the corrective terms of multilevel estimators, its counterpart in our ergodic setting will require a contraction-type assumption [see $(\mathbf{C_w})$ and $(\mathbf{C_s})$ below], which guarantees a *mean confluence result* between the diffusion and its Euler scheme with decreasing step of the following form:

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

It says that the (γ, γ)-empirical measure of the couple $(X, \bar{X})$ concentrates on the diagonal of $\mathbb{R}^d$. Furthermore, under Assumption $(\mathbf{C_s})$, the rate of concentration on the diagonal can be quantified leading to a long-time setting roughly corresponding to the case $\beta = 2$ (even with the Euler–Maruyama scheme). Such properties hold in particular when the diffusion itself is *exponentially confluent* (typically like a mean-reverting Ornstein–Uhlenbeck process) as it is the case under Assumption $(\mathbf{C_s})$.

The methods of proof heavily rely on limit theorems for martingale borrowed from stability theory for nonhomogeneous discrete time Markov chains (see, e.g., [4]) and radically differ from those used to analyze multilevel methods in finite horizon. For convenience, we will work in one dimension, but the extension to the multidimensional setting would be essentially a matter of notation.

Outline of the paper and main results. 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 $(\gamma_n)_{n\geq 1}$ 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 to establish 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. In Section 6, we carry out several numerical experiments, first on an Orsnstein–Uhlenbeck toy model, then on more involved diffusion related to a double-well potential and finally on a statistical example (sparse regression learning inspired by [3]). Finally, in Section 7, we briefly compare with a more classical multilevel method based on finite horizon simulations combined with a convergence of $(X_t)_{t\geq 0}$ toward its invariant distribution (at an exponential rate), close in spirit to that recently introduced in [29] [in which a complexity at the order of $\varepsilon^{-2}(\log(\frac{1}{\varepsilon}))^3$ is attained].

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 $(\bar{X}^{(r)},\bar{Y}^{(r,M)})$, $r=2,\dots,R$, 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

$$\forall m \in \{1,\dots,M\} \qquad \widetilde{\gamma}^{(r,M)}_{M(n-1)+m} = \frac{\gamma_n^{(r,M)}}{M} = \frac{\gamma_n}{M^{r-2}}, \qquad n \geq 1. \tag{2.1}$$

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

$$\Gamma_n^{(\ell,r)} = \sum_{k=1}^{n} \big(\gamma_k^{(r,M)}\big)^{\ell} = M^{-(r-2)\ell} \sum_{k=1}^{n} \gamma_k^{\ell} = M^{-(r-2)\ell}\Gamma_n^{(\ell)}, \tag{2.2}$$

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 $r = 2,\dots,R$ the sequence of difference of the empirical measures of the two schemes by

$$\begin{aligned} &\mu_n^{(r,M)}(dx) \\ &\quad = \frac{1}{\Gamma_n^{(1,r)}} \sum_{k=1}^{n} \left(\left(\sum_{m=0}^{M-1} \widetilde{\gamma}^{(r)}_{M(k-1)+m} \delta_{\bar{Y}^{(r)}_{M(k-1)+m}} \right) - \gamma_k^{(r)} \delta_{\bar{X}^{(r)}_{k-1}} \right), \qquad n \geq 1 \\ &\quad = \frac{1}{\Gamma_n^{(1,r)}} \sum_{k=1}^{n} \frac{\gamma_n}{M^{r-2}} \left(\frac{1}{M} \sum_{m=0}^{M-1} \delta_{\bar{Y}^{(r)}_{M(k-1)+m}} - \delta_{\bar{X}^{(r)}_{k-1}} \right), \qquad n \geq 1. \end{aligned} \tag{2.3}$$

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

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

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

$$n_r = \lfloor q_r n \rfloor, \qquad r = 1,\dots,R.$$

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 3.1(b) and 3.2(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:

$$\begin{aligned} \nu_{n_1}(f) &= \nu(f) + \sum_{\ell=2}^{R+1} \frac{\Gamma_{n_1}^{(\ell)}}{\Gamma_{n_1}} \nu(\Psi_\ell) + \frac{\mathbf{M}_n}{\Gamma_n} \\ &\quad + o\left(\frac{1}{\sqrt{\Gamma_{n_1}}} \wedge \frac{\Gamma_{n_1}^{(R+1)}}{\Gamma_{n_1}} \right), \end{aligned} \tag{2.4}$$

$$\mu_{n_r}^{(r,M)}(f) = \sum_{\ell=2}^{R+1} M^{(r-2)(1-\ell)}\big(M^{1-\ell}-1\big)\frac{\Gamma_{n_r}^{(\ell)}}{\Gamma_{n_r}}\nu(\Psi_\ell) + o\left(\frac{1}{\sqrt{\Gamma_{n_r}}}\wedge\frac{\Gamma_{n_r}^{(R+1)}}{\Gamma_{n_r}}\right), \tag{2.5}$$

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 $\mu_{n_r}^{(r,M)}(f)$, $r=1,\dots,R$, we will be able to kill the bias error without growing the asymptotic variance. But a numerical computation holds in a finite (nonasymptotic) 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.

REMARK. Variants in the definitions of the schemes are possible. Thus, an alternative to the independence of the levels is to make the two schemes of each level (the coarse and the fine one start from the terminal value of the fine scheme of the former level and, for the second level, from the terminal value of the coarse scheme). Doing so, one only needs the increments of a single underlying Brownian motion in the spirit of the original Langevin Monte Carlo simulation. However, it turns out that the theoretical computation the asymptotic variance of the resulting multilevel estimator—on which is based the optimization of its design—seems less tractable.

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

$$\widetilde{\nu}_n^{(R,\mathbf{W})} = \mathbf{W}_1\nu_{n_1} + \sum_{r=2}^{R}\mathbf{W}_r\mu_{n_r}^{(r,M)}, \qquad n\geq 1, \tag{2.6}$$

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. Also, let us remark that the weights $\mathbf{W}_r$ clearly depend on R and will sometimes be denoted $\mathbf{W}_r^{(R)}$ in order to recall this dependence when necessary. Let us now specify $\mathbf{W}$. First, by (2.4) and (2.5), one remarks that it is necessary to assume that $\mathbf{W}_1 = 1$ in order to ensure the convergence toward ν.

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

$$\gamma_k = \gamma_1 k^{-a} \qquad \text{with } \gamma_1 > 0, a \in (0,1). \tag{2.7}$$

Then by plugging the expansions of the bias resulting from (2.4) and (2.5) in the definition (2.6) of the **ML2Rgodic** estimator we derive that

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

where the notation $\approx$ is used to keep in mind that one implicitly assumes that $\frac{\Gamma_{n_r}^{(\ell)}}{\Gamma_{n_r}} - q_r^{-a(\ell-1)} \frac{\Gamma_n^{(\ell)}}{\Gamma_n}$ 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: $\mathbf{W}_1 = 1$ and

$$\forall \ell \in [\![2, R]\!]$$

$$\mathbf{W}_1 q_1^{-a(\ell-1)} + \big(M^{1-\ell} - 1\big) \sum_{r=2}^{R} \mathbf{W}_r M^{-(r-2)(\ell-1)} q_r^{-a(\ell-1)} = 0, \tag{2.8}$$

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

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

where we set, more generally,

$$\begin{aligned}
\widetilde{\mathbf{W}}_{R+i} &= \mathbf{W}_1 q_1^{-a(R+i)} \\
&\quad + \big(M^{-R-i+1} - 1\big) \sum_{r=2}^{R} \mathbf{W}_r M^{-(r-2)(R+i-1)} q_r^{-a(R+i)}, \qquad i \ge 0.
\end{aligned} \tag{2.9}$$

The main difference at this stage with the 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 (2.8) 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 $q_r = \frac{1}{R}$, $r = 1, \dots, R$ where we retrieve exactly the weights of the regular Monte Carlo *ML2R* introduced in [16]. For a given depth $R > 1$, the closed form of $(\mathbf{W}_i)_{i=2}^R$ (keeping in mind that $\mathbf{W}_1 = 1$) is given by the following lemma.

LEMMA 2.1. (a) General re-sizers: *If* $\mathbf{q} := (q_1, \dots, q_R) \in \mathcal{S}_R := \{(x_1, \dots, x_R) \in (0, +\infty)^R, \sum_{i=1}^R x_i = 1\}$ *and satisfies* $\frac{q_r}{M^{r/a}} \neq \frac{q_s}{M^{s/a}}$, $s \neq r$, *then the above system* (2.8) *has a unique solution given by*

$$\mathbf{W}_r^{(R)} = M^{r-2}\left(\frac{q_r}{q_1}\right)^a \sum_{k\geq 0} \frac{1}{M^k} \prod_{s=2, s\neq r}^{R} \frac{1 - M^{s-2-k}(q_s/q_1)^a}{1 - M^{s-r}(q_s/q_r)^a}, \quad r = 2, \dots, R. \tag{2.10}$$

Moreover, the coefficients $\widetilde{\mathbf{W}}_{R+i}^{(R)}$, $i = 1, 2$, *as defined in* (2.9) *read*

$$\widetilde{\mathbf{W}}_{R+1}^{(R)} = \frac{(1 - M^{-R})}{q_1^{aR}} \sum_{k\geq 0} \frac{1}{M^{kR}} \prod_{r=0}^{R-2} \left(1 - M^{k-r}\left(\frac{q_1}{q_{r+2}}\right)^a\right) \tag{2.11}$$

and

$$\widetilde{\mathbf{W}}_{R+2}^{(R)} = \frac{1 - M^{-R-1}}{q_1^{a(R+1)}} \sum_{k\geq 0} \left(\frac{1 + \sum_{r=0}^{R-2} M^{k-r}(\frac{q_1}{q_{r+2}})^a}{M^{k(R+1)}}\right) \times \left(1 + \prod_{r=0}^{R-2}\left(1 - M^{k-r}\left(\frac{q_1}{q_{r+2}}\right)^a\right)\right). \tag{2.12}$$

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

$$\mathbf{W}_r^{(R)} = \mathbf{w}_r^{(R)} + \cdots + \mathbf{w}_R^{(R)}, \qquad r = 1, \dots, R \tag{2.13}$$

with

$$
\mathbf{w}_r^{(R)} = \prod_{s=1, s\neq r}^{R} \frac{M^{-(s-1)}}{M^{-(s-1)} - M^{-(r-1)}} = \prod_{s=1, s\neq r}^{R} \frac{1}{1 - M^{s-r}}, \qquad r = 1, \dots, R. \tag{2.14}
$$

The weights $\mathbf{W}_r^R$ *are uniformly bounded, that is,* $\sup_{r\in[\![1,R]\!], R\geq 1} |\mathbf{W}_r^{(R)}| < +\infty$. *Furthermore,*

$$
(\widetilde{\mathbf{W}}_{R+1}^{(R)}, \widetilde{\mathbf{W}}_{R+2}^{(R)}) = (-1)^{R-1} R^{aR} M^{-\frac{R(R-1)}{2}} \left(1, -R^a \frac{1 - M^R}{1 - M^{-1}}\right). \tag{2.15}
$$

The proof is postponed to [21], Section 1.

EXAMPLES. • $R = 2$: $\mathbf{W}_1^{(2)} = 1$, $\mathbf{W}_2^{(2)} = \frac{M}{M-1}(\frac{q_2}{q_1})^a$.
• $R = 3$:

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

When there is no ambiguity, the superscript $^{(R)}$ will be dropped in the notation $\mathbf{W}^{(R)}$, $\mathbf{w}_r^{(R)}$ and $\mathbf{W}_{R+1}^{(R)}$. In the sequel, $\widetilde{\nu}_n^{(R,\mathbf{W})}$ will be always defined with $\mathbf{W}$ satisfying (2.8) or (2.10).

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:

(**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 $\lambda_0 > 0$], when Assumption (**S**) is in force and f, b and σ are smooth have polynomial growth as well as their derivatives, then (**P**) holds true. Actually, we first recall that under the ellipticity and Lyapunov assumptions, the semi-group converges exponentially fast toward 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., [23]). Then, by [7], 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 [23], Theorem 1. The property is obtained through the a priori estimate, see equation (9.40) in [7], 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 $u = \partial_{i_1,\dots,i_{k-1}} g$ 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 [7] for the second-order partial derivatives of u yields the polynomial growth of the partial derivatives $\partial_{i_1,\dots,i_{k+1}} g$.

The second additional assumption has been introduced in [20] and deeply studied [17]: it requires the diffusion to be *weakly confluent*, that is, 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

$$
\begin{cases} dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t, \\ dY_t = b(Y_t)\,dt + \sigma(Y_t)\,dW_t, \end{cases} \tag{2.16}
$$

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 diffusion (*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 is referred to as

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

REMARK 2.1. ▷ Under slight additional assumptions on the stability of (*SDE*), it can be shown (see [17]) that, if ($\mathbf{C_w}$) holds, the diffusion is statistically confluent in the sense that

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

▷ For the empirical measure $\widetilde{\nu}_n^{(R,\mathbf{W})}$, the role 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* (**S**), (**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* (2.10). *Let* $q =$

$(q_r)_{1\le r\le R} \in \mathcal{S}_R$ *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* $\mathbf{W} = (\mathbf{W}_r)_{r=1,\dots,R}$ *be defined by* (2.10).

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

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

with

$$\sigma_f^2(a,q,R) = \frac{1-a}{\gamma_1}\frac{\sigma_1^2(f)}{q_1^{1-a}} \qquad \text{with } \sigma_1^2(f) = \nu\big(|\sigma^*\nabla g|^2\big). \tag{2.17}$$

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

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

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* (2.9) *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* (3.4)].

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

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

with

$$m_f(a,q,R) := \frac{1-a}{1-a(R+1)}\gamma_1^R \widetilde{\mathbf{W}}_{R+1} c_{R+1}. \tag{2.18}$$

REMARK 2.2. 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 $a = \frac{1}{2R+1}$.

REMARK 2.3. It is worth noting that in this long run setting where we manage jointly Monte Carlo, long-time and discretization errors, the interest of **ML2Rgodic** and especially of the multilevel Richardson–Romberg strategy is to increase the order of the rate of convergence of the procedure. In particular, such a property would not be satisfied with a MLMC alternative.

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 nonasymptotic 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 reformulation 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 $x, y \in \mathbb{R}^d$,

$$\big(b(x) - b(y)|x - y\big)_S + \frac{1}{2}\|\sigma(x) - \sigma(y)\|_S^2 \leq -\alpha\|x - y\|_S^2,$$

where $(\cdot|\cdot)_S$ and by $|\cdot|_S$ 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 2.1 of [21], 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*

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

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

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

where, *on the one hand*

$$\widetilde{\sigma}_f^2(q, R) = \frac{1}{q_1}\sigma_{2,1}^2(f) + \left(1 - \frac{1}{M}\right)\Psi(R, M)\sigma_{2,2}^2(f) \tag{2.19}$$

with

$$\Psi(R, M) = \frac{4R^2}{4R^2 - 1} \sum_{r=2}^{R} (\mathbf{W}_r^{(R)})^2 \tag{2.20}$$

and $\sigma_{2,1}^2(f)$ *and* $\sigma_{2,2}^2(f)$ *are some variance terms explicitly defined further on by* (4.1) *and* (4.4) *in Propositions* 4.1 *and* 4.2, *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* (2.14) *and* $\widetilde{\mathbf{W}}_{R+1}^{(R)}$ *and* $\widetilde{\mathbf{W}}_{R+2}^{(R)}$ *by* (2.15) *so that*

$$\widetilde{m}_f(\bar{q}, R) = -\frac{4R}{R-1} c_{R+1} c_{R+2} \gamma_1^{2R+1} R M^{-R(R-1)} \frac{1 - M^{-R}}{1 - M^{-1}}. \tag{2.21}$$

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 thus start from Theorem 2.2(b) with

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

Then the weights $\mathbf{W}_r$, $r = 1, \dots, R$ and $\widetilde{\mathbf{W}}_{R+1}$ are given by (2.14) and (2.15) (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(ϖ) reads

$$\begin{aligned} \|\nu_n^{R,\mathbf{W}} - \nu(f)\|_2^2 &= \frac{1}{n^{\frac{2R}{2R+1}}} \big(\sigma_f^2(\bar{a}, \bar{q}, R) + m_f^2(\bar{a}, \bar{q}, R)\big) \\ &\quad + \frac{1}{n} \big(\widetilde{\sigma}_f^2(\bar{a}, \bar{q}, R) + \widetilde{m}_f(q, R) + o(1)\big) \end{aligned} \tag{2.22}$$

as n goes to ∞ where, owing to (2.18), (2.21), (2.17) and (2.19),

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

On the other hand, the complexity $K(\varpi, n, M)$ of the multilevel Langevin estimator devised in (2.6) reads

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

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:

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

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 $(\mathbf{W}_r^{(R)})_{1\le r\le R, R\ge 2}$ is uniformly bounded. Consequently, M being fixed, $\Psi(R, M) = O(R)$ as $R \to +\infty$ [where Ψ is defined by (2.20)]. This leads us to define

$$\boldsymbol{\Psi}(M) = \sup_{R\ge 1} \frac{\Psi(R, M)}{R}. \tag{2.23}$$

We refer to Table 2 for some numerical values of Ψ and $\boldsymbol{\Psi}$.

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

(a)

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

where

$$K(f, M) = \frac{2\kappa_0(M+1)}{\log M}\left(\frac{(M-1)\boldsymbol{\Psi}(M)}{\widetilde{c}\theta_1(f)} + 1\right)\widetilde{c}\sigma_1^2(f) \tag{2.24}$$

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*

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

TABLE 1
Values of $x(\varepsilon, M)$

	$\varepsilon = 10^{-1}$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-3}$	$\varepsilon = 10^{-4}$
$M = 2$	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

Furthermore, as $\varepsilon \to 0$,

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

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

$$n(\varepsilon, M) \precsim \frac{2}{\log M}\left(\frac{(M-1)\boldsymbol{\Psi}(M)}{\widetilde{c}\theta_1(f)} + 1\right)\sigma_1^2(f)\varepsilon^{-2}\log\left(\frac{1}{\varepsilon}\right) \qquad \text{as } \varepsilon \to 0. \tag{2.25}$$

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.

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 Poisson 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 $(\nu_n^\gamma(f))_{n\ge1}$ (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 (2.3) 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 $n\ge1$, for every sequence $(v_n)_{n\ge1}$, we set $\Delta v_n=v_n-v_{n-1}$. We will also use the following notation:

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

LEMMA 3.1. *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 $n\ge1$,*

$$\begin{aligned}\Delta g(\bar X_n)&=-\gamma_n\big(f(\bar X_{n-1})-\nu(f)\big)+\left[\sum_{\ell=2}^{L+1}\gamma_n^\ell\varphi_\ell(f)(\bar X_{n-1})\right]\\&\quad+\sum_{i=1}^{3}\Delta M_n^{(i,g)}+\Delta R_{n,L}^{(1,g)}+\Delta R_{n,L}^{(2,g)}+\Delta R_{n,L}^{(3,g)},\end{aligned}\tag{3.1}$$

where, setting $\mathcal{I}_s=\{(m_1,m_2)\in\mathbb{N}^2,m_1+\frac{m_2}{2}=s\}$,

$$\begin{aligned}\varphi_\ell(f)(x)&=\sum_{\mathcal{I}_\ell}g^{(m_1+m_2)}(x)\frac{\rho_{m_2}}{m_1!m_2!}b^{m_1}(x)\sigma^{m_2}(x),\\
\Delta M_n^{(1,g)}&=\sqrt{\gamma_n}\big(g'\sigma\big)(\bar X_{n-1})U_n,\\
\Delta M_n^{(2,g)}&=\frac12\gamma_n g''(\bar X_{n-1})\sigma^2(\bar X_{n-1})\big[U_n^2-1\big],\\
\Delta M_n^{(3,g)}&=\gamma_n^{\frac32}\bigg(\frac12 g''(\bar X_{n-1})b(\bar X_{n-1})\sigma(\bar X_{n-1})U_n\\
&\quad+\frac16 g^{(3)}(\bar X_{n-1})\sigma^3(\bar X_{n-1})U_n^3\bigg),\\
\Delta R_{n,L}^{(1,g)}&=\sum_{\ell=2}^{2L+1}\gamma_n^{\ell+\frac12}\sum_{\mathcal{I}_{\ell+\frac12}}g^{(m_1+m_2)}(\bar X_{n-1})\frac{1}{m_1!m_2!}b^{m_1}(\bar X_{n-1})\sigma^{m_2}(\bar X_{n-1})U_n^{m_2}\\
&\quad+\sum_{\ell=2}^{2L+1}\gamma_n^\ell\sum_{\mathcal{I}_\ell}g^{(m_1+m_2)}(\bar X_{n-1})\end{aligned}$$

$$\times \frac{1}{m_1!m_2!} b^{m_1}(\bar{X}_{n-1})\sigma^{m_2}(\bar{X}_{n-1})\big[U_n^{m_2} - \rho_{m_2}\big],$$

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

$$\Delta R_{n,L}^{(3,g)} = g^{(2L+3)}(\xi_n)\big(\gamma_n b(\bar{X}_{n-1}) + \sqrt{\gamma_n}\sigma(\bar{X}_{n-1})U_n\big)^{2L+3},$$

$$\xi_n \in [\bar{X}_{n-1}, \bar{X}_n].$$

As a consequence,

$$\begin{aligned} &\nu_n^{\eta,\gamma}(f) - \nu(f) \\ &\quad = -\frac{1}{H_n}\sum_{k=1}^n \frac{\eta_k}{\gamma_k}\Delta g(\bar{X}_k) + \sum_{\ell=2}^{L+1} \frac{\sum_{k=1}^n \eta_k\gamma_k^{\ell-1}}{H_n} \nu_n^{\eta\gamma^{\ell-1},\gamma}\big(\varphi_\ell(f)\big) \\ &\qquad + \frac{1}{H_n}\sum_{k=1}^n \frac{\eta_k}{\gamma_k}\left(\sum_{i=1}^3 \Delta M_k^{(i,g)} + \Delta R_{k,L}^{(1,g)} + \Delta R_{k,L}^{(2,g)} + \Delta R_{k,L}^{(3,g)}\right). \end{aligned} \tag{3.2}$$

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

$$\begin{aligned} g'(x)y + \frac{1}{2}g''(x)y^2 &= -\gamma\mathcal{L}g(x) + \sqrt{\gamma}\sigma(x)u + \frac{1}{2}\gamma\sigma^2(x)\big(u^2-1\big) \\ &\quad + \frac{1}{2}g''(x)\big(\gamma^2 b^2(x) + 2\gamma^{\frac{3}{2}}\sigma(x)u\big). \end{aligned}$$

Since

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

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 $f - \nu(f) = -\mathcal{L}(\mathcal{Q}f)$ satisfying $\nu(\mathcal{Q}f) = 0$. [Under Assumption (**P**), $\mathcal{Q}f$ is well defined.]

DEFINITION 3.1. (a) Under Assumption (**P**), one may define a mapping $\varphi_\ell^{[1]}(\cdot)$ from $\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ into itself defined for every $f\in\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ by

$$\varphi_\ell^{[1]}(f)(\cdot)=\sum_{(m_1,m_2),m_1+\frac{m_2}{2}=\ell}\frac{\rho_{m_2}}{m_1!m_2!}b^{m_1}(\cdot)\sigma^{m_2}(\cdot)(\mathcal{Q}f)^{(m_1+m_2)}(\cdot),\tag{3.3}$$

where $h^{(k)}$ denotes the kth 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 notation, 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, $\ell\in\mathbb{N}^*$,

$$\Psi_\ell=\sum_{k=1}^{\ell-1}\sum_{\substack{(m_1,\dots,m_k)\in[\![2,\ell]\!]^k,\\ m_1+\cdots+m_k=\ell+k-1}}\varphi_{m_1}\circ\cdots\circ\varphi_{m_k}.\tag{3.4}$$

For example, note that

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

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

PROPOSITION 3.1 (Bias error expansion for the coarse level). *Assume* (**S**), (**P**) (*and uniqueness of the invariant distribution* ν). *Let* $R\in\mathbb{N}$, $R\ge 2$ *and let* $f\in\mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ *with polynomial growth and* $g=\mathcal{Q}f$.

(a) *If* $(\gamma_n^\ell,\gamma_n)_{n\ge1}$ *is averaging for every* $\ell\in\{1,\dots,R\}$,

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

(b) *If, furthermore, the pair* $(\gamma_n^{R+1},\gamma_n)_{n\ge1}$ *is averaging*

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

(c) *The following sharper expansion also holds when* $(\gamma_n^{R+2},\gamma_n)_{n\ge1}$ *is averaging*

$$\nu_n^\gamma(\omega,f)-\nu(f)-\sum_{\ell=2}^{R}\frac{\Gamma_n^{(\ell)}}{\Gamma_n}\nu\big(\Psi_\ell(f)\big)$$

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

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

REMARK 3.1. The first expansion is adapted to the proof of Theorem 2.1(a), the second one to 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 4.1 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 (3.1) in Lemma 3.1. When $(\gamma_n)_{n\geq 1} = \eta = (\gamma_n)_{n\geq 1}$, $L = R$ and $g = \mathcal{Q}f$, we get

$$\begin{aligned}
&\nu_n^{\gamma}(f) - \nu(f) - \sum_{\ell=1}^{R}\frac{\Gamma_n^{(\ell)}}{\Gamma_n}\nu(\varphi_\ell(f)) \\
&\quad = \frac{g(\bar{X}_0) - g(\bar{X}_n)}{\Gamma_n} + \sum_{\ell=2}^{R}\frac{\Gamma_n^{(\ell)}}{\Gamma_n}\left(\nu_n^{\gamma^\ell,\gamma}(\varphi_\ell(f)) - \nu(\varphi_\ell(f))\right) \\
&\qquad + \frac{\Gamma_n^{(R+1)}}{\Gamma_n}\nu_n^{\gamma^{R+1},\gamma}(\varphi_{R+1}(f)) + \frac{M_n^{1,g}}{\Gamma_n} \\
&\qquad + \frac{1}{\Gamma_n}\sum_{k=1}^{n}\left(\sum_{i=2}^{3}\Delta M_k^{(i,g)} + \Delta R_{k,R}^{(1,g)} + \Delta R_{k,R}^{(2,g)} + \Delta R_{k,R}^{(3,g)}\right).
\end{aligned} \tag{3.5}$$

By Lemma 3.2(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 3.2(ii) applied for different choices of (θ_n), h and $(Z_n)_{n\geq 1}$, we have

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

Finally, Lemma 3.2(iii) and (iv) are adapted to manage $\Delta R_{k,R}^{(2,g)}$ and $\Delta R_{k,R}^{(3,g)}$, respectively. This yields

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

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 (3.7) that there exists $C > 0$ such that

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

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,

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

But using again (3.7), one checks that $(\|\nu_n^{\gamma^{R+1},\gamma}(\varphi_{R+1}(f))\|_{2+\delta})_n$ is a bounded sequence for a positive δ. Thus, an uniform integrability argument yields that

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

But for any ℓ, φ_ℓ is the component corresponding to $k = 1$ in the definition (3.4) of Ψ_ℓ. In (b), $\nu(\varphi_{R+1}(f))$ will thus contribute to $\nu(\Psi_{R+1})$. As well, the terms $\nu(\varphi_\ell)$, $\ell = 2, \dots, R$ exhibited in this first expansion will certainly contribute to $\nu(\Psi_\ell)$, $\ell = 2, \dots, R$.

Now, we focus on the second bias term of (3.5). More precisely, for each $\ell \in \{2, \dots, R\}$, we have to repeat the previous procedure: we apply the expansion (3.1) of Lemma 3.1 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

$$
\begin{aligned}
&\frac{\Gamma_n^{(\ell)}}{\Gamma_n}\big(\nu_n^{\gamma^\ell,\gamma}(f_\ell)-\nu(f_\ell)\big)-\sum_{m=2}^{R-\ell+1}\frac{\Gamma_n^{(\ell+m-1)}}{\Gamma_n}\nu\big(\varphi_m^{[1]}\circ\varphi_\ell^{[1]}(f)\big)\\
&\qquad=-\frac{1}{\Gamma_n}\sum_{k=1}^{n}\gamma_k^{\ell-1}\Delta\mathcal{Q}\varphi_\ell(\bar{X}_k)\\
&\qquad\quad+\sum_{m=2}^{R-\ell+1}\frac{\Gamma_n^{(\ell+m-1)}}{\Gamma_n}\big(\nu_n^{\gamma^{\ell+m-1},\gamma}-\nu\big)\big(\varphi_m\circ\varphi_\ell(f)\big)\\
&\qquad\quad+\frac{\Gamma_n^{(R+1)}}{\Gamma_n}\nu_n^{\gamma^{R+1},\gamma}\big(\varphi_{R-\ell+2}\circ\varphi_\ell(f)\big)\\
&\qquad\quad+\frac{1}{\Gamma_n}\sum_{k=1}^{n}\gamma_k^{\ell-1}\Bigg(\sum_{i=1}^{3}\Delta M_k^{(i,g_\ell)}+\Delta R_{k,R-\ell+1}^{(i,g_\ell)}\Bigg).
\end{aligned}
\tag{3.6}
$$

Applying again Lemma 3.2 allows us to control the L^2-norm of the negligible terms:

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

and

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

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:

$$
\frac{\Gamma_n}{\Gamma_n^{(R+1)}}\bigg(\frac{\Gamma_n^{(R+1)}}{\Gamma_n}\nu_n^{\gamma,\gamma^{L+1}}\big(\varphi_{R-\ell+2}^{[1]}\circ\varphi_\ell^{[1]}(f)\big)\bigg)\xrightarrow{n\to+\infty}\nu\big(\varphi_{L-\ell+2}^{[1]}\circ\varphi_\ell^{[1]}(f)\big),
$$

a.s. and in L^2. This brings a second "contribution" to $\nu(\Psi_{R+1})$.

Finally, it remains to consider for every $\ell\in\{2,\dots,R\}$ each term of (3.6). Setting $\ell=m_1$, $m=m_2$, the sequel of the proof consists in repeating the procedure until $k:=\inf\{i:m_1+\cdots+m_i=R+i\}$. 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. $\square$

LEMMA 3.2. *Assume* (**S**). *Let h be a smooth function with polynomial growth. We know from Proposition 1.1 that, for every $p \in (0, +\infty)$,*

$$C_{h,p} = \sup_{n\geq 1} \|h(X_n)\|_p < +\infty. \tag{3.7}$$

Then:

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

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

(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 $(\theta_k)_{k\geq 0}$*,

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

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

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

(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*

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

PROOF. Using that $(\eta_n/\gamma_n)_{n\geq 1}$ is a nonincreasing 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 $x, y \in \mathbb{R}^d$,

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

Using that b and σ are sublinear functions and Minkowski's inequality,

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

The last statement follows using again Minkowski'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

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

By the definition (2.3), one first notes that for $r = 2,\dots,R$, $\mu_n^{(r,M)} = \mu_n^{M,\gamma^{(r)}}$ 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 3.1 but for the sequence $(\mu_n^{M,\gamma}(f))_{n\geq 1}$.

PROPOSITION 3.2 (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* $\ell \in \{1,\dots,R\}$, *the pair* $(\gamma_n^\ell,\gamma_n)_{n\geq 1}$ *is averaging. Then,*

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

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

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

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

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

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

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

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

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

PROOF. With the notation introduced in (2.1), 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 $n \geq 1$,

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

For (a) and (b), it remains now to apply Proposition 3.1(a) and (b) to both terms on 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 3.1(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 3.2(ii), one easily checks 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 3.2. The fact that we keep less martingale terms in expansion (c) can be understood as follows: in Section (4.2), 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 3.1) will be also negligible at a second order.

4. Rate of convergence for the dominating martingales. In the continuity of Propositions 3.1 and 3.2, 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 3.1. 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 4.1. *Assume* (**S**) *and* (**P**). *Let* $g = \mathcal{Q}f$. *Then*

(a)

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

(b)

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

(c) *If* (γ_n, γ_n^2) *is averaging*,

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

where

$$\sigma_{2,1}^2(f) = \int_{\mathbb{R}} \Big[\varphi_2\big((\sigma g')^2\big) + \frac{1}{2}(\sigma^2 g'')^2 + (\sigma g')\big(g^{(3)}\sigma^3 + 2(\sigma g_2')\big)\Big] \, d\nu, \tag{4.1}$$

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

REMARK 4.1. 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,

$$\frac{\langle M^{(1,g)}\rangle_n}{\Gamma_n} = \nu_n^\gamma\big((\sigma g')^2\big) \xrightarrow{n\to+\infty} \nu\big((\sigma g')^2\big) \qquad \text{a.s.} \tag{4.2}$$

Furthermore, by the Cauchy–Schwarz inequality and (3.7), we have for every $\varepsilon > 0$,

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

This second convergence implies that the so-called *Lindeberg condition* is fulfilled. Then (a) is a consequence of the CLT for martingale arrays (see [10], Corollary 3.1).

(b) By the 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 from the fact that $\sigma g'$ has (at most) polynomial growth that

$$\sup_n \mathbb{E}\big[\big(\nu_n^{\gamma}\big((\sigma g')^2\big)\big)^2\big] \le \sup_n \mathbb{E}\big[1 + |\bar{X}_n|^r\big] < +\infty \tag{4.3}$$

owing to (**S**) and (3.7). As a consequence, $(\nu_n^{\gamma}((\sigma g')^2))_{n\ge 1}$ is a uniformly integrable sequence so that the convergence of $(\nu_n^{\gamma}((\sigma g')^2))$ toward $\nu((\sigma g')^2)$ also holds in L^1. The second statement then follows from (4.2).

(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

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

where

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

On the one hand, since $(\gamma_n^2, \gamma_n)_{n\ge 1}$ is averaging, we deduce from Proposition 1.1 that

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

But using uniform integrability arguments similar to (4.3), 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 3.1(a) (and the fact that $\Psi_2 = \varphi_2$) with $R = 2$, we have

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

By (4.2), we deduce that

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

The last statement follows. □

4.2. *The dominating martingale in the error expansion of* $(\mu_n^{M,\gamma}(f))_{n\ge 1}$. 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 $(\mathcal{M}_n)_{n\ge 1}$ defined in Proposition 3.2 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 4.2. *Assume* (**S**) *and* (**P**). *Let* h_1 *and* h_2 *be locally Lipschitz functions with polynomial growth*:

(a) *If* $(\mathbf{C_w})$ *holds, then* $(\frac{\mathcal{M}_n(h_1)}{\sqrt{\Gamma_n}})_{n\geq 1}$ *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*

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

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*

$$\begin{aligned}\sigma_{2,2}^2(f) &= \left[\frac{1}{2}\int_{\mathbb{R}}(h_1'\sigma)^2\,d\nu + 2\int h_2^2\,d\nu\right]\\ &= \int \sigma^2\left((\sigma g'')^2 + \sigma\sigma' g' g'' + \frac{1}{2}(\sigma' g')^2\right)d\nu.\end{aligned} \tag{4.4}$$

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

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

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 $m \in \{0,\dots,M-1\}$, $(\hat{\nu}_n^{\gamma,m})_n$ 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

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

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 $\mathbb{E}[\frac{\langle \mathcal{M}(\varphi)\rangle_n}{\Gamma_n}] \xrightarrow{n\to+\infty} 0$.

(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 4.1, 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 4.2, 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 4.1. *Under the assumptions of Proposition 4.2*(b),

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

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

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

Let $A^{(2)}$ be the infinitesimal generator of the duplicated diffusion $(X_t^x, X_t^{x'})_{t\geq 0}$ 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

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

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

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 $\frac{\sqrt{\Gamma_n}}{\Gamma_n^{(2)}}\to 0$, for every smooth function $f:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$,

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

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 $\nabla S(x,y)=2\binom{x-y}{y-x}$, $D^{(2)}S(x,y)=2\big[\begin{smallmatrix}1 & -1\\ -1 & 1\end{smallmatrix}\big]$ and $D^{(\ell)}S\equiv 0$, $\ell\geq 3$. Thus, taking advantage of the strong confluence, we derive that

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

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

$$\mathbb{E}\bigg[\frac{\langle\mathcal{M}^{(1)}\rangle_n}{\Gamma_n^{(2)}}\bigg]\longrightarrow 0\qquad\text{as } n\to+\infty.$$

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

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

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

Then we decompose

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

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 $\mathbb{R}^d$,

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

Thus,

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

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

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

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

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

For $(a)_k$, we write

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

where

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

It is clear, owing to Doob's inequality that

$$
\begin{aligned}
\mathbb{E}(\widetilde{a})_k^2 &\le \|\varphi'\|_{\sup}^2 \Big(\gamma_k^2 \sup_{t\ge 0} \mathbb{E}\big|b(X_t)\big|^2 \\
&\qquad + \gamma_k [\sigma]_{\mathrm{Lip}}^2 \mathbb{E}\Big(\sup_{t\in[X_{\Gamma_{k-1}}, \Gamma_{k-\frac{1}{2}})} |X_s - X_{\Gamma_{k-1}}|^2\Big)\Big) \\
&\le C_{b,\sigma,\varphi}\gamma_k^2.
\end{aligned}
$$

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

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

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

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

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

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

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

The result then follows from the orthogonality of the martingales $\mathcal{M}^{(3,m)}$, $m = 1,\dots,M-1$ (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

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

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

LEMMA 4.2. *Under the assumptions of Proposition* 4.2(b),

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

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 $Z\sim\mathcal{N}(0;1)$, $\mathbb{E}((Z^2-1)^2) = 2$, we get since $(\gamma_n,\gamma_n^2)_{n\geq1}$ is averaging,

$$\frac{\langle\mathcal{N}^1\rangle_n}{\Gamma_n^{(2)}} = \frac{2}{\Gamma_n^{(2)}}\sum_{k=1}^n \gamma_k^2\varphi^2(\bar{X}_{k-1}) \longrightarrow 2\int_{\mathbb{R}}\varphi^2\,d\nu \qquad \text{a.s. as } n\to+\infty \tag{4.5}$$

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

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

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 $m \neq m'$,

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

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 $m \in \{0, \dots, M-1\}$ 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:

$$\begin{aligned} W_{\Gamma_k} - W_{\Gamma_{k-1}} = {} & (W_{\Gamma_k} - W_{\Gamma_{k-1+\frac{m+1}{M}}}) + (W_{\Gamma_{k-1+\frac{m+1}{M}}} - W_{\Gamma_{k-1+\frac{m}{M}}}) \\ & + (W_{\Gamma_{k-1+\frac{m}{M}}} - W_{\Gamma_{k-1}}) \end{aligned}$$

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

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

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 $\ell = 1, \dots, m$, combined with the fact that $\sup_j \mathbb{E}[|\bar{Y}_j|^r] < +\infty$ for every $r > 0$, one derives

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

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 (2.16) 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

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

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 (2.6), 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 $n_r = \lfloor q_r n\rfloor$, $r = 2,\dots,R$. For $\nu_{n_1}(f)$ and $\mu_{n_2}^{(2,M)}(f)$, we will make use of the expansions given in Propositions 3.1(b) and 3.2(b), respectively. For $\mu_{n_r}^{(r,M)}(f)$, $r = 3,\dots,R$, as defined by (2.5), we apply Proposition 3.2(b) with step sequence $(\gamma_n/M^{r-2})_{n\ge1}$. More precisely, by (2.2),

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

so that by Proposition 3.2(b), we have for every $r \in \{2,\dots,R\}$,

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

where $(\mathcal{M}_{n_r}^{(r)})_{n\ge1}$ is defined similar to $\mathcal{M}_n$ but with the step sequence $(\gamma_n/M^{r-2})_{n\ge1}$. In particular, $(\bar{X}_n, \bar{Y}_{Mn+m})$ is now a couple of Euler schemes with step sequences $(\gamma_n/M^{r-2})_{n\ge1}$ and $(\gamma_n/M^{r-1})_{n\ge1}$, respectively.

It follows from the expansions of *order* $R+1$ of each term of $\widetilde{\nu}_n^{(R,\mathbf{W})}$ established in Propositions 3.1(b) and 3.2(b), respectively, that

$$\begin{aligned}
&\widetilde{\nu}_n^{(R,\mathbf{W})}(f) - \nu(f) \\
&\quad = \nu_{n_1}(f) - \nu(f) + \sum_{r=2}^{R} \mathbf{W}_r \mu_{n_r}^{(r,M)}(f) \\
&\quad = c_{R+1}\widetilde{\mathbf{W}}_{R+1}\frac{\Gamma_n^{(R+1)}}{\Gamma_n} + \frac{M_{n_1}^{(1,g)}}{\Gamma_{n_1}} - \sum_{r=2}^{R}\mathbf{W}_r\frac{M^{r-2}\mathcal{M}_{n_r}^{(r)}(\sigma g')}{\Gamma_{n_r}} \\
&\qquad + \mathrm{Bias}^{(1)}(a,R,q,n) + \mathrm{Bias}^{(2)}(a,R,q,n) + o_{L^2}\left(\frac{\sqrt{\Gamma_n}\vee\Gamma_n^{(R+1)}}{\Gamma_n}\right),
\end{aligned} \tag{5.1}$$

where $\mathrm{Bias}^{(1)}(a,R,q,n)$ is defined in Lemma 2.2(b) of [21] and

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

By Lemma 2.2 of [21],

$$\left|\mathrm{Bias}^{(1)}(a,R,q,n)\right| + \left|\mathrm{Bias}^{(2)}(a,R,q,n)\right| \le \frac{C}{n^{1-a}} = o\left(\frac{1}{\sqrt{\Gamma_n}}\right). \tag{5.2}$$

As concerns the martingale components, one deduces from Propositions 4.1(a) and 4.2(a) that

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

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

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

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 $(\widetilde{\nu}_n^{(R,\mathbf{W})})_{n\ge1}$ converges in probability toward $m_f(a,q,R)$. Finally, the proof of Theorem 2.1(a) follows the same lines but with the help of the expansions of Propositions 3.1(a) and 3.2(a).

5.2. *Proof of Theorem* 2.2. Claim (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 (5.1) and (5.2) combined with Propositions 4.1(b) and 4.2(a).

Claim (b) is based on the (sharper) second expansions of Propositions 3.1(c) and 3.2(c) up to order $R+2$. More precisely, using the same strategy as in (5.1), one obtains

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

where $\widetilde{\mathbf{W}}_{R+2}$ is defined by (2.9) [and explicitly given by (2.12)], $\mathrm{Bias}^{(3)}$ is given by

$$\begin{aligned}
\mathrm{Bias}^{(3)}(a,R,q,n) &= c_{R+2}\mathbf{W}_1\left(\frac{\Gamma_{n_1}^{(R+2)}}{\Gamma_{n_1}}-q_1^{-aR}\frac{\Gamma_n^{(R+2)}}{\Gamma_n}\right)\\
&\quad + c_{R+2}\sum_{r=2}^{R}\mathbf{W}_r m_{r,R+2}\left(\frac{\Gamma_{n_r}^{(R+2)}}{\Gamma_{n_r}}-q_r^{-aR}\frac{\Gamma_n^{(R+2)}}{\Gamma_n}\right)
\end{aligned}$$

and $\eta_n^{(1)}$ (resp., $\eta_n^{(2)}$) denotes a remainder term induced by the coarse level (resp., by the levels $r=2,\dots,R$). By Propositions 3.1(c) and 3.2(c), one obtains when $a=1/(2R+1)$,

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

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 3.2(c) by separating the martingale component and the bias component in the o_{L_2}.]

On the other hand, we obtain similar to (5.2):

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

With the help of these properties (and from the independence of the strata), we deduce that

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

The result is then a consequence of Propositions 4.1(c) and 4.2(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 equation (2.6) of [21]). 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 (2.22). It amounts to solving the elementary minimization problem:

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

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

LEMMA 5.1. *Let $A,B,R>0$. Then*

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

and

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

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

$$\gamma_1^* = \left(\frac{2R}{2R+1}\right)^{\frac{1}{2R+1}} (8R)^{-\frac{1}{2R+1}} |c_{R+1}|^{-\frac{2}{2R+1}} \sigma_1(f)^{\frac{2}{2R+1}} M^{\frac{R(R-1)}{2R+1}}. \tag{5.3}$$

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 (2.22) are $\rho\varepsilon^2$ and $(1-\rho)\varepsilon^2$, respectively. The first of these two equalities reads

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

where the step parameter $\gamma_1 = \gamma_0^*(R, M)$ is given by (5.3). One straightforwardly derives that

$$n = n(\varepsilon, R, M, \rho) = \left\lceil \rho^{-(1+\frac{1}{2R})} \mu(R) R \sigma_1^2(f) M^{-\frac{R-1}{2}} \varepsilon^{-2-\frac{1}{R}} \right\rceil, \tag{5.4}$$

where

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

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 (2.22). 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 (5.3) of the optimal step size $\gamma_1^*(R, M)$ which yields

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

Consequently, using the function $\mathbf{\Psi}$ introduced in (2.23) and the obvious fact that $1 - M^{-R} \le 1$, this second term will be upper-bounded by $(1-\rho)\varepsilon^2$ as soon as

$$\begin{aligned} &\frac{R}{n(\varepsilon)} \left(\eta(f, R, M)\sigma_1^2(f) + \sigma_{2,1}^2(f) + \mathbf{\Psi}(M) R \left(1 - \frac{1}{M}\right) \sigma_{2,2}^2(f)\right) \\ &\qquad \le (1-\rho)\varepsilon^2, \end{aligned} \tag{5.5}$$

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 $(c_r)_{r \ge 1}$.

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

$$\begin{aligned} &\sigma_{2,1}^2(f) + \mathbf{\Psi}(M) R \left(1 - \frac{1}{M}\right) \sigma_{2,2}^2(f) \\ &\qquad \le (1-\rho)\rho^{-(1+\frac{1}{2R})} \mu(R) \sigma_1^2(f) M^{-\frac{R-1}{2}} \varepsilon^{-\frac{1}{R}}, \end{aligned}$$

or equivalently

$$\varepsilon^{\frac{1}{R}} M^{\frac{R-1}{2}} R \le \frac{1-\rho}{\rho}\rho^{-\frac{1}{2R}} \frac{\mu(R)\theta_1(f)}{(1-\frac{1}{M})\boldsymbol{\Psi}(M) + R^{-1}(\theta_2(f)+\eta(f,R,M))}, \tag{5.6}$$

where

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

In order to ensure the above condition, we begin by rewriting the left-hand side as follows:

$$\varepsilon^{\frac{1}{R}} M^{\frac{R-1}{2}} R = \exp\left(\frac{1}{R}\left(\frac{\log M}{2} R(R-1) + R\log R + \log\varepsilon\right)\right) \tag{5.7}$$

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

LEMMA 5.2. *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*

$$\lim_{\varepsilon\to 0} x(\varepsilon) = +\infty, \qquad x(\varepsilon) \le \frac{1}{2} + \sqrt{\frac{\log(\frac{1}{\varepsilon})}{\delta} + \frac{1}{4}} \tag{5.8}$$

and

$$x(\varepsilon) = \sqrt{\frac{\log(\frac{1}{\varepsilon})}{\delta}} - \frac{\log_{(2)}(\frac{1}{\varepsilon})}{4\delta} + \frac{1}{2} + \frac{\log\delta}{4\delta} + O\left(\frac{\log_{(2)}(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}}\right) \qquad \text{as } \varepsilon \to 0, \tag{5.9}$$

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, $h(\varepsilon,1) = \log\varepsilon \le 0$ 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 (5.8) is a consequence of the fact that $\delta x(\varepsilon)^2 - \delta x(\varepsilon) - \log(\frac{1}{\varepsilon}) \le 0$ as $x(\varepsilon) \ge 1$. For the expansion, we first note that $x(\varepsilon)$ satisfies the second-order equation

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

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

$$x(\varepsilon) = \sqrt{\frac{\log(\frac{1}{\varepsilon})}{\delta}}\left(\sqrt{1 + \frac{(\log(x(\varepsilon)/\alpha))^2}{4\delta\log(\frac{1}{\varepsilon})}} - \frac{\log(x(\varepsilon)/\alpha)}{2\sqrt{\delta\log(\frac{1}{\varepsilon})}}\right). \tag{5.10}$$

We derive from the inequality in equation (5.8) that, for small enough ε,

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

Consequently, we derive from (5.10) that

$$x(\varepsilon) = \sqrt{\frac{\log(\frac{1}{\varepsilon})}{\delta}}\left(1 + O\left(\frac{\log_{(2)}(1/\varepsilon)}{\sqrt{\log(1/\varepsilon)}}\right)\right) \tag{5.11}$$

so that

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

Plugging this back into (5.11) yields

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

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

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

Now, we set

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

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

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

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

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

where h is defined in the proof of the previous lemma. Plugging these inequalities into (5.7) yield

$$1 \le \varepsilon^{\frac{1}{R(\varepsilon)}} M^{\frac{R(\varepsilon)-1}{2}} R(\varepsilon) \le M\left(1 - \frac{1}{R(\varepsilon)}\right)^{-1+\frac{1}{R(\varepsilon)}}\left(\frac{R(\varepsilon)}{M}\right)^{\frac{1}{R(\varepsilon)}}. \tag{5.12}$$

The above inequality on the right-hand side implies that (5.6) will be true as soon as $\rho = \rho(\varepsilon, M) \in (0, 1)$ satisfies

$$\begin{aligned}\frac{1-\rho}{\rho}\rho^{-\frac{1}{2R(\varepsilon)}} &\geq M\left(1-\frac{1}{R(\varepsilon)}\right)^{-1+\frac{1}{R(\varepsilon)}} \\ &\quad\times\left(\frac{R(\varepsilon)}{M}\right)^{\frac{1}{R(\varepsilon)}}\left(\frac{R^{-1}(\varepsilon)\theta_2(f)+(1-\frac{1}{M})\boldsymbol{\Psi}(M)}{\mu(R(\varepsilon))\theta_1(f)}\right).\end{aligned}$$

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

$$\begin{aligned}\frac{1-\rho(\varepsilon,M)}{\rho(\varepsilon,M)}\rho(\varepsilon,M)^{-\frac{1}{2R(\varepsilon)}} &= M\left(1-\frac{1}{R(\varepsilon)}\right)^{-1+\frac{1}{R(\varepsilon)}} \\ &\quad\times\left(\frac{R(\varepsilon)}{M}\right)^{\frac{1}{R(\varepsilon)}}\left(\frac{R(\varepsilon)^{-1}\theta_2(f)+(1-\frac{1}{M})\boldsymbol{\Psi}(M)}{\mu(R(\varepsilon))\theta_1(f)}\right).\end{aligned}$$

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

$$\frac{1-\rho(\varepsilon,M)}{\rho(\varepsilon,M)} \sim \frac{(M-1)\boldsymbol{\Psi}(M)}{\widetilde{c}\theta_1(f)}$$

that is,

$$\rho(\varepsilon,M) \sim \frac{1}{1+\frac{(M-1)\boldsymbol{\Psi}(M)}{\widetilde{c}\theta_1(f)}} \qquad \text{as } \varepsilon \to 0.$$

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 (5.4) in Step 1 combined with the left inequality in equation (5.12) that

$$\begin{aligned}&n(\varepsilon, R(\varepsilon), M, \rho(\varepsilon)) \\ &\quad\sim\left(1+\frac{(M-1)\boldsymbol{\Psi}(M)}{\widetilde{c}\theta_1(f)}\right)\sigma_1^2(f)\widetilde{c}R(\varepsilon)\left(M^{\frac{R(\varepsilon)-1}{2}}\varepsilon^{\frac{1}{R(\varepsilon)}}\right)^{-1}\varepsilon^{-2}\end{aligned}$$

$$\precsim \left(1+\frac{(M-1)\boldsymbol{\Psi}(M)}{\widetilde{c}\theta_1(f)}\right)\sigma_1^2(f)\widetilde{c}R(\varepsilon)^2\varepsilon^{-2}$$

$$\overset{\varepsilon\to 0}{\sim} \frac{2}{\log M}\left(\widetilde{c}+\frac{(M-1)\boldsymbol{\Psi}(M)}{\theta_1(f)}\right)\sigma_1^2(f)\varepsilon^{-2}\log\left(\frac{1}{\varepsilon}\right).$$

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

$$K\big(n(\varepsilon, M), M\big) = n(\varepsilon, M)\big(1+(M+1)\big(R(\varepsilon)-1\big)\big)\kappa_0 \overset{\varepsilon\to 0}{\sim} (M+1)R(\varepsilon)n(\varepsilon, M)\kappa_0,$$

so that

$$K\big(n(\varepsilon, M), M\big) \precsim \frac{2\kappa_0(M+1)}{\log M}\left(\widetilde{c}+\frac{(M-1)\boldsymbol{\Psi}(M)}{\theta_1(f)}\right)\sigma_1^2(f)\varepsilon^{-2}\log\left(\frac{1}{\varepsilon}\right)$$

as $\varepsilon \to 0$.

$\triangleright$ *Initialization of the step*: it follows from (5.3), the assumption made on c_{R+1} and the convergence of $R(\varepsilon) \to +\infty$ that

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

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

$$\gamma_1^*(\varepsilon) \overset{\varepsilon\to 0}{\sim} \widetilde{c}^{-1} \underbrace{M^{-\frac{3}{4}+\frac{\lceil x(\varepsilon,M)\rceil - x(\varepsilon,M)}{2}}}_{\in(M^{-\frac{1}{4}}, M^{-\frac{3}{4}}]}\left(\frac{\log M}{2}\right)^{\frac{1}{4}}$$

$$\times \exp\left(\sqrt{\frac{\log M \log(\frac{1}{\varepsilon})}{2}}\right)\left(\log\left(\frac{1}{\varepsilon}\right)\right)^{-\frac{1}{4}}.$$

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 aim at computing $\nu(f)$ for a given function f such that $f - \nu(f)$ is supposed to be a smooth enough coboundary.

$\triangleright$ *The weights* $\mathbf{W}_r^{(R)})_{r=1,\dots,R}$. When the re-sizers are uniform, they are computed by an instant closed form (2.14). Otherwise, they are given in full generality by the R-tuple of series (2.10) 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"

$$R(\varepsilon) \sim \sqrt{\frac{2\log(1/\varepsilon)}{\log M}},$$

one has

$$\lim_{\varepsilon\to 0} R(\varepsilon) - \sqrt{\frac{2\log(1/\varepsilon)}{\log M}} = -\infty.$$

▷ *Values for* $\boldsymbol{\Psi}(M)$ *and choice of* M. The quantity $\boldsymbol{\Psi}(M)$ appears in the size parameter $n(\varepsilon, M)$ [and in the complexity parameter $K(f, M)$ given by (2.24)]. Going back to the optimization procedure of the previous section, one remarks that for some fixed R and M, one can replace $\boldsymbol{\Psi}(M)$ 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 (6.4) below and what precedes]. Consequently, in Table 2, we give some values of $\boldsymbol{\Psi}(M)$, but also of $\frac{\Psi(R,M)}{R}$, 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 $\boldsymbol{\Psi}(M)$. The conclusion is that $\boldsymbol{\Psi}(M)$ is an acceptable proxy of $\frac{\Psi(R,M)}{R}$.

▷ *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

TABLE 2
Values of $\Psi(R, M)$ *and* $\boldsymbol{\Psi}(M)$

$\frac{\boldsymbol{\Psi}(R,M)}{R}$	$R=2$	$R=3$	$R=4$	$\boldsymbol{\Psi}(M)$
$M=2$	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

$C = \tilde{c}\theta_1(f)$ is known. Going back to the sharper upper-bound of at our disposal, namely (2.24), it suggests to minimize, for fixed C, the function

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

Without going further, let us just note that $2\boldsymbol{\Psi}(3) \le \boldsymbol{\Psi}(2)$ so that $g_C(3) \le g_C(2)$ 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 (the first paragraph is devoted to a "toy" Ornstein–Uhlenbeck setting), a sharper study of the complexity involving $\frac{\Psi(R,M)}{R}$ leads to temper the answer.

$\triangleright$ *Calibration of the parameters*. This calibration can be performed as a preprocessing 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

$$\frac{1}{\Gamma_n}\sum_{\ell=1}^{L}\mathbb{E}\big[\big(\nu_n^{(\ell)}(f) - \bar{\nu}_n^{(L)}(f)\big)^2\big)\big] \to \sigma_1^2(f) \qquad \text{as } L, n \to +\infty, \tag{6.1}$$

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 $(\mu_n^{M,\gamma})_{n\ge 1}$. We thus may use the same strategy as above. More precisely, under Assumption $(\mathbf{C_s})$, we deduce from Propositions 3.2 and 4.2 that

$$\frac{1}{\Gamma_n^{(2)}}\sum_{\ell=1}^{L}\mathbb{E}\big[\big(\mu_n^{(\ell)}(f) - \bar{\mu}_n^{(L)}(f)\big)^2\big] \xrightarrow{n\to+\infty} \sigma_{2,2}^2(f) \tag{6.2}$$

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 (not investigated here).

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 much dependent on the magnitude of the diffusion coefficient [in fact it can be noted that the same property holds for $\theta_1(f)$].

REMARK 6.1. 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 set $\theta_2(f) = \widetilde{c} = 1$. In the following simulations, the rough estimations of $\sigma_1^2(f)$ and $\theta_1(f)$ [using (6.1)] and (6.2) are achieved with $n = 10^4$ and $L = 20$.

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 all 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 cannot 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 (5.4):

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

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

$$\varepsilon^{\frac{1}{R}} M^{\frac{R-1}{2}} R = \frac{1-\rho}{\rho} \rho^{-\frac{1}{2R}} \frac{\mu(R)\theta_1(f)}{R^{-1}\theta_2(f) + (1 - \frac{1}{M}) R^{-1} \Psi(R, M)}, \tag{6.3}$$

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

$$K(\varepsilon, R, M) = \left(1 + M\left(1 - \frac{1}{R}\right)\right) n(\varepsilon, R, M, \rho^\star). \tag{6.4}$$

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 $K(\varepsilon) = 6.93 \times 10^6$ and $K(\varepsilon) = 1.77 \times 10^9$ 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 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?

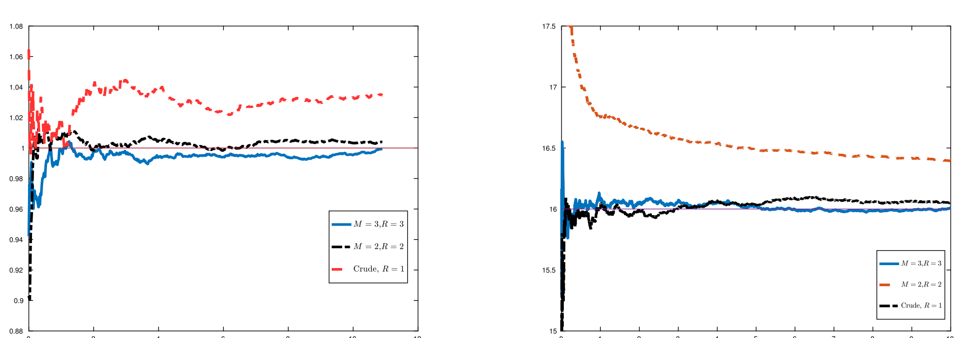

FIG. 1. *Comparison of the evolution in terms of the complexity of the ML2Rgodic with the crude algorithm.*

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

$\sigma = 1$	$R = 2$	$R = 3$	$R = 4$	$\sigma = 4$	$R = 2$	$R = 3$	$R = 4$
$M = 2$	$\mathbf{1.09 \times 10^6}$	1.58×10^6	2.55×10^6	$M = 2$	7.02×10^8	5.23×10^8	7.34×10^8
$M = 3$	1.11×10^6	1.43×10^6	2.05×10^6	$M = 3$	7.17×10^8	$\mathbf{4.76 \times 10^8}$	6.10×10^8
$M = 4$	1.21×10^6	1.57×10^6	2.27×10^6	$M = 4$	7.56×10^8	4.99×10^8	6.55×10^8

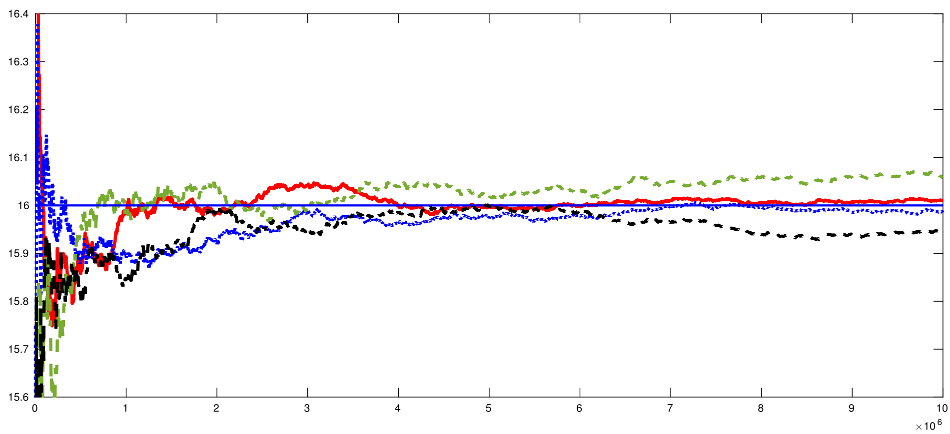

FIG. 2. *Orstein–Uhlenbeck process*: *Evolution of the algorithm in terms of the estimation of the parameters*, *exact value*: $\nu(f) = 16$. $M = 3$, $R = 3$ *and* $M = 2$, $R = 2$ *with exact parameters in red* (*continuous*) *and green* (*dashed*), $M = 3$, $R = 3$ *with estimation of* $\sigma_1^2(f)$ *only and without estimation in blue* (*dot*) *and black* (*dashed*).

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.

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 nonconvex 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 [17], Theorem 2.1). 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 $Z_{V_1} = \int_{\mathbb{R}} \exp(-\frac{V_1(x)}{2\sigma^2})\lambda(dx)$.

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.

Statistical example (*sparse regression learning*). In [3], the authors consider the problem of sparse regression learning by aggregation. For the sake of simplicity, we only recall here 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 $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_N, \mathbf{Y}_N)$ where the vector $\mathbf{X}_i = (X_i^1, \dots, X_i^p)$ is the *predictor* and the scalar $\mathbf{Y}_i$ is the *response*. Suppose that there exists $\theta_0 \in \mathbb{R}^p$ such that

$$\forall i \in \{1, \dots, N\}, \qquad \mathbf{Y}_i = \mathbf{X}_i\theta_0 + \xi_i,$$

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:

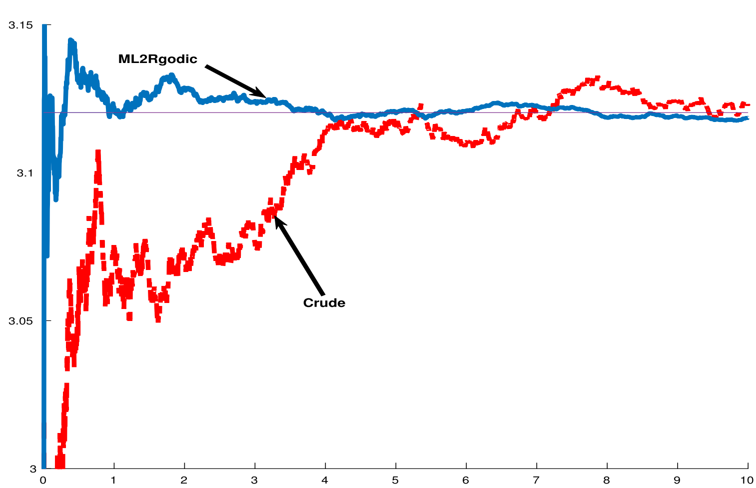

FIG. 3. *Double-well potential*: *approximation of* $\nu(f)$ *with* $f(x) = x^2$, $\sigma = 2$, *exact value*: 3.1207.

how to estimate θ_0 ? When $p \gg N$, the classical methods (such as the least-square method) do not work and it is necessary to introduce some alternative procedures. The estimator of θ_0 proposed by Dalalyan and Tsybakov—called *EWA* (for exponentially weighted aggregate)—is designed as follows:

$$\hat{\theta} = \int_{\mathbb{R}^p} \theta \pi_{V_2}(d\theta),$$

where π_{V_2} is the Gibbs probability measure defined by

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

and Z_{V_2} is a normalizing coefficient and $V_2 : \mathbb{R}^p \mapsto \mathbb{R}$ is the potential defined for some given positive numbers α, β and τ by

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

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

As mentioned (and already numerically tested) in [3], $\hat{\theta}$ is but the expectation related to the invariant distribution of the following SDE:

$$d\theta_t = -\nabla V_2(\theta_t)\, dt + \sqrt{2}\, dW_t. \tag{6.5}$$

It can subsequently 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. More precisely, in order to avoid an explosion of the Euler scheme, we need to impose the step to be not to large for small values of n. We thus assume in what follows that

$$\gamma_n = \min\left(\frac{\gamma_1^\star}{n^a}, \frac{1}{p}\right).$$

Below, we test our **ML2Rgodic** estimator on a *compressed sensing* example given in [3] (see Example 1) with the parameters given in this paper. We fix[1]

$$\alpha = 0, \qquad \beta = 4\sigma^2, \qquad \tau = \frac{4\sigma}{\mathrm{Tr}(\mathbf{X}^t\mathbf{X})^{\frac{1}{2}}}$$

and the computations are achieved with $p = 500$, $N = 100$ and $S = 15$ where S denotes the sparsity parameter, that is, the number S of nonzero components of θ_0 (of course we do not know which ones). Then the matrices $\mathbf{X}$ and $\mathbf{Y}$ are generated from simulated data as follows: in this compressed sensing setting, the matrix $\mathbf{X}$ has independent Rademacher entries with parameter $1/2$. The unknown

[1] From a theoretical point of view, α should be a positive number such that $\alpha \leq 1/(4p\tau)$.

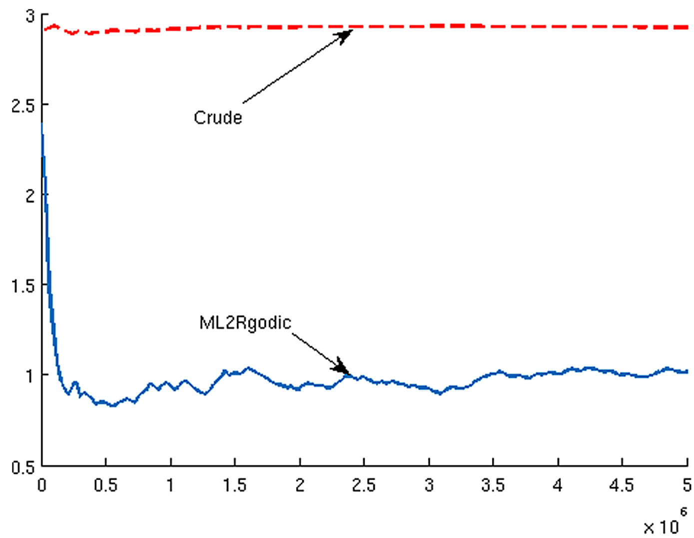

FIG. 4. *Sparse regression learning*: $n \mapsto \|\hat{\theta}_n - \theta_0\|_2$ *for the crude and* **ML2Rgodic** $(R = 3)$ *procedures*.

θ_0 is defined simply by $\theta_0(j) = \mathbf{1}_{j \le S}$, for every $j \in \{1, \dots, p\}$. Finally, following again the parameters given in [3], we set $\sigma^2 = S/9$.

Denoting by $\hat{\theta}_n$ the approximation of $\hat{\theta}$ obtained after n iterations of the scheme, we depict in Figure 4 the evolution of $n \mapsto \|\hat{\theta}_n - \theta_0\|_2$. Note that $\|\hat{\theta}_n - \theta_0\|_2$ converges to $\|\hat{\theta} - \theta_0\|_2$ (which is not equal to 0). We compare it with the crude procedure [taken with $a = 1/3$ whereas for the ML2Rgodic procedure, $a = 1/(2R+1)$ as usual]. We can remark that the correction on the bias involved by the weighted multilevel Langevin procedure strongly improves the estimation of θ_0. This remark is emphasized if we compare with the results of [3] based on an Euler scheme with constant step where the corresponding quantity is equal to 8.917 [in this case, the constant step is about $(Np)^{-1}$].

7. About multilevel finite horizon approach for approximation of invariant distribution. In this paper, we have chosen to develop a multilevel-type estimator of the invariant distribution based on a time discretization of the occupation measure $t^{-1} \int_0^t \delta_{X_s} ds$. As mentioned before, such an ergodic like approach only requires the simulation of one path of the process for the regular procedure (and $2R+1$ in our weighted multilevel setting). However, alternative approaches based on *spatial averaging*, that is, regular Monte Carlo simulations can be considered: thus, one may take advantage of the weak convergence of $\mathcal{L}(X_t)$ toward the invariant distribution. The simplest way to proceed is to fix a large enough horizon T so that $|\mathbb{E}_x[f(X_T)] - \nu(f)|$ is small and to approximate $\mathbb{E}_x[f(X_T)]$ by a Monte-Carlo simulation.

Let us conclude this paper by a rough study of the complexity of such an approach when the above Monte Carlo method is a standard multilevel procedure and by some comparisons with our algorithm. To begin with, let us consider a continuous-time discretization scheme $(\xi_t^h)_{t\geq 0}$ with constant step h.

For a given T, we recall that the standard MLMC method consists in considering the following type-estimator:

$$\Upsilon(T, R, \mathbf{N}) = \frac{1}{N_1}\sum_{k=1}^{N_1} Y_k^{(1)} + \sum_{r=2}^{R}\frac{1}{N_r}\sum_{k=1}^{N_r} Y_k^{(r)},$$

where $\mathbf{N} = (N_1, \dots, N_R)$, $(Y_k^r)_{k,r}$ is a sequence of independent random variables with $Y_k^{(1)} \sim f(\xi_T^h)$, $k = 1, \dots, N_1$ whereas for $r = 2, \dots, R$, Y_k^r is based on a consistent[2] coupling of schemes with step $h/2^{r-2}$ and $h/2^{r-1}$: $Y_k^{(r)} \sim (f(\xi_T^{h2^{1-r}}) - f(\xi_T^{h2^{2-r}}))$.

In the following proposition, we show that under $(\mathbf{C_s})$, a complexity proportional to $\varepsilon^{-2}\log(1/\varepsilon)$ can also be attained.

PROPOSITION 7.1. (i) *Let $f : \mathbb{R}^d \mapsto \mathbb{R}$ be a given function and $x \in \mathbb{R}^d$. Let $(\xi_t^{h,x})_{t\geq 0}$ be a (continuous-time) discretization scheme with constant step h starting from x. Assume that the following properties hold*:

(a) (*Rate of convergence*) *$\exists \rho > 0$ and $c_1 > 0$ such that for any $T \geq 0$, $|\mathbb{E}[f(X_T^x)] - \nu(f)| \leq c_1 e^{-\rho T}$.*

(b) (*Weak error*) *There exists $c_2 > 0$ such that $\sup_{T\geq 0}|\mathbb{E}[f(X_T^x)] - \mathbb{E}[f(\xi_T^h)]| \leq c_2 h$.*

(c) (*L^2-error*) *There exists $\beta \geq 1$ and a real constant c_3 such that for every $T \geq 0$,*

$$\|X_T^x - \xi_T^h\|_2 \leq c_3 h^{\frac{\beta}{2}}.$$

Then there exists a real constant C depending on c_1, c_2 and c_3 such that for any $\varepsilon > 0$, the choice

$$T = \frac{1}{\rho}\log\left(\frac{1}{\varepsilon}\right), \qquad R = \left\lfloor \frac{\log(\frac{1}{\varepsilon})}{\log 2} \right\rfloor,$$

$$N_r = \begin{cases} \lfloor 2^{-r\frac{\beta+1}{2}}\varepsilon^{-2}\log(1/\varepsilon)\rfloor & \text{if } \beta = 1, \\ \lfloor 2^{-r\frac{\beta+1}{2}}\varepsilon^{-2}\rfloor & \text{if } \beta > 1, \end{cases}$$

for $r = 1, \dots, R$, leads to the following properties:

$$\|\Upsilon(T, R, \mathbf{N}) - \nu(f)\|_2 \leq C\varepsilon$$

---

[2]By consistent, we mean that the discretization schemes are built with the same Brownian motion.

and the complexity of the discretization scheme is proportional to

$$\begin{cases} \varepsilon^{-2}\log^3\left(\frac{1}{\varepsilon}\right) & \text{if } \beta = 1, \\ \varepsilon^{-2}\log\left(\frac{1}{\varepsilon}\right) & \text{if } \beta > 1. \end{cases} \tag{7.1}$$

(ii) *Assume that* f *is a Lipschitz continuous function with Lipschitz constant* $[f]_1$. *Let* ξ^h *denote the Euler–Maruyama scheme with constant step* h. *If* $(\mathbf{C_s})$ *holds*, *assumptions* (a), (b) *and* (c) *of* (i) *hold true with with* $\rho = \alpha$ *and* $\beta = 2$ (*and* $c_1 = [f]_{1,S}\int \|x-y\|_S\nu(dy)$). *As a consequence*, *the complexity is proportional to* $\varepsilon^{-2}\log(\frac{1}{\varepsilon})$.

The proof of this proposition is postponed in [21], Section 3. Let us conclude by a series of comments about this alternative approach.

Comments. In the previous result, we thus obtained that the spatial averaging has also the capacity to lead to the complexity $\varepsilon^{-2}\log(\frac{1}{\varepsilon})$ Furthermore, it is worth noting that the strong convexity-type assumption $(\mathbf{C_s})$ is a guarantee to obtain $\beta = 2$ for the classical Euler–Maruyama scheme. In other words, one does not require the use of Milstein scheme or antithetic schemes (see [9]) to get $\beta = 2$ under the contraction assumption $(\mathbf{C_s})$. In fact, one retrieves a similar property in Proposition 4.2 where the order $(\Gamma_n^{(2)})^{\frac{1}{2}}$ of the rate of convergence in the CLT (related to the coupling of consistent coarse and refined Euler schemes) does not depend on the variability of σ. In the same spirit, one can cite the recent paper [29] where an alternative finite horizon approach is developed under $(\mathbf{C_s})$ but leading to a less competitive bound $\varepsilon^{-2}\log^3(\frac{1}{\varepsilon})$ (which corresponds to the case $\beta = 1$ in our previous result).

However, let us insist on the fact that the previous result is rough in the sense that there is no precision about what we mean by "proportional to" and about the constant C. In fact, as in our setting, a precise implementation would need to estimate the parameters c_1, c_2, c_3 and to optimize the choice of the parameters (including certainly the choice of the step h). Such developments should be considered in a future paper to really compare the performances.

Nevertheless, one can objectively consider that the pathwise approach has an important advantage with respect to the finite horizon approach: there is only one asymptotic and it does not directly depend on the parameter ρ [which corresponds to the parameter α in $(\mathbf{C_s})$]. Actually, it seems that in contrast with our method, the "finite horizon" procedure really requires $(\mathbf{C_s})$ to be implemented. In fact, for the occupation measure, the order of the rate of convergence is about $\sqrt{t}$ in a very general setting (including nonconvex settings). Then, as explained before, we only need to get a rough estimation of the limiting variance $\sigma_1^2(f)$ whereas in the finite-horizon approach, one really needs to have a precise idea of ρ to fix the value of T.

For instance, for a gradient-diffusion, $dX_t = -\nabla U(X_t)\,dt + \sigma\,dW_t$ with a $\mathcal{C}^2$ strictly convex potential $U : \mathbb{R}^d \to \mathbb{R}$, we can fix $\rho = \alpha = \min_{x\in\mathbb{R}^d} \underline{\lambda}_{D^2U(x)}$ where for a symmetric matrix A, $\underline{\lambda}_A$ is the lowest eigenvalue of A. This estimate is satisfying for the Ornstein–Uhlenbeck process since D^2U is constant. However, in some less regular settings (where for instance, $\underline{\lambda}_{D^2U(x)}$ is close to 0 in some areas of the space), this parameter ρ can be estimated but may be very pessimistic. Finally, in the nonconvex example (6.5), ρ is clearly unknown and there is no natural way to fix T so that the application of the previous result is not really possible.

In the same way, the long-time control of the L^2-error and the constant c_3 strongly depend on $(\mathbf{C_s})$ and on the corresponding parameter α. Actually, let us first recall that in a general nonergodic setting, this constant may dramatically increase with T. The existence of a c_3 independent of T under $(\mathbf{C_s})$ means that the true and discretized Euler schemes get closer when T increases with the help of the contraction involved by the drift term. In fact, the parameter c_3 is the finite-horizon counterpart to $\sigma_2^2(f)$. Without going into details, let us remark that an explicit computation of c_3 would involve the pessimistic parameter α whereas $\sigma_2^2(f)$ seems to be a more robust and realistic parameter since it is averaged over the invariant distribution.

SUPPLEMENTARY MATERIAL

Supplement to "Weighted multilevel Langevin simulation of invariant measures". (DOI: 10.1214/17-AAP1364SUPP; .pdf). In order to improve the readability of the current article, several technical proofs have been postponed in a supplementary document. In the case in point, the precise reference is given at the end of the proposition.

REFERENCES

[1] BHATTACHARYA, R. N. (1982). On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete* **60** 185–201. MR0663900

[2] BILLINGSLEY, P. (1978). *Ergodic Theory and Information*. Robert E. Krieger Publishing Co., Huntington, NY. Reprint of the 1965 original. MR0524567

[3] DALALYAN, A. S. and TSYBAKOV, A. B. (2012). Sparse regression learning by aggregation and Langevin Monte-Carlo. *J. Comput. System Sci.* **78** 1423–1443. MR2926142

[4] DUFLO, M. (1997). *Random Iterative Models. Applications of Mathematics (New York)* **34**. Springer, Berlin. Translated from the 1990 French original by Stephen S. Wilson and revised by the author. MR1485774

[5] FRIKHA, N. (2016). Multi-level stochastic approximation algorithms. *Ann. Appl. Probab.* **26** 933–985. MR3476630

[6] GARCÍA TRILLOS, C. A. (2015). A decreasing step method for strongly oscillating stochastic models. *Ann. Appl. Probab.* **25** 986–1029. MR3313761

[7] GILBARG, D. and TRUDINGER, N. S. (1983). *Elliptic Partial Differential Equations of Second Order*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **224**. Springer, Berlin. MR0737190

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

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

[10] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

[11] KRENGEL, U. (1985). *Ergodic Theorems*. *De Gruyter Studies in Mathematics* **6**. de Gruyter, Berlin. MR0797411

[12] LAMBERTON, D. and PAGÈS, G. (2002). Recursive computation of the invariant distribution of a diffusion. *Bernoulli* **8** 367–405. MR1913112

[13] LAMBERTON, D. and PAGÈS, G. (2003). Recursive computation of the invariant distribution of a diffusion: The case of a weakly mean reverting drift. *Stoch. Dyn.* **3** 435–451.

[14] LEMAIRE, V. (2005). Estimation récursive de la mesure invariante d'un processus de diffusion. Thèse de doctorat, Univ. Marne-la-Vallée, France.

[15] LEMAIRE, V. (2007). Behavior of the Euler scheme with decreasing step in a degenerate situation. *ESAIM Probab. Stat.* **11** 236–247.

[16] LEMAIRE, V. and PAGÈS, G. (2017). Multilevel Richardson–Romberg extrapolation. *Bernoulli* **23** 2643–2692. MR3648041

[17] LEMAIRE, V., PAGÈS, G. and PANLOUP, F. (2015). Invariant measure of duplicated diffusions and application to Richardson–Romberg extrapolation. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** 1562–1596.

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

[19] PAGÈS, G. and PANLOUP, F. (2009). Approximation of the distribution of a stationary Markov process with application to option pricing. *Bernoulli* **15** 146–177. MR2546802

[20] PAGÈS, G. and PANLOUP, F. (2014). A mixed-step algorithm for the approximation of the stationary regime of a diffusion. *Stochastic Process. Appl.* **124** 522–565. MR3131304

[21] PAGÈS, G. and PANLOUP, F. (2018). Supplement to "Weighted multilevel Langevin simulation of invariant measures." DOI:10.1214/17-AAP1364SUPP.

[22] PANLOUP, F. (2008). Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. *Ann. Appl. Probab.* **18** 379–426. MR2398761

[23] PARDOUX, E. and VERETENNIKOV, A. YU. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** 1061–1085. MR1872736

[24] PICCIONI, M. and SCARLATTI, S. (1994). An iterative Monte Carlo scheme for generating Lie group-valued random variables. *Adv. in Appl. Probab.* **26** 616–628.

[25] RICHARDSON, L. F. (1911). The approximate arithmetical solution by finite differences of physical problems including differential equations, with an application to the stresses in a masonry dam. *Philos. Trans. R. Soc. Lond. Ser. A* **210** 307–357.

[26] RICHARDSON, L. F. (1927). The deferred approach to the limit. *Philos. Trans. R. Soc. Lond. Ser. A* **226** 299–349.

[27] ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268.

[28] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273

[29] SZPRUCH, L., VOLLMER, S., ZYGALAKIS, K. and GILES, M. B. (2016). Multi Level Monte Carlo methods for a class of ergodic stochastic differential equations. ArXiv e-print. Available at arXiv:1605.01384 [math.NA].

[30] TALAY, D. (1990). Second order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stoch. Stoch. Rep.* **29** 13–36.

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

LAREMA
Université d'Angers
2 Bd Lavoisier
49045 Angers Cedex 01
France
E-mail: fabien.panloup@univ-angers.fr