The Annals of Applied Probability
2016, Vol. 26, No. 2, 933–985
DOI: 10.1214/15-AAP1109

MULTI-LEVEL STOCHASTIC APPROXIMATION ALGORITHMS

BY NOUFEL FRIKHA

Université Paris Diderot

This paper studies multi-level stochastic approximation algorithms. Our aim is to extend the scope of the multi-level Monte Carlo method recently introduced by Giles [*Oper. Res.* **56** (2008) 607–617] to the framework of stochastic optimization by means of stochastic approximation algorithm. We first introduce and study a two-level method, also referred as statistical Romberg stochastic approximation algorithm. Then its extension to a multi-level method is proposed. We prove a central limit theorem for both methods and give optimal parameters. Numerical results confirm the theoretical analysis and show a significant reduction in the initial computational cost.

1. Introduction. In this paper we propose and analyze a multi-level paradigm for stochastic optimization problems by means of stochastic approximation schemes. The multi-level Monte Carlo method introduced by Heinrich [18] and popularized in numerical probability by [21] and [15] allows one to significantly increase the computational efficiency of the expectation of an $\mathbb{R}$-valued nonsimulatable random variable Y that can only be strongly approximated by a sequence $(Y^n)_{n\geq 1}$ of easily simulatable random variables (all defined on the same probability space) as the *bias parameter* n goes to infinity with a *weak error* or *bias* $\mathbb{E}[Y] - \mathbb{E}[Y^n]$ of order $n^{-\alpha}$, $\alpha > 0$. Let us be more specific. In this context, the standard Monte Carlo method uses the statistical estimator $M^{-1} \times \sum_{j=1}^{M} Y^{n,j}$ where the $(Y^{n,j})_{j\in[\![1,M]\!]}$ are M independent copies of Y^n. Given the order of the weak error, a natural question is how to find the optimal choice of the sample size M to achieve a global error. If the weak error is of order $n^{-\alpha}$, then for a total error of order $n^{-\alpha}$ ($\alpha \in [1/2, 1]$), the minimal computation necessary for the standard Monte Carlo algorithm is obtained for $M = n^{2\alpha}$; see [8]. So if the computational cost required to simulate one sample of Y^n is of order n, then the optimal computational cost of the Monte Carlo method is $C_{\text{MC}} = C \times n^{2\alpha+1}$, for a positive constant $C > 0$.

In order to reduce the complexity of the computation, the principle of the multi-level Monte Carlo method, introduced by Giles [15] as a generalization of Ke-

Received October 2013; revised February 2015.

MSC2010 subject classifications. 60F05, 62K12, 65C05, 60H35.

Key words and phrases. Multi-level Monte Carlo methods, stochastic approximation, Ruppert–Polyak averaging principle, Euler scheme.

baier's approach [21], consists of using the telescopic sum

$$\mathbb{E}[Y^{m^L}] = \mathbb{E}[Y^1] + \sum_{\ell=1}^{L} \mathbb{E}[Y^{m^\ell} - Y^{m^{\ell-1}}],$$

where $m \in \mathbb{N}^* \setminus \{1\}$ satisfies $m^L = n$. For each level $\ell \in \{1, \dots, L\}$ the numerical computation of $\mathbb{E}[Y^{m^\ell} - Y^{m^{\ell-1}}]$ is achieved by the standard Monte Carlo method using N_ℓ independent samples of $(Y^{m^{\ell-1}}, Y^{m^\ell})$. An important point is that the random samples Y^{m^ℓ} and $Y^{m^{\ell-1}}$ are perfectly correlated. Then the expectation $\mathbb{E}[Y^n]$ is approximated by the following multi-level estimator:

$$\frac{1}{N_0} \sum_{j=1}^{N_0} Y^{1,j} + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{j=1}^{N_\ell} (Y^{m^\ell,j} - Y^{m^{\ell-1},j}),$$

where for each level ℓ, $(Y^{m^\ell,j})_{j \in [\![1,N_\ell]\!]}$ is a sequence of i.i.d. random variables with the same law as Y^{m^ℓ}.

Based on an analysis of the variance, Giles [15] proposed an optimal choice for the sequence $(N_\ell)_{1 \leq \ell \leq L}$ which minimizes the total complexity of the algorithm. More recently, Ben Alaya and Kebaier [6] proposed a different analysis to obtain the optimal choice of the parameters that relies on a Lindeberg–Feller central limit theorem (CLT) for the multi-level Monte Carlo algorithm. To obtain a global error of order $n^{-\alpha}$, both approaches allow one to achieve a complexity of order $n^{2\alpha}(\log n)^2$ if the $L^2(\mathbb{P})$ strong approximation rate of Y by Y^n, namely $\mathbb{E}[|Y^n - Y|^2]$, is of order $1/n$. Hence the multi-level Monte Carlo method is significantly more effective than the crude Monte Carlo and the statistical Romberg methods. Originally introduced for the computation of expectations involving stochastic differential equation (SDE), it has been widely applied to various problems of numerical probability; see Giles [14], Dereich [7], Giles, Higham and Mao [16], among others. We refer the interested reader to the web page http://people.maths.ox.ac.uk/gilesm/mlmc_community.html for further developments.

In the present paper, we are interested in broadening the scope of the multi-level Monte Carlo method to the framework of stochastic approximation (SA) algorithm. Introduced by Robbins and Monro [26], these recursive, simulation-based algorithms are effective procedures that are widely used to solve inverse problems. To be more specific, their aim is to find a zero of a continuous function $h : \mathbb{R}^d \to \mathbb{R}^d$ which is unknown to the experimenter but can only be estimated through experiments. Successfully and widely investigated from both a theoretical and applied point of view since this seminal work, such procedures are now commonly used in various contexts such as convex optimization since minimizing a function amounts to finding a zero of its gradient. In the general Robbins–Monro procedure, the function h writes $h(\theta) := \mathbb{E}[H(\theta, U)]$ where $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$

and U is an $\mathbb{R}^q$-valued random vector. To estimate the zero of h, they proposed the algorithm

$$\theta_{p+1} = \theta_p - \gamma_{p+1} H(\theta_p, U^{p+1}), \qquad p \geq 0, \tag{1.1}$$

where $(U^p)_{p\geq 1}$ is an i.i.d. sequence of copies of U defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, θ_0 is independent of the innovation of the algorithm with $\mathbb{E}[|\theta_0|^2] < +\infty$ and $\gamma = (\gamma_p)_{p\geq 1}$ is a sequence of nonnegative deterministic and decreasing steps satisfying the assumption

$$\sum_{p\geq 1} \gamma_p = +\infty \quad \text{and} \quad \sum_{p\geq 1} \gamma_p^2 < +\infty. \tag{1.2}$$

When the function h is the gradient of a convex potential, the recursive procedure (1.1) is a stochastic gradient algorithm. Indeed, replacing $H(\theta_p, U^{p+1})$ by $h(\theta_p)$ in (1.1) leads to the usual deterministic descent gradient procedure. When $h(\theta) = k(\theta) - \ell$, $\theta \in \mathbb{R}$, where k is a monotone function, say increasing, which writes $k(\theta) = \mathbb{E}[K(\theta, U)]$, $K : \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}$ being a Borel function and ℓ a given desired level, then setting $H = K - \ell$, the recursive procedure (1.1) aims to compute the value $\bar{\theta}$ such that $k(\bar{\theta}) = \ell$.

As in the case of the Monte Carlo method described above, the random vector U is not directly simulatable (at a reasonable cost) but can only be approximated by another sequence of easily simulatable random vectors $((U^n)^p)_{p\geq 1}$, which strongly approximates U as $n \to +\infty$ with a standard weak discretization error (or bias) $\mathbb{E}[f(U)] - \mathbb{E}[f(U^n)]$ of order $n^{-\alpha}$ for a specific class of functions. The computational cost required to simulate one sample of U^n is assumed to be of order n, that is, $\mathrm{Cost}(U^n) = K \times n$ for some positive constant K. One standard situation corresponds to the case of a discretization of an SDE by means of an Euler–Maruyama scheme with n time steps.

Some typical applications are the computations of the implied volatility or the implied correlation, both of which boil down to finding the zero of a function that writes as an expectation. Computing the value-at-risk and the conditional value-at-risk of a financial portfolio when the dynamics of the underlying assets are given by an SDE also appears as an inverse problem for which a SA scheme may be devised; see, for example, [2, 3]. The risk minimization of a financial portfolio by means of SA has been investigated in [4, 12]. For more applications and a complete overview in the theory of stochastic approximation, the reader may refer to [9, 22] and [5].

The important point here is that the function h is generally neither known nor computable (at least at reasonable cost), and since the random variable U cannot be simulated, estimating θ^* using the recursive scheme (1.1) is not possible. Therefore, the following two steps are needed to compute θ^*:

– The first step consists of approximating the zero θ^* of h by the zero $\theta^{*,n}$ of h^n defined by $h^n(\theta) := \mathbb{E}[H(\theta, U^n)]$, $\theta \in \mathbb{R}^d$. It induces *an implicit weak error* which writes

$$\mathcal{E}_D(n) := \theta^* - \theta^{*,n}.$$

Let us note that $\theta^{*,n}$ appears as a proxy of θ^*, and one would naturally expect that $\theta^{*,n} \to \theta^*$ as the bias parameter n tends to infinity.

– The second step consists of approximating $\theta^{*,n}$ by $M \in \mathbb{N}^*$ steps of the following SA scheme:

$$\theta^n_{p+1} = \theta^n_p - \gamma_{p+1} H\big(\theta^n_p, (U^n)^{p+1}\big), \qquad p \in [\![0, M-1]\!], \tag{1.3}$$

where $((U^n)^p)_{p\in[\![1,M]\!]}$ is an i.i.d. sequence of random variables with the same law as U^n, θ^n_0 is independent of the innovation of the algorithm with $\sup_{n\geq 1} \mathbb{E}[|\theta^n_0|^2] < +\infty$ and $\gamma = (\gamma_p)_{p\geq 1}$ is a sequence of nonnegative deterministic and decreasing steps satisfying (1.2). This induces a *statistical error* which writes

$$\mathcal{E}_S(n, M, \gamma) := \theta^{*,n} - \theta^n_M.$$

The global error between the quantity to estimate θ^* and its implementable approximation θ^n_M can be decomposed as follows:

$$\mathcal{E}_{\text{glob}}(n, M, \gamma) = \theta^* - \theta^{*,n} + \theta^{*,n} - \theta^n_M := \mathcal{E}_D(n) + \mathcal{E}_S(n, M, \gamma).$$

The first step of our analysis consists of investigating the behavior of the *implicit weak error* $\mathcal{E}_D(n)$. Under mild assumptions on the functions h and h^n, namely the local uniform convergence of $(h^n)_{n\geq 1}$ toward h and a mean reverting assumption of h and h^n, we prove that $\lim_n \mathcal{E}_D(n) = 0$. We next show that under additional assumption, namely the local uniform convergence of $(Dh^n)_{n\geq 1}$ toward Dh and the nonsingularity of $Dh(\theta^*)$, the rate of convergence of the *standard weak error* $h^n(\theta) - h(\theta)$, for a fixed $\theta \in \mathbb{R}^d$, transfers to the *implicit weak error* $\mathcal{E}_D(n) = \theta^* - \theta^{*,n}$.

Regarding the *statistical error* $\mathcal{E}_S(n, M, \gamma) := \theta^{*,n} - \theta^n_M$, it is well known that under standard assumptions, that is, a mean reverting assumption on h^n and a growth control of the $L^2(\mathbb{P})$-norm of the noise of the algorithm, the Robbins–Monro theorem guarantees that $\lim_M \mathcal{E}_S(n, M, \gamma) = 0$ for each fixed $n \in \mathbb{N}^*$; see Theorem 2.3 below. Moreover, under mild technical conditions, a CLT holds at rate $\gamma^{-1/2}(M)$; that is, for each fixed $n \in \mathbb{N}^*$, $\gamma^{-1/2}(M)\mathcal{E}_S(n, M, \gamma)$ converges in distribution to a normally distributed random variable with mean zero and finite covariance matrix; see Theorem 2.4 below. The reader may also refer to [10, 13] for some recent developments on nonasymptotic deviation bounds for the statistical error. In particular, if we set $\gamma(p) = \gamma_0/p$, $\gamma_0 > 0$, $p \geq 1$, the weak convergence rate is $\sqrt{M}$, provided that $2\mathcal{R}e(\lambda_{\min})\gamma_0 > 1$ where $\lambda_{\min}$ denotes the eigenvalue of $Dh(\theta^*)$ with the smallest real part. However, this local condition on the Jacobian matrix of h at the equilibrium is difficult to handle in practical situations.

To circumvent such a difficulty, it is fairly well known that the key idea is to carefully smooth the trajectories of a converging SA algorithm by averaging according to the *Ruppert–Polyak averaging principle*; see, for example, [24, 27]. It consists of devising the original SA algorithm (1.3) with a slow decreasing step and simultaneously computing the empirical mean $(\bar{\theta}_p^n)_{p\geq 1}$ (which a.s. converges to $\theta^{*,n}$) of the sequence $(\theta_p^n)_{p\geq 0}$ by setting

$$\bar{\theta}_p^n = \frac{\theta_0^n + \theta_1^n + \cdots + \theta_p^n}{p+1} = \bar{\theta}_{p-1}^n - \frac{1}{p+1}(\bar{\theta}_{p-1}^n - \theta_p^n). \tag{1.4}$$

The statistical error now writes $\mathcal{E}_S(n, M, \gamma) := \theta^{*,n} - \bar{\theta}_M^n$, and under mild assumptions a CLT holds at rate $\sqrt{M}$ without any stringent condition on γ_0.

Given the order of the implicit weak error and a step sequence γ satisfying (1.2), a natural question is how to find the optimal balance between the value of n and the number M of steps in (1.3) in order to achieve a given global error. This problem was originally investigated in [8] for the standard Monte Carlo method. The error between θ^* and the approximation θ_M^n writes $\theta_M^n - \theta^* = \theta_M^n - \theta^{*,n} + \theta^{*,n} - \theta^*$, suggesting the selection of $M = \gamma^{-1}(1/n^{2\alpha})$, where γ^{-1} is the inverse function of γ, when the *weak error* is of order $n^{-\alpha}$. However, due to the nonlinearity of the SA algorithm (1.3), the methodology developed in [8] does not apply in our context. The key tool that is necessary to tackle this question consists of linearizing the dynamic of $(\theta_p^n)_{p\in[\![1,M]\!]}$ around its target $\theta^{*,n}$, quantifying the contribution of the nonlinearities in the space variable θ_p^n and the innovations and finally exploiting stability arguments from SA schemes. Optimizing with respect to the usual choice of the step sequence, the minimal computational cost (to achieve an error of order $n^{-\alpha}$) is given by $C_{\mathrm{SA}} = K \times n \times \gamma^{-1}(1/n^{2\alpha})$ and is optimal for $\gamma(p) = \gamma_0/p$, $p \geq 1$, provided that the constant γ_0 satisfies a stringent condition involving h^n, leading to a complexity of order $n^{2\alpha+1}$. Considering the empirical mean sequence $(\bar{\theta}_p^n)_{p\in[\![1,n^{2\alpha}]\!]}$ instead of the crude SA estimate also allows one to reach the optimal complexity for free, without any condition on γ_0.

To increase the computational efficiency for the estimation of θ^* by means of SA algorithm, we investigate in a second part, multi-level SA algorithms. The first one is a two-level method, also referred as the statistical Romberg SA procedures. It consists of approximating the unique zero θ^* of h by $\Theta_n^{\mathrm{sr}} = \theta_{M_1}^{n^\beta} + \theta_{M_2}^n - \theta_{M_2}^{n^\beta}$, $\beta \in (0, 1)$. The couple $(\theta_{M_2}^n, \theta_{M_2}^{n^\beta})$ is computed using M_2 independent copies of (U^n, U^{2n}). Moreover the random samples used to obtain $\theta_{M_1}^{n^\beta}$ are independent of those used for the computation of $(\theta_{M_2}^n, \theta_{M_2}^{n^\beta})$. For an implicit weak error of order $n^{-\alpha}$, we prove a CLT for the sequence $(\Theta_n^{\mathrm{sr}})_{n\geq 1}$ through which we are able to optimally set M_1, M_2 and β with respect to n and the step sequence γ. The intuitive idea is that when n is large, $(\theta_p^n)_{p\in[\![0,M_2]\!]}$ and $(\theta_p^{n^\beta})_{p\in[\![0,M_2]\!]}$ are close to the SA scheme $(\theta_p)_{p\in[\![0,M_2]\!]}$ devised with the innovation variables $(U^p)_{p\geq 1}$ so that the correction term writes $\theta_{M_2}^n - \theta_{M_2} - (\theta_{M_2}^{n^\beta} - \theta_{M_2})$. Then we quantify the two main contributions in this decomposition, namely the one due to the nonlinearity in the space variables $(\theta_p^{n^\beta}, \theta_p^n, \theta_p)_{p\in[\![0,M_2]\!]}$ and the one due to the nonlinearity in the innovation variables $(U^{n^\beta,p}, U^{n,p}, U^p)_{p\geq 1}$. Under mild smoothness assumption on the function H, the weak rate of convergence is ruled by the nonlinearity in the innovation variables for which we use the weak convergence of the normalized error $n^\rho(U^n - U)$, $\rho \in (0, 1/2]$. The optimal choice of the step sequence is again $\gamma_p = \gamma_0/p$, $p \geq 1$ and induces a complexity for the procedure given by $C_{\text{SA-SR}} = K \times n^{2\alpha+1/(1+\rho)}$, provided that γ_0 satisfies again a condition involving h^n which is difficult to handle in practice. By considering the empirical mean sequence $\bar{\Theta}_n^{\text{sr}} = \bar{\theta}_{M_3}^{n^\beta} + \bar{\theta}_{M_4}^{n} - \bar{\theta}_{M_4}^{n^\beta}$, where $(\bar{\theta}_p^{n^\beta})_{p\in[\![0,M_3]\!]}$ and $(\bar{\theta}_p^{n}, \bar{\theta}_p^{n^\beta})_{p\in[\![0,M_4]\!]}$ are, respectively, the empirical means of the sequences $(\theta_p^{n^\beta})_{p\in[\![0,M_3]\!]}$ and $(\theta_p^{n}, \theta_p^{n^\beta})_{p\in[\![0,M_4]\!]}$ devised with the same slow decreasing step sequence, this optimal complexity is reached for free by setting $M_3 = n^{2\alpha}$, $M_4 = n^{2\alpha-1/(1+\rho)}$ without any condition on γ_0.

Moreover, we generalize this approach to the case of the multi-level SA method. In the spirit of [15] for Monte Carlo path simulation, the multi-level SA scheme estimates $\theta^{*,n}$ by computing the quantity $\Theta_n^{\text{ml}} = \theta_{M_0}^1 + \sum_{\ell=1}^L \theta_{M_\ell}^{m^\ell} - \theta_{M_\ell}^{m^{\ell-1}}$ where for every ℓ, the couple $(\theta_{M_\ell}^{m^\ell}, \theta_{M_\ell}^{m^{\ell-1}})$ is obtained using M_ℓ independent copies of $(U^{m^{\ell-1}}, U^{m^\ell})$. Here again to establish a CLT for this estimator (in the spirit of [6] for the Monte Carlo path simulation), our analysis follows the lines of the methodology developed so far. The optimal computational cost to achieve an accuracy of order $1/n$ is reached by setting $M_0 = \gamma^{-1}(1/n^2)$, $M_\ell = \gamma^{-1}(m^\ell \log(m)/(n^2 \log(n)(m-1)))$, $\ell = 1, \dots, L$ in the case $\rho = 1/2$. Once again the step sequence $\gamma(p) = \gamma_0/p$, $p \geq 1$ is optimal among the usual choices, and it induces an asymptotic complexity of order $n^2(\log(n))^2$. We thus recover the rates as in the multi-level Monte Carlo path simulation for SDE obtained in [15] and [6].

The paper is organized as follows. In the next section we state our main results and list the assumptions. Section 3 is devoted to the proofs. In Section 4 numerical results are presented to confirm the theoretical analysis. Finally, Section 5 is devoted to technical results which are useful throughout the paper.

2. Main results.

In the present paper, we make no attempt to provide an exhaustive discussion related to convergence results of SA schemes. We refer the interested readers to [9, 22] and [5], among others, for developments and a more complete overview in SA theory. In the next section, we first recall some basic facts concerning stable convergence (following the notations of Jacod and Protter [20]) and list classical results of SA theory.

2.1. *Preliminaries.*

2.1.1. *Stable convergence.*

For a sequence of E-valued (E being a Polish space) random variables $(X_n)_{n\geq 1}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say that $(X_n)_{n\geq 1}$ converges in law stably to X defined on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$

of $(\Omega, \mathcal{F}, \mathbb{P})$ and write $X_n \overset{\text{stably}}{\Longrightarrow} X$ if for all bounded random variables Λ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and for all $h : E \to \mathbb{R}$ bounded continuous, one has

$$\mathbb{E}[\Lambda h(X_n)] \to \tilde{\mathbb{E}}[\Lambda h(X)], \qquad n \to +\infty.$$

This convergence is obviously stronger than convergence in law, which we denote by "$\Longrightarrow$." Stable convergence was introduced in [25] and notably investigated in [1]. The following lemma is a basic result on stable convergence that will be useful throughout the paper. We refer to [20], Lemma 2.1 for a proof. Here, E and F will denote two Polish spaces. We consider a sequence $(X_n)_{n\geq 1}$ of E-valued random variables defined on $(\Omega, \mathcal{F})$.

LEMMA 2.1. *Let $(Y_n)_{n\geq 1}$ be a sequence of F-valued random variables defined on $(\Omega, \mathcal{F})$, satisfying*

$$Y_n \xrightarrow{\mathbb{P}} Y,$$

where Y is defined on $(\Omega, \mathcal{F})$. If $X_n \overset{\text{stably}}{\Longrightarrow} X$ where X is defined on an extension of $(\Omega, \mathcal{F})$, then we have

$$(X_n, Y_n) \overset{\text{stably}}{\Longrightarrow} (X, Y).$$

Let us note that this result remains valid when $Y_n = Y$, for all $n \geq 1$.

2.1.2. *Application*: *Euler–Maruyama discretization of diffusion processes*. We illustrate this notion by the Euler–Maruyama discretization scheme of a diffusion process X solution of an SDE. The following results will be useful in the sequel in order to illustrate multi-level SA methods. We first introduce some notation, namely for $x \in \mathbb{R}^q$,

$$f(x) = \begin{pmatrix} b_1(x) & \sigma_{11}(x) & \cdots & \sigma_{1q'}(x) \\ b_2(x) & \sigma_{21}(x) & \cdots & \sigma_{2q'}(x) \\ \vdots & \vdots & \cdots & \vdots \\ b_q(x) & \sigma_{q1}(x) & \cdots & \sigma_{qq'}(x) \end{pmatrix}$$

and $dY_t = (dt \;\; dW_t^1 \;\; \cdots \;\; dW_t^{q'})^T$ where $b : \mathbb{R}^q \to \mathbb{R}^q$, $\sigma : \mathbb{R}^q \to \mathbb{R}^q \times \mathbb{R}^{q'}$. Here, as below, u^T denotes the transpose of the vector u. The dynamic of X will be written in the compact form

$$\forall t \in [0, T], \qquad X_t = x + \int_0^t f(X_s)\, dY_s$$

with its Euler–Maruyama scheme with time step $\Delta = T/n$, $t_i = i\Delta$, $i = 0, \ldots, n$, $\phi_n(s) = \sup\{t_i : t_i \leq s\}$

$$X_t^n = x + \int_0^t f\big(X^n_{\phi_n(s)}\big)\, dY_s.$$

We introduce the following smoothness assumption on the coefficients:

(HS) The coefficients b, σ are uniformly Lipschitz continuous.
(HD) The coefficients b, σ satisfy (HS) and are continuously differentiable.

The following result is due to [20], Theorem 3.2, page 276 and Theorem 5.5, page 293.

THEOREM 2.1. *Assume that* (HD) *holds. Then the process* $V^n := X^n - X$ *satisfies*

$$\sqrt{\frac{n}{T}} V^n \overset{stably}{\Longrightarrow} V \qquad as\ n \to +\infty,$$

the process V *being defined by* $V_0 = 0$ *and*

$$dV_t^i = \sum_{j=1}^{q'+1} \sum_{k=1}^{q} f_k^{\prime ij}(X_t) \left[V_t^k \, dY_t^j - \sum_{\ell=1}^{q'+1} f^{k\ell}(X_t) \, dZ_t^{\ell j} \right], \tag{2.1}$$

where $f_k^{\prime ij}$ *is the kth partial derivative of* f^{ij} *and*

$$\forall (i,j) \in [\![2, q'+1]\!] \times [\![2, q'+1]\!],$$
$$Z_t^{ij} = \frac{1}{\sqrt{2}} \sum_{1 \le k, \ell \le q} \int_0^t \sigma^{ik}(X_s) \sigma^{j\ell}(X_s) \, dB_s^{k\ell},$$
$$\forall j \in [\![1, q'+1]\!], \qquad Z^{1j} = 0,$$
$$\forall i \in [\![1, q'+1]\!], \qquad Z^{i1} = 0,$$

where B *is a standard* $(q')^2$*-dimensional Brownian motion defined on an extension* $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t \ge 0}, \tilde{\mathbb{P}})$ *of* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \ge 0}, \mathbb{P})$ *and independent of* W.

We will also use the following result which is due to [6], Theorem 4.

THEOREM 2.2. *Let* $m \in \mathbb{N}^* \setminus \{1\}$. *Assume that* (HD) *holds. Then we have*

$$\sqrt{\frac{m^\ell}{(m-1)T}} \left(X^{m^\ell} - X^{m^{\ell-1}} \right) \overset{\text{stably}}{\Longrightarrow} V \qquad as\ \ell \to +\infty.$$

2.1.3. *On some basic results related to stochastic approximation*. We now turn our attention to SA. There are various theorems that guarantee the a.s. and/or L^p convergence of SA algorithms. We provide below a general result in order to derive the a.s. convergence of such procedures. It is also known as the *Robbins–Monro theorem* and covers most situations; see the remark below.

THEOREM 2.3 (Robbins–Monro theorem). *Let* $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$ *be a Borel function and* U *be an* $\mathbb{R}^q$*-valued random vector with law* μ. *Define*

$$\forall \theta \in \mathbb{R}^d, \qquad h(\theta) = \mathbb{E}[H(\theta, U)],$$

and denote by θ^* *the (unique) solution to* $h(\theta) = 0$. *Suppose that* h *is a continuous function that satisfies the mean-reverting assumption*

$$\forall \theta \in \mathbb{R}^d, \theta \neq \theta^*, \qquad \langle \theta - \theta^*, h(\theta) \rangle > 0. \tag{2.2}$$

Let $\gamma = (\gamma_p)_{p \geq 1}$ *be a sequence of gain parameters satisfying* (1.2). *Suppose that*

$$\forall \theta \in \mathbb{R}^d, \qquad \mathbb{E}\big[|H(\theta, U)|^2\big] \leq C\big(1 + |\theta - \theta^*|^2\big). \tag{2.3}$$

Let $(U_p)_{p \geq 1}$ *be an i.i.d. sequence of random vectors with common law* μ *and* θ_0 *a random vector independent of* $(U_p)_{p \geq 1}$ *satisfying* $\mathbb{E}[|\theta_0|^2] < +\infty$. *Then the recursive procedure defined by*

$$\theta_{p+1} = \theta_p - \gamma_{p+1} H(\theta_p, U_{p+1}), \qquad p \geq 0 \tag{2.4}$$

satisfies

$$\theta_p \xrightarrow{\text{a.s.}} \theta^* \qquad \text{as } p \to +\infty.$$

Let us point out that the Robbins–Monro theorem also covers the framework of stochastic gradient algorithms. Indeed, if the function h is the gradient of a convex potential L, namely $h = \nabla L$ where $L \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}_+)$, satisfying that ∇L is Lipschitz, $|\nabla L|^2 \leq C(1 + L)$ and $\lim_{|\theta| \to +\infty} L(\theta) = +\infty$, then $\operatorname{Argmin} L$ is nonempty, and according to the standard lemma $\theta \mapsto \frac{1}{2}|\theta - \theta^*|^2$, it is a Lyapunov function so that the sequence $(\theta_n)_{n \geq 1}$ defined by (2.4) converges a.s. to θ^*.

LEMMA 2.2. *Let* $L \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}_+)$ *be a convex function. Then*

$$\forall \theta, \theta' \in \mathbb{R}^d, \qquad \langle \nabla L(\theta) - \nabla L(\theta'), \theta - \theta' \rangle \geq 0.$$

Moreover, if $\operatorname{Argmin} L$ *is nonempty, then one has*

$$\forall \theta \in \mathbb{R}^d \setminus \operatorname{Argmin} L, \forall \theta^* \in \operatorname{Argmin} L, \qquad \langle \nabla L(\theta), \theta - \theta^* \rangle > 0.$$

Now, we provide a result on the weak rate of convergence of the SA algorithm. In standard situations, it is well known that a stochastic algorithm $(\theta_p)_{p \geq 1}$ converges to its target at a rate $\gamma_p^{-1/2}$. More precisely, the sequence $(\gamma_p^{-1/2}(\theta_p - \theta^*))_{p \geq 1}$ converges in distribution to some normal distribution with a covariance matrix based on $\mathbb{E}[H(\theta^*, U)H(\theta^*, U)^T]$ where U is the noise of the algorithm. The following result is due to [23] (see also [9], page 161, Theorem 4.III.5) and has the advantage to be local, in the sense that a CLT holds on the set of convergence of the algorithm to an equilibrium which makes possible a straightforward application to multi-target algorithms.

THEOREM 2.4. *Let* $\theta^* \in \{h = 0\}$. *Suppose that* h *is twice continuously differentiable in a neighborhood of* θ^* *and that* $Dh(\theta^*)$ *is a stable* $d \times d$ *matrix; that* *is, all its eigenvalues have strictly positive real parts. Assume that the function H satisfies the following regularity and growth control property*:

$$\theta \mapsto \mathbb{E}\big[H(\theta, U)H(\theta, U)^T\big] \qquad \text{is continuous on } \mathbb{R}^d,$$

$$\exists \varepsilon > 0 \text{ s.t.} \qquad \theta \mapsto \mathbb{E}\big[\big|H(\theta, U)\big|^{2+\varepsilon}\big] \qquad \text{is locally bounded on } \mathbb{R}^d.$$

Assume that the noise of the algorithm is not degenerated at the equilibrium; that is, $\Gamma(\theta^*) := \mathbb{E}[H(\theta^*, U)H(\theta^*, U)^T]$ is a positive definite deterministic matrix.

The step sequence of procedure (2.4) *is given by $\gamma_p = \gamma(p)$, $p \geq 1$, where γ is a positive function defined on $[0, +\infty[$ decreasing to zero. We assume that γ satisfies one of the following assumptions*:

- *γ varies regularly with exponent $(-a)$, $a \in [0, 1)$; that is, for any $x > 0$, $\lim_{t\to+\infty} \gamma(tx)/\gamma(t) = x^{-a}$. In this case, set $\zeta = 0$.*
- *For $t \geq 1$, $\gamma(t) = \gamma_0/t$ and γ_0 satisfy $2\mathcal{R}e(\lambda_{\min})\gamma_0 > 1$, where $\lambda_{\min}$ denotes the eigenvalue of $Dh(\theta^*)$ with the lowest real part. In this case, set $\zeta = 1/(2\gamma_0)$.*

Then, on the event $\{\theta_p \to \theta^*\}$, one has

$$\gamma(p)^{-1/2}(\theta_p - \theta^*) \Longrightarrow \mathcal{N}(0, \Sigma^*),$$

where $\Sigma^* := \int_0^\infty \exp(-s(Dh(\theta^*) - \zeta I_d))^T \Gamma(\theta^*) \exp(-s(Dh(\theta^*) - \zeta I_d))\, ds$.

REMARK 2.1. In SA theory it is also said that $-Dh(\theta^*)$ is a Hurwitz matrix; that is, all its eigenvalues have strictly negative real parts. The assumption on the step sequence $(\gamma_n)_{n\geq1}$ is quite general and includes polynomial step sequences. In practical situations, the above theorem is often applied to the usual gain $\gamma_p = \gamma(p) = \gamma_0 p^{-a}$, with $1/2 < a \leq 1$, which notably satisfies (1.2).

Hence we clearly see that the optimal weak rate of convergence is achieved by choosing $\gamma_p = \gamma_0/p$ with $2\mathcal{R}e(\lambda_{\min})\gamma_0 > 1$. However, the main drawback of this choice is that the constraint on γ_0 is difficult to handle in practical implementation. Moreover it is well known that in this case the asymptotic covariance matrix is not optimal; see, for example, [9] or [5], among others.

As mentioned in the Introduction, a solution consists of devising the original SA algorithm (2.4) with a slow decreasing step $\gamma = (\gamma_p)_{p\geq1}$, where γ varies regularly with exponent $(-a)$, $a \in (1/2, 1)$, and to simultaneously compute the empirical mean $(\bar{\theta}_p)_{p\geq1}$ of the sequence $(\theta_p)_{p\geq0}$ by setting

$$\bar{\theta}_p = \frac{\theta_0 + \theta_1 + \cdots + \theta_p}{p+1} = \bar{\theta}_{p-1} - \frac{1}{p+1}(\bar{\theta}_{p-1} - \theta_p). \tag{2.5}$$

The following result states the weak rate of convergence for the sequence $(\bar{\theta}_p)_{p\geq1}$. In particular, it shows that the optimal weak rate of convergence and the optimal asymptotic covariance matrix can be obtained without any condition on γ_0. For a proof, the reader may refer to [9], page 169.

THEOREM 2.5. *Let $\theta^* \in \{h = 0\}$. Suppose that h is twice continuously differentiable in a neighborhood of θ^* and that $Dh(\theta^*)$ is a stable $d \times d$ matrix; that is, all its eigenvalues have positive real parts. Assume that the function H satisfies the following regularity and growth control property*:

$$\theta \mapsto \mathbb{E}\big[H(\theta, U)H(\theta, U)^T\big] \qquad \textit{is continuous on } \mathbb{R}^d,$$

$$\exists b > 0 \textit{ s.t.} \qquad \theta \mapsto \mathbb{E}\big[\big|H(\theta, U)\big|^{2+b}\big] \qquad \textit{is locally bounded on } \mathbb{R}^d.$$

Assume that the noise of the algorithm is not degenerated at the equilibrium; that is, $\Gamma(\theta^*) := \mathbb{E}[H(\theta^*, U)H(\theta^*, U)^T]$ is a positive definite deterministic matrix.

The step sequence of procedure (2.4) is given by $\gamma_p = \gamma(p)$, $p \geq 1$, where γ varies regularly with exponent $(-a)$, $a \in (1/2, 1)$. Then, on the event $\{\theta_p \to \theta^*\}$, one has

$$\sqrt{p}\big(\bar{\theta}_p - \theta^*\big) \Longrightarrow \mathcal{N}\big(0, Dh\big(\theta^*\big)^{-1}\Gamma\big(\theta^*\big)\big(Dh\big(\theta^*\big)^{-1}\big)^T\big).$$

2.2. *Main assumptions*. We list here the required assumptions in our framework to derive our asymptotic results and make some remarks.

(HWR1) There exists $\rho \in (0, 1/2]$,

$$n^{\rho}\big(U^n - U\big) \overset{\text{stably}}{\Longrightarrow} V \qquad \text{as } n \to +\infty,$$

where V is an $\mathbb{R}^q$-valued random variable eventually defined on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$.

(HWR2) There exists $\rho \in (0, 1/2]$,

$$m^{\ell\rho}\big(U^{m^{\ell}} - U^{m^{\ell-1}}\big) \overset{\text{stably}}{\Longrightarrow} V^m \qquad \text{as } \ell \to +\infty,$$

where V^m is an $\mathbb{R}^q$-valued random variable eventually defined on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ of $(\Omega, \mathcal{F}, \mathbb{P})$.

(HSR) There exists $\delta > 0$,

$$\sup_{n \geq 1} \mathbb{E}\big[\big|n^{\rho}\big(U^n - U\big)\big|^{2+\delta}\big] < +\infty.$$

(HR) There exists $b \in (0, 1]$,

$$\sup_{n \in \mathbb{N}^*, (\theta, \theta') \in (\mathbb{R}^d)^2} \frac{\mathbb{E}[|H(\theta, U^n) - H(\theta', U^n)|^2]}{|\theta - \theta'|^{2b}} < +\infty.$$

(HDH) For all $\theta \in \mathbb{R}^d$, $\mathbb{P}(U \notin \mathcal{D}_{H,\theta}) = 0$ with $\mathcal{D}_{H,\theta} := \{x \in \mathbb{R}^q : x \mapsto H(\theta, x) \text{ is differentiable at } x\}$.

(HLH) For all $(\theta, \theta', x) \in (\mathbb{R}^d)^2 \times \mathbb{R}^q$, $|H(\theta, x) - H(\theta', x)| \leq C(1 + |x|^r)|\theta - \theta'|$, for some $C, r > 0$.

(HI) There exists $\delta > 0$ such that for all $R > 0$, we have

$$\sup_{\{\theta : |\theta| \le R, n \in \mathbb{N}^*\}} \mathbb{E}\big[\big|H\big(\theta, U^n\big)\big|^{2+\delta}\big] < +\infty.$$

The sequence $(\theta \mapsto \mathbb{E}[H(\theta, U^n)H(\theta, U^n)^T])_{n\ge1}$ converges locally uniformly toward $\theta \mapsto \mathbb{E}[H(\theta, U)H(\theta, U)^T]$. The function $\theta \mapsto \mathbb{E}[H(\theta, U)H(\theta, U)^T]$ is continuous, and $\mathbb{E}[H(\theta^*, U)H(\theta^*, U)^T]$ is a positive deterministic matrix.

(HMR) There exists $\underline{\lambda} > 0$ such that $\forall n \ge 1$

$$\forall \theta \in \mathbb{R}^d, \qquad \big\langle \theta - \theta^{*,n}, h^n(\theta)\big\rangle \ge \underline{\lambda}\big|\theta - \theta^{*,n}\big|^2.$$

We will denote by λ_m the lowest real part of the eigenvalues of $Dh(\theta^*)$. We will assume that the step sequence is given by $\gamma_p = \gamma(p)$, $p \ge 1$, where γ is a positive function defined on $[0, +\infty[$ decreasing to zero and satisfying one of the following assumptions:

(HS1) γ varies regularly with exponent $(-a)$, $a \in [0, 1)$; that is, for any $x > 0$, $\lim_{t\to+\infty} \gamma(tx)/\gamma(t) = x^{-a}$.

(HS2) For $t \ge 1$, $\gamma(t) = \gamma_0/t$ and γ_0 satisfy $2\underline{\lambda}\gamma_0 > 1$.

REMARK 2.2. Assumption (HR) is trivially satisfied when $\theta \mapsto H(\theta, x)$ is Hölder-continuous with modulus having polynomial growth in x. However, it is also satisfied when H is less regular. For instance, it holds for $H(\theta, x) = \mathbf{1}_{\{x\le\theta\}}$ under the additional assumption that U^n has a bounded density (uniformly in n).

REMARK 2.3. Assumption (HMR) already appears in [9] and [5]; see also [13] and [10] in another context. It allows one to control the L^2-norm $\mathbb{E}[|\theta_p^n - \theta^{*,n}|^2]$ with respect to the step $\gamma(p)$ uniformly in n; see Lemma 5.2 in Section 5. As discussed in [22], Chapter 10, Section 5, if one considers the projected version of algorithm (1.3) on a bounded convex set D (e.g., a hyperrectangle $\prod_{i=1}^d [a_i, b_i]$) containing $\theta^{*,n}$, $\forall n \ge 1$, as very often happens from a practical point of view, this assumption can be localized on D; that is, it holds on D instead of $\mathbb{R}^d$. In this case, a sufficient condition is $\inf_{\theta\in D, n\in\mathbb{N}^*} \lambda_{\min}((Dh^n(\theta) + Dh^n(\theta)^T)/2) > 0$, where $\lambda_{\min}(A)$ denotes the lowest eigenvalue of the matrix A.

We also want to point out that if (HMR) holds, then one has $\lambda_m \ge \underline{\lambda}$. Indeed, writing $h^n(\theta) = \int_0^1 Dh^n(t\theta + (1-t)\theta^{*,n})(\theta - \theta^{*,n})\,dt$, for all $\theta \in \mathbb{R}^d$, we clearly have

$$\begin{aligned}
&\big\langle \theta - \theta^{*,n}, h^n(\theta)\big\rangle \\
&\quad = \int_0^1 \Big\langle \theta - \theta^{*,n}, \\
&\qquad\qquad \frac{Dh^n(t\theta + (1-t)\theta^{*,n}) + Dh^n(t\theta + (1-t)\theta^{*,n})^T}{2}\big(\theta - \theta^{*,n}\big)\Big\rangle\,dt \\
&\quad \ge \underline{\lambda}\big|\theta - \theta^{*,n}\big|^2.
\end{aligned}$$

Using the local uniform convergence of $(Dh^n)_{n\geq 1}$ and the convergence of $(\theta^{*,n})_{n\geq 1}$ toward θ^*, by passing to the limit $n \to +\infty$ in the above inequality, we obtain

$$\begin{aligned}&\int_0^1 \left\langle \theta - \theta^*, \frac{Dh(t\theta + (1-t)\theta^*) + Dh(t\theta + (1-t)\theta^*)^T}{2}(\theta - \theta^*)\right\rangle dt\\ &\qquad \geq \underline{\lambda}|\theta - \theta^*|^2 \qquad \forall \theta \in K,\end{aligned}$$

where K is a compact set such that $\theta^* + u_m \in K$, u_m being the eigenvector associated to the eigenvalue of $Dh(\theta^*)$ with the lowest real part. Hence, selecting $\theta = \theta^* + \varepsilon u_m$ in the previous inequality and passing to the limit $\varepsilon \to 0$, we get $\lambda_m \geq \underline{\lambda}$.

REMARK 2.4. Assumptions (HWR1), (HWR2) and (HSR) allow us to establish a CLT for the multi-level SA estimators presented in Sections 2.5 and 2.6. They include the case of the value at time T of an SDE, namely, $U = X_T$ approximated by its continuous Euler–Maruyama scheme $U^n = X_T^n$ with n time steps. Under (HD) one has $\rho = 1/2$. Moreover, U may depend on the whole path of an SDE. For instance, one may have $U = L_T$, the local time at level 0 of a one-dimensional continuous and adapted diffusion process, and the approximations may be given by

$$U^n = \sum_{i=1}^{[nt]} f\big(u_n X_{(i-1)/n}, \sqrt{n}(X_{i/n} - X_{(i-1)/n})\big).$$

Then under some assumptions on the function f and the coefficients b, σ, the weak and strong rate of convergence is $\rho = 1/4$; see [19] for more details. Let us note that we do not know what happens when $\rho > 1/2$, which includes the case of higher order schemes for discretization schemes of SDE.

2.3. *On the implicit weak error.* As we have already observed, the approximation of θ^* is affected by two errors: the *implicit discretization error* and the *statistical error*. Our first results concern the convergence of $\theta^{*,n}$ toward θ^* and its convergence rate as $n \to +\infty$. The proof of the next theorem is postponed to Section 3.1.

THEOREM 2.6. *For all $n \in \mathbb{N}^*$, assume that h and h^n satisfy the mean reverting assumption* (2.2) *of Theorem* 2.3. *Moreover, suppose that $(h^n)_{n\geq 1}$ converges locally uniformly toward h. Then one has*

$$\theta^{*,n} \to \theta^* \qquad \text{as } n \to +\infty.$$

Moreover, suppose that h and h^n, $n \geq 1$ are continuously differentiable and that $Dh(\theta^*)$ is nonsingular. Assume that $(Dh^n)_{n\geq 1}$ converges locally uniformly to Dh. If there exists $\alpha \in \mathbb{R}^*$ such that

$$\forall \theta \in \mathbb{R}^d, \qquad \lim_{n\to+\infty} n^{\alpha}\big(h^n(\theta) - h(\theta)\big) = \mathcal{E}(h, \alpha, \theta),$$

then one has

$$\lim_{n\to+\infty} n^{\alpha}(\theta^{*,n}-\theta^*) = -Dh^{-1}(\theta^*)\mathcal{E}(h,\alpha,\theta^*).$$

2.4. *On the optimal tradeoff between the implicit error and the statistical error.* Given the order of the implicit weak error, a natural question is how to find the optimal balance between the value of n in the approximation of U and the number M of steps in (1.3) for the computation of $\theta^{*,n}$ in order to achieve a given global error ε.

THEOREM 2.7. *Suppose that the assumptions of Theorem 2.6 are satisfied and that h satisfies the assumptions of Theorem 2.4. Assume that* (HR), (HI) *and* (HMR) *hold and that h^n is twice continuously differentiable with Dh^n Lipschitz continuous uniformly in n. If* (HS1) *or* (HS2) *is satisfied, then one has*

$$n^{\alpha}(\theta^n_{\gamma^{-1}(1/n^{2\alpha})}-\theta^*) \Longrightarrow -Dh^{-1}(\theta^*)\mathcal{E}(h,\alpha,\theta^*)+\mathcal{N}(0,\Sigma^*),$$

where

$$\begin{aligned}\Sigma^* := \int_0^{\infty} &\exp(-s(Dh(\theta^*)-\zeta I_d))^T\mathbb{E}[H(\theta^*,U)H(\theta^*,U)^T] \\ &\times \exp(-s(Dh(\theta^*)-\zeta I_d))\,ds\end{aligned} \tag{2.6}$$

with $\zeta = 0$ if (HS1) *holds and $\zeta = 1/2\gamma_0$ if* (HS2) *holds.*

The proof of the following lemma is carried out in Section 3.2

LEMMA 2.3. *Let $\delta > 0$. Under the assumptions of Theorem 2.7, one has*

$$n^{\alpha}(\theta^{n^\delta}_{\gamma^{-1}(1/n^{2\alpha})}-\theta^{*,n^\delta}) \Longrightarrow \mathcal{N}(0,\Sigma^*), \qquad n\to+\infty.$$

PROOF OF THEOREM 2.7. We decompose the error as follows:

$$\theta^n_{\gamma^{-1}(1/n^{2\alpha})}-\theta^* = \theta^n_{\gamma^{-1}(1/n^{2\alpha})}-\theta^{*,n}+\theta^{*,n}-\theta^*$$

and analyze each term of the above sum. By Lemma 2.3, we have

$$n^{\alpha}(\theta^n_{\gamma^{-1}(1/n^{2\alpha})}-\theta^{*,n}) \Longrightarrow \mathcal{N}(0,\Sigma^*),$$

and using Theorem 2.6, we also obtain

$$n^{\alpha}(\theta^{*,n}-\theta^*) \to -Dh^{-1}(\theta^*)\mathcal{E}(h,\alpha,\theta^*). \qquad \square$$

The result of Theorem 2.7 could be construed as follows. For a total error of order $1/n^{\alpha}$, it is necessary to achieve at least $M=\gamma^{-1}(1/n^{2\alpha})$ steps of the SA scheme defined by (1.3). Hence in this case, the complexity (or computational cost) of the algorithm is given by

$$C_{\mathrm{SA}}(\gamma) = C \times n \times \gamma^{-1}(1/n^{2\alpha}), \tag{2.7}$$

where C is some positive constant. We now investigate the impact of the step sequence $(\gamma_n)_{n\geq 1}$ on the complexity by considering the two following basic step sequences:

- if we choose $\gamma(p) = \gamma_0/p$ with $2\underline{\lambda}\gamma_0 > 1$, then $C_{\mathrm{SA}} = C \times n^{2\alpha+1}$;
- if we choose $\gamma(p) = \gamma_0/p^\rho$, $\frac{1}{2} < \rho < 1$, then $C_{\mathrm{SA}} = C \times n^{2\alpha/\rho+1}$.

Hence we clearly see that the minimal complexity is achieved by choosing $\gamma_p = \gamma_0/p$ with $2\underline{\lambda}\gamma_0 > 1$. In this latter case, we see that the computational cost is similar to the one achieved by the classical Monte Carlo algorithm for the computation of $\mathbb{E}_x[f(X_T)]$. However, the main drawback of this choice of step sequence comes from the constraint on γ_0. Our next result shows that the optimal complexity can be reached for free through the smoothing of procedure (1.3), according to the Ruppert–Polyak averaging principle.

THEOREM 2.8. *Suppose that the assumptions of Theorem 2.6 are satisfied and that h satisfies the assumptions of Theorem 2.4. Assume that* (HR), (HI) *and* (HMR) *hold and that h^n is twice continuously differentiable with Dh^n Lipschitz continuous uniformly in n. Define the empirical mean sequence $(\bar{\theta}_p^n)_{p\geq 1}$ of the sequence $(\theta_p^n)_{p\geq 1}$ by setting*

$$\bar{\theta}_p^n = \frac{\theta_0 + \theta_1^n + \cdots + \theta_p^n}{p+1} = \bar{\theta}_{p-1}^n - \frac{1}{p+1}(\bar{\theta}_{p-1}^n - \theta_p^n),$$

where the step sequence $\gamma = (\gamma_p)_{p\geq 1}$ satisfies (HS1) *with $\rho \in (1/2, 1)$. Then one has*

$$\begin{aligned} n^\alpha(\bar{\theta}_{n^{2\alpha}}^n - \theta^*) \Longrightarrow & -Dh^{-1}(\theta^*)\mathcal{E}(h, \alpha, \theta^*) \\ & + \mathcal{N}(0, Dh(\theta^*)^{-1}\mathbb{E}[H(\theta^*, U)H(\theta^*, U)^T](Dh(\theta^*)^{-1})^T). \end{aligned}$$

We omit the proof of the following lemma since it can be done in a similar manner to that of the SA literature. We refer to [11] for a proof.

LEMMA 2.4. *Let $\delta > 0$. Under the assumptions of Theorem 2.8, one has*

$$n^\alpha(\bar{\theta}_{n^{2\alpha}}^{n^\delta} - \theta^{*,n^\delta}) \Longrightarrow \mathcal{N}(0, Dh(\theta^*)^{-1}\mathbb{E}[H(\theta^*, U)H(\theta^*, U)^T](Dh(\theta^*)^{-1})^T), \quad n \to +\infty.$$

PROOF OF THEOREM 2.8. Similarly to the proof of Theorem 2.7, we decompose the error as follows:

$$\bar{\theta}^n_{n^{2\alpha}} - \theta^* = \bar{\theta}^n_{n^{2\alpha}} - \theta^{*,n} + \theta^{*,n} - \theta^*.$$

Applying successively Theorem 2.6 and Lemma 2.4, we obtain

$$n^\alpha(\bar{\theta}^n_{n^{2\alpha}} - \theta^*) \Longrightarrow -Dh^{-1}(\theta^*)\mathcal{E}(h, \alpha, \theta^*) + \mathcal{N}(0, \Sigma^*). \qquad \square$$

The result of Theorem 2.8 shows that for a total error of order $1/n^\alpha$, it is necessary to achieve at least $M = n^{2\alpha}$ steps of the SA scheme defined by (1.3) with step sequence satisfying (HS1) and to simultaneously compute its empirical mean, which represents a negligible part of the total cost. As a consequence, we see that in this case the complexity of the algorithm is given by

$$C_{\text{SA-RP}}(\gamma) = C \times n^{2\alpha+1}.$$

Therefore, the optimal complexity is reached for free, without any condition on γ_0, thanks to the Ruppert–Polyak averaging principle.

2.5. *The statistical Romberg stochastic approximation method.* In this section we present a two-level SA scheme that will be also referred to as the statistical Romberg SA method, which allows us to minimize the complexity of the SA algorithm $(\theta^n_p)_{p\in[\![0,\gamma^{-1}(1/n^{2\alpha})]\!]}$ for the numerical computation of θ^* solution to $h(\theta) = \mathbb{E}[H(\theta, U)] = 0$. It is clearly apparent that

$$\theta^{*,n} = \theta^{*,n^\beta} + \theta^{*,n} - \theta^{*,n^\beta}, \qquad \beta \in (0, 1).$$

The statistical Romberg SA scheme independently estimates each of the solutions appearing on the right-hand side in a way that minimizes the computational complexity. Let $\theta^{n^\beta}_{M_1}$ be an estimator of θ^{*,n^β} using M_1 independent samples of U^{n^β} and $\theta^n_{M_2} - \theta^{n^\beta}_{M_2}$ be an estimator of $\theta^{*,n} - \theta^{*,n^\beta}$ using M_2 independent copies of (U^{n^β}, U^n). Using the above decomposition, we estimate θ^* by the quantity

$$\Theta^{\text{sr}}_n = \theta^{n^\beta}_{M_1} + \theta^n_{M_2} - \theta^{n^\beta}_{M_2}.$$

It is important to point out here that the couple $(\theta^n_{M_2}, \theta^{n^\beta}_{M_2})$ is computed using i.i.d. copies of (U^{n^β}, U^n), the random variables U^{n^β} and U^n being perfectly correlated. Moreover, the random variables used to obtain $\theta^{n^\beta}_{M_1}$ are independent of those used for the computation of $(\theta^n_{M_2}, \theta^{n^\beta}_{M_2})$.

We also establish a central limit theorem for the statistical Romberg-based empirical sequence according to the Ruppert–Polyak averaging principle. It consists of estimating θ^* by

$$\bar{\Theta}^{\text{sr}}_n = \bar{\theta}^{n^\beta}_{M_3} + \bar{\theta}^n_{M_4} - \bar{\theta}^{n^\beta}_{M_4},$$

where $(\bar{\theta}_p^{n^\beta})_{p\in[\![0,M_3]\!]}$ and $(\bar{\theta}_p^n, \bar{\theta}_p^{n^\beta})_{p\in[\![0,M_4]\!]}$ are, respectively, the empirical means of the sequences $(\theta_p^{n^\beta})_{p\in[\![0,M_3]\!]}$ and $(\theta_p^n, \theta_p^{n^\beta})_{p\in[\![0,M_4]\!]}$ devised with the same slow decreasing step, that is, a step sequence $(\gamma(p))_{p\geq 1}$ where γ varies regularly with exponent $(-a)$, $a \in (1/2, 1)$.

THEOREM 2.9. *Suppose that h and h^n satisfy the assumptions of Theorem 2.6 with $\alpha \in (\rho \vee 2\rho\beta, 1]$ and that h satisfies the assumptions of Theorem 2.4. Assume that* (HWR1), (HSR), (HD), (HMR), (HDH) *and* (HLH) *hold and that h^n are twice continuously differentiable in a neighborhood of θ^*, with Dh^n Lipschitz-continuous uniformly in n, satisfying*

$$\forall \theta \in \mathbb{R}^d, \qquad n^\rho \| Dh^n(\theta) - Dh(\theta)\| \to 0 \qquad \text{as } n \to +\infty.$$

Suppose that $\tilde{\mathbb{E}}[(D_x H(\theta^*, U)V)(D_x H(\theta^*, U)V)^T]$ is a positive definite matrix. Assume that the step sequence is given by $\gamma_p = \gamma(p)$, $p \geq 1$, where γ is a positive function defined on $[0, +\infty[$ decreasing to zero, satisfying one of the following assumptions:

- *γ varies regularly with exponent $(-a)$, $a \in (1/2, 1)$; that is, for any $x > 0$, $\lim_{t\to+\infty} \gamma(tx)/\gamma(t) = x^{-a}$.*
- *For $t \geq 1$, $\gamma(t) = \gamma_0/t$ and γ_0 satisfy $\underline{\lambda}\gamma_0 > \alpha/(2\alpha - 2\rho\beta)$.*

Then, for $M_1 = \gamma^{-1}(1/n^{2\alpha})$ and $M_2 = \gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))$, one has

$$n^\alpha(\Theta_n^{\mathrm{sr}} - \theta^*) \Longrightarrow Dh^{-1}(\theta^*)\mathcal{E}(h, \alpha, \theta^*) + \mathcal{N}(0, \Sigma^*), \qquad n \to +\infty$$

with

$$\begin{aligned}\Sigma^* := \int_0^\infty \big(e^{-s(Dh(\theta^*)-\zeta I_d)}\big)^T \big(&\mathbb{E}\big[H(\theta^*, U)H(\theta^*, U)^T\big] \\ &+ \tilde{\mathbb{E}}\big[\big(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}\big[D_x H(\theta^*, U)V\big]\big) \\ &\quad \times \big(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}\big[D_x H(\theta^*, U)V\big]\big)^T\big]\big) \\ \times\, e^{-s(Dh(\theta^*)-\zeta I_d)}\, ds.&\end{aligned}$$

REMARK 2.5. Let us note that in the above theorem the condition on the convergence of the discretization error on the Jacobian matrix of h has been strengthened, compared to the standard SA algorithm appearing in Theorem 2.7. This is due to the presence of annoying second-order terms when we deal with the correction term $\theta_{M_2}^n - \theta_{M_2}^{n^\beta}$. Thus this assumption appears as a typical feature of multi-level SA methods; see also Theorems 2.10 and 2.11.

The proof of the following lemma is carried out in Section 3.3.

LEMMA 2.5. *Let $(\theta_p)_{p\geq 0}$ be the procedure defined for $p \geq 0$ by*

$$\theta_{p+1} = \theta_p - \gamma_{p+1} H\big(\theta_p, (U)^{p+1}\big), \tag{2.8}$$

where $((U^n)^p, (U)^p)_{p\geq 1}$ *is an i.i.d. sequence of random variables with the same law as* (U^n, U), $(\gamma_p)_{p\geq 1}$ *is the step sequence of the procedure* $(\theta_p^{n^\beta})_{p\geq 0}$ *and* $(\theta_p^n)_{p\geq 0}$ *and* θ_0 *is independent of the innovation satisfying* $\mathbb{E}[|\theta_0|^2] < +\infty$. *Under the assumptions of Theorem* 2.9, *one has*

$$n^\alpha\big(\theta^{n^\beta}_{\gamma^{-1}(1/(n^{2\alpha-\beta}))} - \theta_{\gamma^{-1}(1/(n^{2\alpha-\beta}))} - \big(\theta^{*,n^\beta} - \theta^*\big)\big) \Longrightarrow \mathcal{N}(0, \Theta^*), \qquad n \to +\infty,$$

with

$$\Theta^* := \int_0^\infty \big(e^{-s(Dh(\theta^*)-\zeta I_d)}\big)^T \tilde{\mathbb{E}}\big[\big(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V]\big) \times \big(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V]\big)^T\big] \times e^{-s(Dh(\theta^*)-\zeta I_d)}\, ds$$

and

$$n^\alpha\big(\theta^n_{\gamma^{-1}(1/(n^{2\alpha-\beta}))} - \theta_{\gamma^{-1}(1/(n^{2\alpha-\beta}))} - \big(\theta^{*,n} - \theta^*\big)\big) \xrightarrow{\mathbb{P}} 0, \qquad n \to +\infty.$$

PROOF OF THEOREM 2.9. We first write the following decomposition:

$$\Theta_n^{\mathrm{sr}} - \theta^* = \theta^{n^\beta}_{\gamma^{-1}(1/n^{2\alpha})} - \theta^{*,n^\beta} + \theta^n_{\gamma^{-1}(1/n^{2\alpha-2\rho\beta})} - \theta^{n^\beta}_{\gamma^{-1}(1/n^{2\alpha-2\rho\beta})} - \big(\theta^{*,n} - \theta^{*,n^\beta}\big) + \theta^{*,n} - \theta^*.$$

For the last term of the above sum, we use Theorem 2.6 to directly deduce

$$n^\alpha\big(\theta^{*,n} - \theta^*\big) \to -Dh^{-1}(\theta^*)\mathcal{E}(h, \alpha, \theta^*) \qquad \text{as } n \to +\infty.$$

For the first term, from Lemma 2.3 it follows that

$$n^\alpha\big(\theta^{n^\beta}_{\gamma^{-1}(1/n^{2\alpha})} - \theta^{*,n^\beta}\big) \Longrightarrow \mathcal{N}(0, \Gamma^*),$$

with

$$\Gamma^* := \int_0^\infty \exp\big(-s\big(Dh(\theta^*) - \zeta I_d\big)\big)^T \mathbb{E}\big[H(\theta^*, U)H(\theta^*, U)^T\big] \times \exp\big(-s\big(Dh(\theta^*) - \zeta I_d\big)\big)\, ds.$$

We decompose the last remaining term, namely $\theta^n_{\gamma^{-1}(1/n^{2\alpha-2\rho\beta})} - \theta^{n^\beta}_{\gamma^{-1}(1/n^{2\alpha-2\rho\beta})} - (\theta^{*,n} - \theta^{*,n^\beta})$, as follows:

$$\begin{aligned}
&\theta^n_{\gamma^{-1}(1/n^{2\alpha-2\rho\beta})} - \theta^{n^\beta}_{\gamma^{-1}(1/n^{2\alpha-2\rho\beta})} - \big(\theta^{*,n} - \theta^{*,n^\beta}\big) \\
&\quad = \theta^n_{\gamma^{-1}(1/n^{2\alpha-2\rho\beta})} - \theta_{\gamma^{-1}(1/n^{2\alpha-2\rho\beta})} - \big(\theta^{*,n} - \theta^*\big) \\
&\qquad - \big(\theta^{n^\beta}_{\gamma^{-1}(1/n^{2\alpha-2\rho\beta})} - \theta_{\gamma^{-1}(1/n^{2\alpha-2\rho\beta})} - \big(\theta^{*,n^\beta} - \theta^*\big)\big),
\end{aligned}$$

and we use Lemma 2.5 to complete the proof. $\square$

THEOREM 2.10. *Suppose that* h *and* h^n *satisfy the assumptions of Theorem* 2.6 (*with* $\alpha \in (\rho \vee 2\rho\beta, 1]$) *and that* h *satisfies the assumptions of Theorem* 2.4. *Assume that the step sequence* $\gamma = (\gamma_p)_{p\geq 1}$ *satisfies* (HS1) *with* $a \in (1/2, 1)$ *and* $a > \frac{\alpha}{2\alpha - 2\rho\beta} \vee \frac{\alpha(1-\beta)}{(\alpha-\rho\beta)}$. *Suppose that* (HWR1), (HSR), (HD), (HMR), (HDH) *and* (HLH) *hold and that* h^n *is twice continuously differentiable in a neighborhood of* θ^*, *with* Dh^n *Lipschitz-continuous uniformly in* n *satisfying*

$$\forall \theta \in \mathbb{R}^d, \qquad n^{\alpha-(\alpha-\rho\beta)a}\|Dh(\theta) - Dh^{n^\beta}(\theta)\| \to 0 \qquad \text{as } n \to +\infty. \tag{2.9}$$

Suppose that $\tilde{\mathbb{E}}[(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V])(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V])^T]$ *is a positive definite matrix. Then, for* $M_3 = n^{2\alpha}$ *and* $M_4 = n^{2\alpha-2\rho\beta}$, *one has*

$$n^\alpha(\bar{\Theta}_n^{\mathrm{sr}} - \theta^*) \Longrightarrow Dh^{-1}(\theta^*)\mathcal{E}(h, \alpha, \theta^*) + \mathcal{N}(0, \bar{\Sigma}^*), \qquad n \to +\infty,$$

where

$$\begin{aligned}\bar{\Sigma}^* := Dh(\theta^*)^{-1}\big(&\mathbb{E}\big[H(\theta^*, U)H(\theta^*, U)^T\big] \\ &+ \tilde{\mathbb{E}}\big[\big(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}\big[D_x H(\theta^*, U)V\big]\big) \\ &\quad \times \big(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}\big[D_x H(\theta^*, U)V\big]\big)^T\big]\big)\big(Dh(\theta^*)^{-1}\big)^T.\end{aligned}$$

We omit the proof of the following lemma since it can be done in a similar manner to that in the SA literature. For a proof the reader may refer to [11].

LEMMA 2.6. *Let* $(\bar{\theta}_p)_{p\geq 1}$ *be the empirical mean sequence associated to* $(\theta_p)_{p\geq 1}$ *defined by* (2.8). *Under the assumptions of Theorem* 2.10, *one has*

$$n^\alpha\big(\bar{\theta}^{n^\beta}_{n^{2\alpha-2\rho\beta}} - \bar{\theta}_{n^{2\alpha-2\rho\beta}} - \big(\theta^{*,n^\beta} - \theta^*\big)\big) \Longrightarrow \mathcal{N}(0, \bar{\Theta}^*)$$

with

$$\bar{\Theta}^* = Dh(\theta^*)^{-1}\tilde{\mathbb{E}}\big[\big(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}\big[D_x H(\theta^*, U)V\big]\big)^T\big]\big(Dh(\theta^*)^{-1}\big)^T$$

and

$$n^\alpha\big(\bar{\theta}^n_{n^{2\alpha-2\rho\beta}} - \bar{\theta}_{n^{2\alpha-2\rho\beta}} - \big(\theta^{*,n} - \theta^*\big)\big) \xrightarrow{\mathbb{P}} 0.$$

PROOF OF THEOREM 2.10. We decompose the error as follows:

$$\bar{\Theta}_n^{\mathrm{sr}} - \theta^* = \bar{\theta}^{n^\beta}_{n^{2\alpha}} - \theta^{*,n^\beta} + \bar{\theta}^n_{n^{2\alpha-2\rho\beta}} - \bar{\theta}^{n^\beta}_{n^{2\alpha-2\rho\beta}} - \big(\theta^{*,n} - \theta^{*,n^\beta}\big) + \theta^{*,n} - \theta^*.$$

For the first term, from Lemma 2.4 it follows that

$$n^\alpha\big(\bar{\theta}^{n^\beta}_{n^{2\alpha}} - \theta^{*,n^\beta}\big) \Longrightarrow \mathcal{N}\big(0, Dh(\theta^*)^{-1}\mathbb{E}\big[H(\theta^*, U)H(\theta^*, U)^T\big]\big(Dh(\theta^*)^{-1}\big)^T\big).$$

For the last term using Theorem 2.6, we have $n^{\alpha}(\theta^{*,n} - \theta^*) \to -Dh^{-1}(\theta^*) \times \mathcal{E}(h, \alpha, \theta^*)$. We now focus on the last remaining term, namely $\bar{\theta}^{n}_{n^{2\alpha-2\rho\beta}} - \bar{\theta}^{n^\beta}_{n^{2\alpha-2\rho\beta}} - (\theta^{*,n} - \theta^{*,n^\beta})$. We decompose it as follows:

$$\begin{aligned}
&\bar{\theta}^{n}_{n^{2\alpha-2\rho\beta}} - \bar{\theta}^{n^\beta}_{n^{2\alpha-2\rho\beta}} - (\theta^{*,n} - \theta^{*,n^\beta}) \\
&\qquad = \bar{\theta}^{n}_{n^{2\alpha-2\rho\beta}} - \bar{\theta}_{n^{2\alpha-2\rho\beta}} - (\theta^{*,n} - \theta^*) - (\bar{\theta}^{n^\beta}_{n^{2\alpha-2\rho\beta}} - \bar{\theta}_{n^{2\alpha-2\rho\beta}} - (\theta^{*,n^\beta} - \theta^*)),
\end{aligned}$$

where $(\bar{\theta}_p)_{p\geq 1}$ is the empirical mean sequence associated to $(\theta_p)_{p\geq 1}$, and we use Lemma 2.6 to complete the proof. $\square$

2.6. *The multi-level stochastic approximation method.* As we mentioned in the Introduction, the multi-level SA method uses $L+1$ stochastic schemes with a sequence of bias parameters $(m^\ell)_{\ell \in [\![0,L]\!]}$, for a fixed integer $m \geq 2$, that satisfies $m^L = n$, that is, $L = \log(n)/\log(m)$ and estimates θ^* by computing the quantity

$$\Theta_n^{\mathrm{ml}} = \theta^1_{M_0} + \sum_{\ell=1}^{L} (\theta^{m^\ell}_{M_\ell} - \theta^{m^{\ell-1}}_{M_\ell}).$$

It is important to point out here that for each level ℓ, the couple $(\theta^{m^\ell}_{M_\ell}, \theta^{m^{\ell-1}}_{M_\ell})$ is computed using i.i.d. copies of $(U^{m^{\ell-1}}, U^{m^\ell})$. Moreover the random variables $U^{m^{\ell-1}}$ and U^{m^ℓ} use two different bias parameter but are perfectly correlated. Finally, for two different levels, the SA schemes are based on independent samples.

THEOREM 2.11. *Suppose that h and $(h^n)_{n\in\mathbb{N}}$ satisfy the assumptions of Theorem 2.6. Assume that* (HWR2), (HSR), (HD), (HMR), (HDH) *and* (HLH) *hold and that h^n is twice continuously differentiable in a neighborhood of θ^*, with Dh^n Lipschitz-continuous uniformly in n. Suppose that $\tilde{\mathbb{E}}[(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V])(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V])^T]$ is a positive definite matrix. Assume that the step sequence is given by $\gamma_p = \gamma(p)$, $p \geq 1$, where γ is a positive function defined on $[0, +\infty[$ decreasing to zero, satisfying one of the following assumptions*:

- *γ varies regularly with exponent $(-a)$, $a \in (1/2, 1)$; that is, for any $x > 0$, $\lim_{t\to+\infty} \gamma(tx)/\gamma(t) = x^{-a}$.*
- *For $t \geq 1$, $\gamma(t) = \gamma_0/t$ and γ_0 satisfy $\underline{\lambda}\gamma_0 > 1$.*

Suppose that ρ satisfies one of the following assumptions:

- *If $\rho \in (0, 1/2)$, then assume that $\alpha > 2\rho$, $\underline{\lambda}\gamma_0 > \alpha/(\alpha - 2\rho)$ [if $\gamma(t) = \gamma_0/t$] and*

$$\exists \beta > \rho, \forall \theta \in \mathbb{R}^d, \qquad \sup_{n\geq 1} n^\beta \|Dh^n(\theta) - Dh(\theta)\| < +\infty.$$

In this case we set $M_0 = \gamma^{-1}(1/(n^{2\alpha}\log(n)))$ and $M_l = \gamma^{-1}(m^{\ell((1+2\rho)/2)} \times (m^{(1-2\rho)/2} - 1)/(n^{2\alpha}(n^{(1-2\rho)/2} - 1)))$, $\ell = 1, \ldots, L$.

- *If* $\rho = 1/2$, *then assume that* $\alpha = 1$, $\theta_0^{m^\ell} = \theta_0$, $\ell = 1, \ldots, L$, *with* $\mathbb{E}[|\theta_0|^2] < +\infty$ *and*

$$\exists \beta > 1/2, \forall \theta \in \mathbb{R}^d, \qquad \sup_{n \geq 1} n^\beta \| Dh^n(\theta) - Dh(\theta) \| < +\infty.$$

In this case we set $M_0 = \gamma^{-1}(1/(n^2 \log(n)))$ *and* $M_l = \gamma^{-1}(m^\ell \log(m)/(n^2 \log(n)(m-1)))$, $\ell = 1, \ldots, L$.

Then one has

$$n^\alpha(\Theta_n^{\mathrm{ml}} - \theta^*) \Longrightarrow -Dh^{-1}(\theta^*)\mathcal{E}(h, 1, \theta^*) + \mathcal{N}(0, \Sigma^*), \qquad n \to +\infty$$

with

$$\Sigma^* := \int_0^\infty (e^{-s(Dh(\theta^*)-\zeta I_d)})^T \tilde{\mathbb{E}}[(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V]) \times (D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V])^T] \times e^{-s(Dh(\theta^*)-\zeta I_d)}\, ds.$$

PROOF. We first write the following decomposition:

$$\Theta_n^{\mathrm{ml}} - \theta^* = \theta^1_{\gamma^{-1}(1/n^2)} - \theta^{*,1} + \sum_{\ell=1}^L (\theta_{M_\ell}^{m^\ell} - \theta_{M_\ell}^{m^{\ell-1}} - (\theta^{*,m^\ell} - \theta^{*,m^{\ell-1}})) + \theta^{*,n} - \theta^*.$$

For the last term of the above sum, we use Theorem 2.6 to directly deduce

$$n^\alpha(\theta^{*,n} - \theta^*) \to -Dh^{-1}(\theta^*)\mathcal{E}(h, 1, \theta^*) \qquad \text{as } n \to +\infty.$$

For the first term, from Lemma 5.2 we get

$$n^\alpha(\theta^1_{\gamma^{-1}(1/(n^{2\alpha}\log(n)))} - \theta^{*,1}) \xrightarrow{\mathbb{P}} 0.$$

Finally to deal with the last remaining term, namely $n^\alpha \sum_{\ell=1}^L (\theta_{M_\ell}^{m^\ell} - \theta_{M_\ell}^{m^{\ell-1}} - (\theta^{*,m^\ell} - \theta^{*,m^{\ell-1}}))$, we will need the following lemma, whose proof is carried out in Section 3.4. □

LEMMA 2.7. *Under the assumptions of Theorem* 2.9, *one has*

$$n^\alpha \sum_{\ell=1}^L (\theta_{M_\ell}^{m^\ell} - \theta_{M_\ell}^{m^{\ell-1}} - (\theta^{*,m^\ell} - \theta^{*,m^{\ell-1}})) \Longrightarrow \mathcal{N}(0, \Theta^*), \qquad n \to +\infty,$$

with

$$\Theta^* := \int_0^\infty \left(e^{-s(Dh(\theta^*)-\zeta I_d)}\right)^T \tag{2.10}$$

$$\times \tilde{\mathbb{E}}\left[\left(D_x H(\theta^*, U)V^m - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V^m]\right)\right.$$

$$\left.\times \left(D_x H(\theta^*, U)V^m - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V^m]\right)^T\right] \tag{2.11}$$

$$\times e^{-s(Dh(\theta^*)-\zeta I_d)}\,ds.$$

REMARK 2.6. The value of M_0 in Theorem 2.11 seems arbitrary and is asymptotically suboptimal. Indeed, one observes that the key point is to choose M_0 such that $\gamma(M_0) = o(1/n^{2\alpha})$ and $M_0 = \mathcal{O}(\gamma^{-1}(1/n^{2\alpha+1-2\rho}))$ if $\rho \in (0, 1/2)$ or $M_0 = \mathcal{O}(\gamma^{-1}(1/(n^2\log(n)))$ if $\rho = 1/2$ so that $n^\alpha(\theta^1_{M_0} - \theta^{*,1}) \xrightarrow{\mathbb{P}} 0$, and M_0 does not influence the asymptotic complexity; see Section 2.7 below. From an asymptotic point of view, any choice satisfying this condition would lead to the same complexity. From a nonasymptotic point of view, one may choose $M_0 = \gamma^{-1}(1/n^{2\alpha})$ so that the nonasymptotic computational complexity is smaller than in the other case. However, now one has $n^\alpha(\theta^1_{\gamma^{-1}(1/n^{2\alpha})} - \theta^{*,1}) \Longrightarrow \mathcal{N}(0, \Gamma^*)$ with

$$\Gamma^* := \int_0^\infty \exp\left(-s\left(Dh(\theta^*) - \zeta I_d\right)\right)^T \mathbb{E}\left[H(\theta^*, U^1)H(\theta^*, U^1)^T\right]$$

$$\times \exp\left(-s\left(Dh(\theta^*) - \zeta I_d\right)\right) ds$$

so that the new asymptotic covariance matrix $\Sigma^* := \Theta^* + \Gamma^*$ is higher than that of Theorem 2.11.

REMARK 2.7. The previous result shows that a CLT for the multi-level stochastic approximation estimator of θ^* holds if the standard weak error (and thus the implicit weak error) is of order $1/n^\alpha$, and the strong rate error is of order $1/n^\rho$ with $\alpha > 2\rho$ or $\alpha = 1$ and $\rho = 1/2$. Due to the nonlinearity of the procedures, which leads to annoying remainder terms in the Taylor expansions, these results do not seem to easily extend to a weak discretization error of order $1/n^\alpha$ with $\alpha < 1$ and $\rho = 1/2$ or a faster strong convergence rate $\rho > 1/2$. Moreover, for the same reason, this result does not seem to extend to the empirical sequence associated to the multi-level estimator, according to the Ruppert–Polyak averaging principle.

2.7. *Complexity analysis*. The result of Theorem 2.9 can be interpreted as follows. For a total error of order $1/n^\alpha$, it is necessary to set $M_1 = \gamma^{-1}(1/n^{2\alpha})$ steps of a stochastic algorithm with time step n^β and $M_2 = \gamma^{-1}(1/n^{2\alpha-2\rho\beta})$ steps of two stochastic algorithms with time step n and n^β using the same Brownian motion, the samples used for the first M_1 steps being independent of those used for the second scheme. Hence the complexity of the statistical Romberg SA method is given by

$$C_{\text{SR-SA}}(\gamma) = C \times \left(n^\beta \gamma^{-1}(1/n^{2\alpha}) + (n + n^\beta)\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))\right) \tag{2.12}$$

under the constraint $\alpha > 2\rho\beta \vee \rho$. Consequently, concerning the impact of the step sequence $(\gamma_n)_{n\geq 1}$ on the complexity of the procedure, we have the two following cases:

- If we choose $\gamma(p) = \gamma_0/p$, then simple computations show that $\beta^* = 1/(1 + 2\rho)$ is the optimal choice leading to a complexity
$$C_{\text{SR-SA}}(\gamma) = C' n^{2\alpha+1/(1+2\rho)},$$
under the constraint $\underline{\lambda}\gamma_0 > \alpha(1+2\rho)/(2\alpha(1+2\rho)-2\rho)$ and $\alpha > 2\rho/(1+2\rho)$. Let us note that this computational cost is similar to the one achieved by the statistical Romberg Monte Carlo method for the computation of $\mathbb{E}_x[f(X_T)]$.
- If we choose $\gamma(p) = \gamma_0/p^a$, $\frac{1}{2} < a < 1$, then the computational cost is given by
$$C_{\text{SR-SA}}(\gamma) = C'\big(n^{(2\alpha/a)+\beta} + n^{(2\alpha/a)-(\beta/a)+1}\big),$$
which is minimized for $\beta^* = a/(2\rho + a)$, leading to an optimal complexity,
$$C_{\text{SR-SA}}(\gamma) = C' n^{(2\alpha/a)+(a/(2\rho+a))},$$
under the constraint $\alpha > 2\rho a/(a + 2\rho) \vee \rho$. Observe that this complexity decreases with respect to a and that it is minimal for $a \to 1$, leading to the optimal computational cost obtained in the previous case. Let us also point out that contrary to the case $\gamma(p) = \gamma_0/p$, $p \geq 1$, there is no constraint on the choice of γ_0. Moreover, such a condition is difficult to handle in practical implementation, so a blind choice often has to be made.

The CLT proved in Theorem 2.10 shows that for a total error of order $1/n^\alpha$, it is necessary to set $M_1 = n^{2\alpha}$, $M_2 = n^{2\alpha-2\rho\beta}$ and to simultaneously compute its empirical mean, which represents a negligible part of the total cost. Both SA algorithm are devised with a step γ satisfying (HS1) with $a \in (1/2, 1)$ and $a > \frac{\alpha}{2\alpha-2\rho\beta} \vee \frac{\alpha(1-\beta)}{\alpha-\rho\beta}$. It is plain to see that $\beta^* = 1/(1 + 2\rho)$ is the optimal choice leading to a complexity given by
$$C_{\text{SR-RP}}(\gamma) = C \times n^{2\alpha+(1/(1+2\rho))},$$
provided that $a > \frac{\alpha(1+2\rho)}{2\alpha+2\rho(2\alpha-1)}$ and $\forall\theta \in \mathbb{R}^d, n^{\alpha-(\alpha-(\rho/(1+2\rho)))a}\|Dh(\theta) - Dh^{n^{1/(1+2\rho)}}(\theta)\| \to 0$ as $n \to +\infty$. (Note that when $a \to 1$, this condition is the same as in Theorem 2.9.) For instance, if $\alpha = 1$ and $\rho = 1/2$, then this condition writes $a > 2/3$ and $n^{1-(3/4)a}\|Dh(\theta) - Dh^{n^{1/2}}(\theta)\| \to 0$, and a should be selected sufficiently close to 1, according to the weak discretization error of the Jacobian matrix of h. Therefore, the optimal complexity is reached for free without any condition on γ_0, thanks to the Ruppert–Polyak averaging principle. Let us also note that although we do not intend to develop this point, it is possible to prove that averaging allows us to achieve the optimal asymptotic covariance matrix, as with standard SA algorithms.

Finally, concerning the CLT provided in Theorem 2.11, we show that in order to obtain an error of order $1/n^\alpha$, one has to set $M_0 = \gamma^{-1}(1/(n^{2\alpha}\log(n)))$ and $M_l = \gamma^{-1}(m^{\ell(1+2\rho)/2}(m^{(1-2\rho)/2} - 1)/(n^{2\alpha}(n^{(1-2\rho)/2} - 1)))$, if $\rho \in (0, 1/2)$ or $M_0 =$

$\gamma^{-1}(1/(n^2\log(n)))$ and $M_l = \gamma^{-1}(m^\ell \log(m)/(n^2\log(n)(m-1)))$ if $\rho = 1/2$, $\ell = 1, \dots, L$ with $L = \log(n)/\log(m)$. In both cases the complexity of the multi-level SA method is given by

$$C_{\text{ML-SA}}(\gamma) = C \times \left(\gamma^{-1}\big(1/\big(n^{2\alpha}\log(n)\big)\big) + \sum_{\ell=1}^{L} M_\ell\big(m^\ell + m^{\ell-1}\big)\right). \tag{2.13}$$

As for the Statistical Romberg SA method, we distinguish the two following cases:

- If $\gamma(p) = \gamma_0/p$, then the optimal complexity is given by

$$\begin{aligned} C_{\text{ML-SA}}(\gamma) &= C\Bigg(n^{2\alpha}\log(n) \\ &\qquad + \frac{n^2(n^{(1-2\rho)/2}-1)}{m^{(1-2\rho)/2}-1}\sum_{\ell=1}^{L} m^{-((1+2\rho)/2)\ell}\big(m^\ell + m^{\ell-1}\big)\Bigg) \\ &= \mathcal{O}\big(n^{2\alpha}n^{1-2\rho}\big), \end{aligned}$$

  if $\rho \in (0, 1/2)$ under the constraint $\underline{\lambda}\gamma_0 > \alpha/(\alpha - 2\rho)$ and

$$C_{\text{ML-SA}}(\gamma) = C\left(n^2\log(n) + n^2(\log n)^2\frac{m^2-1}{m(\log m)^2}\right) = \mathcal{O}\big(n^2\big(\log(n)\big)^2\big),$$

  if $\rho = 1/2$ under the constraint $\underline{\lambda}\gamma_0 > 1$. These computational costs are similar to those achieved by the multi-level Monte Carlo method for the computation of $\mathbb{E}_x[f(X_T)]$; see [15] and [6]. As discussed in [15], this complexity attains a minimum near $m = 7$.
- If we choose $\gamma(p) = \gamma_0/p^a$, $\frac{1}{2} < a < 1$, then simple computations show that the computational cost is given by

$$\begin{aligned} C_{\text{ML-SA}}(\gamma) &= C\Bigg(n^{2\alpha/a}\log^{1/a}(n) \\ &\qquad + n^{2/a}\big(n^{1-2\rho}-1\big)^{1/a}\sum_{\ell=1}^{L} m^{-((1+2\rho)/a)\ell}\big(m^\ell + m^{\ell-1}\big)\Bigg) \\ &= \mathcal{O}\big(n^{2\alpha/a}n^{(1-2\rho)/a}\big), \end{aligned}$$

  if $\rho \in (0, 1/2)$ and

$$\begin{aligned} C_{\text{ML-SA}}(\gamma) &= C\Bigg(n^{2/a}\log^{1/a}(n) \\ &\qquad + n^{2/a}(\log n)^{1/a}\frac{(m-1)^{1/a}(m+1)}{m(\log m)^{1/a}}\sum_{\ell=1}^{L} m^{-\ell((1/a)-1)}\Bigg) \\ &= \mathcal{O}\big(n^{2/a}(\log n)^{1/a}\big) \end{aligned}$$

if $\rho = 1/2$. Observe that once again these computational costs decrease with respect to a and that they are minimal for $a \to 1$ leading to the optimal computational cost obtained in the previous case. In this last case, the optimal choice for the parameter m depends on the value of a.

3. Proofs of main results.

3.1. *Proof of Theorem* 2.6. We first prove that $\theta^{*,n} \to \theta^*$, $n \to +\infty$. Let $\epsilon > 0$. The mean-reverting assumption (2.2) and the continuity of $u \mapsto \langle u, h(\theta^* + \epsilon u)\rangle$ on the (compact) set $\mathcal{S}_d := \{u \in \mathbb{R}^d, |u| = 1\}$ yield

$$\eta := \inf_{u \in \mathcal{S}_d} \langle u, h(\theta^* + \epsilon u)\rangle > 0.$$

The local uniform convergence of $(h^n)_{n\geq 1}$ implies

$$\exists n_\eta \in \mathbb{N}^*, \forall n \geq n_\eta, \qquad \theta \in \bar{B}(\theta^*, \epsilon) \quad \Rightarrow \quad |h^n(\theta) - h(\theta)| \leq \eta/2.$$

Then using the decomposition

$$\langle \theta - \theta^*, h^n(\theta)\rangle = \langle \theta - \theta^*, h(\theta)\rangle + \langle \theta - \theta^*, h^n(\theta) - h(\theta)\rangle,$$

one has for $\theta = \theta^* \pm \epsilon u$, $u \in \mathcal{S}_d$,

$$\begin{aligned} \epsilon\langle u, h^n(\theta^* + \epsilon u)\rangle &\geq \langle \epsilon u, h(\theta^* + \epsilon u)\rangle - \epsilon\eta/2 \geq \epsilon\eta - \epsilon\eta/2 = \epsilon\eta/2, \\ -\epsilon\langle u, h^n(\theta^* - \epsilon u)\rangle &\geq \langle -\epsilon u, h(\theta^* - \epsilon u)\rangle - \epsilon\eta/2 \geq \epsilon\eta - \epsilon\eta/2 = \epsilon\eta/2 \end{aligned}$$

so that $\langle u, h^n(\theta^* + \epsilon u)\rangle > 0$ and $\langle u, h^n(\theta^* - \epsilon u)\rangle < 0$, which combined with the intermediate value theorem applied to the continuous function $x \mapsto \langle u, h^n(\theta^* + xu)\rangle$ on the interval $[-\epsilon, \epsilon]$ yields

$$\langle u, h^n(\theta^* + \tilde{x}u)\rangle = 0$$

for some $\tilde{x} = \tilde{x}(u) \in]-\epsilon, \epsilon[$. Now we set $u = \theta^* - \theta^{*,n}/|\theta^* - \theta^{*,n}|$ as soon as possible. (Otherwise, the proof is complete.) Hence there exists $x^* \in]-\epsilon, \epsilon[$ such that

$$\left\langle \frac{\theta^* - \theta^{*,n}}{|\theta^* - \theta^{*,n}|}, h^n\left(\theta^* + x^* \frac{\theta^* - \theta^{*,n}}{|\theta^* - \theta^{*,n}|}\right)\right\rangle = 0$$

so that multiplying the previous equality by $x^* + |\theta^* - \theta^{*,n}|$, we get

$$\begin{aligned} &\left\langle \theta^{*,n} + \left(\frac{x^*}{|\theta^* - \theta^{*,n}|} + 1\right)(\theta^* - \theta^{*,n}) - \theta^{*,n}, \right. \\ &\qquad \left. h^n\left(\theta^{*,n} + \left(\frac{x^*}{|\theta^* - \theta^{*,n}|} + 1\right)(\theta^* - \theta^{*,n})\right)\right\rangle = 0. \end{aligned}$$

Consequently, by the very definition of $\theta^{*,n}$, we deduce that $x^* = -|\theta^* - \theta^{*,n}|$ and finally $|\theta^* - \theta^{*,n}| < \epsilon$ for $n \geq n_\eta$. Hence we conclude that $\theta^{*,n} \to \theta^*$. We now derive a convergence rate. A Taylor expansion yields for all $n \geq 1$,

$$h^n(\theta^*) = h^n(\theta^{*,n}) + \left(\int_0^1 Dh^n(\lambda\theta^{*,n} + (1-\lambda)\theta^*)\,d\lambda\right)(\theta^* - \theta^{*,n}).$$

Combining the local uniform convergence of $(Dh^n)_{n\geq 1}$ to Dh, the convergence of $(\theta^{*,n})_{n\geq 1}$ to θ^* and the nonsingularity of $Dh(\theta^*)$, one clearly gets that for n large enough, $\int_0^1 Dh^n(\lambda\theta^{*,n} + (1-\lambda)\theta^*)\,d\lambda$ is nonsingular and that

$$\left(\int_0^1 Dh^n(\lambda\theta^{*,n} + (1-\lambda)\theta^*)\,d\lambda\right)^{-1} \to Dh^{-1}(\theta^*), \qquad n \to +\infty.$$

Consequently, recalling that $h(\theta^*) = 0$ and $h^n(\theta^{*,n}) = 0$, it is plain to see

$$\begin{aligned} n^\alpha(\theta^{*,n} - \theta^*) &= -\left(\int_0^1 Dh^n(\lambda\theta^{*,n} + (1-\lambda)\theta^*)\,d\lambda\right)^{-1} n^\alpha(h^n(\theta^*) - h(\theta^*)) \\ &\to -Dh^{-1}(\theta^*)\mathcal{E}(h,\alpha,\theta^*). \end{aligned}$$

3.2. *Proofs of Lemma* 2.3. We define for all $p \geq 1$, $\Delta M_p^{n^\delta} := h^{n^\delta}(\theta_{p-1}^{n^\delta}) - H(\theta_{p-1}^{n^\delta}, (U^{n^\delta})^p) = \mathbb{E}[H(\theta_{p-1}^{n^\delta}, (U^{n^\delta})^p)|\mathcal{F}_{p-1}] - H(\theta_{p-1}^{n^\delta}, (U^{n^\delta})^p)$. Recalling that $((U^{n^\delta})^p)_{p\geq 1}$ is a sequence of i.i.d. random variables, we have that $(\Delta M_p^{n^\delta})_{p\geq 1}$ is a sequence of martingale increments w.r.t. the natural filtration $\mathcal{F} := (\mathcal{F}_p := \sigma(\theta_0^{n^\delta}, (U^{n^\delta})^1, \dots, (U^{n^\delta})^p); p \geq 1)$. From the dynamic (1.3), one clearly gets for $p \geq 0$,

$$\begin{aligned} \theta_{p+1}^{n^\delta} - \theta^{*,n^\delta} &= \theta_p^{n^\delta} - \theta^{*,n^\delta} - \gamma_{p+1} Dh^{n^\delta}(\theta^{*,n^\delta})(\theta_p^{n^\delta} - \theta^{*,n^\delta}) \\ &\quad + \gamma_{p+1}\Delta M_{p+1}^{n^\delta} + \gamma_{p+1}\zeta_p^{n^\delta} \end{aligned}$$

with $\zeta_p^{n^\delta} := Dh^{n^\delta}(\theta^{*,n^\delta})(\theta_p^{n^\delta} - \theta^{*,n^\delta}) - h^{n^\delta}(\theta_p^{n^\delta})$. Moreover, since Dh^{n^δ} is Lipschitz-continuous (uniformly in n), by Taylor's formula one gets $\zeta_p^{n^\delta} = \mathcal{O}(|\theta_p^{n^\delta} - \theta^{*,n^\delta}|^2)$. Hence by a simple induction, we obtain

$$\begin{aligned} &\theta_n^{n^\delta} - \theta^{*,n^\delta} \\ &\quad = \Pi_{1,n}(\theta_0^{n^\delta} - \theta^{*,n^\delta}) + \sum_{k=1}^n \gamma_k \Pi_{k+1,n}\Delta M_k^{n^\delta} \\ &\qquad + \sum_{k=1}^n \gamma_k \Pi_{k+1,n}(\zeta_{k-1}^{n^\delta} + (Dh(\theta^*) - Dh^{n^\delta}(\theta^{*,n^\delta}))(\theta_{k-1}^{n^\delta} - \theta^{*,n^\delta})), \end{aligned} \tag{3.1}$$

where $\Pi_{k,n} := \prod_{j=k}^n (I_d - \gamma_j Dh(\theta^*))$, with the convention that $\Pi_{n+1,n} = I_d$. We now investigate the asymptotic behavior of each term in the above decomposition. Actually, in steps 1 and 2 we will prove that the first and third terms on the right-hand side of the above equality converge in probability to zero at a faster rate than $n^{-\alpha}$. We will then prove in step 3 that the second term satisfies a CLT at rate n^{α}.

Step 1: *study of the sequence* $\{n^{\alpha}\Pi_{1,\gamma^{-1}(1/n^{2\alpha})}(\theta_0^{n^\delta}-\theta^{*,n^\delta}), n\geq 0\}$.

First, since $-Dh(\theta^*)$ is an Hurwitz matrix, $\forall\lambda\in[0,\lambda_m)$, there exists $C>0$ such that for any $k\leq n$, $\|\Pi_{k,n}\|\leq C\prod_{j=k}^n(1-\lambda\gamma_j)\leq C\exp(-\lambda\sum_{j=k}^n\gamma_j)$. We refer to [9] and [5] for more details. Hence one has for all $\eta\in(0,\lambda_m)$,

$$n^{\alpha}\mathbb{E}\big|\Pi_{1,\gamma^{-1}(1/n^{2\alpha})}\big(\theta_0^{n^\delta}-\theta^{*,n^\delta}\big)\big|$$
$$\leq C\Big(\sup_{n\geq 1}\mathbb{E}\big|\theta_0^n\big|+1\Big)n^{\alpha}\exp\Bigg(-(\lambda_m-\eta)\sum_{k=1}^{\gamma^{-1}(1/n^{2\alpha})}\gamma_k\Bigg).$$

Selecting η such that $2(\lambda_m-\eta)\gamma_0>2(\underline{\lambda}-\eta)\gamma_0>1$, under (HS2) and any $\eta\in(0,\lambda_m)$ under (HS1), we derive the convergence to zero of the right-hand side of the last but one inequality.

Step 2: *study of* $\{n^{\alpha}\sum_{k=1}^{\gamma^{-1}(1/n^{2\alpha})}\gamma_k\Pi_{k+1,\gamma^{-1}(1/n^{2\alpha})}(\zeta_{k-1}^{n^\delta}+(Dh(\theta^*)-Dh^{n^\delta}(\theta^{*,n^\delta}))(\theta_{k-1}^{n^\delta}-\theta^{*,n^\delta})), n\geq 0\}$.

We focus on the last term of (3.1). Using Lemma 5.2 we get

$$\mathbb{E}\Bigg|\sum_{k=1}^{n}\gamma_k\Pi_{k+1,n}\big(\zeta_{k-1}^{n^\delta}+\big(Dh(\theta^*)-Dh^{n^\delta}(\theta^{*,n^\delta})\big)\big(\theta_{k-1}^{n^\delta}-\theta^{*,n^\delta}\big)\big)\Bigg|$$
$$\leq C\sum_{k=1}^{n}\|\Pi_{k+1,n}\|\big(\gamma_k^2+\gamma_k^{3/2}\|Dh(\theta^*)-Dh^{n^\delta}(\theta^{*,n^\delta})\|\big),$$

so that by Lemma 5.1 (see also remark 2.3), the local uniform convergence of $(Dh^n)_{n\geq 1}$ and the continuity of Dh at θ^*, we derive

$$\limsup_{n} n^{\alpha}\mathbb{E}\Bigg|\sum_{k=1}^{\gamma^{-1}(1/n^{2\alpha})}\gamma_k\Pi_{k+1,\gamma^{-1}(1/n^{2\alpha})}$$
$$\times\big(\zeta_{k-1}^{n^\delta}+\big(Dh(\theta^*)-Dh^{n^\delta}(\theta^{*,n^\delta})\big)\big(\theta_{k-1}^{n^\delta}-\theta^{*,n^\delta}\big)\big)\Bigg|=0.$$

Step 3: *study of the sequence* $\{n^{\alpha}\sum_{k=1}^{\gamma^{-1}(1/n^{2\alpha})}\gamma_k\Pi_{k+1,\gamma^{-1}(1/n^{2\alpha})}\Delta M_k^{n^\delta}, n\geq 0\}$.

We use the following decomposition:

$$\sum_{k=1}^{n}\gamma_k\Pi_{k+1,n}\Delta M_k^{n^\delta}$$
$$=\sum_{k=1}^{n}\gamma_k\Pi_{k+1,n}\big(h^{n^\delta}(\theta_k^{n^\delta})-h^{n^\delta}(\theta^{*,n^\delta})\big)$$

$$
\begin{aligned}
&\qquad\qquad - \big(H\big(\theta_k^{n^\delta}, (U^{n^\delta})^{k+1}\big) - H\big(\theta^{*,n^\delta}, (U^{n^\delta})^{k+1}\big)\big)\big) \\
&\quad + \sum_{k=1}^{n} \gamma_k \Pi_{k+1,n}\big(h^{n^\delta}(\theta^{*,n^\delta}) - H\big(\theta^{*,n^\delta}, (U^{n^\delta})^{k+1}\big)\big) \\
&:= R_n + M_n.
\end{aligned}
$$

Now, using that $\mathbb{E}[H(\theta_k^{n^\delta}, (U^{n^\delta})^{k+1})|\mathcal{F}_k] = h^{n^\delta}(\theta_k^{n^\delta})$, $\mathbb{E}[H(\theta^{*,n^\delta}, (U^{n^\delta})^{k+1})|\mathcal{F}_k] = h^{n^\delta}(\theta^{*,n^\delta})$ and (HR), we have

$$
\mathbb{E}|R_n|^2 \leq \sum_{k=1}^{n} \gamma_k^2 \|\Pi_{k+1,n}\|^2 \mathbb{E}\big[\big|\theta_k^{n^\delta} - \theta^{*,n^\delta}\big|^{2a}\big] \leq \sum_{k=1}^{n} \gamma_k^{2+a} \|\Pi_{k+1,n}\|^2,
$$

where we use Lemma 5.2 and Jensen's inequality for the last inequality. Moreover, according to Lemma 5.1, we have

$$
\limsup_n n^{2\alpha} \sum_{k=1}^{\gamma^{-1}(1/n^{2\alpha})} \gamma_k^{2+a} \|\Pi_{k+1,\gamma^{-1}(1/n^{2\alpha})}\|^2 = 0
$$

so that $n^\alpha \sum_{k=1}^n \gamma_k \Pi_{k+1,n}(h^{n^\delta}(\theta_k^{n^\delta}) - h^{n^\delta}(\theta^{*,n^\delta}) - (H(\theta_k^{n^\delta}, (U^{n^\delta})^{k+1}) - H(\theta^{*,n^\delta}, (U^{n^\delta})^{k+1}))) \overset{L^2(\mathbb{P})}{\longrightarrow} 0$.

To conclude we prove that the sequence $\{\gamma^{-1/2}(n)M_n, n \geq 0\}$ satisfies a CLT. In order to do this we apply standard results on CLT for martingale arrays. More precisely, we will apply Theorem 3.2 and Corollary 3.1, page 58, in [17] so that we need to prove that the conditional Lindeberg assumption is satisfied, that is, $\lim_n \sum_{k=1}^n \mathbb{E}[|\gamma^{-1/2}(n)\gamma_k \Pi_{k+1,n}(h^{n^\delta}(\theta^{*,n^\delta}) - H(\theta^{*,n^\delta}, (U^{n^\delta})^{k+1}))|^p] = 0$, for some $p > 2$ and that the conditional variance $(S_n)_{n\geq 1}$ defined by

$$
\begin{aligned}
S_n &:= \frac{1}{\gamma(n)} \sum_{k=1}^{n} \gamma_k^2 \Pi_{k+1,n} \mathbb{E}_k\big[\big(h^{n^\delta}(\theta^{*,n^\delta}) - H\big(\theta^{*,n^\delta}, (U^{n^\delta})^{k+1}\big)\big) \\
&\qquad\qquad \times \big(h^{n^\delta}(\theta^{*,n^\delta}) - H\big(\theta^{*,n^\delta}, (U^{n^\delta})^{k+1}\big)\big)^T\big]\Pi_{k+1,n}^T \\
&= \frac{1}{\gamma(n)} \sum_{k=1}^{n} \gamma_k^2 \Pi_{k+1,n} \Gamma_n \Pi_{k+1,n}^T
\end{aligned}
$$

with $\Gamma_n := \mathbb{E}[H(\theta^{*,n^\delta}, U^{n^\delta})(H(\theta^{*,n^\delta}, U^{n^\delta}))^T]$, since $h^{n^\delta}(\theta^{*,n^\delta}) = 0$, satisfies $S_n \overset{\text{a.s.}}{\longrightarrow} \Sigma^*$ as $n \to +\infty$. We also set $\Gamma^* := \mathbb{E}[H(\theta^*, U)(H(\theta^*, U))^T]$.

By (HI), it holds for some $R > 0$ such that $\forall n \geq 1$, $\theta^{*,n} \in B(0, R)$

$$
\begin{aligned}
&\sum_{k=1}^{n} \mathbb{E}\big|\gamma^{-1/2}(n)\gamma_k \Pi_{k+1,n}\big(h^{n^\delta}(\theta^{*,n^\delta}) - H\big(\theta^{*,n^\delta}, (U^{n^\delta})^{k+1}\big)\big)\big|^{2+\delta} \\
&\qquad \leq C \sup_{\{\theta : |\theta| \leq R, n \in \mathbb{N}^*\}} \mathbb{E}\big[\big|H(\theta, U^n)\big|^{2+\delta}\big] \gamma^{-1+\delta/2}(n) \sum_{k=1}^{n} \gamma_k^{2+\delta} \|\Pi_{k+1,n}\|^{2+\delta}.
\end{aligned}
$$

By Lemma 5.1, we have $\limsup_n \gamma^{-1+\delta/2}(n)\sum_{k=1}^{n}\gamma_k^{2+\delta}\|\Pi_{k+1,n}\|^{2+\delta} \leq \limsup_n \gamma^{\delta/2}(n) = 0$, so that the conditional Lindeberg condition (see [17], Corollary 3.1) is satisfied. Now we focus on the conditional variance. By the local uniform convergence of $(\theta \mapsto \mathbb{E}[H(\theta, U^{n^\delta})(H(\theta, U^{n^\delta}))^T])_{n\geq 0}$, the continuity of $\theta \mapsto \mathbb{E}[H(\theta, U)(H(\theta, U))^T]$ at θ^* and since $\theta^{*,n^\delta} \to \theta^*$, we have $\Gamma_n \to \Gamma^*$, so that from Lemma 5.1, it follows that

$$\limsup_n \left\| \frac{1}{\gamma(n)} \sum_{k=1}^{n} \gamma_k^2 \Pi_{k+1,n}(\Gamma_n - \Gamma^*)\Pi_{k+1,n}^T \right\| \leq \limsup_n \|\Gamma_n - \Gamma^*\| = 0.$$

Hence we see that $\lim_n S_n = \lim_n \frac{1}{\gamma(n)}\sum_{k=1}^{n}\gamma_k^2\Pi_{k+1,n}\Gamma^*\Pi_{k+1,n}^T$ if this latter limit exists. Let us note that Σ^* given by (2.6) is the (unique) matrix A solution to the Lyapunov equation

$$\Gamma^* - (Dh(\theta^*) - \zeta I_d)A - A(Dh(\theta^*) - \zeta I_d)^T = 0.$$

We aim at proving that $S_n \xrightarrow{\text{a.s.}} \Sigma^*$. In order to do this, we define

$$A_{n+1} := \frac{1}{\gamma(n+1)} \sum_{k=1}^{n+1} \gamma_k^2 \Pi_{k+1,n}\Gamma^*\Pi_{k+1,n}^T,$$

which can be written in the following recursive form:

$$\begin{aligned} A_{n+1} &= \gamma_{n+1}\Gamma^* + \frac{\gamma_n}{\gamma_{n+1}}(I_d - \gamma_{n+1}Dh(\theta^*))A_n(I_d - \gamma_{n+1}Dh(\theta^*))^T \\ &= A_n + \gamma_n(\Gamma^* - Dh(\theta^*)A_n - A_n Dh(\theta^*)^T) \\ &\quad + (\gamma_{n+1} - \gamma_n)\Gamma^* + \gamma_n\gamma_{n+1}Dh(\theta^*)A_n Dh(\theta^*)^T + \frac{\gamma_n - \gamma_{n+1}}{\gamma_{n+1}}A_n. \end{aligned}$$

Under the assumptions made on the step sequence $(\gamma_n)_{n\geq 1}$, we have $\frac{\gamma_n - \gamma_{n+1}}{\gamma_{n+1}} = 2\zeta\gamma_n + o(\gamma_n)$ and $\gamma_{n+1} - \gamma_n = \mathcal{O}(\gamma_n^2)$. Consequently, introducing $Z_n = A_n - \Sigma^*$, simple computations from the previous equality yield

$$\begin{aligned} Z_{n+1} = Z_n &- \gamma_n((Dh(\theta^*) - \zeta I_d)Z_n + Z_n(Dh(\theta^*) - \zeta I_d)^T) \\ &+ \gamma_n\gamma_{n+1}Dh(\theta^*)Z_n Dh(\theta^*)^T + \left(\frac{\gamma_n - \gamma_{n+1}}{\gamma_{n+1}} - 2\zeta\gamma_n\right)Z_n \\ &+ \gamma_n\gamma_{n+1}Dh(\theta^*)\Sigma^* Dh(\theta^*)^T + (\gamma_{n+1} - \gamma_n)\Gamma^* \\ &+ \left(\frac{\gamma_n - \gamma_{n+1}}{\gamma_{n+1}} - 2\zeta\gamma_n\right)\Sigma^*. \end{aligned}$$

Let us note that by the very definition of ζ and assumptions (HS1) and (HS2), the matrix $Dh(\theta^*) - \zeta I_d$ is stable, so that taking the norm in the previous equality, there exists $\lambda > 0$ such that

$$\|Z_{n+1}\| \leq (1 - \lambda\gamma_n + o(\gamma_n))\|Z_n\| + o(\gamma_n)$$

for $n \geq n_0$, n_0 large enough. By a simple induction, it holds for $n \geq N \geq n_0$,

$$\|Z_n\| \leq C\|Z_N\| \exp(-\lambda s_{N,n}) + C \exp(-\lambda s_{N,n}) \sum_{k=N}^{n} \exp(\lambda s_{N,k}) \gamma_k \|e_k\|,$$

where $e_n = o(1)$, and we set $s_{N,n} := \sum_{k=N}^{n} \gamma_k$. From assumption (1.2), it follows that for $N \geq n_0$,

$$\limsup_n \|Z_n\| \leq C \sup_{k \geq N} \|e_k\|,$$

and passing to the limit as N goes to infinity, it clearly yields $\limsup_n \|Z_n\| = 0$. Hence $S_n \xrightarrow{\text{a.s.}} \Theta^*$, and the proof is complete.

3.3. *Proof of Lemma* 2.5. We will just prove the first assertion of the lemma; the second will readily follow. When the exact value of a constant is not important, we may repeat the same symbol for constants that may change from one line to next. We come back to the decomposition used in the proof of Lemma 2.3, and we consequently use the same notation. Let us note that the procedure $(\theta_p)_{p \geq 0}$ a.s. converges to θ^* and satisfies a CLT according to Theorem 2.4.

From the dynamics of $(\theta_p^{n^\delta})_{p \geq 0}$ and $(\theta_p)_{p \geq 0}$, we write for $p \geq 0$,

$$\begin{aligned}\theta_{p+1}^{n^\beta} - \theta^{*,n^\beta} &= \theta_p^{n^\beta} - \theta^{*,n^\beta} - \gamma_{p+1} Dh^{n^\beta}(\theta^{*,n^\beta})(\theta_p^{n^\beta} - \theta^{*,n^\beta}) \\ &\quad + \gamma_{p+1} \Delta M_{p+1}^n + \gamma_{p+1} \zeta_p^{n^\beta}, \\ \theta_{p+1} - \theta^* &= \theta_p - \theta^* - \gamma_{p+1} Dh(\theta^*)(\theta_p - \theta^*) + \gamma_{p+1} \Delta M_{p+1} + \gamma_{p+1} \zeta_p,\end{aligned}$$

with $\Delta M_{p+1} = h(\theta_p) - H(\theta_p, (U)^{p+1})$, $p \geq 0$ and $\zeta_p^{n^\beta} := Dh^{n^\beta}(\theta^{*,n^\beta})(\theta_p^{n^\beta} - \theta^{*,n^\beta}) - h^{n^\beta}(\theta_p^{n^\beta})$, $\zeta_p = Dh(\theta^*)(\theta_p - \theta^*) - h(\theta_p)$. Since Dh^n and Dh are Lipschitz-continuous, by Taylor's formula one gets $\zeta_p^{n^\beta} = \mathcal{O}(|\theta_p^{n^\beta} - \theta^{*,n^\beta}|^2)$ and $\zeta_p = \mathcal{O}(|\theta_p - \theta^*|^2)$. Therefore, defining $z_p^{n^\beta} = \theta_p^{n^\beta} - \theta_p - (\theta^{*,n^\beta} - \theta^*)$, $p \geq 0$, with $z_0^{n^\beta} = \theta^* - \theta^{*,n^\beta}$, by a simple induction argument one has

$$\begin{aligned} z_n^{n^\beta} &= \Pi_{1,n} z_0^{n^\beta} + \sum_{k=1}^{n} \gamma_k \Pi_{k+1,n} \Delta N_k^{n^\beta} + \sum_{k=1}^{n} \gamma_k \Pi_{k+1,n} \Delta R_k^{n^\beta} \\ &\quad + \sum_{k=1}^{n} \gamma_k \Pi_{k+1,n} \big(\zeta_{k-1}^n - \zeta_{k-1} \\ &\qquad + \big(Dh(\theta^*) - Dh^{n^\beta}(\theta^{*,n^\beta})\big)(\theta_{k-1}^{n^\beta} - \theta^{*,n^\beta})\big), \end{aligned} \tag{3.2}$$

where $\Pi_{k,n} := \prod_{j=k}^{n} (I_d - \gamma_j Dh(\theta^*))$, with the convention that $\Pi_{n+1,n} = I_d$ and $\Delta N_k^{n^\beta} := h^{n^\beta}(\theta^*) - h(\theta^*) - (H(\theta^*, (U^{n^\beta})^{k+1}) - H(\theta^*, U^{k+1}))$, $\Delta R_k^{n^\beta} =$

$h^{n^\beta}(\theta_k^{n^\beta}) - h^{n^\beta}(\theta^*) - (H(\theta_k^{n^\beta}, (U^{n^\beta})^{k+1}) - H(\theta^*, (U^{n^\beta})^{k+1})) + H(\theta_k, U^{k+1}) - H(\theta^*, U^{k+1}) - (h(\theta_k) - h(\theta^*))$ for $k \geq 1$. We will now investigate the asymptotic behavior of each term in the above decomposition. We will see that the second term, which represents the nonlinearity in the innovation variables (U^{n^β}, U), provides the announced weak rate of convergence.

Step 1: study of the sequence $\{n^\alpha \Pi_{1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} z_0^{n^\beta}, n \geq 0\}$.

Under the assumptions on the step sequence γ, one has for all $\eta \in (0, \lambda_m)$,

$$\begin{aligned} &n^\alpha \mathbb{E}\big[\big|\Pi_{1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} z_0^{n^\beta}\big|\big] \\ &\quad \leq n^\alpha \|\Pi_{1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))}\| \Big(\mathbb{E}|\theta_0| + \sup_{n\geq 1} \mathbb{E}|\theta_0^n| + \big|\theta^{*,n^\beta} - \theta^*\big|\Big) \\ &\quad \leq C n^\alpha \exp\left(-(\lambda_m - \eta) \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k\right) \to 0, \end{aligned}$$

by selecting η s.t. $(\lambda_m - \eta)\gamma_0 > (\underline{\lambda} - \eta)\gamma_0 > \alpha/(2\alpha - 2\rho\beta)$ if $\gamma(p) = \gamma_0/p$, $p \geq 1$.

Step 2: study of the sequence $\{n^\alpha \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k \Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \times (\zeta_{k-1}^n - \zeta_{k-1} + (Dh(\theta^*) - Dh^{n^\beta}(\theta^{*,n^\beta}))(\theta_{k-1}^{n^\beta} - \theta^{*,n^\beta})), n \geq 0\}$.

By Lemma 5.2, one has

$$\begin{aligned} &\mathbb{E}\left|\sum_{k=1}^{n} \gamma_k \Pi_{k+1,n} \big(\zeta_{k-1}^{n^\beta} - \zeta_{k-1} + \big(Dh(\theta^*) - Dh^{n^\beta}(\theta^{*,n^\beta})\big)\big(\theta_{k-1}^{n^\beta} - \theta^{*,n^\beta}\big)\big)\right| \\ &\quad \leq C \sum_{k=1}^{n} \|\Pi_{k+1,n}\| \big(\gamma_k^2 + \gamma_k^{3/2} \big\|Dh(\theta^*) - Dh^{n^\beta}(\theta^{*,n^\beta})\big\|\big), \end{aligned}$$

so that by Lemma 5.1, we easily derive that [if $\gamma(p) = \gamma_0/p$—recall that $\underline{\lambda}\gamma_0 > \alpha/(2\alpha - 2\rho\beta)$] $\sum_{k=1}^n \gamma_k^2 \|\Pi_{k+1,n}\| = o(\gamma^{\alpha/(2\alpha-2\rho\beta)}(n))$ and [note that $\underline{\lambda}\gamma_0 > \alpha/(2\alpha - 2\rho\beta) > 1/2$] $\sum_{k=1}^n \gamma_k^{3/2} \|\Pi_{k+1,n}\| = \mathcal{O}(\gamma^{1/2}(n))$ which in turn yields

$$\limsup_n n^\alpha \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k^2 \|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))}\| = 0.$$

Moreover, since Dh^{n^β} is Lipschitz-continuous (uniformly in n), we clearly have

$$\begin{aligned} &\sum_{k=1}^{n} \gamma_k^{3/2} \|\Pi_{k+1,n}\| \big\|Dh(\theta^*) - Dh^{n^\beta}(\theta^{*,n^\beta})\big\| \\ &\quad \leq \sum_{k=1}^{n} \gamma_k^{3/2} \|\Pi_{k+1,n}\| \big(\big\|Dh(\theta^*) - Dh^{n^\beta}(\theta^*)\big\| + \big|\theta^{*,n^\beta} - \theta^*\big|\big), \end{aligned}$$

which, combined with $n^{\rho\beta}\|Dh(\theta^*) - Dh^{n^\beta}(\theta^*)\| \to 0$ and $n^{\rho\beta}|\theta^{*,n^\beta} - \theta^*| \to 0$ (recall that $\alpha > \rho$), implies that

$$\limsup_n n^\alpha \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k^{3/2}\|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))}\| \|Dh(\theta^*) - Dh^{n^\beta}(\theta^{*,n^\beta})\| = 0.$$

Hence we conclude that

$$n^\alpha \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k \Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \times \left(\zeta_{k-1}^{n^\beta} - \zeta_{k-1} + \left(Dh(\theta^*) - Dh^{n^\beta}(\theta^{*,n^\beta})\right)\left(\theta_{k-1}^{n^\beta} - \theta^{*,n^\beta}\right)\right) \xrightarrow{L^1(\mathbb{P})} 0.$$

Step 3: *study of the sequence* $\{n^\alpha \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k \Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \times \Delta R_k^{n^\beta}, n \geq 0\}$.

Regarding the third term of (3.2), namely $\sum_{k=1}^n \gamma_k \Pi_{k+1,n} \Delta R_k^{n^\beta}$, we decompose as follows:

$$\begin{aligned}
\sum_{k=1}^n \gamma_k \Pi_{k+1,n} \Delta R_k^{n^\beta}
&= \sum_{k=1}^n \gamma_k \Pi_{k+1,n}\big(h^{n^\beta}(\theta_k^{n^\beta}) - h^{n^\beta}(\theta^*) \\
&\qquad - \big(H(\theta_k^{n^\beta}, (U^{n^\beta})^{k+1}) - H(\theta^*, (U^{n^\beta})^{k+1})\big)\big) \\
&\quad + \sum_{k=1}^n \gamma_k \Pi_{k+1,n}\big(H(\theta_k, U^{k+1}) - H(\theta^*, U^{k+1}) - \big(h(\theta_k) - h(\theta^*)\big)\big) \\
&= A_n + B_n.
\end{aligned}$$

Now, using that $\mathbb{E}[H(\theta_k^{n^\beta}, (U^{n^\beta})^{k+1}) - H(\theta^*, (U^{n^\beta})^{k+1})|\mathcal{F}_k] = h^{n^\beta}(\theta_k^{n^\beta}) - h^{n^\beta}(\theta^*)$ and (HLH), it follows that

$$\begin{aligned}
\mathbb{E}|A_n|^2 &\leq C \sum_{k=1}^n \gamma_k^2 \|\Pi_{k+1,n}\|^2 \big(\mathbb{E}|\theta_k^{n^\beta} - \theta^{*,n^\beta}|^2 + |\theta^{*,n^\beta} - \theta^*|^2\big) \\
&\leq C\left(\sum_{k=1}^n \gamma_k^3 \|\Pi_{k+1,n}\|^2 + \sum_{k=1}^n \gamma_k^2 \|\Pi_{k+1,n}\|^2 |\theta^{*,n^\beta} - \theta^*|^2\right) \\
&:= A_n^1 + A_n^2.
\end{aligned}$$

From Lemma 5.1 we get

$$\sum_{k=1}^{n} \gamma_k^3 \|\Pi_{k+1,n}\|^2 = o\big(\gamma_n^{2\alpha/(2\alpha-2\rho\beta)}\big) \quad \text{and} \quad \sum_{k=1}^{n} \gamma_k^2 \|\Pi_{k+1,n}\|^2 = \mathcal{O}(\gamma_n).$$

Consequently, we derive

$$\limsup_n n^{2\alpha} A^1_{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} = 0 \quad \text{and} \quad \limsup_n n^{2\alpha} A^2_{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} = 0.$$

Similarly using (HLH) and Lemma 5.2, we derive $n^\alpha B_{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \xrightarrow{L^2(\mathbb{P})} 0$ as $n \to +\infty$ so that

$$n^\alpha \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k \Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \Delta R_k^{n^\beta} \xrightarrow{\mathbb{P}} 0, \qquad n \to +\infty.$$

Step 4: *study of the sequence* $\{n^\alpha \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k \Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \times \Delta N_k^{n^\beta}, n \geq 0\}$.

We now prove a CLT for the sequence

$$\left\{ n^\alpha \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k \Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \Delta N_k^{n^\beta}, n \geq 0 \right\}.$$

It holds

$$\begin{aligned} &\sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \mathbb{E}\big|n^\alpha \gamma_k \Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \Delta N_k^{n^\beta}\big|^{2+\delta} \\ &\quad \leq \sup_{n\geq 1} \sup_{k \in [\![1,n]\!]} \mathbb{E}\big|n^{\rho\beta} \Delta N_k^{n^\beta}\big|^{2+\delta} \\ &\qquad \times n^{(2+\delta)(\alpha-\rho\beta)} \left(\sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k^{2+\delta} \|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))}\|^{2+\delta} \right). \end{aligned}$$

By Lemma 5.1, we have the following bound: $\sum_{k=1}^n \gamma_k^{2+\delta} \|\Pi_{k+1,n}\|^{2+\delta} = o(\gamma^{(2+\delta)(\alpha-\rho\beta)/(2\alpha-2\rho\beta)}(n))$, which implies

$$\limsup_n n^{(2+\delta)(\alpha-\rho\beta)} \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k^{2+\delta} \|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))}\|^{2+\delta} = 0.$$

Moreover simple computations lead to

$$\begin{aligned} &\mathbb{E}\big|n^{\rho\beta} \Delta N_k^{n^\beta}\big|^{2+\delta} \\ &\quad \leq C\big(\big|n^{\rho\beta}(h^{n^\beta}(\theta^*) - h(\theta^*))\big|^{2+\delta} + \mathbb{E}\big(n^{\rho\beta}\big|H(\theta^*, U^{n^\beta}) - H(\theta^*, U)\big|\big)^{2+\delta}\big). \end{aligned}$$

For the first term in the above inequality, we have $\sup_{n\geq 1}|n^{\rho\beta}(h^{n^\beta}(\theta^*) - h(\theta^*))|^{2+\delta} < +\infty \Leftrightarrow \alpha \geq \rho$. For the second term, using assumptions (HLH) and (HSR), we get $\sup_{n\geq 1}\mathbb{E}[(n^{\rho\beta}|H(\theta^*, U^{n^\beta}) - H(\theta^*, U)|)^{2+\delta}] < +\infty$. Hence we conclude that

$$\sup_{n\geq 1}\sup_{k\in[\![1,n]\!]}\mathbb{E}\big|n^{\rho\beta}\Delta N_k^{n^\beta}\big|^{2+\delta} < +\infty$$

so that the conditional Lindeberg condition holds. Now, we focus on the conditional variance. We set

$$\begin{aligned}S_n := n^{2\alpha}\sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} & \gamma_k^2\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))}\mathbb{E}_k\big[\Delta N_k^{n^\beta}\big(\Delta N_k^{n^\beta}\big)^T\big] \\ & \times \Pi^T_{k+1,\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))}\end{aligned} \tag{3.3}$$

and $V^{n^\beta} = U^{n^\beta} - U$.

A Taylor expansion yields

$$n^{\rho\beta}\big(H\big(\theta^*, U^{n^\beta}\big) - H\big(\theta^*, U\big)\big) = D_x H\big(\theta^*, U\big)n^{\rho\beta}V^{n^\beta} + \psi\big(\theta^*, U, V^{n^\beta}\big)n^{\rho\beta}V^{n^\beta}$$

with $\psi(\theta^*, U, V^{n^\beta}) \xrightarrow{\mathbb{P}} 0$. From the tightness of $(n^{\rho\beta}V^{n^\beta})_{n\geq 1}$, we get $\psi(\theta^*, U, V^{n^\beta})n^{\rho\beta}V^{n^\beta} \xrightarrow{\mathbb{P}} 0$ so that using Theorem 2.1 and Lemma 2.1 yields

$$n^{\rho\beta}\big(H\big(\theta^*, U^{n^\beta}\big) - H\big(\theta^*, U\big)\big) \Longrightarrow D_x H\big(\theta^*, U\big)V.$$

Moreover, from assumptions (HLH) and (HSR) it follows that

$$\sup_{n\geq 1}\mathbb{E}\big[\big|n^{\rho\beta}\big(H\big(\theta^*, U^{n^\beta}\big) - H\big(\theta^*, U\big)\big)\big|^{2+\delta}\big] < +\infty,$$

which, combined with (HDH), implies

$$\begin{aligned}&\mathbb{E}\big[n^{\rho\beta}\big(H\big(\theta^*, U^{n^\beta}\big) - H\big(\theta^*, U\big)\big)\big] \to \tilde{\mathbb{E}}\big[D_x H\big(\theta^*, U\big)V\big],\\ &\mathbb{E}\big[n^{\rho\beta}\big(H\big(\theta^*, U^{n^\beta}\big) - H\big(\theta^*, U\big)\big)\big(n^{\rho\beta}\big(H\big(\theta^*, U^{n^\beta}\big) - H\big(\theta^*, U\big)\big)\big)^T\big]\\ &\qquad \to \tilde{\mathbb{E}}\big[\big(D_x H\big(\theta^*, U\big)V\big)\big(D_x H\big(\theta^*, U\big)V\big)^T\big].\end{aligned}$$

Hence we have

$$\begin{aligned}\Gamma_n \to \Gamma^* := \tilde{\mathbb{E}}\big[&\big(D_x H\big(\theta^*, U\big)V - \tilde{\mathbb{E}}\big[D_x H\big(\theta^*, U\big)V\big]\big)\\ &\times \big(D_x H\big(\theta^*, U\big)V - \tilde{\mathbb{E}}\big[D_x H\big(\theta^*, U\big)V\big]\big)^T\big],\end{aligned}$$

where for $n \geq 1$ we set

$$\Gamma_n := n^{2\rho\beta}\mathbb{E}_k\big[\Delta N_k^{n^\beta}\big(\Delta N_k^{n^\beta}\big)^T\big].$$

Consequently, using the decomposition

$$\frac{1}{\gamma(n)}\sum_{k=1}^{n}\gamma_k^2\Pi_{k+1,n}\Gamma_n\Pi_{k+1,n}^T$$
$$=\frac{1}{\gamma(n)}\sum_{k=1}^{n}\gamma_k^2\Pi_{k+1,n}\Gamma^*\Pi_{k+1,n}^T+\frac{1}{\gamma(n)}\sum_{k=1}^{n}\gamma_k^2\Pi_{k+1,n}(\Gamma_n-\Gamma^*)\Pi_{k+1,n}^T$$

with

$$\limsup_n \frac{1}{\gamma(n)}\left\|\sum_{k=1}^{n}\gamma_k^2\Pi_{k+1,n}(\Gamma_n-\Gamma^*)\Pi_{k+1,n}^T\right\|\le C\limsup_n\|\Gamma_n-\Gamma^*\|=0,$$

which is a consequence of Lemma 5.1, we clearly see that

$$\lim_n S_n=\lim_n\frac{1}{\gamma(n)}\sum_{k=1}^{n}\gamma_k^2\Pi_{k+1,n}\Gamma^*\Pi_{k+1,n}^T$$

if this latter limit exists. Let us note that Θ^* is the (unique) matrix A solution to the Lyapunov equation

$$\Gamma^*-\big(Dh(\theta^*)-\zeta I_d\big)A-A\big(Dh(\theta^*)-\zeta I_d\big)^T=0.$$

Following the lines of the proof of Lemma 2.3, step 3, we have $S_n\xrightarrow{\text{a.s.}}\Theta^*$. We leave the computational details to the reader.

3.4. *Proof of Lemma* 2.7. We come back to the decomposition used in the proof of Lemma 2.3, and we consequently use the same notation. We will not go into the computational detail. We deal with the case $\rho\in(0,1/2)$. The case $\rho=1/2$ can be handled in a similar fashion.

We first write for $p\ge 0$,

$$\theta_{p+1}^{m^\ell}-\theta^{*,m^\ell}=\theta_p^{m^\ell}-\theta^{*,m^\ell}-\gamma_{p+1}Dh^{m^\ell}(\theta^{*,m^\ell})(\theta_p^{m^\ell}-\theta^{*,m^\ell})$$
$$+\gamma_{p+1}\Delta M_{p+1}^{m^\ell}+\gamma_{p+1}\zeta_p^{m^\ell}$$

with $\Delta M_{p+1}^{m^\ell}=h^{m^\ell}(\theta_p^{m^\ell})-H(\theta_p^{m^\ell},(U^{m^\ell})^{p+1})$ and $\zeta_p^{m^\ell}=\mathcal{O}(|\theta_{p+1}^{m^\ell}-\theta^{*,m^\ell}|^2)$, $p\ge 0$. Therefore, defining $z_p^\ell=\theta_p^{m^\ell}-\theta_p^{m^{\ell-1}}-(\theta^{*,m^\ell}-\theta^{*,m^{\ell-1}})$, $p\ge 0$, with $z_0^\ell=\theta_0^{m^\ell}-\theta_0^{m^\ell}-(\theta^{*,m^\ell}-\theta^{*,m^{\ell-1}})$, by a simple induction argument, one has

$$\begin{aligned}z_{M_\ell}^\ell&=\Pi_{1,M_\ell}z_0^\ell+\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}\Delta N_k^\ell+\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}\Delta R_k^\ell\\&\quad+\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}\\&\qquad\times\big(\zeta_{k-1}^\ell-\zeta_{k-1}^{\ell-1}+\big(Dh(\theta^*)-Dh^{m^\ell}(\theta^{*,m^\ell})\big)(\theta_{k-1}^{m^\ell}-\theta^{*,m^\ell})\\&\qquad\quad-\big(Dh(\theta^*)-Dh^{m^{\ell-1}}(\theta^{*,m^{\ell-1}})\big)(\theta_{k-1}^{m^{\ell-1}}-\theta^{*,m^{\ell-1}})\big),\end{aligned}\tag{3.4}$$

where $\Pi_{k,n} := \prod_{j=k}^{n}(I_d - \gamma_j Dh(\theta^*))$, with the convention that $\Pi_{n+1,n} = I_d$, and $\Delta N_k^\ell := h^{m^\ell}(\theta^*) - h^{m^{\ell-1}}(\theta^*) - (H(\theta^*, (U^{m^\ell})^{k+1}) - H(\theta^*, (U^{m^{\ell-1}})^{k+1}))$, $\Delta R_k^\ell = h^{m^\ell}(\theta_k^{m^\ell}) - h^{m^\ell}(\theta^*) - (H(\theta_k^{m^\ell}, (U^{m^\ell})^{k+1}) - H(\theta^*, (U^{m^\ell})^{k+1})) + H(\theta_k^{m^{\ell-1}}, (U^{m^{\ell-1}})^{k+1}) - H(\theta^*, (U^{m^{\ell-1}})^{k+1}) - (h^{m^{\ell-1}}(\theta_k^{m^{\ell-1}}) - h^{m^{\ell-1}}(\theta^*))$ for $k \geq 0$. We follow the same methodology developed so far and quantify the contribution of each term. Once again the weak rate of convergence will be ruled by the second term which involves the nonlinearity in the innovation variable $(U^{m^{\ell-1}}, U^{m^\ell})$, for which we prove a CLT.

Step 1: *study of* $\{n^\alpha \sum_{\ell=1}^{L} \Pi_{1,M_\ell} z_0^\ell, n \geq 0\}$.

Under the assumptions on the step sequence γ, for all $\eta \in (0, \lambda_m)$ we have

$$\begin{aligned}\|\Pi_{1,M_\ell}\| &\leq \exp\left(-(\lambda_m - \eta)\sum_{k=1}^{M_\ell}\gamma_k\right)\\ &= Cm^{\ell((1+2\rho)/2)(\lambda_m-\eta)\gamma_0}/\big(\big(n^{2\alpha+((1-2\rho)/2)}\big)^{(\lambda_m-\eta)\gamma_0}\big)\end{aligned}$$

if $\gamma(p) = \gamma_0/p$ or $\|\Pi_{1,M_\ell}\| = \mathcal{O}(\gamma(M_\ell))$ otherwise. Therefore, if $\gamma(p) = \gamma_0/p$, we select $\eta > 0$ such that $\gamma_0(\lambda_m - \eta) > \alpha/(\alpha - 2\rho)$, and then one has

$$\begin{aligned}\mathbb{E}\left|n^\alpha \sum_{\ell=1}^{L}\Pi_{1,M_\ell} z_0^\ell\right| &\leq Cn^\alpha \sum_{\ell=1}^{L}\|\Pi_{1,M_\ell}\|\\ &\leq \frac{C}{n^{(\lambda_m-\eta)\gamma_0(2\alpha+((1-2\rho)/2))-\alpha}}\sum_{\ell=1}^{L} m^{\ell(\lambda_m-\eta)\gamma_0((1+2\rho)/2)}\\ &\leq \frac{C}{n^{2(\lambda_m-\eta)\gamma_0(\alpha-2\rho)-\alpha}} \to 0\end{aligned}$$

as $n \to +\infty$. Otherwise one has

$$\mathbb{E}\left|n^\alpha \sum_{\ell=1}^{L}\Pi_{1,M_\ell} z_0^\ell\right| \leq Cn^\alpha \sum_{\ell=1}^{L}\gamma(M_\ell) \leq C\frac{n^{(1+2\rho)/2}}{n^{\alpha+((1-2\rho)/2)}} = \frac{C}{n^{\alpha-2\rho}} \to 0.$$

Step 2: *study of* $\{n^\alpha \sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k \Pi_{k+1,M_\ell}(\zeta_{k-1}^\ell - \zeta_{k-1}^{\ell-1}), n \geq 0\}$.

By Lemma 5.2, one has

$$\mathbb{E}\left|n^\alpha \sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k \Pi_{k+1,M_\ell}\big(\zeta_{k-1}^\ell - \zeta_{k-1}^{\ell-1}\big)\right| \leq Cn^\alpha \sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k^2\|\Pi_{k+1,M_\ell}\|.$$

However, from Lemma 5.1 [if $\gamma(p) = \gamma_0/p$—recall that $\lambda_m\gamma_0 > 1$] we easily derive $\limsup_n \frac{1}{\gamma(n)}\sum_{k=1}^{n}\gamma_k^2\|\Pi_{k+1,n}\| \leq 1$, so that

$$n^\alpha \sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k^2\|\Pi_{k+1,M_\ell}\| \leq Cn^\alpha \sum_{\ell=1}^{L}\gamma(M_\ell) \to 0, \qquad n \to +\infty.$$

Step 3: *study of* $\{n^{\alpha}\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}((Dh(\theta^*)-Dh^{m^\ell}(\theta^{*,m^\ell}))(\theta_{k-1}^{m^\ell}-\theta^{*,m^\ell})), n\geq 0\}$ *and* $\{n^{\alpha}(\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}(Dh(\theta^*)-Dh^{m^{\ell-1}}(\theta^{*,m^{\ell-1}}))\times(\theta_{k-1}^{m^{\ell-1}}-\theta^{*,m^{\ell-1}})), n\geq 0\}$.

By Lemma 5.2 and since Dh^{m^ℓ} is a Lipschitz function uniformly in m, we clearly have

$$
\begin{aligned}
&\mathbb{E}\left|n^{\alpha}\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k^{3/2}\Pi_{k+1,M_\ell}\big(Dh(\theta^*)-Dh^{m^\ell}(\theta^{*,m^\ell})\big)\big(\theta_{k-1}^{m^\ell}-\theta^{*,m^\ell}\big)\right| \\
&\quad\leq n^{\alpha}\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k^{3/2}\|\Pi_{k+1,n}\|\times\big(\|Dh(\theta^*)-Dh^{m^\ell}(\theta^*)\|+|\theta^{*,m^\ell}-\theta^*|\big) \\
&\quad\leq Cn^{\alpha}\sum_{\ell=1}^{L}\gamma^{1/2}(M_\ell)\big(\|Dh(\theta^*)-Dh^{m^\ell}(\theta^*)\|+|\theta^{*,m^\ell}-\theta^*|\big),
\end{aligned}
$$

which, combined with $\sup_{n\geq 1}n^{\beta}\|Dh(\theta^*)-Dh^{n}(\theta^*)\|<+\infty$ with $\beta>\rho$ and $\sup_{n\geq 1}n^{\alpha}|\theta^{*,n}-\theta^*|<+\infty$, implies that

$$
\begin{aligned}
&\mathbb{E}\left|n^{\alpha}\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}\big(Dh(\theta^*)-Dh^{m^\ell}(\theta^{*,m^\ell})\big)\big(\theta_{k-1}^{m^\ell}-\theta^{*,m^\ell}\big)\right| \\
&\quad\leq \frac{C}{n^{(1-2\rho)/4}}\sum_{\ell=1}^{L}m^{\ell(1+2\rho)/4}\big(m^{-\ell\alpha}+m^{-\ell\beta}\big) \\
&\quad\leq C\big(n^{\rho-\alpha}+n^{\rho-\beta}\big)
\end{aligned}
$$

so that $n^{\alpha}\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}(Dh(\theta^*)-Dh^{m^\ell}(\theta^{*,m^\ell}))(\theta_{k-1}^{m^\ell}-\theta^{*,m^\ell})\xrightarrow{L^1(\mathbb{P})}0$. By similar arguments, we easily deduce $n^{\alpha}\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}(Dh(\theta^*)-Dh^{m^{\ell-1}}(\theta^{*,m^{\ell-1}}))(\theta_{k-1}^{m^{\ell-1}}-\theta^{*,m^{\ell-1}})\xrightarrow{L^1(\mathbb{P})}0$.

Step 4: *study of* $\{n^{\alpha}\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}\Delta R_k^{\ell}, n\geq 0\}$.

Using the Cauchy–Schwarz inequality we deduce

$$
\begin{aligned}
&\mathbb{E}\left|n^{\alpha}\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}\Delta R_k^{\ell}\right|^2 \\
&\quad\leq n^{\alpha}\sum_{\ell=1}^{L}\Bigg(\sum_{k=1}^{M_\ell}\gamma_k^2\|\Pi_{k+1,M_\ell}\|^2 \\
&\qquad\qquad\times\mathbb{E}\big|H\big(\theta_k^{m^\ell},(U^{m^\ell})^{k+1}\big)-H\big(\theta^*,(U^{m^\ell})^{k+1}\big)\big|^2\Bigg)^{1/2}
\end{aligned}
$$

$$+ n^{\alpha} \sum_{\ell=1}^{L} \Biggl(\sum_{k=1}^{M_\ell} \gamma_k^2 \|\Pi_{k+1,M_\ell}\|^2 \times \mathbb{E}\bigl|H\bigl(\theta_k^{m^{\ell-1}}, (U^{m^{\ell-1}})^{k+1}\bigr) - H\bigl(\theta^*, (U^{m^{\ell-1}})^{k+1}\bigr)\bigr|^2 \Biggr)^{1/2}$$

$$\le C n^{\alpha} \sum_{\ell=1}^{L} \Biggl(\sum_{k=1}^{M_\ell} \gamma_k^3 \|\Pi_{k+1,M_\ell}\|^2 \Biggr)^{1/2},$$

where we use (HLH) and Lemma 5.2. Now from Lemma 5.1 and simple computations, it follows that

$$n^{\alpha} \sum_{\ell=1}^{L} \Biggl(\sum_{k=1}^{M_\ell} \gamma_k^3 \|\Pi_{k+1,M_\ell}\|^2 \Biggr)^{1/2} \le C n^{\alpha} \sum_{\ell=1}^{L} \gamma(M_\ell) \to 0, \qquad n \to +\infty.$$

Therefore, we conclude that

$$n \sum_{\ell=1}^{L} \sum_{k=1}^{M_\ell} \gamma_k \Pi_{k+1,M_\ell} \Delta R_k^{\ell} \xrightarrow{L^2(\mathbb{P})} 0, \qquad n \to +\infty.$$

Step 5: *study of* $\{n^{\alpha} \sum_{\ell=1}^{L} \sum_{k=1}^{M_\ell} \gamma_k \Pi_{k+1,M_\ell} \Delta N_k^{\ell}, n \ge 0\}$.

We now prove a CLT for the sequence $\{n^{\alpha} \sum_{\ell=1}^{L} \sum_{k=1}^{M_\ell} \gamma_k \Pi_{k+1,M_\ell} \Delta N_k^{\ell}, n \ge 0\}$. By Burkhölder's inequality and elementary computations, it holds that

$$\begin{aligned} &\sum_{\ell=1}^{L} \mathbb{E}\Biggl| \sum_{k=1}^{M_\ell} n^{\alpha} \gamma_k \Pi_{k+1,M_\ell} \Delta N_k^{\ell} \Biggr|^{2+\delta} \\ &\quad \le C n^{(2+\delta)\alpha} \sum_{\ell=1}^{L} \mathbb{E} \Biggl(\sum_{k=1}^{M_\ell} \gamma_k^2 \|\Pi_{k+1,M_\ell}\|^2 |\Delta N_k^{\ell}|^2 \Biggr)^{1+\delta/2} \\ &\quad \le C n^{(2+\delta)\alpha} \\ &\qquad \times \sum_{\ell=1}^{L} \Biggl(\sum_{k=1}^{M_\ell} \gamma_k^2 \|\Pi_{k+1,M_\ell}\|^2 \Biggr)^{\delta/2} \Biggl(\sum_{k=1}^{M_\ell} \gamma_k^{2+\delta} \|\Pi_{k+1,M_\ell}\|^{2+\delta} \mathbb{E}|\Delta N_k^{\ell}|^{2+\delta} \Biggr). \end{aligned}$$

Using (HLH) and (HSR) we have $\sup_{\ell \ge 1} \mathbb{E}[(m^{\rho\ell}|H(\theta^*, U^{m^\ell}) - H(\theta^*, U)|)^{2+\delta}] < +\infty$ so that

$$\mathbb{E}|\Delta N_k^{\ell}|^{2+\delta} \le \frac{K}{m^{\ell(2\rho+\rho\delta)}}.$$

Moreover, by Lemma 5.1, we have

$$\limsup_n \bigl(1/\gamma^{(1+\delta)}(n)\bigr) \sum_{k=1}^{n} \gamma_k^{2+\delta} \|\Pi_{k+1,n}\|^{2+\delta} \le 1$$

and

$$\limsup_n \big(1/\gamma(n)\big) \sum_{k=1}^{n} \gamma_k^2 \|\Pi_{k+1,n}\|^2 \le 1.$$

Consequently we deduce

$$\begin{aligned}\sum_{\ell=1}^{L} \mathbb{E}\left|\sum_{k=1}^{M_\ell} n^\alpha \gamma_k \Pi_{k+1,M_\ell} \Delta N_k^\ell\right|^{2+\delta} &\le C n^{(2+\delta)\alpha} \sum_{\ell=1}^{L} \gamma^{1+3\delta/2}(M_\ell) m^{-\ell(2\rho+\rho\delta)} \\ &\le \frac{C}{n^{2\alpha\delta}} n^{2\rho(1+3\delta/2)-2\rho-\rho\delta} = \frac{C}{n^{2\delta(\alpha-\rho)}},\end{aligned}$$

which in turn implies

$$\sum_{\ell=1}^{L} \mathbb{E}\left|\sum_{k=1}^{M_\ell} n^\alpha \gamma_k \Pi_{k+1,M_\ell} \Delta N_k^\ell\right|^{2+\delta} \to 0, \qquad n \to +\infty$$

so that the conditional Lindeberg condition is satisfied. Now, we focus on the conditional variance. We set

$$\begin{aligned} &S_\ell := n^{2\alpha} \sum_{k=1}^{M_\ell} \gamma_k^2 \Pi_{k+1,M_\ell} \mathbb{E}_k\big[\Delta N_k^\ell (\Delta N_k^\ell)^T\big] \Pi_{k+1,M_\ell}^T \quad \text{and} \\ &U^\ell = U^{m^\ell} - U^{m^{\ell-1}}. \end{aligned} \tag{3.5}$$

Observe that by the very definition of M_ℓ one has

$$\begin{aligned} S_\ell = &\frac{1}{\gamma(M_\ell)} \big(m^{(1-2\rho)/2} - 1\big) \frac{m^{\ell((1+2\rho)/2)}}{n^{(1-2\rho)/2} - 1} \\ &\times \sum_{k=1}^{M_\ell} \gamma_k^2 \Pi_{k+1,M_\ell} \mathbb{E}_k\big[\Delta N_k^\ell (\Delta N_k^\ell)^T\big] \Pi_{k+1,M_\ell}^T. \end{aligned}$$

A Taylor expansion yields

$$\begin{aligned} &H\big(\theta^*, U^{m^\ell}\big) - H\big(\theta^*, U^{m^{\ell-1}}\big) \\ &\quad = D_x H\big(\theta^*, U\big) U^\ell + \psi\big(\theta^*, U, U^{m^\ell} - U\big)\big(U^{m^\ell} - U\big) \\ &\qquad + \psi\big(\theta^*, U, U^{m^{\ell-1}} - U\big)\big(U^{m^{\ell-1}} - U\big) \end{aligned}$$

with $(\psi(\theta^*, U, U^{m^\ell} - U), \psi(\theta^*, U, U^{m^{\ell-1}} - U)) \xrightarrow{\mathbb{P}} 0$ as $\ell \to +\infty$. From the tightness of the sequences $(m^{\rho\ell}(U^{m^\ell} - U))_{\ell \ge 1}$ and $(m^{\rho\ell}(U^{m^{\ell-1}} - U))_{\ell \ge 1}$, we get

$$\begin{aligned} &m^{\rho\ell}\big(\psi\big(\theta^*, U, U^{m^\ell} - U\big)\big(U^{m^\ell} - U\big) + \psi\big(\theta^*, U, U^{m^{\ell-1}} - U\big)\big(U^{m^{\ell-1}} - U\big)\big) \\ &\quad \xrightarrow{\mathbb{P}} 0, \qquad \ell \to +\infty. \end{aligned}$$

Therefore using Theorem 2.1 and Lemma 2.1 yields

$$m^{\rho\ell}\big(H(\theta^*, U^{m^\ell}) - H(\theta^*, U^{m^{\ell-1}})\big) \Longrightarrow D_x H(\theta^*, U)V^m.$$

Moreover, from assumption (HLH) and (HRH) it follows that

$$\sup_{\ell\geq 1} \mathbb{E}\big|m^{\rho\ell}\big(H(\theta^*, U^{m^\ell}) - H(\theta^*, U^{m^{\ell-1}})\big)\big|^{2+\delta} < +\infty,$$

which, combined with (HDH), implies

$$\begin{aligned} &m^{\rho\ell}\mathbb{E}\big[H(\theta^*, U^{m^\ell}) - H(\theta^*, U^{m^{\ell-1}})\big] \to \tilde{\mathbb{E}}\big[D_x H(\theta^*, U)V^m\big],\\ &m^{2\rho\ell}\mathbb{E}\big[\big(H(\theta^*, U^{m^\ell}) - H(\theta^*, U^{m^{\ell-1}})\big)\big(H(\theta^*, U^{m^\ell}) - H(\theta^*, U^{m^{\ell-1}})\big)^T\big]\\ &\qquad \to \tilde{\mathbb{E}}\big[\big(D_x H(\theta^*, U)V^m\big)\big(D_x H(\theta^*, U)V^m\big)^T\big] \end{aligned}$$

as $\ell \to +\infty$. Hence, we have

$$\begin{aligned} m^{2\rho\ell}\Gamma_\ell \to \Gamma^* := {}&\tilde{\mathbb{E}}\big[\big(D_x H(\theta^*, U)V^m - \tilde{\mathbb{E}}\big[D_x H(\theta^*, U)V^m\big]\big)\\ &\times \big(D_x H(\theta^*, U)V^m - \tilde{\mathbb{E}}\big[D_x H(\theta^*, U)V^m\big]\big)^T\big], \end{aligned}$$

where for $\ell \geq 1$,

$$\begin{aligned} \Gamma_\ell &:= \mathbb{E}_k\big[\Delta N_k^\ell(\Delta N_k^\ell)^T\big]\\ &= \mathbb{E}\big[\big(H(\theta^*, U^{m^\ell}) - H(\theta^*, U^{m^{\ell-1}})\big)\big(H(\theta^*, U^{m^\ell}) - H(\theta^*, U^{m^{\ell-1}})\big)^T\big]\\ &\quad - \big(h^{m^\ell}(\theta^*) - h^{m^{\ell-1}}(\theta^*)\big)\big(h^{m^\ell}(\theta^*) - h^{m^{\ell-1}}(\theta^*)\big)^T. \end{aligned}$$

Consequently, using the decomposition

$$\begin{aligned} &\frac{1}{\gamma(M_\ell)} m^{2\rho\ell} \sum_{k=1}^{M_\ell} \gamma_k^2 \Pi_{k+1,M_\ell}\Gamma_\ell \Pi_{k+1,M_\ell}^T\\ &\quad = \frac{1}{\gamma(M_\ell)} \sum_{k=1}^{M_\ell} \gamma_k^2 \Pi_{k+1,M_\ell}\Gamma^* \Pi_{k+1,M_\ell}^T\\ &\qquad + \frac{1}{\gamma(M_\ell)} \sum_{k=1}^{M_\ell} \gamma_k^2 \Pi_{k+1,M_\ell}\big(m^{2\rho\ell}\Gamma_\ell - \Gamma^*\big) \Pi_{k+1,M_\ell}^T \end{aligned}$$

with

$$\begin{aligned} &\limsup_\ell \frac{1}{\gamma(M_\ell)} \left\| \sum_{k=1}^{M_\ell} \gamma_k^2 \Pi_{k+1,M_\ell}\big(m^{2\rho\ell}\Gamma_\ell - \Gamma^*\big) \Pi_{k+1,M_\ell}^T \right\|\\ &\qquad \leq C \limsup_\ell \big\|m^{2\rho\ell}\Gamma_\ell - \Gamma^*\big\| = 0, \end{aligned}$$

which is a consequence of Lemma 5.1, we clearly see that $\frac{n^{(1-2\rho)/2}-1}{m^{\ell((1-2\rho)/2)}(m^{(1-2\rho)/2}-1)} \times \lim_\ell S_\ell = \lim_{p\to+\infty} \frac{1}{\gamma(p)} \sum_{k=1}^{p} \gamma_k^2 \Pi_{k+1,p}\Gamma^* \Pi_{k+1,p}^T$ if this latter limit exists. The matrix Θ^* defined by (2.10) is the (unique) matrix A solution to the Lyapunov equation

$$\Gamma^* - (Dh(\theta^*) - \zeta I_d)A - A(Dh(\theta^*) - \zeta I_d)^T = 0.$$

Following along the lines of the proof of step 3, Lemma 2.3, we have $S_\ell \frac{(n^{(1-2\rho)/2}-1)}{m^{\ell(1-2\rho)/2}(m^{(1-2\rho)/2}-1)} \xrightarrow{\text{a.s.}} \Theta^*$ as $\ell \to +\infty$. We leave the computational details to the reader. Finally, from Cesàro's lemma it follows that

$$\sum_{\ell=1}^{L} S_\ell = \left(\frac{m^{(1-2\rho)/2}-1}{n^{(1-2\rho)/2}-1}\right) \sum_{\ell=1}^{L} \left(S_\ell \frac{(n^{(1-2\rho)/2}-1)}{m^{\ell(1-2\rho)/2}(m^{(1-2\rho)/2}-1)}\right) m^{\ell(1-2\rho)/2}$$
$$\xrightarrow[n\to+\infty]{\text{a.s.}} \Theta^*.$$

4. Numerical results. In this section we illustrate the results obtained in Section 2.

4.1. *Computation of quantiles of a one-dimensional diffusion process.* We first consider the problem of the computation of a quantile at level $l \in (0, 1)$ of a one-dimensional diffusion process. This quantity, also referred as the value-at-risk at level l in the practice of risk management, is the lowest amount not exceeded by X_T with probability l, namely

$$q_l(X_T) := \inf\{\theta : \mathbb{P}(X_T \le \theta) \ge l\}.$$

To illustrate the results of Sections 2.3 and 2.4, we consider a simple geometric Brownian motion

$$X_t = x + \int_0^t r X_s\, ds + \int_0^t \sigma X_s\, dW_s, \qquad t \in [0, T] \tag{4.1}$$

for which the quantile is explicitly known at any level l. Hence we have $\rho = 1/2$. Because the distribution function of X_T is increasing, $q_l(X_T)$ is the unique solution of the equation $h(\theta) = \mathbb{E}_x[H(\theta, X_T)] = 0$ with $H(\theta, x) = \mathbf{1}_{\{x \le \theta\}} - l$. A simple computation shows that

$$q_l(X_T) = x_0 \exp\big((r - \sigma^2/2)T + \sigma\sqrt{T}\phi^{-1}(l)\big),$$

where ϕ is the distribution function of the standard normal distribution $\mathcal{N}(0, 1)$. We associate to the SDE (4.1) its Euler like scheme $X^n = (X_t^n)_{t\in[0,T]}$ with time step $\Delta = T/n$. We use the following values for the parameters: $x = 100, r = 0.05, \sigma = 0.4, T = 1, l = 0.7$. The reference Black–Scholes quantile is $q_{0.7}(X_T) = 119.69$.

REMARK 4.1. Let us note that when l is close to 0 or 1 (usually less than 0.05 or more than 0.95), the convergence of the considered SA algorithm is slow and chaotic. This is mainly due to the fact that the procedure obtains few significant samples to update the estimate in this rare event situation. One solution is to combine it with a variance reduction algorithm, such as an adaptive importance sampling procedure, that will generate more samples in the area of interest; see, for example, [2] and [3].

In order to illustrate the result of Theorem 2.6, we plot in Figure 1 the behaviors of $nh^n(\theta^*)$ and $n(\theta^{*,n} - \theta^*)$ for $n = 100, \dots, 500$. Actually, $h^n(\theta^*)$ is approximated by its Monte Carlo estimator, and $\theta^{*,n}$ is estimated by θ_M^n, both estimators

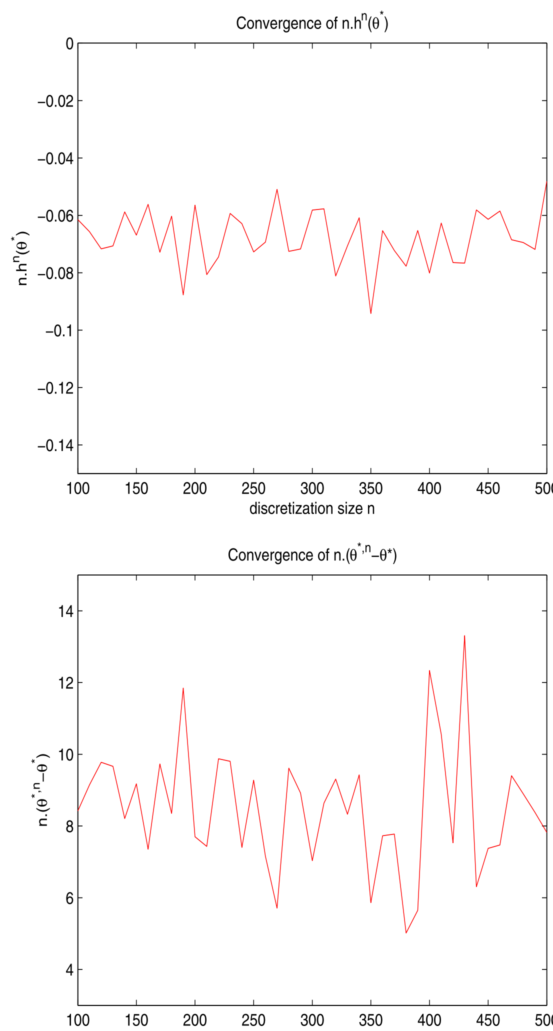

FIG. 1. *On the top*: *weak discretization error* $n \mapsto nh^n(\theta^*)$. *On the bottom*: *implicit discretization error* $n \mapsto n(\theta^{*,n} - \theta^*)$, $n = 100, \dots, 500$.

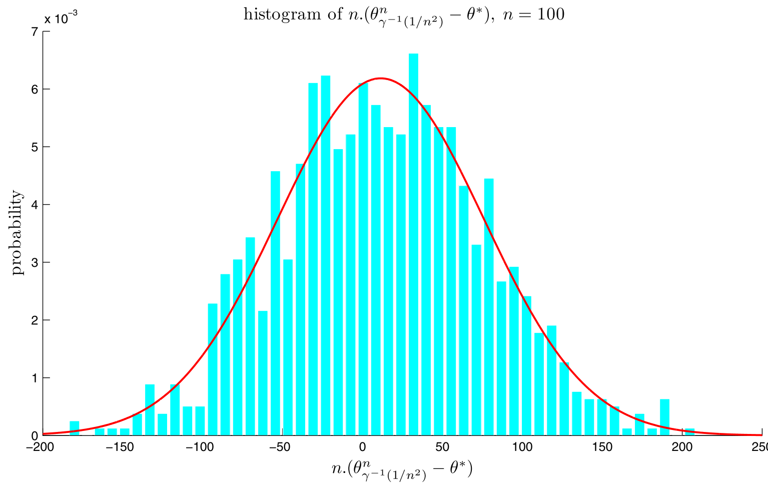

FIG. 2. *Histogram of* $n(\theta^n_{\gamma^{-1}(1/n^2)} - \theta^*)$, $n = 100$, *with* $N = 1000$ *samples.*

being computed with $M = 10^8$ samples. The variance of the Monte Carlo estimator ranges from 2102.4 for $n = 100$ to 53,012.5 for $n = 500$. We set $\gamma_p = \gamma_0/p$ with $\gamma_0 = 200$. We clearly see that $nh^n(\theta^*)$ and $n(\theta^{*,n} - \theta^*)$ are stable with respect to n. The histogram of Figure 2 illustrates Theorem 2.7. The distribution of $n(\theta^n_{\gamma^{-1}(1/n^2)} - \theta^*)$, obtained with $n = 100$ and $N = 1000$ samples, is close to a normal distribution.

4.2. *Computation of the level of an unknown function.* We turn our attention to the computation of the level of the function $\theta \mapsto e^{-rT}\mathbb{E}(X_T - \theta)_+$ (European call option) for which the closed-form formula under the dynamic (4.1) is given by

$$e^{-rT}\mathbb{E}(X_T - \theta)_+ = e^{-rT}x\phi\big(d_+(x, \theta, \sigma)\big) - e^{-rT}\theta\phi\big(d_-(x, \theta, \sigma)\big), \tag{4.2}$$

where $d_\pm(x, y, z) = \log(x/y)/(z\sqrt{T}) \pm z\sqrt{T}/2$. Therefore, we first fix a value θ^* (the target of our procedure) and compute the corresponding level $l = \mathbb{E}(X_T - \theta^*)_+$ by (4.2). Therefore we set $h(\theta) = \mathbb{E}_x[H(\theta, X_T)]$ with $H(\theta, x) = \ell - (x - \theta)_+$. The values of the parameters x, r, σ, T remain unchanged. We plot in Figure 3 the behaviors of $nh^n(\theta^*)$ and $n(\theta^{*,n} - \theta^*)$ for $n = 100, \dots, 500$. As in the previous example, $h^n(\theta^*)$ is approximated by its Monte Carlo estimator, and $\theta^{*,n}$ is estimated by θ^n_M, both estimators being computed with $M = 10^8$ samples. The variance of the Monte Carlo estimator ranges from 9.73×10^6 for $n = 100$ to 9.39×10^7 for $n = 500$.

To compare the three methods in terms of computational costs, we compute the different estimators, namely $\theta^n_{\gamma^{-1}(1/n^2)}$ where $(\theta^n_p)_{p\geq 1}$ is given by (1.3), Θ^{sr}_n and Θ^{ml}_n for a set of $N = 200$ values of the target θ^* equidistributed on the interval $[90, 110]$ and for different values of n. For each value n and for each method we

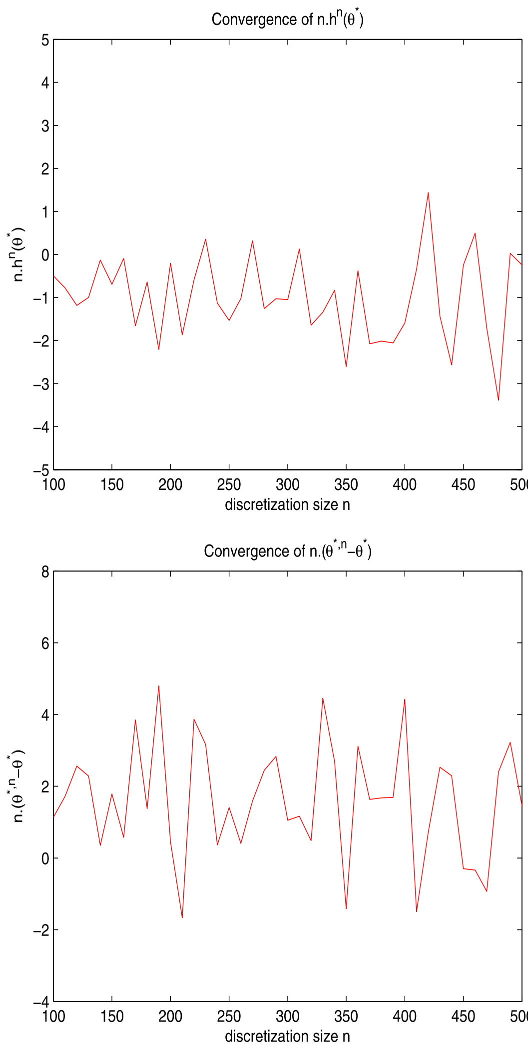

FIG. 3. *On the top*: *weak discretization error* $n \mapsto nh^n(\theta^*)$. *On the bottom*: *implicit discretization error* $n \mapsto n(\theta^{*,n} - \theta^*)$, $n = 100, \ldots, 500$.

compute the complexity given by (2.7), (2.12) and (2.13), respectively, and the root-mean-squared error which is given by

$$\mathrm{RMSE} = \left(\frac{1}{N}\sum_{k=1}^{N}(\Theta_k^n - \theta_k^*)^2\right)^{1/2},$$

where $\Theta_k^n = \theta^n_{\gamma^{-1}(1/n^2)}$, Θ_n^{sr} or Θ_n^{ml} is the considered estimator. For each given n, we provide a couple (RMSE, complexity), which is plotted on Figure 5. Let us note that the multi-level SA estimator has been computed for different values of m

(ranging from $m = 2$ to $m = 7$) and different values of L. We set $\gamma(p) = \gamma_0/p$, with $\gamma_0 = 2$, $p \geq 1$, so that $\beta^* = 1/2$.

From a practical point of view, it is of interest to use the information provided at level 1 by the statistical Romberg SA estimator and at each level by the multi-level SA estimator. More precisely, the initialization point of the SA procedures devised to compute the correction terms $\theta^n_{\gamma_0 n^{3/2}} - \theta^{\sqrt{n}}_{\gamma_0 n^{3/2}}$ (for the statistical Romberg SA) and $\theta^{m^\ell}_{M_\ell} - \theta^{m^{\ell-1}}_{M_\ell}$ (for the multi-level SA) at level ℓ are fixed to $\theta^{\sqrt{n}}_{\gamma_0 n^2}$ and to $\theta^1_{\gamma_0 n^2} + \sum_{\ell=1}^{L-1} \theta^{m^\ell}_{M_\ell} - \theta^{m^{\ell-1}}_{M_\ell}$, respectively. We set $\theta^{n^{1/2}}_0 = \theta^1_0 = x$ for all $k \in \{1, \dots, M\}$ to initialize the procedures. Moreover, by Lemma 5.2, the $L^1(\mathbb{P})$-norm of an increment of a SA algorithm is of order $\sqrt{\gamma_0/p}$ since $\mathbb{E}|\theta^n_{p+1} - \theta^n_p| \leq \mathbb{E}[|\theta^n_{p+1} - \theta^{*,n}|^2]^{1/2} + \mathbb{E}[|\theta^n_p - \theta^{*,n}|^2]^{1/2} \leq C(H, \gamma)\sqrt{\gamma(p)}$. Hence during the first iterations (say, $M/100$ if M denotes the number of samples of the estimator), to ensure that the different procedures do not jump too far ahead in one step, we freeze the value of $\theta^{\sqrt{n}}_{p+1}$ (resp., $\theta^{m^\ell}_{p+1}$) and reset it to the value of the previous step as soon as $|\theta^{\sqrt{n}}_{p+1} - \theta^{\sqrt{n}}_p| \leq K/\sqrt{p}$ (resp., $|\theta^{m^\ell}_{p+1} - \theta^{m^\ell}_p| \leq K/\sqrt{p}$), for a pre-specified value of K. This is just an heuristic approach that notably prevents the algorithm from blowing up during the first steps of the procedure. We select $K = 5$ in the different procedures. Note, however, that this projection-reinitialization step does not lead to additional bias, but slightly increases the complexity of each procedures. In our numerical examples, we observe that it only represents roughly 1–2% of the total complexity.

Now let us interpret Figure 5. The curves of the statical Romberg SA and the multi-level SA methods are displaced below the curve of the SA method. Therefore, for a given error, the complexity of both methods is much lower than that of the crude SA. The difference in terms of computational cost becomes more significant as the RMSE is small, which corresponds to large values of n. The difference between the statistical romberg and the multi-level SA method is not significant for small values of n, that is, for a RMSE between 1 and 0.1. For a RMSE lower than 5×10^{-2}, which corresponds to a number of steps n greater than about 600–700, we observe that the multi-level SA procedure becomes much more effective than both methods. For a RMSE fixed around 1 (which corresponds to $n = 100$ for the SA algorithm and statistical Romberg SA), one divides the complexity by a factor of approximately 5 by using the statistical romberg SA. For a RMSE fixed at 10^{-1}, the computational cost gain is approximately equal to 10 by using either the statistical Romberg SA or the multi-level SA algorithm. Finally, for a RMSE fixed at $5 \times 5 \times 10^{-2}$, the complexity gain achieved by using the multi-level SA procedure instead of the statistical Romberg procedure is approximately equal to 5.

The histograms of Figure 4 illustrate Theorems 2.7, 2.9 and 2.11. The distributions of $n(\theta^n_{\gamma^{-1}(1/n^2)} - \theta^*)$, $n(\Theta^{\mathrm{sr}}_n - \theta^*)$ and $n(\Theta^{\mathrm{ml}}_n - \theta^*)$, obtained with $n = 4^4 = 256$ and $N = 1000$ samples, are close to a normal distribution.

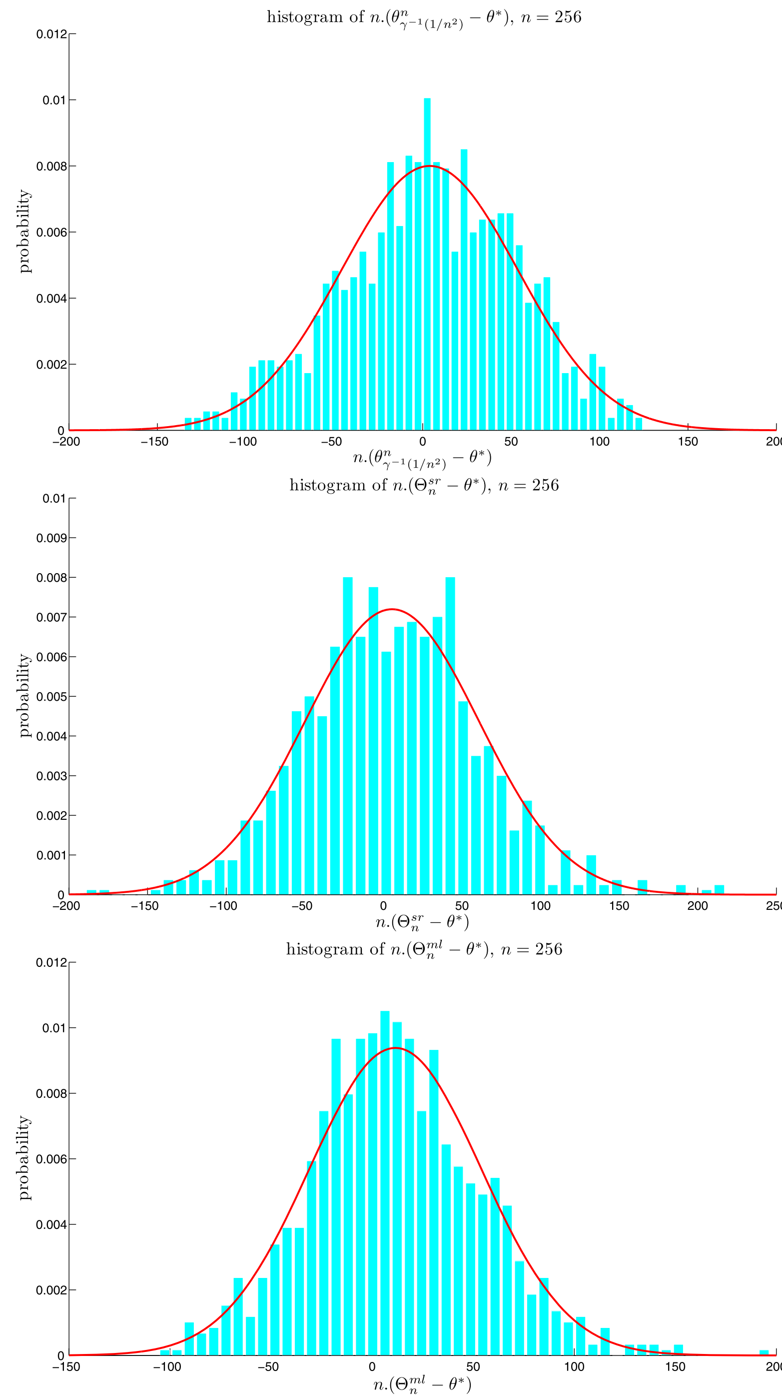

FIG. 4. *Histograms of* $n(\theta_{\gamma^{-1}(1/n^2)} - \theta^*)$, $n(\Theta_n^{\mathrm{sr}} - \theta^*)$ *and* $n(\Theta_n^{\mathrm{ml}} - \theta^*)$ (*from top to bottom*), $n = 256$, *with* $N = 1000$ *samples*.

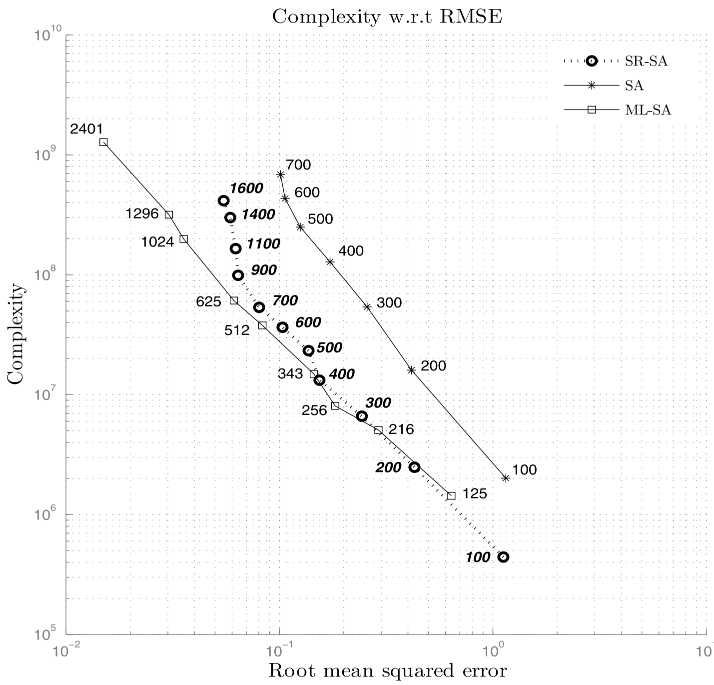

FIG. 5. *Complexity with respect to RMSE.*

5. Technical results. We provide here some useful technical results that are used repeatedly throughout the paper. When the exact value of a constant is not important, we may repeat the same symbol for constants that may change from one line to next.

LEMMA 5.1. *Let H be a stable $d \times d$ matrix, and denote by $\lambda_{\min}$ its eigenvalue with the lowest real part. Let $(\gamma_n)_{n\geq 1}$ be a sequence defined by $\gamma_n = \gamma(n)$, $n \geq 1$, where γ is a positive function defined on $[0, +\infty[$ decreasing to zero and such that $\sum_{n\geq 1} \gamma(n) = +\infty$. Let $a, b > 0$. We assume that γ satisfies one of the following assumptions*:

- *γ varies regularly with exponent $(-c)$, $c \in [0, 1)$; that is, for any $x > 0$, $\lim_{t\to+\infty} \gamma(tx)/\gamma(t) = x^{-c}$;*
- *for $t \geq 1$, $\gamma(t) = \gamma_0/t$ with $b\mathcal{R}e(\lambda_{\min})\gamma_0 > a$.*

Let $(v_n)_{n\geq 1}$ be a nonnegative sequence. Then, for some positive constant C, one has

$$\limsup_n \gamma_n^{-a} \sum_{k=1}^{n} \gamma_k^{1+a} v_k \|\Pi_{k+1,n}\|^b \leq C \limsup_n v_n,$$

where $\Pi_{k,n} := \prod_{j=k}^{n} (I_d - \gamma_j H)$, with the convention $\Pi_{n+1,n} = I_d$.

PROOF. First, from the stability of H, for all $0 < \lambda < \mathcal{R}e(\lambda_{\min})$, there exists a positive constant C such that for any $k \le n$, $\|\Pi_{k+1,n}\| \le C \prod_{j=k}^{n}(1-\lambda\gamma_j)$. Hence we have $\sum_{k=1}^{n} \gamma_k^{1+a} v_k \|\Pi_{k+1,n}\|^b \le C \sum_{k=1}^{n} \gamma_k^{1+a} v_k e^{-\lambda b(s_n - s_k)}$, $n \ge 1$, with $s_n := \sum_{k=1}^{n} \gamma_k$. We set $z_n := \sum_{k=1}^{n} \gamma_k^{1+a} v_k e^{-\lambda b(s_n-s_k)}$. It can be written in the recursive form

$$z_{n+1} = e^{-\lambda b \gamma_{n+1}} z_n + \gamma_{n+1}^{a+1} v_{n+1}, \qquad n \ge 0.$$

Hence a simple induction shows that for any $n > N$, $N \in \mathbb{N}^*$,

$$\begin{aligned} z_n &= z_N \exp\bigl(-\lambda b(s_n - s_N)\bigr) + \exp(-\lambda b s_n) \sum_{k=N+1}^{n} \exp(\lambda b s_k) \gamma_k^{a+1} v_k \\ &\le z_N \exp\bigl(-\lambda b(s_n - s_N)\bigr) + \Bigl(\sup_{k>N} v_k\Bigr) \exp(-\lambda b s_n) \sum_{k=N+1}^{n} \exp(\lambda b s_k)\gamma_k^{a+1}. \end{aligned}$$

We study now the impact of the step sequence $(\gamma_p)_{p\ge1}$ on the above estimate. We first assume that $\gamma_p = \gamma_0/p$ with $b\mathcal{R}e(\lambda_{\min})\gamma_0 > a$. We select $\lambda > 0$ such that $b\mathcal{R}e(\lambda_{\min})\gamma_0 > b\lambda\gamma_0 > a$. Then one has $s_p = \gamma_0 \log(p) + c_1 + r_p$, $c_1 > 0$ and $r_p \to 0$ so that a comparison between the series and the integral yields

$$\exp(-\lambda b s_n) \sum_{k=N+1}^{n} \exp(\lambda b s_k)\gamma_k^{a+1} \le C \frac{1}{n^{b\lambda\gamma_0}} \sum_{k=N+1}^{n} \frac{1}{k^{a-b\lambda\gamma_0+1}} \le \frac{C}{n^a}$$

for some positive constant C (independent of N) so that we clearly have

$$\limsup_n \gamma_n^{-a} z_{n+1} \le C \sup_{k>N} v_k,$$

and we conclude by passing to the limit $N \to +\infty$.

We now assume that γ varies regularly with exponent $-c$, $c \in [0,1)$. Let $s(t) = \int_0^t \gamma(s)\,ds$. We have

$$\begin{aligned} \exp(-\lambda b s_n) \sum_{k=N}^{n} \exp(\lambda b s_k)\gamma_{k+1}^{a+1} &\sim \exp\bigl(-\lambda b s(n)\bigr) \int_0^n \exp\bigl(\lambda b s(t)\bigr)\gamma^{a+1}(t)\,dt \\ &\sim \exp\bigl(-\lambda b s(n)\bigr) \int_0^{s(n)} \exp(\lambda b t)\gamma^a\bigl(s^{-1}(t)\bigr)\,dt, \end{aligned}$$

so that for any x such that $0 < x < 1$, since $t \mapsto \gamma^a(s^{-1}(t))$ is decreasing, we deduce

$$\begin{aligned} &\int_0^{s(n)} \exp(\lambda b t)\gamma^a\bigl(s^{-1}(t)\bigr)\,dt \\ &\quad \le \gamma^a\bigl(s^{-1}(0)\bigr) \int_0^{xs(n)} \exp(\lambda b t)\,dt + \gamma^a\bigl(s^{-1}\bigl(xs(n)\bigr)\bigr) \int_{xs(n)}^{s(n)} \exp(\lambda b t)\,dt \\ &\quad \le \frac{\gamma^a(s^{-1}(0))}{\lambda b} \exp\bigl(\lambda b x s(n)\bigr) + \frac{\gamma^a(s^{-1}(xs(n)))}{\lambda b} \exp\bigl(\lambda b s(n)\bigr). \end{aligned}$$

Hence it follows that

$$\begin{aligned}
&\frac{\exp(-\lambda b s(n))}{\gamma^a(n)} \int_0^{s(n)} \exp(\lambda b t)\gamma^{a+1}(t)\, dt \\
&\quad \leq \frac{\gamma(s^{-1}(0))}{\lambda \gamma^a(n)} \exp\bigl(-\lambda b(1-x)s(n)\bigr) + \frac{\gamma^a(s^{-1}(xs(n)))}{\lambda b \gamma^a(n)},
\end{aligned}$$

and since $t \mapsto \gamma^a(s^{-1}(t))$ varies regular with exponent $-ac/(1-c)$, and $\lim_{n\to+\infty} \frac{1}{\gamma^a(n)} \exp(-\lambda(1-x)s(n)) = 0$,

$$\limsup_{n\to+\infty} \frac{\exp(-\lambda b s(n))}{\gamma^a(n)} \int_0^{s(n)} \exp(\lambda b t)\gamma^{a+1}(t)\, dt \leq \frac{x^{-ac/(1-c)}}{\lambda b}.$$

An argument similar to the previous case completes the proof. □

We omit the proof of the following lemma, which is quite standard, and refer the interested reader to [11].

LEMMA 5.2. *Let $(\theta_p^n)_{p\geq 0}$ be the procedure defined by* (1.3) *where θ_0^n is independent of the innovation of the algorithm with* $\sup_{n\geq 1} \mathbb{E}|\theta_0^n|^2 < +\infty$. *Suppose that the assumptions of Theorem* 2.6 *are satisfied and that the mean-field function h^n satisfies*

$$\exists \underline{\lambda} > 0, \forall n \in \mathbb{N}^*, \forall \theta \in \mathbb{R}^d, \qquad \langle \theta - \theta^{*,n}, h^n(\theta)\rangle \geq \underline{\lambda}|\theta - \theta^{*,n}|^2, \tag{5.1}$$

where $\theta^{*,n}$ is the unique zero of h^n satisfying $\sup_{n\geq 1} |\theta^{*,n}| < +\infty$. *Moreover, we assume that γ satisfies one of the following assumptions*:

- *γ varies regularly with exponent $(-c)$, $c \in [0,1)$; that is, for any $x > 0$,* $\lim_{t\to+\infty} \gamma(tx)/\gamma(t) = x^{-c}$;
- *for $t \geq 1$, $\gamma(t) = \gamma_0/t$ with $2\underline{\lambda}\gamma_0 > 1$.*

Then for some positive constant C (*independent of p and n*), *one has*

$$\forall p \geq 1, \qquad \sup_{n\geq 1} \mathbb{E}\bigl[|\theta_p^n - \theta^{*,n}|^2\bigr] + \mathbb{E}\bigl[|\theta_p - \theta^*|^2\bigr] \leq C\gamma(p).$$

PROPOSITION 5.1. *Assume that the assumptions of Theorem* 2.10 *are satisfied. Then, for all $n \in \mathbb{N}$ there exist two sequences $(\tilde{\mu}_p^n)_{p\in[\![0,n]\!]}$ and $(\tilde{r}_p^n)_{p\in[\![0,n]\!]}$ with $\tilde{r}_0^n = \theta_0^n - \theta_0 - (\theta^* - \theta^{*,n})$ such that*

$$\forall p \in [\![0,n]\!], \qquad z_p^n = \theta_p^n - \theta^{*,n} - \bigl(\theta_p - \theta^*\bigr) = \tilde{\mu}_p^n + \tilde{r}_p^n,$$

and satisfying for all $n \in \mathbb{N}$, for all $p \in [\![1,n]\!]$,

$$\sup_{p\geq 1} \gamma_p^{-1/2} \mathbb{E}|\tilde{\mu}_p^n| < Cn^{-\rho}, \qquad \sup_{n\geq 1, p\geq 0} \gamma_p^{-1} \mathbb{E}\bigl[|\tilde{r}_p^n|\bigr] < +\infty.$$

PROOF. Using (3.2), we define the two sequences $(\tilde{\mu}_p^n)_{p\in[\![0,n]\!]}$ and $(\tilde{r}_p^n)_{p\in[\![0,n]\!]}$ by

$$\begin{aligned}\tilde{\mu}_p^n &= \sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}\Delta N_k^n + \sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}\big(Dh(\theta^*) - Dh^n(\theta^{*,n})\big)\big(\theta_{k-1}^n - \theta^{*,n}\big)\\ &\quad + \sum_{k=1}^{p}\gamma_k\Pi_{k+1,n}\big(h^n(\theta^{*,n}) - h^n(\theta^*)\\ &\qquad\qquad - \big(H(\theta^{*,n},(U^n)^{k+1}) - H(\theta^*,(U^n)^{k+1})\big)\big)\end{aligned}$$

and

$$\begin{aligned}\tilde{r}_p^n &= \Pi_{1,p}z_0^n + \sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}\big(\zeta_{k-1}^n - \zeta_{k-1}\big)\\ &\quad + \sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}\big(h^n(\theta_k^n) - h^n(\theta^{*,n})\\ &\qquad\qquad - \big(H(\theta_k^n,(U^n)^{k+1}) - H(\theta^{*,n},(U^n)^{k+1})\big)\big)\\ &\quad + \sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}\big(H(\theta_k,U^{k+1}) - H(\theta^*,U^{k+1}) - \big(h(\theta_k) - h(\theta^*)\big)\big).\end{aligned}$$

We first focus on the sequence $(\tilde{\mu}_p^n)_{p\in[\![0,n]\!]}$. Moreover, by the definition of the sequence $(\Delta N_k^n)_{k\in[\![1,n]\!]}$ and the Cauchy–Schwarz inequality, we derive

$$\begin{aligned}&\mathbb{E}\left|\sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}\Delta N_k^n\right|\\ &\quad\le C\big(\mathbb{E}\big|H(\theta^*,U^n) - H(\theta^*,U)\big|^2\big)^{1/2}\left(\sum_{k=1}^{p}\gamma_k^2\|\Pi_{k+1,p}\|^2\right)^{1/2}\\ &\quad= \mathcal{O}\big(\gamma_p^{1/2}n^{-\rho}\big).\end{aligned}$$

Taking the expectation for the third term and following along the lines of the proof of Lemma 2.7, we obtain

$$\begin{aligned}&\mathbb{E}\left|\sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}\big(Dh(\theta^*) - Dh^n(\theta^{*,n})\big)\big(\theta_{k-1}^n - \theta^{*,n}\big)\right|\\ &\quad\le C\sum_{k=1}^{p}\gamma_k^{3/2}\|\Pi_{k+1,p}\|\big(\big|\theta^{*,n} - \theta^*\big| + \big\|Dh(\theta^*) - Dh^n(\theta^*)\big\|\big)\\ &\quad= \mathcal{O}\big(\gamma_p^{1/2}n^{-\rho}\big).\end{aligned}$$

Finally we take the square of the L^2-norm of the last term and use Lemma 5.1 to derive

$$\begin{aligned}\mathbb{E}\left|\sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}\big(h^n(\theta^{*,n})-h^n(\theta^*)-\big(H(\theta^{*,n},(U^n)^{k+1})-H(\theta^*,(U^n)^{k+1})\big)\big)\right|^2\\ \le|\theta^*-\theta^{*,n}|^2\sum_{k=1}^{p}\gamma_k^2\|\Pi_{k+1,p}\|^2\\ =\mathcal{O}(\gamma_p n^{-2\rho}).\end{aligned}$$

We now prove the bound concerning the sequence $(\tilde{r}_p^n)_{p\in[\![0,n]\!]}$. Under the assumption on the step sequence, we have

$$\mathbb{E}\big[|\Pi_{1,p}z_0^n|\big]\le\|\Pi_{1,p}\|\big(1+|\theta^*-\theta^{*,n}|\big)=\mathcal{O}(\gamma_p).$$

By Lemma 5.2, we derive

$$\sup_{n\ge1}\mathbb{E}\left|\sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}(\zeta_{k-1}^n-\zeta_{k-1})\right|\le C\sum_{k=1}^{p}\gamma_k^2\|\Pi_{k+1,p}\|=\mathcal{O}(\gamma_p).$$

Concerning the second term, following along the lines of the proof of Lemma 2.7, we simply take the square of its $L^2(\mathbb{P})$-norm to derive

$$\begin{aligned}\sup_{n\ge1}\mathbb{E}\Bigg|\sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}\big(h^n(\theta_k^n)-h^n(\theta^{*,n})\\ -\big(H(\theta_k^n,(U^n)^{k+1})-H(\theta^{*,n},(U^n)^{k+1})\big)\big)\Bigg|^2\\ \le C\sum_{k=1}^{p}\gamma_k^3\|\Pi_{k+1,p}\|^2\\ =\mathcal{O}(\gamma_p^2)\end{aligned}$$

and similarly

$$\begin{aligned}\mathbb{E}\left|\sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}\big(\big(H(\theta_k,U^{k+1})-H(\theta^*,U^{k+1})\big)-\big(h(\theta_k)-h(\theta^*)\big)\big)\right|^2\\ =\mathcal{O}(\gamma_p^2).\end{aligned}$$

□

REFERENCES

[1] ALDOUS, D. J. and EAGLESON, G. K. (1978). On mixing and stability of limit theorems. *Ann. Probab.* **6** 325–331. MR0517416

[2] BARDOU, O., FRIKHA, N. and PAGÈS, G. (2009). Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling. *Monte Carlo Methods Appl*. **15** 173–210. MR2573212

[3] BARDOU, O., FRIKHA, N. and PAGÈS, G. (2009). Recursive computation of value-at-risk and conditional value-at-risk using MC and QMC. In *Monte Carlo and Quasi-Monte Carlo Methods* 2008 193–208. Springer, Berlin. MR2743895

[4] BARDOU, O., FRIKHA, N. and PAGÈS, G. (2010). CVaR hedging using quantization based stochastic approximation algorithm. *Math*. *Finance*. To appear.

[5] BENVENISTE, A., MÉTIVIER, M. and PRIOURET, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. *Applications of Mathematics* (*New York*) **22**. Springer, Berlin. MR1082341

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

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

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

[9] DUFLO, M. (1996). *Algorithmes Stochastiques*. *Mathématiques & Applications* (*Berlin*) [*Mathematics & Applications*] **23**. Springer, Berlin. MR1612815

[10] FATHI, M. and FRIKHA, N. (2013). Transport-entropy inequalities and deviation estimates for stochastic approximations schemes. *Electron*. *J*. *Probab*. **18** 1–36. MR3084653

[11] FRIKHA, N. (2014). Multilevel stochastic approximation. Preprint HAL (hal-00870585).

[12] FRIKHA, N. (2014). Shortfall risk minimization in discrete time financial market models. *SIAM J*. *Financial Math*. **5** 384–414. MR3211032

[13] FRIKHA, N. and MENOZZI, S. (2012). Concentration bounds for stochastic approximations. *Electron*. *Commun*. *Probab*. **17** no. 47, 15. MR2988393

[14] GILES, M. (2008). Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and Quasi-Monte Carlo Methods* 2006 343–358. Springer, Berlin. MR2479233

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

[16] GILES, M. B., HIGHAM, D. J. and MAO, X. (2009). Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff. *Finance Stoch*. **13** 403–413. MR2519838

[17] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*: *Probability and Mathematical Statistics*. Academic Press, New York. MR0624435

[18] HEINRICH, S. (2001). Multilevel Monte Carlo methods. In *Large-Scale Scientific Computing*. Springer, Berlin.

[19] JACOD, J. (1998). Rates of convergence to the local time of a diffusion. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **34** 505–544. MR1632849

[20] JACOD, J. and PROTTER, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann*. *Probab*. **26** 267–307. MR1617049

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

[22] KUSHNER, H. J. and YIN, G. G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **35**. Springer, New York. MR1993642

[23] PELLETIER, M. (1998). Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann*. *Appl*. *Probab*. **8** 10–44. MR1620405

[24] POLYAK, B. T. and JUDITSKY, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J*. *Control Optim*. **30** 838–855. MR1167814

[25] RÉNYI, A. (1963). On stable sequences of events. *Sankhyā Ser. A* **25** 293–302. MR0170385

[26] ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* **22** 400–407. MR0042668

[27] RUPPERT, D. (1991). Stochastic approximation. In *Handbook of Sequential Analysis. Statist. Textbooks Monogr.* **118** 503–529. Dekker, New York. MR1174318

LPMA
UNIVERSITÉ PARIS DIDEROT
BÂT. SOPHIE GERMAIN
5 RUE THOMAS MANN
75205 PARIS, CEDEX 13
FRANCE
E-MAIL: frikha@math.univ-paris-diderot.fr