MULTI-LEVEL STOCHASTIC APPROXIMATION ALGORITHMS

N. Frikha[1]

Abstract. This paper studies multi-level stochastic approximation algorithms. Our aim is to extend the scope of the multilevel Monte Carlo method recently introduced by Giles [Gil08b] 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 multi-level is proposed. We prove a central limit theorem for both methods and describe the possible optimal choices of step size sequence. Numerical results confirm the theoretical analysis and show a significant reduction in the initial computational cost.

1991 Mathematics Subject Classification. 60F05, 62K12, 65C05, 60H35.

October 8, 2013.

1. Introduction

A basic problem in numerical probability is the computation of quantities like $\mathbb{E}_x[f(X_T)]$ for a given function $f:\mathbb{R}^q\to\mathbb{R}$ and where $X:=(X_t)_{t\in[0,T]}$ is a q-dimensional diffusion process defined on a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$, satisfying the usual conditions, and solution to the following stochastic differential equation (SDE)

$$X_t = x+\int_0^t b(X_s)ds+\int_0^t \sigma(X_s)dW_s, \qquad (SDE_{b,\sigma})$$

where $(W_t)_{t\geq 0}$ is a q'-dimensional $(\mathcal{F}_t)_{t\geq 0}$ Brownian motion and the coefficients b,σ are assumed to be Lipschitz-continuous. For instance, it appears in mathematical finance and represents the price of a European option with maturity T when the dynamic of the underlying assets is given by $(SDE_{b,\sigma})$. Under suitable assumptions on the function f and the coefficients b,σ, namely smoothness or non degeneracy, it can also be related to the Feynman-Kac representation of the heat equation associated to the generator of X. In order to do this, the first step consists in discretizing $(SDE_{b,\sigma})$ using the continuous Euler-Maruyama scheme $(X_t^n)_{t\in[0,T]}$ with time step $\Delta=T/n$ and regular points $t_i=i\Delta$, $i=0,\cdots,n$, namely

$$X_t^n = x+\int_0^t b(X^n_{\phi_n(s)})ds+\int_0^t \sigma(X^n_{\phi_n(s)})dW_s,\quad \phi_n(s)=\sup\{t_i:\ t_i\leq s\}\,. \qquad (1.1)$$

This step introduces the so-called *weak-error* $\mathbb{E}_x[f(X_T)]-\mathbb{E}_x[f(X_T^n)]$ which has been widely investigated in the literature. Since the seminal work of [TT90], it is known that, under smoothness assumption on the coefficients b, σ, the continuous Euler scheme produces a weak error of order Δ. In a hypoelliptic setting for the

Keywords and phrases: Multi-level Monte Carlo methods, stochastic approximation, Euler scheme

[1] LPMA, Université Paris Diderot, Bât. Sophie Germain, 5 Rue Thomas Mann, 75205 Paris, Cedex 13, frikha@math.univ-paris-diderot.fr, http://www.proba.jussieu.fr/pageperso/frikha/

coefficients b and σ and for a bounded measurable function f, Bally and Talay [BT96] obtained the expected order using Malliavin calculus. The second step consists in approximating the expectation $\mathbb{E}_x[f(X_T^n)]$ using a Monte Carlo estimator $M^{-1} \times \sum_{k=1}^{M} f((X_T^n)^j)$ where the $((X_T^n)^j)_{j\in[\![1,M]\!]}$ are M independent copies of the Euler-Maruyama scheme starting at the initial value x at time 0. This step gives rise to a statistical error $\mathbb{E}_x[f(X_T^n)] - M^{-1} \times \sum_{j=1}^{M} f((X_T^n)^j)$. Given the order of the weak error, a natural question is to find the optimal choice of the step 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 Monte Carlo algorithm is obtained for $M = n^{2\alpha}$, see [DG95]. So, the computational cost of the algorithm is $C_{MC} = C \times n^{2\alpha+1}$, for a positive constant $C > 0$.

In order to reduce the complexity of the computation, Kebaier [Keb05] introduced a two-level Monte Carlo scheme, originally referred as statistical Romberg method, which uses two Euler schemes with time step T/n and T/n^{β}, $\beta \in (0,1)$ and approximates $\mathbb{E}_x[f(X_T)]$ by

$$\frac{1}{n^{\gamma_1}} \sum_{j=1}^{n^{\gamma_1}} f((\hat{X}_T^{n^\beta})^j) + \frac{1}{n^{\gamma_2} T} \sum_{j=1}^{n^{\gamma_2} T} f((X_T^n)^j) - f((X_T^{n^\beta})^j)$$

where $\hat{X}^{n^\beta}$ is a second Euler-Maruyama scheme with time step T/n^{β} generated with brownian paths which are independent of the ones used to simulate X^{n^β} and X^n. If the weak error is of order $n^{-\alpha}$ then to achieve a global error of order $n^{-\alpha}$, $\alpha \in [1/2, 1]$, the optimal choice, that is the one minimizing the complexity, is obtained for $\gamma_1 = 2\alpha$ and $\gamma_2 = 2\alpha - \beta$ and $\beta = 1/2$ leading to an optimal complexity of order $n^{2\alpha+\frac{1}{2}}$ which is lower than the classical complexity C_{MC}.

Generalizing Kebaier's approach, Giles [Gil08b] proposed a multi-level Monte Carlo algorithm which relies on devising Euler schemes with a geometric sequence of different time steps T/m^{ℓ}, $\ell = 0, \cdots, L$, $m \in \mathbb{N}^* \backslash \{1\}$ s.t. $m^L = n$ and approximates $\mathbb{E}_x[f(X_T)]$ by

$$\frac{1}{n^{\gamma_0}} \sum_{j=1}^{n^{\gamma_0}} f((X_T^1)^j) + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{j=1}^{N_\ell} \left(f((X_T^{m^\ell})^j) - f((X_T^{m^{\ell-1}})^j) \right),$$

where all these $L+1$ empirical mean sequences are based on independent samples. For each level $\ell \in \{1, \cdots, L\}$, and each sample $j \in \{1, \cdots, N_\ell\}$, $(X_T^{m^\ell})^j$ and $(X_T^{m^{\ell-1}})^j$ are based on the same path but with two different time steps. Based on an analysis of the variance, Giles [Gil08b] proposed an optimal choice for the sequence $(N_\ell)_{1\le \ell \le L}$ which minimizes the total complexity of the algorithm. More recently, Ben Alaya and Kebaier [AK12] proposed a different analysis to obtain the optimal choice of the parameters relying on a Lindeberg Feller central limit theorem for the multi-level Monte Carlo algorithm. To achieve a global error of order $n^{-\alpha}$, both approaches lead to a complexity of order $n^{2\alpha}(\log n)^2$ which is significantly lower than the computational costs of the Monte Carlo and the statistical Romberg methods. For further developments on multi-level Monte Carlo methods, we refer to Giles [Gil08a], Dereich [Der11], Giles, Higham and Mao [GHM09] among others.

In the present paper, we are interested in broadening the scope of the multi-level Monte Carlo method to the framework of stochastic approximation algorithm. Introduced by Robbins and Monro [RM51], these recursive simulation based algorithms appear as effective and widely used procedures 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 a $\mathbb{R}^q$-valued random vector. To estimate the zero of h, they proposed the algorithm

$$\theta_{p+1} = \theta_p - \gamma_{p+1} H(\theta_p, U^{p+1}), \quad p \ge 0 \tag{1.2}$$

where $(U^p)_{p\geq 1}$ is an i.i.d. sequence of random variables with the same law as 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 non-negative deterministic and decreasing steps satisfying the assumption

$$\sum_{p\geq 1} \gamma_p = +\infty, \quad \text{and} \quad \sum_{p\geq 1} \gamma_p^2 < +\infty. \tag{1.3}$$

When the function h is the gradient of a convex potential, the recursive procedure (1.2) is a stochastic gradient algorithm. Indeed, replacing $H(\theta_p, U^{p+1})$ by $h(\theta_p)$ in (1.2) 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 (1.2) aims to compute the value $\bar{\theta}$ such that $k(\bar{\theta}) = \ell$.

In many applications, notably in computational finance, we are interested in the computation of the zero θ^* of h given by $h(\theta) := \mathbb{E}_x[H(\theta, X_T)]$, where $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$ is a Borel function and X_T is the value at time T of the SDE given by $(SDE_{b,\sigma})$. For instance, the computations of the implied volatility or the implied correlation boils down to finding the zero of an unknown function. 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 $(SDE_{b,\sigma})$ also appears as an inverse problem for which a stochastic approximation may be devised, see e.g. [BFP09a] and [BFP09b]. Risk minimizing a financial portfolio by means of stochastic approximation has been studied in [BFP10]. For more applications and a complete overview in the theory of stochastic approximation, the reader may refer to [Duf96], [KY03] and [BMP90].

The function h is generally neither known nor computable and since the random variable X_T cannot be simulated in general, estimating θ^* by devising directly the recursive scheme (1.2) is not possible. Therefore, two steps are needed to compute θ^*:

- the first step consists in approximating the dynamic of $(X_t)_{t\in[0,T]}$ by its Euler-Maruyama discretization scheme $(X_t^n)_{t\in[0,T]}$ given by (1.1) that can be easily simulated. Hence, the zero θ^* of h is approximated by the zero $\theta^{*,n}$ of h^n defined by $h^n(\theta) := \mathbb{E}_x[H(\theta, X_T^n)]$, $\theta \in \mathbb{R}^d$. It induces *an implicit discretization error* which writes

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

Let us note that $\theta^{*,n}$ appears as a proxy of θ^* and one would naturally expect that $\theta^{*,n} \to \theta^*$ as the number n of time step in the Euler-Maruyama scheme tends to infinity.

- the second step consists in approximating $\theta^{*,n}$ involving the scheme (1.1) by $M \in \mathbb{N}^*$ steps of the following stochastic approximation scheme

$$\theta_{p+1}^n = \theta_p^n - \gamma_{p+1} H(\theta_p^n, (X_T^n)^{p+1}), \; p \in [\![0, M-1]\!], \tag{1.4}$$

where $((X_T^n)^p)_{p\in[\![1,M]\!]}$ is an i.i.d. sequence of random variables with the same law as X_T^n, θ_0^n is independent of the innovation of the algorithm with $\sup_{n\geq 1} \mathbb{E}|\theta_0^n|^2 < +\infty$ and $\gamma = (\gamma_p)_{p\geq 1}$ is a sequence of non-negative deterministic and decreasing steps satisfying (1.3). This induces a *statistical error* which writes

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

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

$$\begin{aligned}\mathcal{E}_{glob}(M, \gamma, H) &= \theta^* - \theta^{*,n} + \theta^{*,n} - \theta_M^n \\ &:= \mathcal{E}_D(n, T, b, \sigma, H) + \mathcal{E}_S(n, M, \gamma, T, H).\end{aligned}$$

The first step of our analysis consists in investigating the behavior of the *implicit discretization error* $\mathcal{E}_D(n, T, b, \sigma, H)$. Under mild assumptions on the functions h and h^n, namely the local uniform convergence of

$(h^n)_{n\geq 1}$ towards h and a mean reverting assumption of h and h^n, we prove that $\lim_n \mathcal{E}_D(n,T,b,\sigma,H) = 0$. We next show that under stronger assumption, namely the local uniform convergence of $(Dh^n)_{n\geq 1}$ towards Dh and the non-singularity of $Dh(\theta^*)$, the rate of convergence of the *standard weak discretization* $h^n(\theta) - h(\theta)$, for a fixed $\theta \in \mathbb{R}^d$, transfers to the *implicit discretization error* $\mathcal{E}_D(n,T,b,\sigma,H) = \theta^* - \theta^{*,n}$.

Regarding the *statistical error* $\mathcal{E}_S(n,M,\gamma,T,H) := \theta^{*,n} - \theta^n_M$, it is well-known that under standard assumptions, i.e. a mean reverting assumption on h^n and a growth control of the L^2-norm of the noise of the algorithm, the Robbins-Monro theorem guarantees that $\lim_M \mathcal{E}_S(n,M,\gamma,T,H) = 0$ for each fixed $n \in \mathbb{N}^*$, see Theorem 2.3 below. Moreover, under mild technical conditions, a central limit theorem (CLT) holds at rate $\gamma^{-1/2}(M)$, that is, for each fixed $n \in \mathbb{N}^*$, $\gamma^{-1/2}(M)\mathcal{E}_S(n,M,\gamma,T,H)$ 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 [FM12] and [FF13] for some recent developments on non-asymptotic 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 $\gamma_0 > 1/(2\mathcal{R}e(\lambda_{min}))$ where λ_{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 situation.

To circumvent such a difficulty, it is fairly well-known that the key idea is to carefully smooth the trajectories of a converging stochastic approximation algorithm by averaging according to the *Ruppert & Polyak averaging principle*, see e.g. [Rup91] and [PJ92]. It consists in devising the original stochastic approximation algorithm (1.4) with a slow decreasing step and to simultaneously compute the empirical mean $(\bar{\theta}^n_p)_{p\geq 1}$ (which *a.s.* converges to $\theta^{*,n}$) of the sequence $(\theta^n_p)_{p\geq 0}$ by setting

$$\bar{\theta}^n_p = \frac{\theta^n_0 + \theta^n_1 + \cdots + \theta^n_p}{p+1} = \bar{\theta}^n_{p-1} - \frac{1}{p+1}\left(\bar{\theta}^n_{p-1} - \theta^n_p\right). \tag{1.5}$$

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

Given the order of the implicit discretization error and a step sequence γ satisfying (1.3) with which the procedure (1.4) to estimate $\theta^{*,n}$ is devised, a natural question is to find the optimal balance between reducing the time step T/n in the discretization scheme (1.1) and increasing the number M of steps in (1.4) to achieve a given global error. This problem was originally investigated and solved in [DG95] for the Monte Carlo approximation of $\mathbb{E}_x[f(X_T)]$. Their result implies that it is optimal to have $M = n^{2\alpha}$ Monte Carlo simulations when the weak discretization error is of the order $n^{-\alpha}$, $\alpha > 0$. The error between θ^* and the approximation θ^n_M writes $\theta^n_M - \theta^* = \theta^n_M - \theta^{*,n} + \theta^{*,n} - \theta^*$ suggesting to select $M = \gamma^{-1}(1/n^{2\alpha})$, where γ^{-1} is the inverse function of γ, when the *weak discretization error* is of order $n^{-\alpha}$. However, due to the non-linearity of the stochastic approximation algorithm (1.4), the methodology developed in [DG95] does not apply in our context. The key tool to tackle this question consists in linearizing the dynamic of $(\theta^n_p)_{p\in[\![1,M]\!]}$ around its target $\theta^{*,n}$, quantifying the contribution of the non linearities in the space variable θ^n_p and the innovations and finally exploiting stability arguments from stochastic approximation schemes. Optimizing with respect to the usual choice of the step sequence, the minimal computational cost to achieve an error of order $n^{-\alpha}$ which is given by $C_{\mathrm{SA}} = C \times n \times \gamma^{-1}(1/n^{2\alpha})$ is reached by setting $\gamma(p) = \gamma_0/p$, $p \geq 1$, provided that the constant γ_0 satisfies a stringent condition involving h^n, leading to a complexity of order $n^{2\alpha+1}$. We also obtain that the optimal complexity is reached for free without any condition on γ_0 when considering the empirical mean sequence $(\bar{\theta}^n_p)_{p\in[\![1,n^{2\alpha}]\!]}$.

To reduce the computational cost of estimating θ^* by means of stochastic approximation algorithm, we investigate in a second part multi-level stochastic approximation algorithms. The first one is a two-level stochastic approximation scheme, also referred as the statistical Romberg stochastic approximation method, that approximates the unique zero θ^* of h by $\Theta^{sr}_n = \theta^{n^\beta}_{M_1} + \theta^n_{M_2} - \theta^{n^\beta}_{M_2}$, $\beta \in (0,1)$. The couple $(\theta^n_{M_2}, \theta^{n^\beta}_{M_2})$ is computed using two Euler discretization schemes with different time steps but with the same Brownian motions and the Brownian paths used for the computation of $\theta^{n^\beta}_{M_1}$ are independent of those used for the computation of $(\theta^n_{M_2}, \theta^{n^\beta}_{M_2})$. For an implicit discretization error of order $n^{-\alpha}$, we prove a CLT for the sequence $(\Theta^{sr}_n)_{n\geq 1}$

through which we are able to optimally fix 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 $(\theta_p)_{p\in[\![0,M_2]\!]}$ devised with the innovation variables $((X_T)^p)_{p\geq 1}$ so that the correction term writes $\theta^n_{M_2} - \theta_{M_2} - (\theta^{n^\beta}_{M_2} - \theta_{M_2})$. The key idea is then to quantify the two main contributions in this decomposition, namely the one due to the non linearity in the space variables $(\theta_p^{n^\beta}, \theta_p^n, \theta_p)_{p\in[\![0,M_2]\!]}$ on one hand and the one due to the non linearity in the innovation variables $((X_T^{n^\beta})^p, (X_T^n)^p, (X_T)^p)_{p\geq 1}$ in the other hand. Under mild smoothness assumption on the function H, the weak rate of convergence is ruled by the non linearity in the innovation variables for which we use the weak convergence of the normalized error $\sqrt{n/T}(X_T^n - X_T)$ of the Euler scheme for diffusions proved in [JP98]. 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 $C_{\text{SA-SR}} = C \times n^{2\alpha+1/2}$, $C > 0$ provided that γ_0 satisfies again a condition involving h^n which is difficult to handle in practice. By considering the empirical mean sequence $\bar{\Theta}_n^{sr} = \bar{\theta}^{n^\beta}_{M_3} + \bar{\theta}^n_{M_4} - \bar{\theta}^{n^\beta}_{M_4}$, 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 sequence, this optimal complexity is reached for free by setting $M_3 = n^{2\alpha}$, $M_4 = n^{2\alpha-\beta}T$ without any condition on γ_0.

Moreover, we generalize this first approach to the case of the multi-level stochastic approximation method. In the spirit of [Gil08b] for Monte Carlo path simulation, the multi-level stochastic approximation scheme estimates θ^* by computing the quantity $\Theta_n^{ml} = \theta^1_{M_0} + \sum_{\ell=1}^L \theta^{m^\ell}_{M_\ell} - \theta^{m^{\ell-1}}_{M_\ell}$ based on Euler schemes with the same geometric sequence of time steps as for the estimation of $\mathbb{E}_x[f(X_T)]$. Here again to establish a CLT for this estimator (as in [AK12] 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 $M_0 = \gamma^{-1}(1/n^2)$, $M_\ell = \gamma^{-1}(m^\ell \log(m)/(n^2\log(n)(m-1)T))$, $\ell = 1, \cdots, L$. Once again the step sequence $\gamma(p) = \gamma_0/p$, $p \geq 1$, is optimal among the usual choices of step sequence and it induces a complexity for the procedure given by $C_{\text{SA-ML}} = C \times n^2(\log(n))^2$ which is of the same order as the one obtained in [Gil08b] and [AK12].

The paper is organized as follows. Basic results concerning the Euler-Maruyama discretization scheme and stochastic approximation schemes are briefly presented in the next section. We also investigate the behavior of the implicit discretization error and derive the optimal balance between reducing the time step T/n and increasing the number of steps in the stochastic approximation procedure to achieve a given global error. In Section 3 we present and study the multi-level stochastic approximation algorithms. 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. General framework

In this section, we present some basic results concerning the Euler-Maruyama discretization scheme and stochastic approximation schemes. In the present paper, we make no attempt to provide an exhaustive discussion related to convergence results. We refer readers to [Duf96], [KY03] and [BMP90] among others for developments and a more complete overview in the theory of stochastic approximation.

2.1. On some basic results related to the Euler-Maruyama scheme

In the current work, we assume that the coefficients of $(SDE_{b,\sigma})$ satisfy the mild smoothness condition:

(HS) The coefficients b, σ are uniformly Lipschitz continuous.

(HD) The coefficients b, σ satisfy **(HS)** and are continuously differentiable.

Throughout the paper, we will use these well-known properties concerning the Euler-Maruyama scheme which are valid under **(HS)**, namely

$$\forall p \geq 1,\ \exists C := C(p,T,b,\sigma) > 0\ \ \mathbb{E}_x[\sup_{0\leq t\leq T} |X_t - X_t^n|^p]^{1/p} \leq \frac{C}{n^{1/2}}, \tag{2.6}$$

and

$$\forall p \geq 1,\ \exists K := K(p,x,T,b,\sigma) > 0\ \ \mathbb{E}_x[\sup_{0\leq t\leq T}|X_t|^p]^{1/p} + \mathbb{E}_x[\sup_{0\leq t\leq T}|X_t^n|^p]^{1/p} \leq K. \tag{2.7}$$

Now we turn our attention to the weak convergence rate of the Euler scheme. We follow the notation of [JP98]. 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 $X_n \overset{stably}{\Longrightarrow} X$, if for all bounded random variable U defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and for all $h : E \to \mathbb{R}$ bounded continuous, one has

$$\mathbb{E}Uh(X_n) \to \tilde{\mathbb{E}}Uh(X), \quad n \to +\infty.$$

Stable convergence was introduced in [Rén63] and notably investigated in [AE78].The following results will be useful in order to derive a CLT for multi-level stochastic approximation algorithms. We first introduce some notations, namely

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

and $dY_t = (dt\ dW_t^1\ \cdots\ dW_t^{q'})^T$ where here as below u^T denotes the transpose of the vector u. Consequently, the SDE $(SDE_{b,\sigma})$ can be written in the compact form

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

with its continuous Euler-Maruyama scheme with time step $\Delta = T/n$

$$X_t^n = x + \int_0^t f(X_{\phi_n(s)}^n)dW_s.$$

The following result is due to [JP98], Theorem 3.2 p.276 and Theorem 5.5, p.293.

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

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

the process U *being defined by* $U_0 = 0$ *and*

$$dU_t^i = \sum_{j=1}^{q'+1}\sum_{k=1}^{q} f_k^{'ij}(X_t)\left[U_t^k dY_t^j - \sum_{\ell=1}^{q'+1} f^{k\ell}(X_t)dZ_t^{\ell j}\right] \tag{2.8}$$

where $f_k^{'ij}$ *is the* k*th partial derivative of* f^{ij} *and*

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

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

The following lemma is a basic result on stable convergence that will be useful throughout the paper. Here, E and F will denote two Polish spaces. We consider a sequence $(X_n)_{n\geq 1}$ of E-valued random variable defined on $(\Omega, \mathcal{F})$.

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

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

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

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

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

To prove a CLT for the multi-level stochastic approximation method, we will also need the following result which is due to [AK12], Theorem 4.

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

$$\sqrt{\frac{m^\ell}{(m-1)T}}(X^{m^\ell} - X^{m^{\ell-1}}) \stackrel{stably}{\Longrightarrow} U, \quad as \quad \ell \to +\infty.$$

2.2. On some basic results related to stochastic approximation

The stochastic approximation provides various theorems that guarantee the $a.s.$ and/or L^p convergence of stochastic approximation algorithms. We provide below a general result in order to derive the $a.s.$ convergence of such procedures. It is also known as *Robbins-Monro Theorem* and covers most situations (see the remark below).

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

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

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

$$\forall \theta \in \mathbb{R}^d, \theta \neq \theta^*, \quad \langle \theta - \theta^*, h(\theta) \rangle > 0. \tag{2.9}$$

Let $\gamma = (\gamma_p)_{p\geq 1}$ be a sequence of gain parameters satisfying (1.3). Suppose that

$$\forall \theta \in \mathbb{R}^d, \quad \mathbb{E}|H(\theta, U)|^2 \leq C(1 + |\theta - \theta^*|^2) \tag{2.10}$$

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

$$\theta_{p+1} = \theta_p - \gamma_{p+1} H(\theta_p, U_{p+1}), \ p \geq 0 \tag{2.11}$$

satisfies

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

Let us point out that the Robbins-Monro theorem also covers the framework of stochastic gradient algorithm. Indeed, if the function h is the gradient of a convex potential L, namely $h = \nabla L$ where $L \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}_+)$, that satisfies: ∇L is Lipschitz, $|\nabla L|^2 \leq C(1+L)$ and $\lim_{|\theta| \to +\infty} L(\theta) = +\infty$ then, $\operatorname{Argmin} L$ is non-empty and according to the following standard lemma $\theta \mapsto \frac{1}{2}|\theta - \theta^*|^2$ is a Lyapunov function so that the sequence $(\theta_n)_{n\geq 1}$ defined by (2.11) converges *a.s.* to θ^*.

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

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

Moreover, if $\operatorname{Argmin} L$ *is non-empty, then one has*

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

Now, we provide a result on the weak rate of convergence of stochastic approximation algorithm. In standard situations, it is well-known that a stochastic algorithm $(\theta_p)_{p\geq 1}$ converges to its target at a rate $\gamma_p^{-1/2}$. We also refer to [FM12] and [FF13] for some recent developments on non-asymptotic deviation bounds. 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 [Pel98] (see also [Duf96], p.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, i.e. all its eigenvalues have positive real parts. Assume that the function H satisfies the following regularity and growth control property*

$$\theta \mapsto \mathbb{E}H(\theta, U)H(\theta, U)^T \text{ is continuous on } \mathbb{R}^d, \quad \exists b > 0 \text{ s.t. } \theta \mapsto \mathbb{E}|H(\theta, U)|^{2+b} \text{ is locally bounded on } \mathbb{R}^d.$$

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

The step sequence of the procedure (2.11) *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 $(-\rho)$, $\rho \in [0, 1)$, that is, for any $x > 0$, $\lim_{t \to +\infty} \gamma(tx)/\gamma(t) = x^{-\rho}$. In this case, set $\zeta = 0$.*
- *for $t \geq 1$, $\gamma(t) = \gamma_0/t$ and γ_0 satisfies $\gamma_0 > 1/(2\mathcal{R}e(\lambda_{min}))$, where λ_{min} denotes the eigenvalue of $Dh(\theta^*)$ with the lowest real part. In this case, set $\zeta = \frac{1}{2\gamma_0}$.*

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

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

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

Remark 2.1. *The assumption on the step sequence $(\gamma_n)_{n\geq 1}$ is quite general and the above theorem is often applied to the usual gain $\gamma_p = \gamma(p) = \gamma_0 p^{-\rho}$, with $1/2 < \rho \leq 1$, which notably satisfies* (1.3).

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 with 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 e.g. [Duf96] or [BMP90] among others.

As mentioned in the introduction, a solution consists in devising the original stochastic approximation algorithm (2.11) with a slow decreasing step $\gamma = (\gamma_p)_{p\geq 1}$, where γ varies regularly with exponent $(-\rho)$, $\rho \in (1/2, 1)$ and to simultaneously compute the empirical mean $(\bar{\theta}_p)_{p\geq 1}$ of the sequence $(\theta_p)_{p\geq 0}$ by setting

$$\bar{\theta}_p = \frac{\theta_0 + \theta_1 + \cdots + \theta_p}{p+1} = \bar{\theta}_{p-1} - \frac{1}{p+1}\left(\bar{\theta}_{p-1} - \theta_p\right). \tag{2.12}$$

The following result states the weak rate of convergence for the sequence $(\bar{\theta}_p)_{p\geq 1}$. 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 [Duf96], p.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, i.e. all its eigenvalues have positive real parts. Assume that the function H satisfies the following regularity and growth control property*

$$\theta \mapsto \mathbb{E}H(\theta, U)H(\theta, U)^T \text{ is continuous on } \mathbb{R}^d, \quad \exists b > 0 \text{ s.t. } \theta \mapsto \mathbb{E}|H(\theta, U)|^{2+b} \text{ is locally bounded on } \mathbb{R}^d.$$

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

The step sequence of the procedure (2.11) *is given by $\gamma_p = \gamma(p)$, $p \geq 1$, where γ varies regularly with exponent $(-\rho)$, $\rho \in (1/2, 1)$. Then, on the event $\{\theta_p \to \theta^*\}$, one has*

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

2.3. On the implicit discretization error

As already observed the approximation of θ^* solution of $h(\theta) = \mathbb{E}_x[H(\theta, X_T)] = 0$ is affected by two errors: the *implicit discretization error* and the *statistical error*. A first interesting problem concerns the convergence of $\theta^{*,n}$ toward θ^* as $n \to +\infty$ or equivalently the behavior of the discretization error as the number of time step n of the continuous Euler scheme goes to infinity.

Theorem 2.6. *For all $n \in \mathbb{N}^*$, assume that h and h^n satisfy the mean reverting assumption* (2.9) *of Theorem 2.3. Moreover, suppose that $(h^n)_{n\geq 1}$ converges locally uniformly towards h. Then, one has*

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

Proof. Let $\epsilon > 0$. The mean-reverting assumption (2.9) and the continuity of $u \mapsto \langle u, h(\theta^* + \epsilon u)\rangle$ on the (compact) set $\mathcal{S}_d := \left\{u \in \mathbb{R}^d, |u| = 1\right\}$ yields

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

The local uniform convergence of h^n implies

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

Then, using the following 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 it is possible (otherwise the proof is complete). Hence, there exists $x^* \in]-\epsilon, \epsilon[$ such that

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

which clearly implies

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

so that by the very definition of $\theta^{*,n}$, we have $x^* = \pm|\theta^* - \theta^{*,n}|$ and finally $|\theta^* - \theta^{*,n}| < \epsilon$ for $n \geq n_\eta$. This completes the proof. □

Now, we derive a convergence rate.

Theorem 2.7. *Suppose the assumptions of theorem 2.6 hold and that h and h^n, $n \geq 1$, are continuously differentiable and that $Dh(\theta^*)$ is non-singular. Assume that $(Dh^n)_{n\geq 1}$ converges locally uniformly to Dh. If there exists $\alpha \in [0, 1]$ such that*

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

then, one has

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

Proof. 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 non-singularity of $Dh(\theta^*)$, ones clearly gets that for n large enough $\int_0^1 Dh^n(\lambda\theta^{*,n} + (1-\lambda)\theta^*)d\lambda$ is non singular and that

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

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

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

□

2.4. On the optimal tradeoff between the implicit discretization and the statistical errors

Given the order of the implicit discretization error, a natural question is to find the optimal balance between the number of time steps n in the discretization of the process $(X_t)_{0\leq t\leq T}$ and the number M of steps in (1.4) for the computation of θ^* to achieve a given global error ϵ. We suppose that h^n and H satisfy the following assumptions:

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

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

(HI) There exists $b>0$ such that for all $R>0$, we have $\sup_{\{\theta:|\theta|\leq R,\ n\in\mathbb{N}^*\}} \mathbb{E}|H(\theta,X_T^n)|^{2+b} < +\infty$. The sequence $(\theta\mapsto \mathbb{E}H(\theta,X_T^n)H(\theta,X_T^n)^T)_{n\geq 1}$ converges locally uniformly towards $\theta\mapsto\mathbb{E}H(\theta,X_T)H(\theta,X_T)^T$. The function $\theta\mapsto\mathbb{E}H(\theta,X_T)H(\theta,X_T)^T$ is continuous and $\mathbb{E}H(\theta^*,X_T)H(\theta^*,X_T)^T$ is a positive deterministic matrix.

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

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

We will denote λ_m the lowest real part of the eigenvalues of $Dh(\theta^*)$. We will assume that the step sequence is given by $\gamma_p=\gamma(p)$, $p\geq 1$, 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 $(-\rho)$, $\rho\in[0,1)$, that is, for any $x>0$, $\lim_{t\to+\infty}\gamma(tx)/\gamma(t)=x^{-\rho}$.

(HS2) for $t\geq 1$, $\gamma(t)=\gamma_0/t$ and γ_0 satisfies $2\underline{\lambda}\gamma_0>1$.

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 $H(\theta,x)=\mathbf{1}_{\{x\geq\theta\}}$ under the additional assumption that X_T^n has a bounded density (uniformly in n).*

Remark 2.3. *Assumption* **(HMR)** *already appears in [Duf96] and [BMP90], see also [FM12] and [FF13] in another context. It allows 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 [KY03], Chapter 10, Section 5, if one considers the projected version of the algorithm* (1.4) *on a bounded convex set H (for instance an hyperrectangle $\Pi_{i=1}^d[a_i,b_i]$) containing $\theta^{*,n}$, $\forall n\geq 1$, as very often happens from a practical point of view, this assumption can be localized on H, that is it holds on H instead of $\mathbb{R}^d$. In this case, a sufficient condition is $\inf_{\theta\in H,n\in\mathbb{N}^*}\lambda_{min}((Dh^n(\theta)+Dh^n(\theta)^T)/2)>0$, where $\lambda_{min}(A)$ denotes the lowest eigenvalue of the matrix A.*

We also want to point out that if it is satisfied then one has $\lambda_m\geq\underline{\lambda}$. 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

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

Using the local uniform convergence of $(Dh^n)_{n\geq1}$ and the convergence of $(\theta^{*,n})_{n\geq1}$ toward θ^*, by passing to the limit $n\to+\infty$ in the above inequality, we obtain

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

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 $\lambda_m\geq\underline{\lambda}$.

Theorem 2.8. *Suppose that the assumptions of Theorem 2.7 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*

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

where

$$\Sigma^{*}:=\int_{0}^{\infty}\exp\left(-s(Dh(\theta^{*})-\zeta I_{d})\right)^{T}\mathbb{E}_{x}[H(\theta^{*},X_{T})H(\theta^{*},X_{T})^{T}]\exp\left(-s(Dh(\theta^{*})-\zeta I_{d})\right)ds$$

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

Lemma 2.3. *Let $\delta>0$. Under the assumptions of Theorem 2.8, one has*

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

Proof. With the notations of Section 2.2, we define for all $p\geq1$, $\Delta M^{n^{\delta}}_{p}:=h^{n^{\delta}}(\theta^{n^{\delta}}_{p-1})-H(\theta^{n^{\delta}}_{p-1},(X^{n^{\delta}}_{T})^{p})=\mathbb{E}[H(\theta^{n^{\delta}}_{p-1},(X^{n^{\delta}}_{T})^{p})\big|\mathcal{F}_{p-1}]-H(\theta^{n^{\delta}}_{p-1},(X^{n^{\delta}}_{T})^{p})$. Recalling that $((X^{n^{\delta}}_{T})^{p})_{p\geq1}$ is a sequence of i.i.d. random variables we have that $(\Delta M^{n^{\delta}}_{p})_{p\geq1}$ is a sequence of martingale increments w.r.t. the natural filtration $\mathcal{F}:=(\mathcal{F}_p:=\sigma(\theta_0,(X^{n^{\delta}}_{T})^{1},\cdots,(X^{n^{\delta}}_{T})^{p});p\geq1)$.

Using Taylor's formula, we get for $p\geq0$

$$\theta^{n^{\delta}}_{p+1}-\theta^{*,n^{\delta}}=\theta^{n^{\delta}}_{p}-\theta^{*,n^{\delta}}-\gamma_{p+1}Dh^{n^{\delta}}(\theta^{*,n^{\delta}})(\theta^{n^{\delta}}_{p}-\theta^{*,n^{\delta}})+\gamma_{p+1}\Delta M^{n^{\delta}}_{p+1}-\gamma_{p+1}\zeta^{n^{\delta}}_{p}$$

with $\zeta^{n^{\delta}}_{p}:=h^{n^{\delta}}(\theta^{n^{\delta}}_{p})-Dh^{n^{\delta}}(\theta^{n^{\delta}}_{p})(\theta^{n^{\delta}}_{p}-\theta^{*,n^{\delta}})=\mathcal{O}(|\theta^{n^{\delta}}_{p}-\theta^{*,n^{\delta}}|^{2})$ since $Dh^{n^{\delta}}$ is Lipschitz-continuous uniformly in n. Hence, by a simple induction, we obtain

$$\theta^{n^{\delta}}_{n}-\theta^{*,n^{\delta}}=\Pi_{1,n}(\theta^{n^{\delta}}_{0}-\theta^{*,n^{\delta}})+\sum_{k=1}^{n}\gamma_k\Pi_{k+1,n}\Delta M^{n^{\delta}}_{k}+\sum_{k=1}^{n}\gamma_k\Pi_{k+1,n}\left(\zeta^{n^{\delta}}_{k-1}+(Dh(\theta^{*})-Dh^{n^{\delta}}(\theta^{*,n^{\delta}}))(\theta^{n^{\delta}}_{k-1}-\theta^{*,n^{\delta}})\right)\tag{2.13}$$

where $\Pi_{k,n}:=\prod_{j=k}^{n}(I_d-\gamma_jDh(\theta^{*}))$, with the convention that $\Pi_{n+1,n}=I_d$. We now investigate the asymptotic behavior of each term in the above decomposition.

Step 1: study of the sequence $\left\{n^{\alpha}\Pi_{1,\gamma^{-1}(1/n^{2\alpha})}(\theta^{n^{\delta}}_{0}-\theta^{*,n^{\delta}}),n\geq0\right\}$

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

$$n^{\alpha}\mathbb{E}|\Pi_{1,\gamma^{-1}(1/n^{2\alpha})}z^{n^{\delta}}_{0}|\leq C(\sup_{n\geq1}\mathbb{E}|\theta^{n}_{0}|+1)n^{\alpha}\exp\left(-(\lambda_m-\eta)\sum_{k=1}^{\gamma^{-1}(1/n^{2\alpha})}\gamma_k\right).$$

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 the sequence $\left\{n^{\alpha}\sum_{k=1}^{\gamma^{-1}(1/n^{2\alpha})}\gamma_k\Pi_{k+1,\gamma^{-1}(1/n^{2\alpha})}\left(\zeta^{n^{\delta}}_{k-1}+(Dh(\theta^{*})-Dh^{n^{\delta}}(\theta^{*,n^{\delta}}))(\theta^{n^{\delta}}_{k-1}-\theta^{*,n^{\delta}})\right),n\geq0\right\}$

We focus on the last term of (2.13). Using Lemma 5.2 we get

$$\mathbb{E}\left|\sum_{k=1}^{n}\gamma_k\Pi_{k+1,n}(\zeta^{n^{\delta}}_{k-1}+(Dh(\theta^{*})-Dh^{n^{\delta}}(\theta^{*,n^{\delta}}))(\theta^{n^{\delta}}_{k-1}-\theta^{*,n^{\delta}}))\right|\leq C\sum_{k=1}^{n}\|\Pi_{k+1,n}\|(\gamma_k^{2}+\gamma_k^{3/2}\|Dh(\theta^{*})-Dh^{n^{\delta}}(\theta^{*,n^{\delta}})\|),$$

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

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

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

We use the following decomposition

$$\begin{aligned}\sum_{k=1}^{n}\gamma_k\Pi_{k+1,n}\Delta M_k^{n^\delta} &= \sum_{k=1}^{n}\gamma_k\Pi_{k+1,n}(h^{n^\delta}(\theta_k^{n^\delta})-h^{n^\delta}(\theta^{*,n^\delta})-(H(\theta_k^{n^\delta},(X_T^{n^\delta})^{k+1})-H(\theta^{*,n^\delta},(X_T^{n^\delta})^{k+1})))\\ &+\sum_{k=1}^{n}\gamma_k\Pi_{k+1,n}(h^{n^\delta}(\theta^{*,n^\delta})-H(\theta^{*,n^\delta},(X_T^{n^\delta})^{k+1}))\\ &:= R_n+M_n\end{aligned}$$

Now, by **(HR)**, we have

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

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

To conclude we prove that the sequence $\left\{\frac{1}{\gamma^{1/2}(n)}M_n,\ n\geq 0\right\}$, 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, p.58 in [HH80]. By **(HI)**, it holds for some $R>0$ such that $\forall n\geq 1$, $\theta^{*,n}\in B(0,R)$

$$\sum_{k=1}^{n}\mathbb{E}\left|\gamma^{-\frac{1}{2}}(n)\gamma_k\Pi_{k+1,n}(h^{n^\delta}(\theta^{*,n^\delta})-H(\theta^{*,n^\delta},(X_T^{n^\delta})^{k+1}))\right|^{2+b}\leq C(\sup_{\theta:|\theta|\leq R,n\in\mathbb{N}^*}\mathbb{E}|H(\theta,X_T^{n})|^{2+b})\gamma^{-1+\frac{b}{2}}(n)\sum_{k=1}^{n}\gamma_k^{2+b}\|\Pi_{k+1,n}\|^{2+b}$$

By Lemma 5.1, we have $\limsup_n\gamma^{-1+b/2}(n)\sum_{k=1}^{n}\gamma_k^{2+b}\|\Pi_{k+1,n}\|^{2+b}\leq\limsup_n\gamma^{b/2}(n)=0$, so that the conditional Lindeberg condition, see [HH80], Corollary 3.1 is satisfied. Now we focus on the conditional variance. We set

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

with

$$\Gamma_n:=\mathbb{E}[H(\theta^{*,n^\delta},X_T^{n^\delta})(H(\theta^{*,n^\delta},X_T^{n^\delta}))^T]\quad\text{and}\quad\Gamma^*:=\mathbb{E}[H(\theta^*,X_T))(H(\theta^*,X_T))^T].$$

By the local uniform convergence of $(\theta \mapsto \mathbb{E}H(\theta, X_T^{n^\delta})(H(\theta, X_T^{n^\delta}))^T)_{n\geq 0}$, the continuity of $\theta \mapsto \mathbb{E}H(\theta, X_T)(H(\theta, X_T))^T$ 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 Σ^* be the (unique) matrix 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 $S_n \stackrel{a.s.}{\longrightarrow} \Sigma^*$. 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

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

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

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

Let us note that by the very definition of ζ and assumptions **(HS1)**, **(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 the assumption (1.3), 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, $S_n \xrightarrow{a.s.} \Theta^*$ and the proof is complete.

□

Proof of Theorem 2.8. 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

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

where $\Sigma^* := \int_0^\infty \exp\left(-s(Dh(\theta^*) - \zeta I_d)\right)^T \mathbb{E}_x[H(\theta^*, X_T)H(\theta^*, X_T)^T] \exp\left(-s(Dh(\theta^*) - \zeta I_d)\right) ds$. Moreover, using Theorem 2.7, we obtain

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

This completes the proof.

□

The result of Theorem 2.8 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 stochastic approximation scheme defined by (1.4). Hence, in this case the complexity (or computational cost) of the algorithm is given by

$$C_{SA}(\gamma) = C \times n \times \gamma^{-1}(1/n^{2\alpha}), \tag{2.14}$$

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

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 with this choice of step sequence comes from the constraint on γ_0. Next result shows that the optimal complexity can be reached for free through the smoothing of the procedure (1.4) according to the Ruppert & Polyak averaging principle.

Theorem 2.9. *Suppose that the assumptions of Theorem 2.7 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 $(\bar{\theta}^n_p)_{p\geq 1}$ of the sequence $(\theta^n_p)_{p\geq 1}$ by setting*

$$\bar{\theta}^n_p = \frac{\theta_0 + \theta^n_1 + \cdots + \theta^n_p}{p+1} = \bar{\theta}^n_{p-1} - \frac{1}{p+1}\left(\bar{\theta}^n_{p-1} - \theta^n_p\right),$$

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

$$n^\alpha\left(\bar{\theta}^n_{n^{2\alpha}} - \theta^*\right) \Longrightarrow -Dh^{-1}(\theta^*)\mathcal{E}(h, \alpha, \theta^*) + \mathcal{N}\left(0, Dh(\theta^*)^{-1}\mathbb{E}_x[H(\theta^*, X_T)H(\theta^*, X_T)^T](Dh(\theta^*)^{-1})^T\right),$$

Lemma 2.4. *Let $\delta > 0$. Under the assumptions of Theorem 2.9, one has*

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

Proof. We freely use the notations and the intermediate results of the proof of Lemma 2.3. Using (2.13) in its recursive form, for any $p \geq 0$ and for n large enough, it holds

$$\theta_p^{n^\delta} - \theta^{*,n^\delta} = \frac{1}{\gamma_{p+1}}(Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}(\theta_{p+1}^{n^\delta} - \theta_p^{n^\delta}) - (Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}\Delta M_{p+1}^{n^\delta} - (Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}\zeta_p^{n^\delta}.$$

Hence, using an Abel's transform we derive

$$\begin{aligned}
\theta_{n^{2\alpha}}^{n^\delta} - \theta^{*,n^\delta} &= \frac{1}{n^{2\alpha}+1}\sum_{k=0}^{n^{2\alpha}} \theta_k^{n^\delta} - \theta^{*,n^\delta} = \frac{(Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}}{n^{2\alpha}+1}\sum_{k=0}^{n^{2\alpha}} \frac{1}{\gamma_{k+1}}(\theta_{k+1}^{n^\delta} - \theta_k^{n^\delta}) \\
&\quad - \frac{(Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}}{n^{2\alpha}+1}\sum_{k=0}^{n^{2\alpha}} \Delta M_{k+1}^{n^\delta} - \frac{(Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}}{n^{2\alpha}+1}\sum_{k=0}^{n^{2\alpha}} \zeta_k^{n^\delta} \\
&= \frac{(Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}}{n^{2\alpha}+1}\left(\frac{\theta_{n^{2\alpha}+1}^{n^\delta} - \theta^{*,n^\delta}}{\gamma_{n^{2\alpha}+1}} - \frac{\theta_0^{n^\delta} - \theta^{*,n^\delta}}{\gamma_1}\right) + \frac{(Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}}{n^{2\alpha}+1}\sum_{k=1}^{n^{2\alpha}}\left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k+1}}\right)(\theta_k^{n^\delta} - \theta^{*,n^\delta}) \\
&\quad - \frac{(Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}}{n^{2\alpha}+1}\sum_{k=0}^{n^{2\alpha}} \Delta M_{k+1}^{n^\delta} - \frac{(Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}}{n^{2\alpha}+1}\sum_{k=0}^{n^{2\alpha}} \zeta_k^{n^\delta}
\end{aligned}$$

We now study each term of the above decomposition.

Step 1: study of the sequence $\left\{\frac{n^\alpha}{n^{2\alpha}+1}\left(\frac{\theta_{n^{2\alpha}+1}^{n^\delta} - \theta^{*,n^\delta}}{\gamma_{n^{2\alpha}+1}} - \frac{\theta_0^{n^\delta} - \theta^{*,n^\delta}}{\gamma_1}\right), n \geq 0\right\}$

For the first term, by Lemma 5.2, Proposition 5.1 and **(HS1)** it follows

$$\begin{aligned}
\mathbb{E}\left|\frac{n^\alpha}{n^{2\alpha}+1}\left(\frac{\theta_{n^{2\alpha}+1}^{n^\delta} - \theta^{*,n^\delta}}{\gamma_{n^{2\alpha}+1}} - \frac{\theta_0^{n^\delta} - \theta^{*,n^\delta}}{\gamma_1}\right)\right| &\leq C\left(\frac{1}{n^\alpha \gamma_{n^{2\alpha}+1}^{\frac{1}{2}}}\gamma_{n^{2\alpha}+1}^{-\frac{1}{2}}\mathbb{E}|\theta_{n^{2\alpha}+1}^n - \theta^{*,n}| + \frac{1}{n^\alpha}(\sup_{n\geq 1}\mathbb{E}|\theta_0^n| + 1)\right) \\
&\leq C\left(\frac{1}{n^\alpha \gamma_{n^{2\alpha}+1}^{\frac{1}{2}}} + \frac{1}{n^\alpha}\right) \longrightarrow 0,
\end{aligned}$$

since $n\gamma_n \to 0$, $n \to +\infty$.

Step 2: study of the sequence $\left\{\frac{n^\alpha}{n^{2\alpha}+1}\sum_{k=1}^{n^{2\alpha}}\left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k+1}}\right)(\theta_k^{n^\delta} - \theta^{*,n^\delta}), n \geq 0\right\}$

Similarly for the second term, we have

$$\begin{aligned}
\mathbb{E}\left|\frac{n^\alpha}{n^{2\alpha}+1}\sum_{k=1}^{n^{2\alpha}}\left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k+1}}\right)(\theta_k^{n^\delta} - \theta^{*,n^\delta})\right| &\leq C\frac{1}{n^\alpha}\sum_{k=1}^{n^{2\alpha}}\gamma_k^{1/2}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)\gamma_k^{-1/2}\mathbb{E}|\theta_k^{n^\delta} - \theta^{*,n^\delta}| \\
&\leq C\frac{1}{n^\alpha}\sum_{k=1}^{n^{2\alpha}}\gamma_k^{1/2}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right) \to 0, \quad n \to +\infty.
\end{aligned}$$

where we used Lemma 5.2 for the last inequality.

Step 3: study of the sequence $\left\{\frac{n^\alpha}{n^{2\alpha}+1}\sum_{k=0}^{n^{2\alpha}}\Delta M_{k+1}^{n^\delta}, n \geq 0\right\}$

As in the proof of Lemma 2.3, we decompose this sequence as follows

$$\begin{aligned}\frac{n^{\alpha}}{n^{2\alpha}+1}\sum_{k=0}^{n^{2\alpha}}\Delta M_{k+1}^{n^\delta} &= \frac{n^{\alpha}}{n^{2\alpha}+1}\sum_{k=1}^{n^{2\alpha}}(h^{n^\delta}(\theta_k^{n^\delta})-h^{n^\delta}(\theta^{*,n^\delta})-(H(\theta_k^{n^\delta},(X_T^{n^\delta})^{k+1})-H(\theta^{*,n^\delta},(X_T^{n^\delta})^{k+1})))\\ &+\frac{n^{\alpha}}{n^{2\alpha}+1}\sum_{k=1}^{n^{2\alpha}}(h^{n^\delta}(\theta^{*,n^\delta})-H(\theta^{*,n^\delta},(X_T^{n^\delta})^{k+1}))\\ &:= R_n+M_n\end{aligned}$$

For the sequence $(R_n)_{n\geq 1}$ we use **(HR)** to write

$$\mathbb{E}|R_n|^2\leq \frac{C}{n^{2\alpha}}\sum_{k=0}^{n^{2\alpha}}\mathbb{E}|H(\theta_k^{n^\delta},(X_T^{n^\delta})^{k+1})-H(\theta^{*,n^\delta},X_T^{n^\delta})|^2=\frac{C}{n^{2\alpha}}\sum_{k=1}^{n^{2\alpha}}\gamma_k^{2a}\to 0,$$

owing to Cesàro's Lemma. We now prove a CLT for the sequence $(M_n)_{n\geq 1}$ by applying Theorem 3.2 and Corollary 3.1, p.58 in [HH80]. Since $\theta^{*,n^\delta}\to\theta^*$, it holds for some $R>0$

$$\sum_{k=0}^{n^{2\alpha}}\mathbb{E}\left|\frac{n^{\alpha}}{n^{2\alpha}+1}(h^{n^\delta}(\theta^{*,n^\delta})-H(\theta^{*,n^\delta},(X_T^{n^\delta})^{k+1}))\right|^{2+b}\leq\frac{C}{n^{\alpha b}}(\sup_{\theta:|\theta|\leq R,\ n\in\mathbb{N}^*}\mathbb{E}|H(\theta,X_T^n)|^{2+b})\to 0,\quad n\to+\infty,$$

so that the conditional Lindeberg condition is satisfisfied, see [HH80] Corollary 3.1. Now, we focus on the conditional variance. For convenience, we set

$$\begin{aligned}S_n &:= \frac{n^{2\alpha}}{(n^{2\alpha}+1)^2}\sum_{k=1}^{n^{2\alpha}}\mathbb{E}_k[(h^{n^\delta}(\theta^{*,n^\delta})-H(\theta^{*,n^\delta},(X_T^{n^\delta})^{k+1}))(h^{n^\delta}(\theta^{*,n^\delta})-H(\theta^{*,n^\delta},(X_T^{n^\delta})^{k+1}))^T]\\ &=\frac{n^{2\alpha}}{(n^{2\alpha}+1)^2}\sum_{k=1}^{n^{2\alpha}}\mathbb{E}[H(\theta^{*,n^\delta},X_T^{n^\delta})(H(\theta^{*,n^\delta},X_T^{n^\delta}))^T]\\ &=\frac{n^{4\alpha}}{(n^{2\alpha}+1)^2}\mathbb{E}[H(\theta^{*,n^\delta},X_T^{n^\delta})(H(\theta^{*,n^\delta},X_T^{n^\delta}))^T],\end{aligned}$$

so that we clearly have $S_n\to\mathbb{E}[H(\theta^*,X_T)(H(\theta^*,X_T))^T]$ by the local uniform convergence of $(\theta\mapsto\mathbb{E}[H(\theta,X_T^n)(H(\theta,X_T^n))^T])_{n\geq 1}$, the continuity of $\theta\mapsto\mathbb{E}[H(\theta,X_T)(H(\theta,X_T))^T]$ at θ^* and the convergence of $(\theta^{*,n^\delta})_{n\geq 1}$ towards θ^*. Therefore, since $(Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}\to(Dh(\theta^*))^{-1}$, we conclude that

$$(Dh^{n^\delta}(\theta^{*,n^\delta}))^{-1}\frac{n^{\alpha}}{n^{2\alpha}+1}\sum_{k=0}^{n^{2\alpha}}\Delta M_{k+1}^{n^\delta}\Longrightarrow\mathcal{N}(0,Dh(\theta^*)^{-1}\mathbb{E}_x[H(\theta^*,X_T)H(\theta^*,X_T)^T](Dh(\theta^*)^{-1})^T).$$

Step 4: study of the sequence $\left\{\frac{n^\alpha}{n^{2\alpha}+1}\sum_{k=0}^{n^{2\alpha}}\zeta_k^{n^\delta},n\geq 0\right\}$

Now, observe that by Lemma 5.2 the last term is bounded in L^1-norm by

$$\frac{n^{\alpha}}{n^{2\alpha}+1}\sum_{k=0}^{n^{2\alpha}}\mathbb{E}|\zeta_k^{n^\delta}|\leq\frac{C}{n^{\alpha}}\sum_{k=0}^{n^{2\alpha}}\gamma_k\to 0,\ n\to+\infty$$

since γ varies regularly with exponent $-\rho$, $\rho\in(1/2,1)$. □

Proof of Theorem 2.9. Similarly to the proof of Theorem 2.8 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 2.7 and Lemma 2.4, we obtain

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

□

The result of Theorem 2.9 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 stochastic approximation scheme defined by (1.4) 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.

3. Multi-level stochastic approximation algorithms

3.1. The statistical Romberg stochastic approximation method

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

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

The statistical Romberg stochastic approximation 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 samples and $\theta^n_{M_2} - \theta^{n^\beta}_{M_2}$ be an estimator of $\theta^{*,n} - \theta^{*,n^\beta}$ using M_2 paths. Using the above decomposition, we estimate θ^* by the quantity

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

It is important to point out here that the couple $(\theta^n_{M_2}, \theta^{n^\beta}_{M_2})$ is computed using two Euler approximation schemes with different time steps but with the same Brownian path. Moreover, the quantity $\theta^{n^\beta}_{M_1}$ comes from Brownian paths which are independent to 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 in estimating θ^* by

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

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

To establish the rate of convergence of the two-level stochastic approximation scheme, we require smoothness assumptions on H:

(HDH) For all $\theta \in \mathbb{R}^d$, $\mathbb{P}(X_T \notin \mathcal{D}_{H,\theta}) = 0$ 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 $C,r>0$.

Theorem 3.1. *Suppose that h and h^n satisfy the assumptions of Theorem 2.7 with $\alpha\in(1/2\vee\beta,1]$ and that h satisfies the assumptions of Theorem 2.4. Assume that* **(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:*

$$\forall\theta\in\mathbb{R}^d,\ n^{1/2}\|Dh^n(\theta)-Dh(\theta)\|\to 0,\quad \text{as } n\to+\infty.$$

Suppose that $\tilde{\mathbb{E}}(D_xH(\theta^*,X_T)U_T)(D_xH(\theta^*,X_T)U_T)^T$ is a positive definite matrix. Assume that the step sequence is given by $\gamma_p=\gamma(p)$, $p\geq1$, where γ is a positive function defined on $[0,+\infty[$ decreasing to zero, satisfying one of the following assumptions:

- *γ varies regularly with exponent $(-\rho)$, $\rho\in(1/2,1)$, that is, for any $x>0$, $\lim_{t\to+\infty}\gamma(tx)/\gamma(t)=x^{-\rho}$.*
- *for $t\geq1$, $\gamma(t)=\gamma_0/t$ and γ_0 satisfies $\underline{\lambda}\gamma_0>1$.*

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

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

with

$$\Sigma^*:=\int_0^\infty\left(e^{-s(Dh(\theta^*)-\zeta I_d)}\right)^T\left(\mathbb{E}_x[H(\theta^*,X_T)H(\theta^*,X_T)^T]+\tilde{\mathbb{E}}\left(D_xH(\theta^*,X_T)U_T\right)\left(D_xH(\theta^*,X_T)U_T\right)^T\right)e^{-s(Dh(\theta^*)-\zeta I_d)}ds$$

and U_T is the value at time T of the process (2.8) defined on $(\tilde{\Omega},\tilde{\mathcal{F}},(\tilde{\mathcal{F}}_t)_{t\geq0},\tilde{\mathbb{P}})$.

Lemma 3.1. *Let $(\theta_p)_{p\geq0}$ be the procedure defined for $p\geq0$ by*

$$\theta_{p+1}=\theta_p-\gamma_{p+1}H(\theta_p,(X_T)^{p+1}),\quad \theta_0=\theta_0^n,\tag{3.15}$$

where $((X_T^n)^p,(X_T)^p)_{p\geq1}$ is an i.i.d sequence of random variables with the same law as (X_T^n,X_T) and $(\gamma_p)_{p\geq1}$ is the step sequence of the procedure $(\theta_p^{n^\beta})_{p\geq0}$ and $(\theta_p^n)_{p\geq0}$. Under the assumptions of Theorem 3.1, one has

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

with $\Theta^*:=\int_0^\infty\exp\left(-s(Dh(\theta^*)-\zeta I_d)\right)^T\tilde{\mathbb{E}}\left(D_xH(\theta^*,X_T)U_T\right)\left(D_xH(\theta^*,X_T)U_T\right)^T\exp\left(-s(Dh(\theta^*)-\zeta I_d)\right)ds$, and

$$n^{\alpha}\left(\theta^{n}_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))}-\theta_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))}-(\theta^{*,n}-\theta^*)\right)\xrightarrow{\mathbb{P}}0,\quad n\to+\infty.$$

Proof. We will just prove the first assertion of the Lemma. The second one 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. We consequently use the same notations. Let us note that the procedure $(\theta_p)_{p\geq0}$ *a.s.* converges to θ^* and satisfies a CLT according to Theorem 2.4.

A Taylor's expansion yields for $p\geq0$

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

with $\Delta M_{p+1} = h(\theta_p) - H(\theta_p, (X_T)^{p+1})$, $p \geq 0$. 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

$$z_n^{n^\beta} = \Pi_{1,n} z_0^{n^\beta} + \sum_{k=1}^{n} \gamma_k \Pi_{k+1,n} \Delta N_k^{n^\beta} + \sum_{k=1}^{n} \gamma_k \Pi_{k+1,n} \Delta R_k^{n^\beta}$$
$$+ \sum_{k=1}^{n} \gamma_k \Pi_{k+1,n} \left(\zeta_{k-1}^n - \zeta_{k-1} + (Dh(\theta^*) - Dh^{n^\beta}(\theta^{*,n^\beta}))(\theta_{k-1}^{n^\beta} - \theta^{*,n^\beta}) \right) \tag{3.16}$$

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 $\Delta N_k^{n^\beta} := h^{n^\beta}(\theta^*) - h(\theta^*) - (H(\theta^*, (X_T^{n^\beta})^{k+1}) - H(\theta^*, (X_T)^{k+1}))$, $\Delta R_k^{n^\beta} = h^{n^\beta}(\theta_k^{n^\beta}) - h^{n^\beta}(\theta^*) - (H(\theta_k^{n^\beta}, (X_T^{n^\beta})^{k+1}) - H(\theta^*, (X_T^{n^\beta})^{k+1})) + H(\theta_k, (X_T)^{k+1}) - H(\theta^*, (X_T)^{k+1}) - (h(\theta_k) - h(\theta^*))$ for $k \geq 1$.

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

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

$$n^\alpha |\Pi_{1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))} z_0^{n^\beta}| \leq n^\alpha \|\Pi_{1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))}\| |\theta^{*,n^\beta} - \theta^*| \leq C n^{(1-\beta)\alpha} \exp(-(\lambda_m - \eta) \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k) \to 0,$$

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

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

By Lemma 5.2, one has

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

so that by Lemma 5.1, we easily derive that (if $\gamma(p) = \gamma_0/p$ recall that $\underline{\lambda}\gamma_0 > 1$) $\sum_{k=1}^{n} \gamma_k^2 \|\Pi_{k+1,n}\| = \mathcal{O}(\gamma(n))$ and $\sum_{k=1}^{n} \gamma_k^{3/2} \|\Pi_{k+1,n}\| = \mathcal{O}(\gamma^{1/2}(n))$ so that

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

Moreover, since Dh^{n^β} is a Lipschitz function uniformly in n we clearly have

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

which combined with $n^{\beta/2}\|Dh(\theta^*) - Dh^{n^\beta}(\theta^*)\| \to 0$ and $n^{\beta/2}|\theta^{*,n^\beta} - \theta^*| \to 0$ (recall that $\alpha > 1/2$) imply that $\limsup_n n^\alpha \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k^{3/2} \|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))}\| \|Dh(\theta^*) - Dh^{n^\beta}(\theta^{*,n^\beta})\| = 0$.

Using the notations of Proposition 5.1 and the inequality $|\theta_k - \theta^*|^2 \leq |\mu_k|^2 + 2|\theta_k - \theta^*||r_k|$, we get

$$\left| \sum_{k=1}^{n} \gamma_k \Pi_{k+1,n} \zeta_{k-1} \right| \leq C \sum_{k=1}^{n} \gamma_k^2 \|\Pi_{k+1,n}\| \gamma_{k-1}^{-1} |\mu_{k-1}|^2 + C \sum_{k=1}^{n} \gamma_k^{1+b} \|\Pi_{k+1,n}\| |\theta_k - \theta^*| \gamma_{k-1}^{-b} |r_{k-1}|.$$

Now since $\alpha > \beta$, we clearly derive

$$\limsup_n n^{\alpha} \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k^2 \|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))}\| \gamma_{k-1}^{-1}\mathbb{E}|\mu_{k-1}|^2$$
$$\leq (\sup_{k\geq 1} \gamma_k^{-1}\mathbb{E}|\mu_k|^2) \limsup_n n^{\alpha} \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k^2 \|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))}\| = 0.$$

For the second term, we first write

$$\sum_{k=1}^{n} \gamma_k^{1+b} \|\Pi_{k+1,n}\| |\theta_k - \theta^*| \gamma_{k-1}^{-b} |r_{k-1}| \leq X_1 (\sup_{k\geq 1} |\theta_k - \theta^*|) \sum_{k=1}^{n} \gamma_k^{1+b} \|\Pi_{k+1,n}\| Y_{k-1},$$

with $X_1 < +\infty$, $\sup_{k\geq 1} \mathbb{E} Y_k < +\infty$. Now observe that from Lemma 5.2, it follows $\limsup_n \gamma^{-b}(n) \sum_{k=1}^{n} \gamma_k^{1+b} \|\Pi_{k+1,n}\| \leq 1$ under our assumptions on the step sequence (if $\gamma(p) = \gamma_0/p$ it is valid for any $b \in (0,1)$ since $\lambda_m \gamma_0 > 1$ and we select b such that $(2\alpha - \beta)b > \alpha$ otherwise $b = 1$). Therefore, we clearly have

$$\limsup_n n^{\alpha} (\sup_{k\geq 1} \mathbb{E}|Y_{k-1})|) \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k^{1+b} \|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))}\| = 0,$$

which in turn implies

$$n^{\alpha} \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k \Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \zeta_{k-1} \xrightarrow{\mathbb{P}} 0.$$

Hence, we finally conclude that

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

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

Regarding the third term of (3.16), namely $\sum_{k=1}^{n} \gamma_k \Pi_{k+1,n} \Delta R_k^{n^{\beta}}$, we decompose it as follows

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

Now, by **(HLH)** it follows that

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

Similar computations to those of Lemma 5.1 show that there are two cases to distinguish:

- If γ varies regularly with exponent $(-\rho)$, $\rho \in [0,1)$, then $\sum_{k=1}^{n} \gamma_k^3 \|\Pi_{k+1,n}\|^2 = \mathcal{O}(\gamma^2(n))$.
- if $\gamma(p) = \gamma_0/p$ then a comparison between series and integrals show that:
  - if $\lambda_m \gamma_0 < 1$ then $\sum_{k=1}^{n} \gamma_k^3 \|\Pi_{k+1,n}\|^2 = \mathcal{O}(n^{-2\lambda_m \gamma_0})$,
  - if $\lambda_m \gamma_0 = 1$ then $\sum_{k=1}^{n} \gamma_k^3 \|\Pi_{k+1,n}\|^2 = \mathcal{O}(\log(n) n^{-2})$,
  - $\sum_{k=1}^{n} \gamma_k^3 \|\Pi_{k+1,n}\|^2 = \mathcal{O}(n^{-2})$ otherwise.

Consequently, under our assumptions on the step sequence (if $\gamma(p) = \gamma_0/p$ recall that $\lambda_m \gamma_0 > 1$) we have $\limsup_n n^{2\alpha} A^1_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} = 0$. Moreover since $n^{\beta/2}|\theta^{*,n^\beta} - \theta^*| \to 0$ as $n \to +\infty$, we also derive $\limsup_n n^{2\alpha} A^2_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} = 0$. We now focus on the sequence $(B_n)_{n\geq 1}$. We freely use the notations of Proposition 5.1. Let $\epsilon > 0$. We write

$$\mathbb{P}(n^\alpha B_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} > \epsilon) \leq \mathbb{P}(n^\alpha B^1_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} > \epsilon/2) + \mathbb{P}(n^\alpha B^2_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} > \epsilon/2)$$

with $B^1_n := \sum_{k=1}^{n} \gamma_k \Pi_{k+1,n} \left\{H(\theta_k, (X_T)^{k+1}) - H(\theta^*, (X_T)^{k+1}) - (h(\theta_k) - h(\theta^*))\right\} \mathbf{1}_{\left\{\gamma_k^{-b}(\sup_{k\geq 1}|\theta_k - \theta^*|)|r_k| < KY_k\right\}}$ and $B^2_n := B_n - B^1_n$ for all $K > 0$. Using the Chebyshev inequality with the trivial inequality $|\theta_k - \theta^*|^2 \leq |\mu_k|^2 + 2|\theta_k - \theta^*||r_k|$ and **(HLH)** we deduce from the previous computations that

$$\begin{aligned}
\mathbb{P}(n^\alpha B^1_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} > \epsilon) &\leq \frac{n^{2\alpha}}{\epsilon^2} \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k^2 \|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))}\|^2 \\
&\times \quad \mathbb{E}|H(\theta_k, (X_T)^{k+1}) - H(\theta^*, (X_T)^{k+1})|^2 \mathbf{1}_{\left\{\gamma_k^{-b}(\sup_{k\geq 1}|\theta_k - \theta^*|)|r_k| < KY_k\right\}} \\
&\leq \frac{Cn^{2\alpha}}{\epsilon^2} \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k^{2+b} \|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))}\|^2 \\
&\times \gamma_k^{-b}(\mathbb{E}[|\mu_k|^2 + (\sup_{k\geq 1}|\theta_k - \theta^*|)|r_k| \mathbf{1}_{\left\{\gamma_k^{-b}(\sup_{k\geq 1}|\theta_k - \theta^*|)|r_k| < KY_k\right\}}]) \\
&\leq \frac{C(K)n^{2\alpha}}{\epsilon^2} \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k^{2+b} \|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))}\|^2 \sup_{k\geq 1}(\gamma_k^{-b}\mathbb{E}|\mu_k|^2 + \mathbb{E}|Y_k|)
\end{aligned}$$

where $C(K)$ is a constant depending on K only. Therefore using Lemma 5.1 we derive (if $\gamma(p) = \gamma_0/p$ take b such that $1 > b > \beta/(2\alpha - \beta)$ otherwise take $b = 1$) that

$$\lim_n \mathbb{P}(n^\alpha B^1_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} > \epsilon) = 0.$$

Moreover, since for all $k \in [\![1, n]\!]$, $\gamma_k^{-b}|r_k| \leq XY_k$ it follows

$$\begin{aligned}
\mathbb{P}(n^\alpha B^2_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} > \epsilon/2) &= \mathbb{P}(n^\alpha B^2_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} > \epsilon/2, X(\sup_{k\geq 1}|\theta_k - \theta^*|) \geq K) \\
&\quad + \mathbb{P}(n^\alpha B^2_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} > \epsilon/2, X(\sup_{k\geq 1}|\theta_k - \theta^*|) < K) \\
&\leq \mathbb{P}(X(\sup_{k\geq 1}|\theta_k - \theta^*|) \geq K)
\end{aligned}$$

which in turn implies

$$\lim_n \mathbb{P}(n^\alpha B^2_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} > \epsilon) \leq \mathbb{P}(X(\sup_{k\geq 1}|\theta_k - \theta^*|) \geq K).$$

Letting K goes to infinity in the previous inequality we conclude that $n^{\alpha} B^2_{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \xrightarrow{\mathbb{P}} 0$ as $n \to +\infty$ so that

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

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

We now prove a CLT for the sequence $\left\{ n^{\alpha} \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k \Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \Delta N_k^{n^{\beta}},\ n \geq 0 \right\}$. Let $\epsilon > 0$. It holds

$$\sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \mathbb{E} \left| n^{\alpha} \gamma_k \Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \Delta N_k^{n^{\beta}} \right|^{2+\epsilon} \leq \sup_{n\geq 1} \sup_{k \in [\![1,n]\!]} \mathbb{E} \left| n^{\beta/2} \Delta N_k^{n^{\beta}} \right|^{2+\epsilon}$$
$$\times\, n^{(2+\epsilon)(\alpha-\beta/2)} \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k^{2+\epsilon} \|\Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))}\|^{2+\epsilon}.$$

By Lemma 5.1, we have the following bound: $\sum_{k=1}^{n} \gamma_k^{2+\epsilon} \|\Pi_{k+1,n}\|^{2+\epsilon} = \mathcal{O}(\gamma^{1+\epsilon}(n))$ so that we have

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

Moreover simple computations lead

$$\mathbb{E} \left| n^{\beta/2} \Delta N_k^{n^{\beta}} \right|^{2+\epsilon} \leq C(|n^{\beta/2}(h^{n^{\beta}}(\theta^*) - h(\theta^*))|^{2+\epsilon} + \mathbb{E}(n^{\beta/2}|H(\theta^*, X_T^{n^{\beta}}) - H(\theta^*, X_T)|)^{2+\epsilon}).$$

For the first term in the above inequality we have $\sup_{n\geq 1} |n^{\beta/2}(h^{n^{\beta}}(\theta^*) - h(\theta^*))|^{2+\,\epsilon} < +\infty \Leftrightarrow \alpha \geq 1/2$. For the second term, using assumption **(HLH)**, properties (2.6) and (2.7) we have $\sup_{n\geq 1} \mathbb{E}(n^{\beta/2}|H(\theta^*, X_T^{n^{\beta}}) - H(\theta^*, X_T)|)^{2+\epsilon} < +\infty$. Hence we conclude that

$$\sup_{n\geq 1} \sup_{k \in [\![1,n]\!]} \mathbb{E} \left| n^{\beta/2} \Delta N_k^{n^{\beta}} \right|^{2+\epsilon} < +\infty,$$

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

$$S_n := n^{2\alpha} \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \gamma_k^2 \Pi_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))} \mathbb{E}_k[\Delta N_k^{n^{\beta}} (\Delta N_k^{n^{\beta}})^T] \Pi^T_{k+1,\gamma^{-1}(1/(n^{2\alpha-\beta}T))}, \text{ and } U_T^{n^{\beta}} = X_T^{n^{\beta}} - X_T. \tag{3.17}$$

A Taylor's expansion yields

$$\sqrt{\frac{n^{\beta}}{T}} \left(H(\theta^*, X_T^{n^{\beta}}) - H(\theta^*, X_T) \right) = D_x H(\theta^*, X_T) \sqrt{\frac{n^{\beta}}{T}} U_T^{n^{\beta}} + \psi(\theta^*, X_T, U_T^{n^{\beta}}) \sqrt{\frac{n^{\beta}}{T}} U_T^{n^{\beta}}$$

with $\psi(\theta^*, X_T, U_T^{n^{\beta}}) \xrightarrow{\mathbb{P}} 0$. From the tightness of $(\sqrt{\frac{n^{\beta}}{T}} U_T^{n^{\beta}})_{n\geq 1}$, we get $\psi(\theta^*, X_T, U_T^{n^{\beta}}) \sqrt{\frac{n^{\beta}}{T}} U_T^{n^{\beta}} \xrightarrow{\mathbb{P}} 0$ so that using Theorem 2.1 and Lemma 2.1 yield

$$\sqrt{\frac{n^{\beta}}{T}} \left(H(\theta^*, X_T^{n^{\beta}}) - H(\theta^*, X_T) \right) \Longrightarrow D_x H(\theta^*, X_T) U_T.$$

Moreover, from assumption **(HLH)**, properties (2.6) and (2.7) it follows that

$$\forall p>0,\ \sup_{n\geq 1}\mathbb{E}|\sqrt{\frac{n^\beta}{T}}(H(\theta^*,X_T^{n^\beta})-H(\theta^*,X_T))|^{2+p}<+\infty,$$

which combined with **(HDH)** imply

$$\mathbb{E}\left(\sqrt{\frac{n^\beta}{T}}\left(H(\theta^*,X_T^{n^\beta})-H(\theta^*,X_T)\right)\right)\to\tilde{\mathbb{E}}D_xH(\theta^*,X_T)U_T=0$$

$$\mathbb{E}\left(\sqrt{\frac{n^\beta}{T}}\left(H(\theta^*,X_T^{n^\beta})-H(\theta^*,X_T)\right)\right)\left(\sqrt{\frac{n^\beta}{T}}\left(H(\theta^*,X_T^{n^\beta})-H(\theta^*,X_T)\right)\right)^T\to\tilde{\mathbb{E}}\left(D_xH(\theta^*,X_T)U_T\right)\left(D_xH(\theta^*,X_T)U_T\right)^T$$

where we used $\tilde{\mathbb{E}}D_xH(\theta^*,X_T)U_T=\tilde{\mathbb{E}}[D_xH(\theta^*,X_T)\tilde{\mathbb{E}}[U_T|\mathcal{F}_T]]$ and $\tilde{\mathbb{E}}[U_T|\mathcal{F}_T]=0$ (see e.g. Proposition 2.1, p.2685 in [Keb05]). Hence, we have

$$\Gamma_n\to\Gamma^*:=\tilde{\mathbb{E}}\left(D_xH(\theta^*,X_T)U_T\right)\left(D_xH(\theta^*,X_T)U_T\right)^T$$

where for $n\geq 1$

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

Consequently, using the following 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

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

which is a consequence of by 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. We denote by Θ^* 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 the lines of the proof of Lemma 2.3, step 3, we have $S_n\xrightarrow{a.s.}\Theta^*$. We leave the computational details to the reader. □

Proof of Theorem 3.1. We first write the following decomposition

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

For the last term of the above sum, we use Theorem 2.7 to directly deduce

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

For the first term, from Lemma 2.3 it follows

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

with $\Gamma^* := \int_0^\infty \exp\left(-s(Dh(\theta^*) - \zeta I_d)\right)^T \mathbb{E}_x[H(\theta^*, X_T)H(\theta^*, X_T)^T] \exp\left(-s(Dh(\theta^*) - \zeta I_d)\right) ds$. We decompose the last remaining term, namely $\theta^n_{\gamma^{-1}(1/n^{2\alpha-\beta})} - \theta^{n^\beta}_{\gamma^{-1}(1/n^{2\alpha-\beta})} - (\theta^{*,n} - \theta^{*,n^\beta})$ as follows

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

and use Lemma 3.1 to conclude the proof. □

Theorem 3.2. *Suppose that h and h^n satisfy the assumptions of Theorem 2.7 (with $\alpha \in (1/2 \vee \beta, 1]$) and that h satisfies the assumptions of Theorem 2.4. Assume that* **(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:*

$$\forall \theta \in \mathbb{R}^d,\ n^{\alpha-(\alpha-\beta/2)\rho}\|Dh(\theta) - Dh^{n^\beta}(\theta)\| \to 0, \quad \text{as } n \to +\infty. \tag{3.18}$$

Assume that the step sequence $\gamma = (\gamma_p)_{p\geq 1}$ satisfies **(HS1)** *with $\rho \in (1/2, 1)$ and $\rho > \frac{\alpha}{2\alpha-\beta} \vee \frac{\alpha(1-\beta)}{(\alpha-\beta/2)}$.*
Suppose that $\tilde{\mathbb{E}}(D_x H(\theta^*, X_T)U_T)(D_x H(\theta^*, X_T)U_T)^T$ is a positive definite matrix.
Then, for $M_3 = n^{2\alpha}$ and $M_4 = n^{2\alpha-\beta}T$, one has

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

where

$$\bar{\Sigma}^* := Dh(\theta^*)^{-1}(\mathbb{E}H(\theta^*, X_T)H(\theta^*, X_T)^T + \tilde{\mathbb{E}}\left(D_x H(\theta^*, X_T)U_T\right)\left(D_x H(\theta^*, X_T)U_T\right)^T)(Dh(\theta^*)^{-1})^T.$$

Lemma 3.2. *Let $(\bar{\theta}_p)_{p\geq 1}$ be the empirical mean sequence associated to $(\theta_p)_{p\geq 1}$ defined by* (3.15)*. Under the assumptions of Theorem 3.2, one has*

$$n^\alpha\left(\bar{\theta}^{n^\beta}_{n^{2\alpha-\beta}T} - \bar{\theta}_{n^{2\alpha-\beta}T} - (\theta^{*,n^\beta} - \theta^*)\right) \Longrightarrow \mathcal{N}(0, Dh(\theta^*)^{-1}\,\tilde{\mathbb{E}}\left(D_x H(\theta^*, X_T)U_T\right)\left(D_x H(\theta^*, X_T)U_T\right)^T(Dh(\theta^*)^{-1})^T),$$

and

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

Proof. We will just prove the first assertion. The second one will readily follow. The notation C denotes a constant that may change from one line to the next. Using the notations of Lemma 3.1, the sequence $(\bar{z}^{n^\beta}_p)_{p\in[\![0,n^{2\alpha-\beta}T]\!]}$ can be decomposed as follows:

$$\begin{aligned}\bar{z}^{n^\beta}_{n^{2\alpha-\beta}T} &= \frac{1}{n^{2\alpha-\beta}T+1}\sum_{k=0}^{n^{2\alpha-\beta}T} z^{n^\beta}_k \\ &= (Dh(\theta^*))^{-1}\frac{1}{n^{2\alpha-\beta}T+1}\left(\frac{z^{n^\beta}_{n^{2\alpha-\beta}T+1}}{\gamma_{n^{2\alpha-\beta}T+1}} - \frac{z^{n^\beta}_0}{\gamma_1}\right) + (Dh(\theta^*))^{-1}\frac{1}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k+1}}\right) z^{n^\beta}_k \\ &\quad - (Dh(\theta^*))^{-1}\frac{1}{n^{2\alpha-\beta}T+1}\sum_{k=0}^{n^{2\alpha-\beta}T}(\Delta N^{n^\beta}_{k+1} + \Delta R^{n^\beta}_{k+1}) \\ &\quad - (Dh(\theta^*))^{-1}\frac{1}{n^{2\alpha-\beta}T+1}\sum_{k=0}^{n^{2\alpha-\beta}T}(\zeta^{n^\beta}_k - \zeta_k + (Dh(\theta^*) - Dh^{n^\beta}(\theta^{*,n^\beta}))(\theta^{n^\beta}_k - \theta^{*,n^\beta})).\end{aligned}$$

Our aim is to study the contribution of each term in this decomposition.

Step 1: study of the sequence $\left\{\frac{n^{\alpha}}{n^{2\alpha-\beta}T+1}\left(\frac{z^{n^\beta}_{n^{2\alpha-\beta}T+1}}{\gamma_{n^{2\alpha-\beta}T+1}} - \frac{z^{n^\beta}_0}{\gamma_1}\right), n\geq 0\right\}$:

Using Proposition 5.2 clearly yields

$$\frac{n^{\alpha}}{n^{2\alpha-\beta}T+1}\left|\frac{z^{n^\beta}_{n^{2\alpha-\beta}T+1}}{\gamma_{n^{2\alpha-\beta}T+1}} - \frac{z^{n^\beta}_0}{\gamma_1}\right| \leq \frac{C}{(n^{\alpha-\beta}T)\gamma_{n^{2\alpha-\beta}T+1}}(|\tilde{\mu}^{n^\beta}_{n^{2\alpha-\beta}T}|+|\tilde{r}^{n^\beta}_{1,n^{2\alpha-\beta}T}|+|\tilde{r}_{2,n^{2\alpha-\beta}T}|)+\frac{C}{n^{\alpha-\beta}}|\theta^*-\theta^{*,n^\beta}|.$$

We evaluate each term appearing in the right hand side of the last but one inequality. First we clearly have

$$\frac{1}{(n^{\alpha-\beta}T)\gamma_{n^{2\alpha-\beta}T+1}}\mathbb{E}|\tilde{\mu}^{n^\beta}_{n^{2\alpha-\beta}T}| \leq \frac{C}{\sqrt{(n^{2\alpha-\beta}T)\gamma_{n^{2\alpha-\beta}T+1}}} \to 0, \quad \text{as } n\to+\infty,$$

and

$$\frac{1}{(n^{\alpha-\beta}T)\gamma_{n^{2\alpha-\beta}T+1}}|\tilde{r}^{n^\beta}_{1,n^{2\alpha-\beta}T}| \leq C\frac{1}{n^{\alpha-\beta}}\tilde{X}\tilde{Y}^{n^\beta}_{n^{2\alpha-\beta}T} \xrightarrow{\mathbb{P}} 0, \text{ as } n\to+\infty.$$

We write $\tilde{r}_{2,n^{2\alpha-\beta}T} = c_{1,n}+c_{2,n}$ where for $K>0$ the sequence $(c_{1,n})_{n\geq 0}$ is given by

$$c_{1,n} := \sum_{k=1}^{n^{2\alpha-\beta}T}\gamma_k\Pi_{k+1,n^{2\alpha-\beta}T}(H(\theta_k,(X_T)^{k+1})-H(\theta^*,(X_T)^{k+1})-(h(\theta_k)-h(\theta^*)))\mathbf{1}_{\left\{\gamma_k^{-1}(\sup_{k\geq 1}|\theta_k-\theta^*|)|r_k|\leq KY_k\right\}}$$

satisfying

$$\begin{aligned}\mathbb{E}|c_{1,n}|^2 &\leq C\sum_{k=1}^{n^{2\alpha-\beta}T}\gamma_k^2\|\Pi_{k+1,n^{2\alpha-\beta}T}\|^2\mathbb{E}|\theta_k-\theta^*|^2\mathbf{1}_{\left\{\gamma_k^{-1}(\sup_{k\geq 1}|\theta_k-\theta^*|)|r_k|\leq KY_k\right\}}\\ &\leq C\sum_{k=1}^{n^{2\alpha-\beta}T}\gamma_k^2\|\Pi_{k+1,n^{2\alpha-\beta}T}\|^2(\mathbb{E}|\mu_k|^2+\mathbb{E}(\sup_{k\geq 1}|\theta_k-\theta^*|)|r_k|\mathbf{1}_{\left\{\gamma_k^{-1}(\sup_{k\geq 1}|\theta_k-\theta^*|)|r_k|\leq KY_k\right\}})\\ &\leq C\sum_{k=1}^{n^{2\alpha-\beta}T}\gamma_k^3\|\Pi_{k+1,n^{2\alpha-\beta}T}\|^2 = \mathcal{O}(\gamma^2_{n^{2\alpha-\beta}T})\end{aligned}$$

where we used $|\theta_k-\theta^*|^2\leq|\mu_k|^2+2|\theta_k-\theta^*||r_k|$. Hence, we deduce that $\frac{1}{n^{\alpha-\beta}\gamma_{n^{2\alpha-\beta}T}}c_{1,n}\xrightarrow{L^1(\mathbb{P})}0$. Now observe that from Proposition 5.1 for all $k\in[\![1,n]\!]$, we have $\gamma_k^{-1}r_k\leq KY_k$ so that for all $\epsilon>0$

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n^{\alpha-\beta}\gamma_{n^{2\alpha-\beta}T+1}}c_{2,n}>\epsilon\right) &\leq \mathbb{P}\left(\frac{C}{n^{\alpha-\beta}\gamma_{n^{2\alpha-\beta}T+1}}c_{2,n}>\epsilon,(\sup_{k\geq 1}|\theta_k-\theta^*|)X\leq K\right)+\mathbb{P}\left((\sup_{k\geq 1}|\theta_k-\theta^*|)X>K\right)\\ &= \mathbb{P}\left((\sup_{k\geq 1}|\theta_k-\theta^*|)X>K\right).\end{aligned}$$

Therefore passing to the limit $n\to+\infty$ in the previous inequality we get $\lim_n\mathbb{P}(\frac{1}{n^{\alpha-\beta}\gamma_{n^{2\alpha-\beta}T+1}}c_{2,n}>\epsilon)\leq\mathbb{P}\left((\sup_{k\geq 1}|\theta_k-\theta^*|)X>K\right)$ for all $K>0$. Passing to the limit $K\to+\infty$, we conclude that $\frac{1}{n^{\alpha-\beta}\gamma_{n^{2\alpha-\beta}T+1}}c_{2,n}\xrightarrow{\mathbb{P}}0$ as $n\to+\infty$ which in turn implies

$$\frac{1}{n^{\alpha-\beta}\gamma_{n^{2\alpha-\beta}T+1}}|\tilde{r}_{2,n^{2\alpha-\beta}T}|\xrightarrow{\mathbb{P}}0.$$

Finally since $\alpha > \beta$ and $|\theta^* - \theta^{*,n^\beta}| \to 0$, we clearly have $\frac{1}{n^{\alpha-\beta}}|\theta^* - \theta^{*,n^\beta}| \to 0$ so that

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\left(\frac{z^{n^\beta}_{n^{2\alpha-\beta}T+1}}{\gamma_{n^{2\alpha-\beta}T+1}} - \frac{z^{n^\beta}_0}{\gamma_1}\right) \xrightarrow{\mathbb{P}} 0, \quad \text{as} \quad n \to +\infty.$$

Step 2: study of the sequence $\left\{\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k+1}}\right) z^{n^\beta}_k, n \geq 0\right\}$:

Note that we also have

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\left|\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k+1}}\right) z^{n^\beta}_k\right| \leq \frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)(|\tilde{\mu}^{n^\beta}_k| + |\tilde{r}^{n^\beta}_{1,k}| + |\tilde{r}_{2,k}|).$$

We take the expectation of the first term appearing in the right hand side of the above inequality and since $\rho < 1$ we deduce

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)\mathbb{E}|\tilde{\mu}^{n^\beta}_k| \leq \frac{C}{n^{\alpha-\beta/2}T}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)\gamma_k^{\frac{1}{2}} \to 0.$$

For the second term, we have

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)|\tilde{r}^{n^\beta}_{1,k}| \leq \frac{\tilde{X}}{n^{\alpha-\beta}T}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)\gamma_k \tilde{Y}^n_k,$$

which combined with

$$\frac{1}{n^{\alpha-\beta}T}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)\gamma_k \mathbb{E}\tilde{Y}^n_k \leq (\sup_{n\geq 1, k\geq 1}\mathbb{E}\tilde{Y}^n_k)\frac{1}{n^{\alpha-\beta}T}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)\gamma_k \to 0$$

since $\alpha > \beta$ allow to deduce that

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)|\tilde{r}^{n^\beta}_{1,k}| \xrightarrow{\mathbb{P}} 0.$$

For the third term, we use the decomposition $\tilde{r}_{2,p} = c_{1,p} + c_{2,p}$ as previously done. We clearly have

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)\mathbb{E}|c_{1,k}| \leq \frac{C}{n^{\alpha-\beta}T}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)\gamma_k \to 0,$$

and for all $K > 0$

$$\begin{aligned}
\mathbb{P}\left(\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)|c_{2,k}| > \epsilon\right) &\leq \mathbb{P}\left(\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right)|c_{2,k}| > \epsilon, (\sup_{k\geq 1}|\theta_k - \theta^*|)X \leq K\right) \\
&\quad + \mathbb{P}\left((\sup_{k\geq 1}|\theta_k - \theta^*|)X > K\right) \\
&= \mathbb{P}\left((\sup_{k\geq 1}|\theta_k - \theta^*|)X > K\right)
\end{aligned}$$

so that passing to the limit $n \to +\infty$ we get $\lim_n \mathbb{P}\left(\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}}-\frac{1}{\gamma_k}\right)|c_{2,k}|>\epsilon\right) \leq \mathbb{P}\left((\sup_{k\geq 1}|\theta_k-\theta^*|)X>K\right)$. Passing to the limit $K \to +\infty$ leads to $\lim_n \mathbb{P}\left(\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_{k+1}}-\frac{1}{\gamma_k}\right)|c_{2,k}|>\epsilon\right)=0$ which in turn implies

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\left(\frac{1}{\gamma_k}-\frac{1}{\gamma_{k+1}}\right)z_k^{n^\beta} \xrightarrow{\mathbb{P}} 0.$$

Step 3: study of the sequence $\left\{\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}(\zeta_k^{n^\beta}-\zeta_k+(Dh(\theta^*)-Dh^{n^\beta}(\theta^{*,n^\beta}))(\theta_k^{n^\beta}-\theta^{*,n^\beta})), n\geq 0\right\}$:

Now we focus on the last term. We firstly note that thanks to Lemma 5.2 we clearly have

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\mathbb{E}\left|\sum_{k=0}^{n^{2\alpha-\beta}T}\zeta_k^{n^\beta}\right| \leq \frac{C}{n^{\alpha-\beta}T}\sum_{k=0}^{n^{2\alpha-\beta}T}\gamma_k \to 0$$

since $\rho > \alpha/(2\alpha-\beta)$. Moreover, from Proposition 5.1 it follows

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\left|\sum_{k=0}^{n^{2\alpha-\beta}T}\zeta_k\right| \leq \frac{C}{n^{\alpha-\beta}T}\sum_{k=0}^{n^{2\alpha-\beta}T}|\mu_k|^2+(\sup_{k\geq 1}|\theta_k-\theta^*|)|r_k| \leq \frac{C}{n^{\alpha-\beta}T}\sum_{k=0}^{n^{2\alpha-\beta}T}|\mu_k|^2+(\sup_{k\geq 1}|\theta_k-\theta^*|)X\frac{C}{n^{\alpha-\beta}T}\sum_{k=0}^{n^{2\alpha-\beta}T}\gamma_k Y_k.$$

The first term converges to zero in $L^1(\mathbb{P})$ since $(1/(n^{-(\alpha-\beta)}T)\sum_{k=0}^{n^{2\alpha-\beta}T}\mathbb{E}|\mu_k|^2 \leq (\sup_{k\geq 1}\gamma_k^{-1}\mathbb{E}|\mu_k|^2)\frac{1}{n^{\alpha-\beta}T}\sum_{k=0}^{n^{2\alpha-\beta}}\gamma_k \to 0$ and similarly the second term converges to zero in probability. Now since Dh^{n^β} is Lipschitz-continuous uniformly in n we easily get

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\mathbb{E}\left|\sum_{k=0}^{n^{2\alpha-\beta}T}(Dh(\theta^*)-Dh^{n^\beta}(\theta^{*,n^\beta}))(\theta_k^{n^\beta}-\theta^{*,n^\beta})\right| \leq \frac{C}{n^{\alpha-\beta}T}(\|Dh(\theta^*)-Dh^{n^\beta}(\theta^*)\|+|\theta^*-\theta^{*,n^\beta}|)\sum_{k=0}^{n^{2\alpha-\beta}T}\gamma_k^{\frac{1}{2}},$$

and recalling that $n^{\alpha-(\alpha-\beta/2)\rho}\|Dh(\theta^*)-Dh^{n^\beta}(\theta^*)\| \to 0$ and $\rho > \alpha(1-\beta)/(\alpha-\beta/2)$ which implies $n^{\alpha-(\alpha-\beta/2)\rho}|\theta^*-\theta^{*,n^\beta}| \to 0$ we deduce

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=0}^{n^{2\alpha-\beta}T}(Dh(\theta^*)-Dh^{n^\beta}(\theta^{*,n^\beta}))(\theta_k^{n^\beta}-\theta^{*,n^\beta}) \xrightarrow{L^1(\mathbb{P})} 0.$$

Step 4: study of the sequence $\left\{\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=0}^{n^{2\alpha-\beta}T}(\Delta N_{k+1}^{n^\beta}+\Delta R_{k+1}^{n^\beta}), n\geq 0\right\}$:

Similarly to the proof of Lemma 3.3, we decompose the sequence $\left\{\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=1}^{n^{2\alpha-\beta}T}\Delta R_k^{n^\beta}, n\geq 1\right\}$ as follows

$$\begin{aligned}\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=0}^{n^{2\alpha-\beta}T}\Delta R_k^{n^\beta} &= \frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=0}^{n^{2\alpha-\beta}T}(h^{n^\beta}(\theta_k^{n^\beta})-h^{n^\beta}(\theta^*)-(H(\theta_k^{n^\beta},(X_T^{n^\beta})^{k+1})-H(\theta^*,(X_T^{n^\beta})^{k+1})))\\ &+\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=0}^{n^{2\alpha-\beta}T}(H(\theta_k,(X_T)^{k+1})-H(\theta^*,(X_T)^{k+1})-(h(\theta_k)-h(\theta^*)))\\ &= A_n+B_n.\end{aligned}$$

From the Cauchy-Schwarz inequality and Lemma 5.2 it easily follows

$$\mathbb{E}|A_n| \le \frac{1}{n^{\alpha-\beta}T}\left(\sum_{k=0}^{n^{2\alpha-\beta}T}\gamma_k\right)^{\frac{1}{2}} \to 0$$

since $\rho > \alpha/(2\alpha-\beta) > \beta/(2\alpha-\beta)$. Now we write $B_n = B_{1,n} + B_{2,n}$ with for all $K > 0$

$$B_{1,n} = \frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=0}^{n^{2\alpha-\beta}T}(H(\theta_k,(X_T)^{k+1}) - H(\theta^*,(X_T)^{k+1}) - (h(\theta_k)-h(\theta^*)))\mathbf{1}_{\{\gamma_k^{-1}(\sup_{k\ge1}|\theta_k-\theta^*|)|r_k|\le KY_k\}}.$$

and simple computations similar to that of the sequence $(c_{1,n})_{n\ge1}$ lead to

$$\mathbb{E}|B_{1,n}| \le \frac{C}{n^{\alpha-\beta}T}\left(\sum_{k=0}^{n^{2\alpha-\beta}T}\gamma_k\right)^{1/2} \to 0.$$

Moreover, similarly to the computations done for the sequence $(c_{2,n})_{n\ge1}$, we have $\lim_n \mathbb{P}(B_{2,n} > \epsilon) \le \mathbb{P}\left((\sup_{k\ge1}|\theta_k-\theta^*|)X > K\right)$ and passing to the limit $K \to +\infty$ we obtain $\lim_n \mathbb{P}(B_{2,n} > \epsilon) = 0$ which in turn implies

$$\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=0}^{n^{2\alpha-\beta}T}\Delta R_k^{n^\beta} \xrightarrow{\mathbb{P}} 0, \quad n \to +\infty.$$

We now prove a CLT for the sequence $\left\{\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\sum_{k=0}^{n^{2\alpha-\beta}T}\Delta N_k^{n^\beta},\ n\ge0\right\}$. Let $\epsilon > 0$.

$$\sum_{k=0}^{n^{2\alpha-\beta}T}\mathbb{E}\left|\frac{n^\alpha}{n^{2\alpha-\beta}T+1}\Delta N_k^{n^\beta}\right|^{2+\epsilon} \le \sup_{n\ge1}\sup_{k\in[\![0,n]\!]}\mathbb{E}\left|n^{\beta/2}\Delta N_k^{n^\beta}\right|^{2+\epsilon}\frac{1}{n^{\alpha\epsilon-\beta\epsilon/2}} \to 0 \quad n\to+\infty$$

where we used assumption **(HLH)**, properties (2.6) and (2.7) to derive that $\sup_{n\ge1}\sup_{k\in[\![1,n]\!]}\mathbb{E}\left|n^{\beta/2}\Delta N_k^{n^\beta}\right|^{2+\epsilon} < +\infty$. Therefore the conditional Lindeberg condition is satisfied. Now, we focus on the conditional variance. Recall that (see the the proof of Lemma 3.3) we have

$$\begin{aligned}\frac{n^\beta}{T}\mathbb{E}_k[\Delta N_k^{n^\beta}(\Delta N_k^{n^\beta})^T] &= \frac{n^\beta}{T}\mathbb{E}\Big[(H(\theta^*,X_T^{n^\beta}) - H(\theta^*,X_T) - (h^{n^\beta}(\theta^*)-h(\theta^*)))\\ &\qquad\times(H(\theta^*,X_T^{n^\beta}) - H(\theta^*,X_T) - (h^{n^\beta}(\theta^*)-h(\theta^*)))^T\Big]\\ &\to \tilde{\mathbb{E}}\left(D_xH(\theta^*,X_T)U_T\right)\left(D_xH(\theta^*,X_T)U_T\right)^T,\end{aligned}$$

so that if we set

$$\begin{aligned}S_n &:= \frac{n^{2\alpha}}{(n^{2\alpha-\beta}T+1)^2}\sum_{k=0}^{n^{2\alpha-\beta}T}\mathbb{E}_k[\Delta N_k^{n^\beta}(\Delta N_k^{n^\beta})^T]\\ &= \frac{n^{2\alpha-\beta}T}{n^{2\alpha-\beta}T+1}\frac{n^\beta}{T}\mathbb{E}\left[(H(\theta^*,X_T^{n^\beta}) - H(\theta^*,X_T) - (h^{n^\beta}(\theta^*)-h(\theta^*)))(H(\theta^*,X_T^{n^\beta}) - H(\theta^*,X_T) - (h^{n^\beta}(\theta^*)-h(\theta^*)))^T\right],\end{aligned}$$

we clearly get

$$S_n \to \tilde{\mathbb{E}}\left(D_x H(\theta^*, X_T)U_T\right)\left(D_x H(\theta^*, X_T)U_T\right)^T.$$

This completes the proof. □

Proof of Theorem 3.2. We decompose the error as follows

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

For the first term, from Lemma 2.4 it follows that

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

For the last term using Theorem 2.7, we have $n^\alpha(\theta^{*,n} - \theta^*) \to -Dh^{-1}(\theta^*)\mathcal{E}(h, \alpha, \theta^*)$. We now focus on the last remaining term, namely $\bar{\theta}^{n}_{n^{2\alpha-\beta}} - \bar{\theta}^{n^\beta}_{n^{2\alpha-\beta}} - (\theta^{*,n} - \theta^{*,n^\beta})$. We decompose it as follows

$$\bar{\theta}^{n}_{n^{2\alpha-\beta}} - \bar{\theta}^{n^\beta}_{n^{2\alpha-\beta}} - (\theta^{*,n} - \theta^{*,n^\beta}) = \bar{\theta}^{n}_{n^{2\alpha-\beta}} - \bar{\theta}_{n^{2\alpha-\beta}} - (\theta^{*,n} - \theta^*) - (\bar{\theta}^{n^\beta}_{n^{2\alpha-\beta}} - \bar{\theta}_{n^{2\alpha-\beta}} - (\theta^{*,n^\beta} - \theta^*))$$

where $(\bar{\theta}_p)_{p\geq 1}$ is the empirical mean sequence associated to $(\theta_p)_{p\geq 1}$ and use Lemma 3.2 to conclude the proof. □

3.2. **The multi-level stochastic approximation method**

As mentioned in the introduction the multi-level stochastic approximation scheme uses L Euler approximation schemes with different time steps given by T/m^ℓ, $\ell \in \{1, \cdots, L\}$ for a fixed integer $m \geq 2$ such that $m^L = n$ and estimates θ^* by computing the quantity

$$\Theta^{ml}_n = \theta^1_{M_0} + \sum_{\ell=1}^{L} \theta^{m^\ell}_{M_\ell} - \theta^{m^{\ell-1}}_{M_\ell}.$$

It is important to point out here that each couple $(\theta^{m^\ell}_{M_\ell}, \theta^{m^{\ell-1}}_{M_\ell})$ is computed using two Euler approximation schemes with different time steps but with the same Brownian path. Moreover, for two different levels, the stochastic approximation schemes are based on independent Brownian paths.

Theorem 3.3. *Suppose that h and h^{m^ℓ}, $\ell = 0, \cdots, L$, satisfy the assumptions of Theorem 2.7 (with $\alpha = 1$) and that h satisfies the assumptions of Theorem 2.4. Assume that* **(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:*

$$\exists \beta > 1/2, \quad \forall \theta \in \mathbb{R}^d, \ n^\beta \|Dh^n(\theta) - Dh(\theta)\| \to 0, \quad \text{as } n \to +\infty.$$

Suppose that $\tilde{\mathbb{E}}(D_x H(\theta^*, X_T)U_T)(D_x H(\theta^*, X_T)U_T)^T$ 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 $(-\rho)$, $\rho \in (1/2, 1)$, that is, for any $x > 0$, $\lim_{t\to+\infty} \gamma(tx)/\gamma(t) = x^{-\rho}$.*
- *for $t \geq 1$, $\gamma(t) = \gamma_0/t$ and γ_0 satisfies $\underline{\lambda}\gamma_0 > 1$.*

Then, for $M_0 = \gamma^{-1}(1/n^2)$ and $M_l = \gamma^{-1}(m^\ell \log(m)/(n^2 \log(n)(m-1)T))$, $\ell = 1, \cdots, L$, one has

$$n(\Theta^{ml}_n - \theta^*) \Longrightarrow -Dh^{-1}(\theta^*)\mathcal{E}(h, 1, \theta^*) + \mathcal{N}(0, \Sigma^*), \quad n \to +\infty$$

with

$$\Sigma^* := \int_0^\infty \left(e^{-s(Dh(\theta^*)-\zeta I_d)}\right)^T (\mathbb{E}_x[H(\theta^*, X^1_T)H(\theta^*, X^1_T)^T] + \tilde{\mathbb{E}}\left(D_x H(\theta^*, X_T)U_T\right)\left(D_x H(\theta^*, X_T)U_T\right)^T) e^{-s(Dh(\theta^*)-\zeta I_d)} ds$$

and U_T is the value at time T of the process (2.8) defined on $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t\geq 0}, \tilde{\mathbb{P}})$.

Proof. We first write the following decomposition

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

For the last term of the above sum, we use Theorem 2.7 to directly deduce

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

For the first term, the standard CLT (theorem 2.4) for stochastic approximation leads to

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

with $\Gamma^* := \int_0^\infty \exp\left(-s(Dh(\theta^*) - \zeta I_d)\right)^T \mathbb{E}_x[H(\theta^*, X_T^1)H(\theta^*, X_T^1)^T] \exp\left(-s(Dh(\theta^*) - \zeta I_d)\right) ds$.To deal with the last remaining term, namely $n^\alpha \sum_{\ell=1}^{L} \theta^{m^\ell}_{M_\ell} - \theta^{m^{\ell-1}}_{M_\ell} - (\theta^{*,m^\ell} - \theta^{*,m^{\ell-1}})$ we will need the following lemma. □

Lemma 3.3. *Under the assumptions of Theorem 3.1, one has*

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

with $\Theta^* := \int_0^\infty \exp\left(-s(Dh(\theta^*) - \zeta I_d)\right)^T \tilde{\mathbb{E}}\left(D_x H(\theta^*, X_T)U_T\right)\left(D_x H(\theta^*, X_T)U_T\right)^T \exp\left(-s(Dh(\theta^*) - \zeta I_d)\right) ds.$

Proof. We come back to the decomposition used in the proof of Lemma 2.3. We consequently use the same notations. We will not go into all computational details.

A Taylor's expansion yields for $p \geq 0$

$$\theta^{m^\ell}_{p+1} - \theta^{*,m^\ell} = \theta^{m^\ell}_p - \theta^{*,m^\ell} - \gamma_{p+1} Dh^{m^\ell}(\theta^{*,m^\ell})(\theta^{m^\ell}_p - \theta^{*,m^\ell}) + \gamma_{p+1}\Delta M^{m^\ell}_{p+1} - \gamma_{p+1}\zeta^{m^\ell}_p$$

with $\Delta M^{m^\ell}_{p+1} = h^{m^\ell}(\theta^{m^\ell}_p) - H(\theta^{m^\ell}_p, (X^{m^\ell}_T)^{p+1})$, $p \geq 0$. Therefore, defining $z^\ell_p = \theta^{m^\ell}_p - \theta^{m^{\ell-1}}_p - (\theta^{*,m^\ell} - \theta^{*,m^{\ell-1}})$, $p \geq 0$, with $z^\ell_0 = \theta^{*,m^\ell} - \theta^{*,m^{\ell-1}}$, by a simple induction argument one has

$$\begin{aligned} z^\ell_{M_\ell} &= \Pi_{1,M_\ell} z^\ell_0 + \sum_{k=1}^{M_\ell} \gamma_k \Pi_{k+1,M_\ell} \Delta N^\ell_k + \sum_{k=1}^{M_\ell} \gamma_k \Pi_{k+1,M_\ell} \Delta R^\ell_k \\ &\quad + \sum_{k=1}^{M_\ell} \gamma_k \Pi_{k+1,M_\ell} \Big(\zeta^\ell_{k-1} - \zeta^{\ell-1}_{k-1} + (Dh(\theta^*) - Dh^{m^\ell}(\theta^{*,m^\ell}))(\theta^{m^\ell}_{k-1} - \theta^{*,m^\ell}) \\ &\quad - (Dh(\theta^*) - Dh^{m^{\ell-1}}(\theta^{*,m^{\ell-1}}))(\theta^{m^{\ell-1}}_{k-1} - \theta^{*,m^{\ell-1}}) \Big) \end{aligned} \tag{3.19}$$

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 $\Delta N^\ell_k := h^{m^\ell}(\theta^*) - h^{m^{\ell-1}}(\theta^*) - (H(\theta^*, (X^{m^\ell}_T)^{k+1}) - H(\theta^*, (X^{m^{\ell-1}}_T)^{k+1}))$, $\Delta R^\ell_k = h^{m^\ell}(\theta^{m^\ell}_k) - h^{m^\ell}(\theta^*) - (H(\theta^{m^\ell}_k, (X^{m^\ell}_T)^{k+1}) - H(\theta^*, (X^{m^\ell}_T)^{k+1})) + H(\theta^{m^{\ell-1}}_k, (X^{m^{\ell-1}}_T)^{k+1}) - H(\theta^*, (X^{m^{\ell-1}}_T)^{k+1}) - (h^{m^{\ell-1}}(\theta^{m^{\ell-1}}_k) - h^{m^{\ell-1}}(\theta^*))$ for $k \geq 0$.

Step 1: study of $\left\{ n \sum_{\ell=1}^{L} \Pi_{1,M_\ell} z^\ell_0, n \geq 0 \right\}$

Under the assumptions on the step sequence γ, for all $\eta \in (0, \lambda_m)$ we have $\|\Pi_{1,M_\ell}\| \leq \exp(-(\lambda_m - \eta)\sum_{k=1}^{M_\ell} \gamma_k) \leq C\sqrt{\gamma(M_\ell)}$ for a positive constant C independent of ℓ (select η s.t. $2(\lambda_m - \eta)\gamma_0 > 1$ if $\gamma(p) = \gamma_0/p$,

$p \geq 1$). Therefore, one has

$$\left| n \sum_{\ell=1}^{L} \Pi_{1,M_\ell} z_0^\ell \right| \leq n \sum_{\ell=1}^{L} \|\Pi_{1,M_\ell}\| |\theta^{*,m^\ell} - \theta^{*,m^{\ell-1}}| \leq Cn \sum_{\ell=1}^{L} \frac{1}{m^\ell} \frac{m^{\ell/2}}{n \log^{1/2}(n)} \leq \frac{C}{\log^{1/2}(n)} \to 0, \quad n \to +\infty.$$

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

By Lemma 5.2, one has

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

However, by 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

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

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

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

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

which combined with $\sup_{n \geq 1} n^\beta \|Dh(\theta^*) - Dh^n(\theta^*)\| < +\infty$ with $\beta > 1/2$ and $\sup_{n \geq 1} n|\theta^{*,n} - \theta^*| < +\infty$ imply that

$$\mathbb{E} \left| n \sum_{\ell=1}^{L} \sum_{k=1}^{M_\ell} \gamma_k \Pi_{k+1,M_\ell} (Dh(\theta^*) - Dh^{m^\ell}(\theta^{*,m^\ell}))(\theta_{k-1}^{m^\ell} - \theta^{*,m^\ell}) \right| \leq \frac{C}{\log^{1/2}(n)} \sum_{\ell=1}^{L} m^{\ell/2}(m^{-\ell} + m^{-\ell\beta}) \leq \frac{C}{\log^{1/2}(n)}$$

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

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

We take the square of the $L^2(\mathbb{P})$-norm of this term to deduce

$$\begin{aligned}\mathbb{E}\left|n\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}\Delta R_k^\ell\right|^2 &\leq 2n^2\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k^2\|\Pi_{k+1,M_\ell}\|^2\mathbb{E}|H(\theta_k^{m^\ell},(X_T^{m^\ell})^{k+1})-H(\theta^*,(X_T^{m^\ell})^{k+1})|^2\\ &+2n^2\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k^2\|\Pi_{k+1,M_\ell}\|^2\mathbb{E}|H(\theta_k^{m^{\ell-1}},(X_T^{m^{\ell-1}})^{k+1})-H(\theta^*,(X_T^{m^{\ell-1}})^{k+1})|^2\\ &\leq Cn^2\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k^3\|\Pi_{k+1,M_\ell}\|^2\end{aligned}$$

where we used **(HLH)** and Lemma 5.2. Now from Lemma 5.1 and simple computations it follows

$$n^2\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k^3\|\Pi_{k+1,M_\ell}\|^2\leq Cn^2\sum_{\ell=1}^{L}\gamma^2(M_\ell)\leq\frac{C}{n^2\log^2(n)}\sum_{\ell=1}^{L}m^{2\ell}\leq\frac{C}{\log^2(n)}\to 0,\quad n\to+\infty.$$

Therefore, we conclude that

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

Step 5: study of $\left\{n\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}\Delta N_k^\ell, n\geq 0\right\}$

We now prove a CLT for the sequence $\left\{n\sum_{\ell=1}^{L}\sum_{k=1}^{M_\ell}\gamma_k\Pi_{k+1,M_\ell}\Delta N_k^\ell,\ n\geq 0\right\}$. Let $\epsilon>0$. By Burkholder's inequality and elementary computations, it holds

$$\begin{aligned}\sum_{\ell=1}^{L}\mathbb{E}\left|\sum_{k=1}^{M_\ell}n\gamma_k\Pi_{k+1,M_\ell}\Delta N_k^\ell\right|^{2+\epsilon} &\leq Cn^{(2+\epsilon)}\sum_{\ell=1}^{L}\mathbb{E}\left(\sum_{k=1}^{M_\ell}\gamma_k^2\|\Pi_{k+1,M_\ell}\|^2|\Delta N_k^\ell|^2\right)^{1+\epsilon/2}\\ &\leq Cn^{(2+\epsilon)}\sum_{\ell=1}^{L}(\sum_{k=1}^{M_\ell}\gamma_k^2\|\Pi_{k+1,M_\ell}\|^2)^{\epsilon/2}\sum_{k=1}^{M_\ell}\gamma_k^{2+\epsilon}\|\Pi_{k+1,M_\ell}\|^{2+\epsilon}\mathbb{E}|\Delta N_k^\ell|^{2+\epsilon}.\end{aligned}$$

Using **(HLH)**, properties (2.6) and (2.7) we have $\sup_{\ell\geq 1}\mathbb{E}(m^{\ell/2}|H(\theta^*,X_T^{m^\ell})-H(\theta^*,X_T)|)^{2+\epsilon}<+\infty$ so that

$$\mathbb{E}|\Delta N_k^\ell|^{2+\epsilon}\leq\frac{K}{m^{\ell(1+\epsilon/2)}}.$$

Moreover, by Lemma 5.1, we have

$$\limsup_n(1/\gamma^{(1+\epsilon)}(n))\sum_{k=1}^{n}\gamma_k^{2+\epsilon}\|\Pi_{k+1,n}\|^{2+\epsilon}\leq 1\quad\text{and}\quad\limsup_n(1/\gamma(n))\sum_{k=1}^{n}\gamma_k^2\|\Pi_{k+1,n}\|^2\leq 1$$

so that

$$\sum_{\ell=1}^{L}\mathbb{E}\left|\sum_{k=1}^{M_\ell}n\gamma_k\Pi_{k+1,M_\ell}\Delta N_k^\ell\right|^{2+\epsilon}\leq Cn^{(2+\epsilon)}\sum_{\ell=1}^{L}\gamma^{1+3\epsilon/2}(M_\ell)m^{-\ell(1+\epsilon/2)}\leq\frac{C}{n^{2\epsilon}\log^{1+3\epsilon/2}(n)}\sum_{\ell=1}^{L}m^{\ell\epsilon}\leq\frac{C}{n^{\epsilon}\log^{1+3\epsilon/2}(n)}$$

which in turn implies

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

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

$$S_\ell := n^2 \sum_{k=1}^{M_\ell} \gamma_k^2 \Pi_{k+1,M_\ell}\mathbb{E}_k[\Delta N_k^{\ell}(\Delta N_k^{\ell})^T]\Pi_{k+1,M_\ell}^T, \quad \text{and} \quad U_T^{\ell} = X_T^{m^{\ell}} - X_T^{m^{\ell-1}}. \tag{3.20}$$

Observe that by the very definition of M_ℓ one has

$$S_\ell = \frac{1}{\gamma(M_\ell)}\frac{\log(m)}{\log(n)}\frac{m^{\ell}}{(m-1)T}\sum_{k=1}^{M_\ell}\gamma_k^2 \Pi_{k+1,M_\ell}\mathbb{E}_k[\Delta N_k^{\ell}(\Delta N_k^{\ell})^T]\Pi_{k+1,M_\ell}^T$$

A Taylor's expansion yields

$$H(\theta^*, X_T^{m^{\ell}}) - H(\theta^*, X_T^{m^{\ell-1}}) = D_x H(\theta^*, X_T)U_T^{\ell} + \psi(\theta^*, X_T, X_T^{m^{\ell}} - X_T)(X_T^{m^{\ell}} - X_T) + \psi(\theta^*, X_T, X_T^{m^{\ell-1}} - X_T)(X_T^{m^{\ell-1}} - X_T)$$

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

$$\sqrt{\frac{m^{\ell}}{(m-1)T}}\Big(\psi(\theta^*, X_T, X_T^{m^{\ell}} - X_T)(X_T^{m^{\ell}} - X_T) + \psi(\theta^*, X_T, X_T^{m^{\ell-1}} - X_T)(X_T^{m^{\ell-1}} - X_T)\Big) \xrightarrow{\mathbb{P}} 0, \quad \ell \to +\infty.$$

Therefore using Theorem 2.1 and Lemma 2.1 yield

$$\sqrt{\frac{m^{\ell}}{(m-1)T}}\Big(H(\theta^*, X_T^{m^{\ell}}) - H(\theta^*, X_T^{m^{\ell-1}})\Big) \Longrightarrow D_x H(\theta^*, X_T)U_T.$$

Moreover, from assumption **(HLH)**, properties (2.6) and (2.7) it follows that

$$\forall p>0, \quad \sup_{\ell\geq 1}\mathbb{E}\left|\sqrt{\frac{m^{\ell}}{(m-1)T}}(H(\theta^*, X_T^{m^{\ell}}) - H(\theta^*, X_T^{m^{\ell-1}}))\right|^{2+p} < +\infty,$$

which combined with **(HDH)** imply

$$\sqrt{\frac{m^{\ell}}{(m-1)T}}\mathbb{E}(H(\theta^*, X_T^{m^{\ell}}) - H(\theta^*, X_T^{m^{\ell-1}})) \to \tilde{\mathbb{E}} D_x H(\theta^*, X_T)U_T = 0$$

$$\frac{m^{\ell}}{(m-1)T}\mathbb{E}(H(\theta^*, X_T^{m^{\ell}}) - H(\theta^*, X_T^{m^{\ell-1}}))(H(\theta^*, X_T^{m^{\ell}}) - H(\theta^*, X_T^{m^{\ell-1}}))^T \to \tilde{\mathbb{E}}\left(D_x H(\theta^*, X_T)U_T\right)\left(D_x H(\theta^*, X_T)U_T\right)^T$$

as $\ell\to+\infty$, where we used $\tilde{\mathbb{E}} D_x H(\theta^*, X_T)U_T = \tilde{\mathbb{E}}[D_x H(\theta^*, X_T)\tilde{\mathbb{E}}[U_T|\mathcal{F}_T]]$ and $\tilde{\mathbb{E}}[U_T|\mathcal{F}_T] = 0$. Hence, we have

$$\frac{m^{\ell}}{(m-1)T}\Gamma_\ell \to \Gamma^* := \tilde{\mathbb{E}}\left(D_x H(\theta^*, X_T)U_T\right)\left(D_x H(\theta^*, X_T)U_T\right)^T$$

where for $\ell \geq 1$

$$\Gamma_\ell := \mathbb{E}_k[\Delta N_k^\ell (\Delta N_k^\ell)^T]$$
$$= \mathbb{E}(H(\theta^*, X_T^{m^\ell}) - H(\theta^*, X_T^{m^{\ell-1}}))(H(\theta^*, X_T^{m^\ell}) - H(\theta^*, X_T^{m^{\ell-1}}))^T - (h^{m^\ell}(\theta^*) - h^{m^{\ell-1}}(\theta^*))(h^{m^\ell}(\theta^*) - h^{m^{\ell-1}}(\theta^*))^T.$$

Consequently, using the following decomposition

$$\frac{1}{\gamma(M_\ell)} \frac{m^\ell}{(m-1)T} \sum_{k=1}^{M_\ell} \gamma_k^2 \Pi_{k+1,M_\ell} \Gamma_\ell \Pi_{k+1,M_\ell}^T = \frac{1}{\gamma(M_\ell)} \sum_{k=1}^{M_\ell} \gamma_k^2 \Pi_{k+1,M_\ell} \Gamma^* \Pi_{k+1,M_\ell}^T$$
$$+ \frac{1}{\gamma(M_\ell)} \sum_{k=1}^{M_\ell} \gamma_k^2 \Pi_{k+1,M_\ell} \left(\frac{m^\ell}{(m-1)T} \Gamma_\ell - \Gamma^* \right) \Pi_{k+1,M_\ell}^T$$

with

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

which is a consequence of Lemma 5.1, we clearly see that $\frac{\log(n)}{\log(m)} \lim_\ell S_\ell = \lim_{p \to +\infty} \frac{1}{\gamma(p)} \sum_{k=1}^p \gamma_k^2 \Pi_{k+1,p} \Gamma^* \Pi_{k+1,p}^T$ if this latter limit exists. We denote by Θ^* 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 the lines of the proof of Lemma 2.3, step 3, we have $\frac{\log(n)}{\log(m)} S_\ell \xrightarrow{a.s.} \Theta^*$. We leave the computational details to the reader. Finally, from Cesàro's Lemma it follows that

$$\sum_{\ell=1}^L S_\ell = \frac{1}{L} \sum_{\ell=1}^L S_\ell \frac{\log(n)}{\log(m)} \underset{n \to +\infty}{\longrightarrow} \Theta^*.$$

□

Remark 3.1. *The previous result shows that a CLT for the multi-level stochastic approximation estimator of θ^* holds if the weak discretization, and thus the implicit discretization errors, is of order $1/n$. Due to the non-linearity of the procedure, this result seems not to extend to a weak discretization error of order $1/n^\alpha$ with $\alpha < 1$. 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.*

3.3. **Complexity Analysis**

The result of Theorem 3.1 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 $M_2 = \gamma^{-1}(1/(n^{2\alpha-\beta}T))$ 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 stochastic approximation method is given by

$$C_{\text{SR-SA}}(\gamma) = C \times (n^\beta \gamma^{-1}(1/n^{2\alpha}) + (n + n^\beta)\gamma^{-1}(1/(n^{2\alpha-\beta}T))) \tag{3.21}$$

under the constraint: $\alpha > \beta \vee 1/2$. 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 $\beta^* = 1/2$ is the optimal choice leading to a complexity
$$C_{\text{SR-SA}}(\gamma) = C' n^{2\alpha+1/2},$$
under the constraint $\underline{\lambda}\gamma_0 > 1$ and and $\alpha > 1/2$. 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 $\gamma(p) = \gamma_0/p^\rho$, $\frac{1}{2} < \rho < 1$ then the computational cost is given by
$$C_{\text{SR-SA}}(\gamma) = C'(n^{\frac{2\alpha}{\rho}+\beta} + n^{\frac{2\alpha}{\rho}-\frac{\beta}{\rho}+1})$$
which is minimized for $\beta^* = \rho/(1+\rho)$ leading to an optimal complexity
$$C_{\text{SR-SA}}(\gamma) = C' n^{\frac{2\alpha}{\rho}+\frac{\rho}{1+\rho}}.$$
under the constraint $\alpha > 1/2$. Observe that this complexity decreases with respect to ρ and that it is minimal for $\rho \to 1$ 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 condition is difficult to handle and to check in practical implementation so that a blind choice has often to be made.

The CLT proved in Theorem 3.2 shows that for a total error of order $1/n^\alpha$, it is necessary to set $M_1 = n^{2\alpha}$ steps of the stochastic approximation scheme defined by (1.4) with time step T/n^β, $M_2 = n^{2\alpha-\beta}T$ steps of stochastic approximation scheme defined by (1.4) with time step T/n and T/n^β and to simultaneously compute its empirical mean, which represents a negligible part of the total cost. Both stochastic approximation algorithm are devised with a step γ satisfying **(HS1)** with $\rho \in (1/2, 1)$ and $\rho > \frac{\alpha}{2\alpha-\beta} \vee \frac{\alpha(1-\beta)}{\alpha-\beta/2}$. It is plain to see that $\beta^* = 1/2$ is the optimal choice leading to a complexity given by
$$C_{\text{SR-RP}}(\gamma) = C \times n^{2\alpha+1/2},$$
provided that $\rho > \frac{\alpha}{2\alpha-1/2}$ and $\forall \theta \in \mathbb{R}^d$, $\quad n^{\alpha-(\alpha-\frac{1}{4})\rho}\|Dh(\theta) - Dh^{n^{1/2}}(\theta)\| \to 0$ as $n \to +\infty$ (note that when $\rho \to 1$ this condition is the same as in Theorem 3.1). For instance, if $\alpha = 1$, then this condition writes $\rho > 2/3$ and $n^{1-\frac{3}{4}\rho}\|Dh(\theta) - Dh^{n^{1/2}}(\theta)\| \to 0$ and ρ 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 ought we do not intend to develop this point, it is possible to prove that averaging allows to achieve the optimal asymptotic covariance matrix as for standard stochastic approximation algorithms.

Finally, the CLT proved in Theorem 3.3 shows that if the weak discretization error is of order $1/n$, that is $\alpha = 1$, then for a total error of order $1/n$, if we set $M_0 = \gamma^{-1}(1/n^2)$ and $M_l = \gamma^{-1}(m^\ell \log(m)/(n^2\log(n)(m-1)T))$, $\ell = 1, \cdots, L$, the complexity of the multi-level stochastic approximation method is given by
$$C_{\text{ML-SA}}(\gamma) = C \times \left(\gamma^{-1}(1/n^2) + \sum_{\ell=1}^{L} M_\ell(m^\ell + m^{\ell-1})\right). \tag{3.22}$$

As for the statistical Romberg stochastic approximation method, we distinguish the two following cases:

- If $\gamma(p) = \gamma_0/p$ then the optimal complexity is given by
$$C_{\text{ML-SA}}(\gamma) = C\left(n^2 + n^2(\log n)^2\frac{m^2-1}{m(\log m)^2}\right) = \mathcal{O}(n^2(\log(n))^2),$$
under the constraint $\underline{\lambda}\gamma_0 > 1$. This computational cost is similar to the one achieved by the multi-level Monte Carlo method for the computation of $\mathbb{E}_x[f(X_T)]$, see [Gil08b] and [AK12]. As discussed in [Gil08b], this complexity attains a minimum near $m = 7$.

- If we choose $\gamma(p) = \gamma_0/p^\rho$, $\frac{1}{2} < \rho < 1$ then simple computations show that the computational cost is given by

$$C_{\text{ML-SA}}(\gamma) = C \times \left(n^{\frac{2}{\rho}} + n^{\frac{2}{\rho}} (\log n)^{\frac{1}{\rho}} \frac{(m-1)^{\frac{1}{\rho}}(m+1)}{m(\log m)^{\frac{1}{\rho}}} \sum_{\ell=1}^{L} m^{-\ell(\frac{1}{\rho}-1)} \right) = \mathcal{O}(n^{\frac{2}{\rho}} (\log n)^{\frac{1}{\rho}}).$$

Observe that once again this complexity decreases with respect to ρ and that it is minimal for $\rho \to 1$ 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 ρ.

4. Numerical Results

In this section we illustrate the results obtained in sections 2.3 and 2.4 on one hand and those obtained in section 3.

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

$$q_l(X_T) := \inf \left\{ \theta : \mathbb{P}(X_T \leq \theta) \geq l \right\}.$$

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

$$X_t = x_0 + \int_0^t r X_s ds + \int_0^t \sigma X_s dW_s, \quad t \in [0,T] \tag{4.23}$$

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

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

where ϕ is the distribution function of the standard normal distribution $\mathcal{N}(0,1)$. We associate to the SDE (4.23) its Euler like scheme $X^n = (X^n_t)_{t\in[0,T]}$ with time step $\Delta = T/n$. We use the following values for the parameters: $x_0 = 100$, $r = 0.05$, $\sigma = 0.4$, $T = 1$, $l = 0.7$. 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 stochastic approximation 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 e.g. [BFP09a] and [BFP09b].*

In order to illustrate the result of Theorem 2.7, we plot in Figure 1 the behaviors of $nh^n(\theta^*)$ and $n(\theta^{*,n} - \theta^*)$ for $n = 100, \cdots, 500$. Actually, $h^n(\theta^*)$ 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 2102.4 for $n = 100$ to 53012.5 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 Fig 2 illustrates Theorem 2.8. 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.

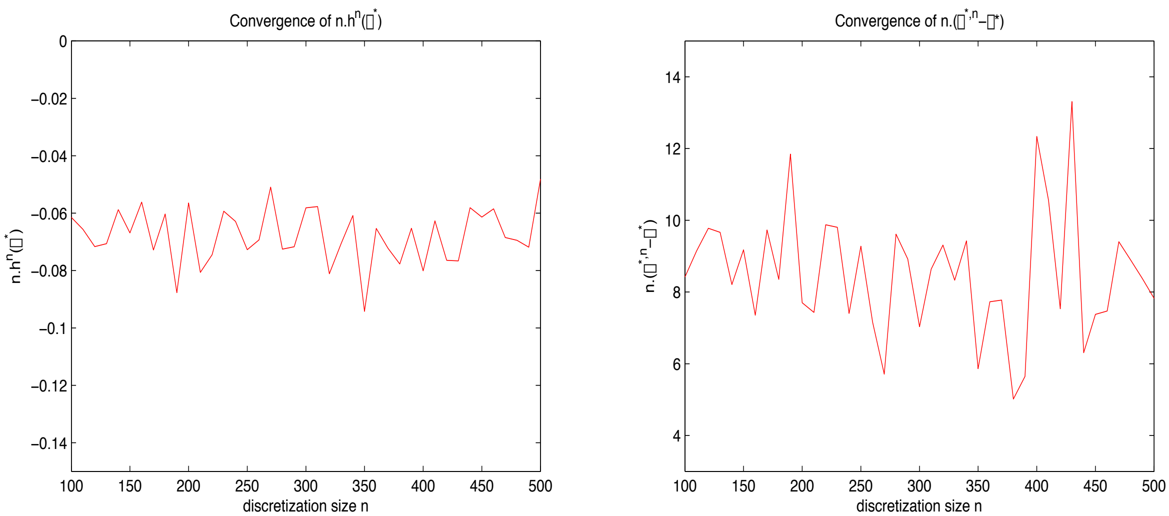

FIGURE 1. On the left: Weak discretization error $n \mapsto nh^n(\theta^*)$. On the right: Implicit discretization error $n \mapsto n(\theta^{*,n} - \theta^*)$, $n = 100, \cdots, 500$.

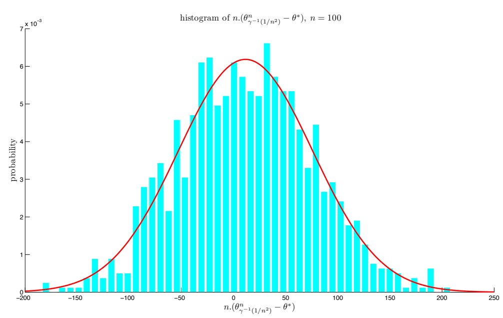

FIGURE 2. Histogram of $n(\theta^n_{\gamma^{-1}(1/n^2)} - \theta^*)$, $n = 100$, with $N = 1000$ samples.

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 (4.23) is given by

$$e^{-rT}\mathbb{E}(X_T - \theta)_+ = e^{-rT}x_0\phi(d_+(x_0, \theta, \sigma)) - e^{-rT}\theta\phi(d_-(x_0, \theta, \sigma)), \tag{4.24}$$

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 (4.24). The values of the parameters x_0, r, σ, T remain unchanged. We plot in Figure 3 the behaviors of $nh^n(\theta^*)$ and $n(\theta^{*,n} - \theta^*)$ for $n = 100, \cdots, 500$. 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 to approximate the solution to $h(\theta) = \mathbb{E}_{x_0}[H(\theta, X_T)] = 0$ with $H(\theta, x) = l - (x - \theta)_+$ in terms of computational costs, we compute the different estimators, namely $\theta^n_{\gamma^{-1}(1/n^2)}$ where

$(\theta_p^n)_{p\geq 1}$ is given by (1.4), Θ_n^{sr} and Θ_n^{ml} 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 compute the complexity given by (2.14), (3.21) and (3.22) respectively and the root-mean-squared error which is given by

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

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 $\beta^* = 1/2$.

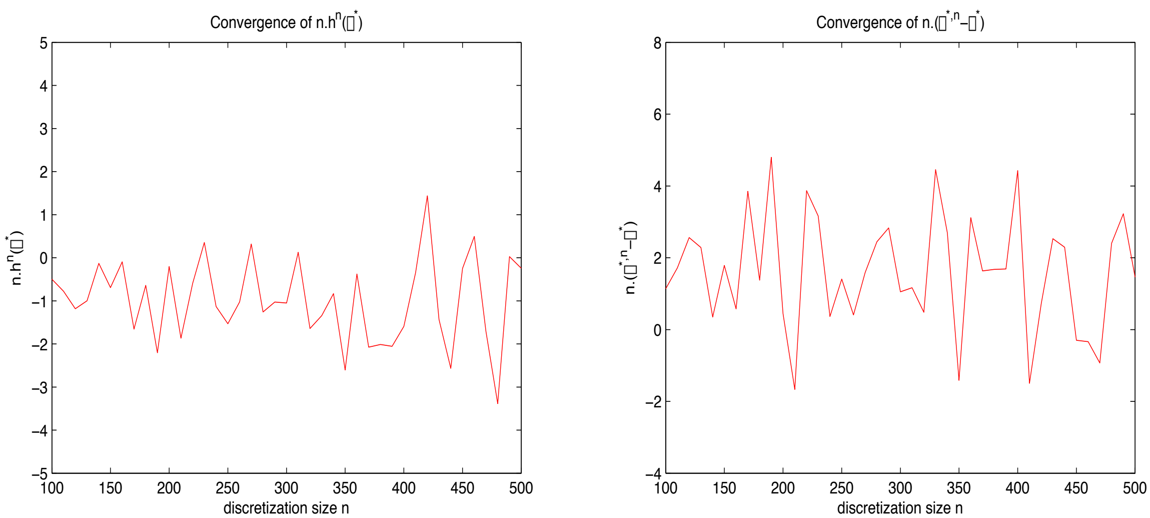

FIGURE 3. On the left: Weak discretization error $n \mapsto nh^n(\theta^*)$. On the right: Implicit discretization error $n \mapsto n(\theta^{*,n} - \theta^*)$, $n = 100, \cdots, 500$.

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 $\theta^n_{\gamma_0 n^{3/2}T} - \theta^{\sqrt{n}}_{\gamma_0 n^{3/2}T}$ (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 $\theta_0^{n^{1/2}} = \theta_0^1 = x_0$ for all $k \in \{1, \cdots, M\}$ to initialize the procedures. Moreover, by Lemma 5.2, the L^1-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, to ensure that the different procedures do not jump too far ahead in one step, we freeze the value of $\theta^{\sqrt{n}}_{p+1}$ (respectively $\theta^{m^\ell}_{p+1}$) and reset it to the value of the previous step as soon as $|\theta^{\sqrt{n}}_{p+1} - \theta^{\sqrt{n}}_p| \leq K/\sqrt{p}$ (respectively $|\theta^{m^\ell}_{p+1} - \theta^{m^\ell}_p| \leq K/\sqrt{p}$), for a pre-specified value of K. It notably prevents the algorithm from blowing up during the first iterates. We select $K = 5$ in the different procedures. Note anyway that this projection-reinitialization step slightly increases the complexity of each procedures. In our numerical examples, we observe that it only represents around 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 are much lower than the SA procedure one. 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, *i.e.* 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 statical 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 algorithm or the multi-level SA one. Finally, for a RMSE fixed at $5.5.10^{-2}$, the complexity gain achieved by using the multi-level SA procedure instead of the statistical romberg one is approximately equal to 5.

The histograms of Fig 4 illustrates Theorems 2.8, 3.1 and 3.3. The distributions of $n(\theta^n_{\gamma^{-1}(1/n^2)} - \theta^*)$, $n(\Theta^{sr}_n - \theta^*)$ and $n(\Theta^{ml}_n - \theta^*)$, obtained with $n = 4^4 = 256$ and $N = 1000$ samples, are close to a normal distribution.

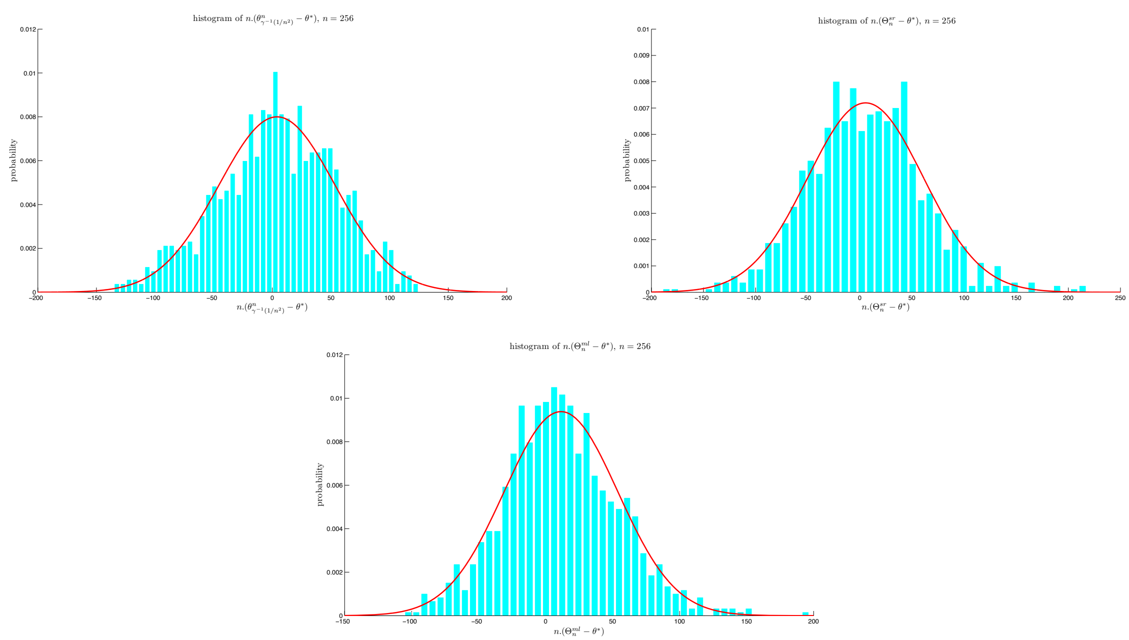

FIGURE 4. Histograms of $n(\theta_{\gamma^{-1}(1/n^2)} - \theta^*)$, $n(\Theta^{sr}_n - \theta^*)$ and $n(\Theta^{ml}_n - \theta^*)$ (from left to right), $n = 256$, with $N = 1000$ samples.

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 λ_{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 $(-\rho)$, $\rho \in [0, 1)$, that is for any $x > 0$, $\lim_{t\to+\infty} \gamma(tx)/\gamma(t) = x^{-\rho}$.*

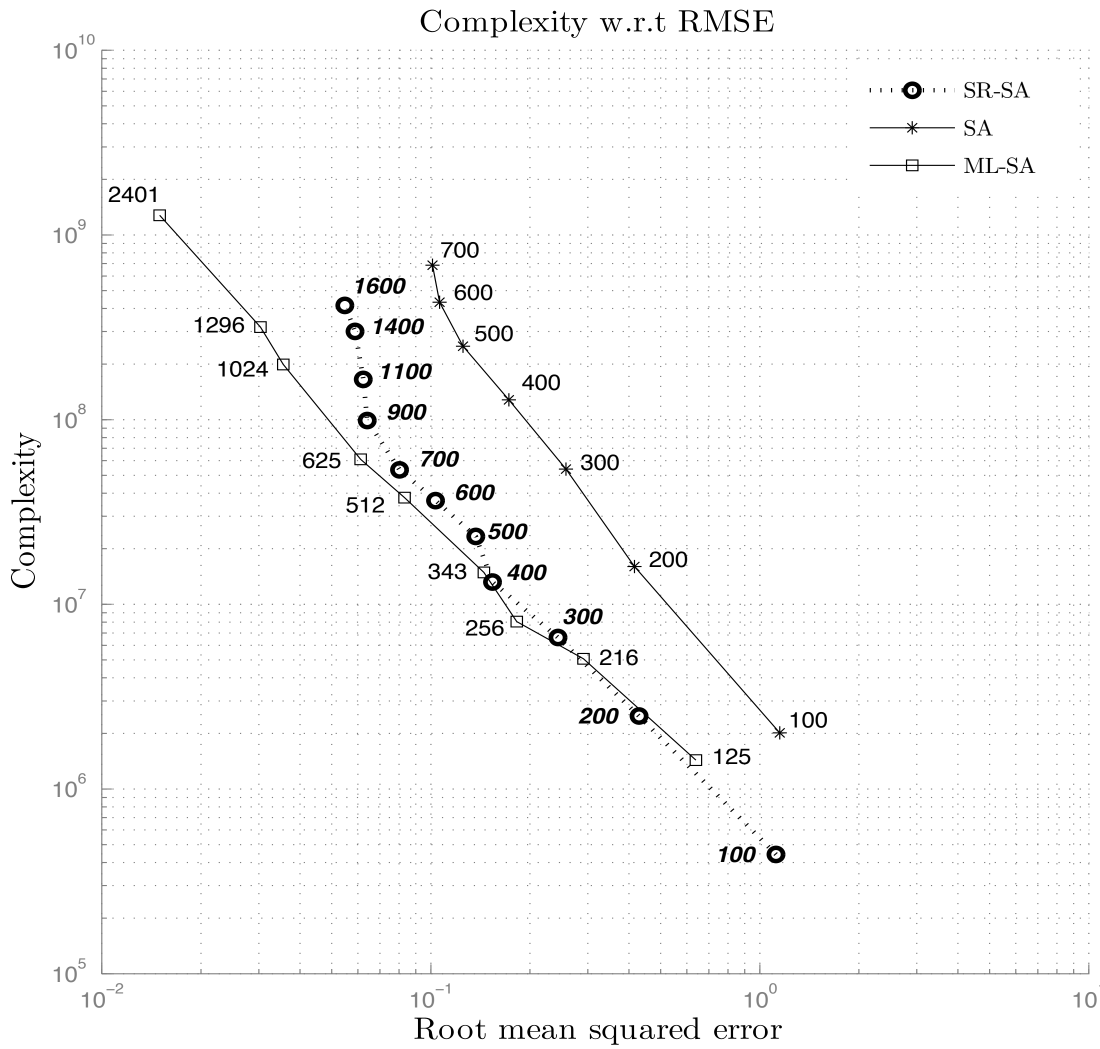

FIGURE 5. Complexity with respect to RMSE.

- *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 non-negative 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 $k \leq n$, $\|\Pi_{k+1,n}\| \leq C \prod_{j=k}^{n}(1-\lambda\gamma_j)$. Hence, we have $\sum_{k=1}^{n} \gamma_k^{1+a} v_k \|\Pi_{k+1,n}\|^b \leq C \sum_{k=1}^{n} \gamma_k^{1+a} v_k e^{-\lambda b(s_n - s_k)}$,

$n \geq 1$, 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 written in the recursive form

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

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

$$\begin{aligned} z_{n+1} &\leq z_N \exp(-\lambda b(s_n - s_{N-1})) + \exp(-\lambda b s_n) \sum_{k=N}^{n} \exp(\lambda b s_k) \gamma_k^{a+1} v_{k+1} \\ &\leq z_N \exp(-\lambda b(s_n - s_{N-1})) + \left(\sup_{k>N} v_k\right) \exp(-\lambda b s_n) \sum_{k=N}^{n} \exp(\lambda b s_k) \gamma_k^{a+1}. \end{aligned}$$

We study now the impact of the step sequence $(\gamma_p)_{p\geq 1}$ 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

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

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

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

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

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

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

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

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

Hence it follows that

$$\frac{\exp(-\lambda b s(n))}{\gamma^a(n)} \int_0^{s(n)} \exp(\lambda b t) \gamma^{a+1}(t) dt \leq \frac{\gamma(s^{-1}(0))}{\lambda \gamma^a(n)} \exp(-\lambda b(1-x)s(n)) + \frac{\gamma^a(s^{-1}(xs(n)))}{\lambda b \gamma^a(n)},$$

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

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

An argument similar to the previous case concludes the proof.

□

Lemma 5.2. *Let $(\theta^n_p)_{p\geq 0}$ be the procedure defined by (1.4) where θ^n_0 is independent of the innovation of the algorithm with $\sup_{n\geq 1}\mathbb{E}|\theta^n_0|^2 < +\infty$. Suppose that the mean-field function h^n satisfies*

$$\exists \underline{\lambda} > 0,\ \forall n \in \mathbb{N}^*,\ \forall \theta \in \mathbb{R}^d,\ \langle \theta - \theta^{*,n}, h^n(\theta)\rangle \geq \underline{\lambda}|\theta - \theta^{*,n}|^2, \tag{5.25}$$

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

Then, one has:

$$\forall p \geq 1, \quad \sup_{n\geq 1}\mathbb{E}[|\theta^n_p - \theta^{*,n}|^2] \leq C\gamma(p)$$

for some positive constant C independent of p and n.

Proof. From the dynamic of $(\theta^n_p)_{p\geq 1}$, we have

$$\begin{aligned}|\theta^n_{p+1} - \theta^{*,n}|^2 = |\theta^n_p - \theta^{*,n}|^2 - 2\gamma_{p+1}\langle \theta^n_p - \theta^{*,n}, h^n(\theta^n_p)\rangle + 2\gamma_{p+1}\langle \theta^n_p - \theta^{*,n}, \Delta M^n_{p+1}\rangle \\ + \gamma^2_{p+1}|H(\theta^n_p, (X^n_T)^{p+1})|^2,\end{aligned}$$

so that taking expectation in the previous equality and using assumptions (2.10) and (5.25), we easily derive

$$\mathbb{E}|\theta^n_{p+1} - \theta^{*,n}|^2 \leq (1 - 2\underline{\lambda}\gamma_{p+1} + C\gamma^2_{p+1})\mathbb{E}|\theta^n_p - \theta^{*,n}|^2 + C\gamma^2_{p+1}.$$

Now a simple induction argument yields

$$\mathbb{E}|\theta^n_p - \theta^{*,n}|^2 \leq \mathbb{E}|\theta^n_0 - \theta^{*,n}|^2\Pi_{1,p} + \sum_{k=1}^{p}\Pi_{k+1,p}\gamma^2_k$$

where we set $\Pi_{k,p} := \prod_{j=k}^{p}(1 - 2\underline{\lambda}\gamma_j + C\gamma^2_j)$ for sake of simplicity. Computations similar to the proof of Lemma 5.1 imply

$$\forall p \geq 1, \quad \mathbb{E}|\theta^n_p - \theta^{*,n}|^2 \leq C\gamma(p).$$

□

Proposition 5.1. *Assume that the assumptions of Theorem 2.4 are satisfied and that θ_0 is independent of the innovation of the algorithm with $\mathbb{E}|\theta_0|^2 < +\infty$.Then, there exists two sequences $(\mu_p)_{p\geq 0}$ and $(r_p)_{p\geq 0}$, with $\mu_0 = \theta_0 - \theta^*$ and $r_0 = 0$ such that*

$$\forall p \geq 0, \quad \theta_p - \theta^* = \mu_p + r_p$$

and satisfying

$$\sup_{p\geq 1}\gamma_p^{-1}\mathbb{E}|\mu_p|^2 < +\infty \quad \text{and} \quad \gamma_p^{-b}|r_p| \leq X_1 Y_p$$

with $b = 1$ under **(HS1)** *or $b \in (1/2, 1 \wedge \lambda_m\gamma_0)$ under* **(HS2)** *and where X_1 is a finite random variable and $(Y_p)_{p\geq 1}$ is a sequence of random variables bounded in $L^1(\mathbb{P})$, that is $\sup_{p\geq 1}\mathbb{E}|Y_p| < +\infty$.*

Proof. We first write

$$\theta_{p+1} - \theta^* = (I_d - \gamma_{p+1}Dh(\theta^*))(\theta_p - \theta^*) + \gamma_{p+1}\Delta M_{p+1} - \gamma_{p+1}(h(\theta_p) - Dh(\theta^*)(\theta_p - \theta^*)).$$

We define the two sequences $(\mu_p)_{p\geq 0}$ and $(r_p)_{p\geq 0}$ by

$$\mu_{p+1} = (I_d - \gamma_{p+1}Dh(\theta^*))\mu_p + \gamma_{p+1}\Delta M_{p+1} \tag{5.26}$$

with $\mu_0 = \theta_0 - \theta^*$ and $r_p = \theta_p - \theta^* - \mu_p$, $p \geq 1$, with $r_0 = 0$. By iterating (5.26), we clearly get

$$\mu_n = \Pi_{1,n}\mu_0 + \sum_{k=1}^{n} \gamma_k \Pi_{k,n} \Delta M_k$$

Since θ_0 is independent of the innovation of the algorithm with $\mathbb{E}|\theta_0|^2 < +\infty$, for all $\epsilon \in (0, \lambda_m)$, we have

$$\begin{aligned}\mathbb{E}|\mu_n|^2 &\leq C\|\Pi_{1,n}\|^2\mathbb{E}|\theta_0|^2 + C\sum_{k=1}^{n}\gamma_k^2\|\Pi_{k,n}\|^2\mathbb{E}|\Delta M_k|^2 \\ &\leq C\exp(-2(\lambda_m - \epsilon)\sum_{k=1}^{n}\gamma_k)\mathbb{E}|\theta_0|^2 + C\sup_{k\geq 1}\mathbb{E}|\theta_k - \theta^*|^2\sum_{k=1}^{n}\gamma_k^2\|\Pi_{k,n}\|^2\end{aligned}$$

where we used (2.10) to derive $\mathbb{E}|\Delta M_k|^2 \leq C\sup_{k\geq 1}\mathbb{E}|\theta_k - \theta^*|^2 < +\infty$ for the last inequality. Consequently, similar computations as those used in Lemma 5.1 (select ϵ s.t. $(\lambda_m - \epsilon)\gamma_0 > 1/2$ under **(HS1)**) allow to derive

$$\sup_{p\geq 1}\gamma_p^{-1}\mathbb{E}|\mu_p|^2 < +\infty.$$

We now prove that $\mu_n \xrightarrow{a.s.} 0$. The convergence to zero of $(\sum_{k=1}^{n}\gamma_k\Pi_{k,n}\Delta M_k)_{n\geq 1}$ will follow from the convergence of the series

$$N_n = \sum_{k=1}^{n}\gamma_k\Delta M_k,\ n \geq 1.$$

The sequence $(N_n)_{n\geq 1}$ is a martingale. By (2.10) and the $a.s.$ convergence of $(\theta_k)_{k\geq 1}$, one has

$$\langle N\rangle_\infty = \sum_{k\geq 1}\gamma_k^2\mathbb{E}[(\Delta M_k)^2 \big| \mathcal{F}_{k-1}] < \infty$$

which yields the $a.s.$ convergence of $(N_n)_{n\geq 1}$ towards a finite r.v. N_∞. Then, using an abel's transform, we get

$$\begin{aligned}\sum_{k=1}^{n}\gamma_k\Pi_{k,n}\Delta M_k &= \Pi_{n,n}N_n - \sum_{k=1}^{n-1}(\Pi_{k+1,n} - \Pi_{k,n})N_k \\ &= \Pi_{n,n}(N_n - N_\infty) + \Pi_{1,n}N_\infty - \sum_{k=1}^{n}\gamma_k\Pi_{k+1,n}Dh(\theta^*)(N_k - N_\infty).\end{aligned}$$

The $a.s.$ convergence of (N_n) towards N_∞ yields the $a.s.$ convergence to zero of the first term. Since $\|\Pi_{1,n}\| \to 0$, the second term $a.s.$ converges to zero. The $a.s.$ convergence to zero of the last term follows from

$$|\sum_{k=1}^{n}\gamma_k\Pi_{k+1,n}Dh(\theta^*)(N_k - N_\infty)| \leq \sum_{k=1}^{n}\gamma_k\|\Pi_{k+1,n}\|\|Dh(\theta^*)\||N_k - N_\infty|$$

and Lemma 5.1.

Now, we focus on the estimates concerning $(r_p)_{p\geq 1}$. Since h is twice differentiable in a neighborhood of θ^*, the line i of the column vector $h(\theta_p) - Dh(\theta^*)(\theta_p - \theta^*)$ is equal to $(\theta_p - \theta^*)^T H_i^p(\theta_p - \theta^*)$ with $(H_i^p)_{k,l} = \int_0^1 \frac{1}{2}(1-t)^2\frac{\partial^2 h_i}{\partial\theta_k\partial\theta_l}(t\theta^* + (1-t)\theta_p)dt$, $(k,l) \in [\![1,d]\!]^2$. Hence, we define H^p such that $h(\theta_p) - Dh(\theta^*)(\theta_p - \theta^*) = (\theta_p - \theta^*)^T H^p(\theta_p - \theta^*)$ and the line i of the column vector $(\theta_p - \theta^*)^T H^p(\theta_p - \theta^*)$ is $(\theta_p - \theta^*)^T H_i^p(\theta_p - \theta^*)$. With these notations, we have

$$\begin{aligned} r_{p+1} &= (I_d - \gamma_{p+1} Dh(\theta^*))r_p + \gamma_{p+1}(\theta_p - \theta^*)^T H^p(\theta_p - \theta^*) \\ &= (I_d - \gamma_{p+1} Dh(\theta^*) + 2\gamma_{p+1}\mu_p^T H^p + \gamma_{p+1} r_p^T H^p) r_p + \gamma_{p+1}\mu_p^T H^p \mu_p. \end{aligned} \tag{5.27}$$

By iterating the above equality we obtain

$$r_p = \sum_{k=1}^{p} \gamma_k A_{k+1,p} \mu_{k-1}^T H^{k-1} \mu_{k-1}$$

with $A_{k,p} := \prod_{j=k}^{p}(I_d - \gamma_j(Dh(\theta^*) - 2\mu_{j-1}^T H^{j-1} - r_{j-1}^T H^{j-1}))$. Let us note that since $\theta_p \xrightarrow{a.s.} \theta^*$ and $\mu_p \xrightarrow{a.s.} 0$, we have $r_p \xrightarrow{a.s.} 0$ so that the sequence $(Dh(\theta^*) - 2\mu_{p-1}^T H^{p-1} - r_{p-1}^T H^{p-1})_{p\geq 1}$ of random matrices converges *a.s.* to the stable matrix $Dh(\theta^*)$. Hence (see e.g. [Duf96]) for all $\delta \in (0, \lambda_m)$ there exists a finite random variable X such that for all $k \in [\![1, p]\!]$,

$$\|A_{k,p}\| \leq X \exp(-(\lambda_m - \delta) \sum_{j=k}^{p} \gamma_j)$$

so that we derive

$$|r_p| \leq X \sup_{k\geq 1} \|H^k\| \sum_{k=1}^{p} \gamma_k \exp(-(\lambda_m - \delta) \sum_{j=k}^{p} \gamma_j) |\mu_{k-1}|^2.$$

where we used the fact that $\sup_{k\geq 1} \|H^k\| < +\infty$ since h is twice continuously differentiable. Hence, there exists a finite random variable that we still denote X such that

$$\gamma_p^{-b} |r_p| \leq X \gamma_p^{-b} \sum_{k=1}^{p} \gamma_k^2 \exp(-(\lambda_m - \delta) \sum_{j=k}^{p} \gamma_j) \gamma_{k-1}^{-1} |\mu_{k-1}|^2.$$

We select δ such that $(\lambda_m - \delta)\gamma_0 > b$ and an analysis along the lines of the proof of Lemma 5.1 shows that the sequence of random variables appearing in the second term in the right hand side of the above inequality is bounded in $L^1(\mathbb{P})$.

□

Proposition 5.2. *Assume that the assumptions of Theorem 3.2 are satisfied. Then, for all $n \in \mathbb{N}$ there exists two sequences $(\tilde{\mu}_p^n)_{p\in[\![0,n]\!]}$ and $(\tilde{r}_p^n)_{p\in[\![0,n]\!]}$ with $\tilde{\mu}_0^n = \theta^* - \theta^{*,n}$ such that*

$$\forall p \in [\![0, n]\!], \quad z_p^n = \theta_p^n - \theta^{*,n} - (\theta_p - \theta^*) = \tilde{\mu}_p^n + \tilde{r}_{1,p}^n + \tilde{r}_{2,p}$$

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

$$\sup_{p\geq 1} \gamma_p^{-1/2} \mathbb{E}|\tilde{\mu}_p^n| < Cn^{-1/2}, \quad \gamma_p^{-1} |\tilde{r}_{1,p}^n| \leq \tilde{X} \tilde{Y}_p^n$$

and

$$\tilde{r}_{2,p} = \sum_{k=1}^{p} \gamma_k \Pi_{k+1,p}(H(\theta_k, (X_T)^{k+1}) - H(\theta^*, (X_T)^{k+1}) - (h(\theta_k) - h(\theta^*)))$$

for some positive constant C independent of p and n and where $\tilde{X}$ is a finite random variable (being independent of n) and $(\tilde{Y}_p^n)_{p\geq 1}$ is a sequence of random variables bounded in $L^1(\mathbb{P})$, that is $\sup_{n\geq 1, p\geq 1} \mathbb{E}|\tilde{Y}_p^n| < +\infty$.

Proof. Using (3.16), we define the two sequences $(\tilde{\mu}_p^n)_{p\in[\![0,n]\!]}$ and $(\tilde{r}_p^n)_{p\in[\![0,n]\!]}$ by

$$\tilde{\mu}_p^n = \Pi_{1,p} z_0^n + \sum_{k=1}^{p} \gamma_k \Pi_{k+1,p} \Delta N_k^n + \sum_{k=1}^{p} \gamma_k \Pi_{k+1,p} (Dh(\theta^*) - Dh^n(\theta^{*,n}))(\theta_{k-1}^n - \theta^{*,n})$$
$$+ \sum_{k=1}^{p} \gamma_k \Pi_{k+1,n} (h^n(\theta^{*,n}) - h^n(\theta^*) - (H(\theta^{*,n}, (X_T^n)^{k+1}) - H(\theta^*, (X_T^n)^{k+1})))$$

and

$$\tilde{r}_{1,p}^n = \sum_{k=1}^{p} \gamma_k \Pi_{k+1,p} (\zeta_{k-1}^n - \zeta_{k-1}) + \sum_{k=1}^{p} \gamma_k \Pi_{k+1,p} (h^n(\theta_k^n) - h^n(\theta^{*,n}) - (H(\theta_k^n, (X_T^n)^{k+1}) - H(\theta^{*,n}, (X_T^n)^{k+1}))).$$

We first focus on the sequence $(\tilde{\mu}_p^n)_{p\in[\![0,n]\!]}$. Under the assumption on the step sequence we have

$$|\Pi_{1,p} z_0^n| \leq \|\Pi_{1,p}\| |\theta^* - \theta^{*,n}| = \mathcal{O}(\gamma_p^{1/2} n^{-1/2}).$$

Moreover, by the definition of the sequence $(\Delta N_k^n)_{k\in[\![1,n]\!]}$ and the Cauchy-Schwarz inequality we derive

$$\mathbb{E}\left|\sum_{k=1}^{p} \gamma_k \Pi_{k+1,p} \Delta N_k^n\right| \leq C(\mathbb{E}|H(\theta^*, X_T^n) - H(\theta^*, X_T)|^2)^{1/2} (\sum_{k=1}^{p} \gamma_k^2 \|\Pi_{k+1,p}\|^2)^{1/2} = \mathcal{O}(\gamma_p^{1/2} n^{-1/2}).$$

Taking the expectation for the third term and following the lines of the proof of Lemma 3.3, we obtain

$$\mathbb{E}\left|\sum_{k=1}^{p} \gamma_k \Pi_{k+1,p} (Dh(\theta^*) - Dh^n(\theta^{*,n}))(\theta_{k-1}^n - \theta^{*,n})\right| \leq C \sum_{k=1}^{p} \gamma_k^{3/2} \|\Pi_{k+1,p}\| (|\theta^{*,n} - \theta^*| + \|Dh(\theta^*) - Dh^n(\theta^*)\|)$$
$$= \mathcal{O}(\gamma_p^{1/2} n^{-1/2}).$$

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

$$\mathbb{E}\left|\sum_{k=1}^{p} \gamma_k \Pi_{k+1,p} (h^n(\theta^{*,n}) - h^n(\theta^*) - (H(\theta^{*,n}, (X_T^n)^{k+1}) - H(\theta^*, (X_T^n)^{k+1})))\right|^2 \leq |\theta^* - \theta^{*,n}|^2 \sum_{k=1}^{p} \gamma_k^2 \|\Pi_{k+1,p}\|^2$$
$$= \mathcal{O}(\gamma_p n^{-1}).$$

We now prove the bound concerning the sequence $(\tilde{r}_{1,p}^n)_{p\in[\![0,n]\!]}$. Observe first that the inequality $|\theta_{k-1} - \theta^*|^2 \leq |\mu_{k-1}|^2 + 2|\theta_{k-1} - \theta^*||r_{k-1}|$ combined with Proposition 5.1 lead to

$$\left|\sum_{k=1}^{p} \gamma_k \Pi_{k+1,p} (\zeta_{k-1}^n - \zeta_{k-1})\right| \leq C \sum_{k=1}^{p} \gamma_k^2 \|\Pi_{k+1,p}\| (\gamma_{k-1}^{-1} |\theta_{k-1}^n - \theta^{*,n}|^2 + \gamma_{k-1}^{-1} |\mu_{k-1}|^2 + |\theta_{k-1} - \theta^*| \gamma_{k-1}^{-1} |r_{k-1}|)$$
$$\leq C(1 + \sup_{k\geq 0} |\theta_k - \theta^*|) \sum_{k=1}^{p} \gamma_k^2 \|\Pi_{k+1,p}\| (\gamma_{k-1}^{-1} |\theta_{k-1}^n - \theta^{*,n}|^2 + \gamma_{k-1}^{-1} |\mu_{k-1}|^2 + Y_{k-1}),$$

so that since $\sup_{n\geq 1} \mathbb{E} \sum_{k=1}^{p} \gamma_k^2 \|\Pi_{k+1,p}\| (\gamma_{k-1}^{-1} |\theta_{k-1}^n - \theta^{*,n}|^2 + \gamma_{k-1}^{-1} |\mu_{k-1}|^2 + Y_{k-1}) = \mathcal{O}(\gamma_p)$ we conclude that the first term appearing in the decomposition of $\tilde{r}_{1,p}^n$ satisfies the desired bound. Concerning the second term, following the lines of the proof of Lemma 3.3 we simply take the square of its L^2-norm to derive

$$\sup_{n\geq 1}\mathbb{E}\left|\sum_{k=1}^{p}\gamma_k\Pi_{k+1,p}(h^n(\theta_k^n)-h^n(\theta^{*,n})-(H(\theta_k^n,(X_T^n)^{k+1})-H(\theta^{*,n},(X_T^n)^{k+1})))\right|^2 \leq C\sum_{k=1}^{p}\gamma_k^3\|\Pi_{k+1,p}\|^2 = \mathcal{O}(\gamma_p^2).$$

□

References

[AE78] D. J. Aldous and G. K. Eagleson. On mixing and stability of limit theorems. *Ann. Probability*, 6(2):325–331, 1978.

[AK12] M. Ben Alaya and A. Kebaier. Central limit theorem for the multilevel monte carlo euler method and applications to asian options. *Preprint*, 2012.

[BFP09a] O. Bardou, N. Frikha, and G. Pagès. Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling. *Monte Carlo Methods Appl.*, 15(3):173–210, 2009.

[BFP09b] O. Bardou, N. Frikha, and G. Pagès. Recursive computation of value-at-risk and conditional value-at-risk using MC and QMC. In *Monte Carlo and quasi-Monte Carlo methods 2008*, pages 193–208. Springer, Berlin, 2009.

[BFP10] O. Bardou, N. Frikha, and G. Pagès. CVaR hedging using quantization based stochastic approximation algorithm. *forthcoming in Mathematical Finance*, 2010.

[BMP90] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990. Translated from the French by Stephen S. Wilson.

[BT96] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function. *Prob. Th. Rel. Fields*, 104-1:43–60, 1996.

[Der11] S. Dereich. Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *Ann. Appl. Probab.*, 21(1):283–311, 2011.

[DG95] D Duffie and P. Glynn. Efficient monte carlo simulation of security prices. *Ann. Appl. Probab.*, 5(4):897–905, 1995.

[Duf96] M. Duflo. *Algorithmes stochastiques*, volume 23 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 1996.

[FF13] M. Fathi and N. Frikha. Transport-entropy inequalities and deviation estimates for stochastic approximation schemes. *Electron. J. Probab.*, 18:no. 67, 1–36, 2013.

[FM12] N. Frikha and S. Menozzi. Concentration bounds for stochastic approximations. *Electron. Commun. Probab.*, 17:no. 47, 15, 2012.

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

[Gil08a] M. B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and quasi-Monte Carlo methods 2006*, pages 343–358. Springer, Berlin, 2008.

[Gil08b] M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

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

[JP98] J. Jacod and P. Protter. Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.*, 26(1):267–307, 1998.

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

[KY03] H. J. Kushner and G. G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.

[Pel98] M. Pelletier. Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann. Appl. Probab.*, 8(1):10–44, 1998.

[PJ92] B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.*, 30(4):838–855, 1992.

[Rén63] A. Rényi. On stable sequences of events. *Sankhyā Ser. A*, 25:293 302, 1963.

[RM51] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statistics*, 22:400–407, 1951.

[Rup91] D. Ruppert. Stochastic approximation. In *Handbook of sequential analysis*, volume 118 of *Statist. Textbooks Monogr.*, pages 503–529. Dekker, New York, 1991.

[TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. and App.*, 8-4:94–120, 1990.