



HAL
open science

Multi-level stochastic approximation algorithms

Noufel Frikha

► **To cite this version:**

Noufel Frikha. Multi-level stochastic approximation algorithms. The Annals of Applied Probability, 2016, 26 (2), pp.933-985. 10.1214/15-AAP1109 . hal-00870585v2

HAL Id: hal-00870585

<https://hal.science/hal-00870585v2>

Submitted on 6 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MULTI-LEVEL STOCHASTIC APPROXIMATION ALGORITHMS

N. FRIKHA¹

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.

August 6, 2014.

1. INTRODUCTION

In this paper we propose and analyze a multi-level paradigm for stochastic optimization problem by means of stochastic approximation schemes. The multi-level Monte Carlo method introduced by Heinrich [Hei01] and popularized in numerical probability by [Keb05] and [Gil08b] allows to significantly increase the computational efficiency of the expectation of an \mathbb{R} -valued non-simulatable random variable Y that can only be strongly approximated by a sequence $(Y^n)_{n \geq 1}$ of easily simulatable random variables (all defined on the same probability space) as the *bias parameter* n goes to infinity with a *weak error* or *bias* $\mathbb{E}[Y] - \mathbb{E}[Y^n]$ of order $n^{-\alpha}$, $\alpha > 0$. Let us be more specific. In this context, the standard Monte Carlo method uses the statistical estimator $M^{-1} \times \sum_{j=1}^M Y^{n,j}$ where the $(Y^{n,j})_{j \in [1, M]}$ are M independent copies of Y^n . Given the order of the weak error, a natural question is to find the optimal choice of the sample size M to achieve a global error. If the weak error is of order $n^{-\alpha}$ then for a total error of order $n^{-\alpha}$ ($\alpha \in [1/2, 1]$), the minimal computation necessary for the standard Monte Carlo algorithm is obtained for $M = n^{2\alpha}$, see [DG95]. So, if the computational cost required to simulate one sample of U^n is of order n then the optimal computational cost of the Monte Carlo method is $C_{MC} = C \times n^{2\alpha+1}$, for a positive constant $C > 0$.

In order to reduce the complexity of the computation, the principle of the multi-level Monte Carlo method introduced by Giles [Gil08b] as a generalization of Kebaier's approach [Keb05] consists in using the telescopic sum

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

Keywords and phrases: Multi-level Monte Carlo methods, stochastic approximation, Ruppert&Polyak averaging principle, 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/>

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

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

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

Based on an analysis of the variance, Giles [Gil08b] proposed an optimal choice for the sequence $(N_\ell)_{1 \leq \ell \leq L}$ which minimizes the total complexity of the algorithm. More recently, Ben Alaya and Kebaier [AK12] proposed a different analysis to obtain the optimal choice of the parameters relying on a Lindeberg Feller central limit theorem (CLT) for the multi-level Monte Carlo algorithm. To obtain a global error of order $n^{-\alpha}$, both approaches allow to achieve a complexity of order $n^{2\alpha}(\log n)^2$ if the $L^2(\mathbb{P})$ strong approximation rate $\mathbb{E}[|U^n - U|^2]$ of U by U^n is of order $1/n$. Hence, the multi-level Monte Carlo method is significantly more effective than the crude Monte Carlo and the statistical Romberg methods. Originally introduced for the computation of expectations involving stochastic differential equation (SDE), it has been widely applied to various problems of numerical probability, see Giles [Gil08a], Dereich [Der11], Giles, Higham and Mao [GHM09] among others. We refer the interested reader to the webpage: http://people.maths.ox.ac.uk/gilesm/mlmc_community.html.

In the present paper, we are interested in broadening the scope of the multi-level Monte Carlo method to the framework of stochastic approximation (SA) algorithm. Introduced by Robbins and Monro [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 $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ 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 $H : \mathbb{R}^d \times \mathbb{R}^q \rightarrow \mathbb{R}^d$ and U is an \mathbb{R}^q -valued random vector. To estimate the zero of h , they proposed the algorithm

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

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

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

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

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

Some typical applications are the computations of the implied volatility or the implied correlation which both boil down to finding the zero of a function which writes as an expectation. Computing the Value-at-Risk and the Conditional Value-at-Risk of a financial portfolio when the dynamics of the underlying assets are given by an SDE also appears as an inverse problem for which a SA scheme may be devised, see e.g. [BFP09a, BFP09b]. The risk minimization of a financial portfolio by means of SA has been investigated in [BFP10, Fri14]. For more applications and a complete overview in the theory of stochastic approximation, the reader may refer to [Duf96], [KY03] and [BMP90].

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

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

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

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

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

$$\theta_{p+1}^n = \theta_p^n - \gamma_{p+1} H(\theta_p^n, (U^n)^{p+1}), \quad p \in \llbracket 0, M-1 \rrbracket, \quad (1.3)$$

where $((U^n)^p)_{p \in \llbracket 1, M \rrbracket}$ is an i.i.d. sequence of random variables with the same law as U^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 (1.2). This induces a *statistical error* which writes

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

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

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

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

Regarding the *statistical error* $\mathcal{E}_S(n, M, \gamma) := \theta^{*,n} - \theta_M^n$, it is well-known that under standard assumptions, i.e. a mean reverting assumption on h^n and a growth control of the $L^2(\mathbb{P})$ -norm of the noise of the algorithm, the Robbins-Monro theorem guarantees that $\lim_M \mathcal{E}_S(n, M, \gamma) = 0$ for each fixed $n \in \mathbb{N}^*$, see Theorem 2.3 below. Moreover, under mild technical conditions, a CLT holds at rate $\gamma^{-1/2}(M)$, that is, for each fixed $n \in \mathbb{N}^*$, $\gamma^{-1/2}(M)\mathcal{E}_S(n, M, \gamma)$ converges in distribution to a normally distributed random variable with mean zero and finite covariance matrix, see Theorem 2.4 below. The reader may also refer to [FM12, 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 $2\mathcal{R}e(\lambda_{min})\gamma_0 > 1$ 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 SA algorithm by averaging according to the *Ruppert & Polyak averaging principle*, see e.g. [Rup91, PJ92]. It consists in devising the original SA algorithm (1.3) with a slow decreasing step and to simultaneously compute the empirical mean $(\bar{\theta}_p^n)_{p \geq 1}$ (which *a.s.* converges to $\theta^{*,n}$) of the sequence $(\theta_p^n)_{p \geq 0}$ by setting

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

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

Given the order of the implicit weak error and a step sequence γ satisfying (1.2) a natural question is to find the optimal balance between the value of n and the number M of steps in (1.3) in order to achieve a given global error. This problem was originally investigated in [DG95] for the standard Monte Carlo method. The error between θ^* and the approximation θ_M^n writes $\theta_M^n - \theta^* = \theta_M^n - \theta^{*,n} + \theta^{*,n} - \theta^*$ suggesting to select $M = \gamma^{-1}(1/n^{2\alpha})$, where γ^{-1} is the inverse function of γ , when the *weak error* is of order $n^{-\alpha}$. However, due to the non-linearity of the SA algorithm (1.3), the methodology developed in [DG95] does not apply in our context. The key tool to tackle this question consists in linearizing the dynamic of $(\theta_p^n)_{p \in [1, M]}$ around its target $\theta^{*,n}$, quantifying the contribution of the non linearities in the space variable θ_p^n and the innovations and finally exploiting stability arguments from SA schemes. Optimizing with respect to the usual choice of the step sequence, the minimal computational cost (to achieve an error of order $n^{-\alpha}$) given by $C_{SA} = K \times n \times \gamma^{-1}(1/n^{2\alpha})$ 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 $n^{2\alpha+1}$. Considering the empirical mean sequence $(\bar{\theta}_p^n)_{p \in [1, n^{2\alpha}]}$ instead of the crude SA estimate also allows to reach the optimal complexity for free without any condition on γ_0 .

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

Moreover, we generalize this approach to the case of multi-level SA method. In the spirit of [Gil08b] for Monte Carlo path simulation, the multi-level SA scheme estimates $\theta^{*,n}$ by computing the quantity $\Theta_n^{ml} = \theta_{M_0}^1 + \sum_{\ell=1}^L \theta_{M_\ell}^{m_\ell} - \theta_{M_\ell}^{m_\ell-1}$ where for every ℓ , the couple $(\theta_{M_\ell}^{m_\ell}, \theta_{M_\ell}^{m_\ell-1})$ is obtained using M_ℓ independent copies of $(U^{m_\ell-1}, U^{m_\ell})$. Here again to establish a CLT for this estimator (in the spirit of [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 $M_0 = \gamma^{-1}(1/n^2)$, $M_\ell = \gamma^{-1}(m^\ell \log(m)/(n^2 \log(n)(m-1)))$, $\ell = 1, \dots, L$ in the case $\rho = 1/2$. Once again the step sequence $\gamma(p) = \gamma_0/p$, $p \geq 1$, is optimal among the usual choices of step sequence and it induces a complexity for the procedure given by $C_{\text{SA-ML}} = K \times n^2(\log(n))^2$. We thus recover the rates as in the multi-level Monte Carlo path simulation for SDE obtained in [Gil08b] and [AK12].

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

2. MAIN RESULTS

In the present paper, we make no attempt to provide an exhaustive discussion related to convergence results of SA schemes. We refer the interested readers to [Duf96], [KY03] and [BMP90] among others for developments and a more complete overview in SA theory. In the next section, we first recall some basic facts concerning stable convergence (following the notations of [JP98]) and list classical results of SA theory.

2.1. Preliminaries

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 $X_n \xrightarrow{\text{stably}} X$, if for all bounded random variable U defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and for all $h : E \rightarrow \mathbb{R}$ bounded continuous, one has

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

This convergence is obviously stronger than convergence in law that we denote by “ \implies ”. Stable convergence was introduced in [Rén63] and notably investigated in [AE78]. The following lemma is a basic result on stable convergence that will be useful throughout the paper. We refer to [JP98], Lemma 2.1 for a proof. Here, E and F will denote two Polish spaces. We consider a sequence $(X_n)_{n \geq 1}$ of E -valued random variable defined on (Ω, \mathcal{F}) .

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

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

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

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

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

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

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

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

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

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

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

We introduce the following smoothness assumption on the coefficients:

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

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

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

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

$$\sqrt{\frac{n}{T}} V^n \xrightarrow{\text{stably}} V, \text{ as } n \rightarrow +\infty$$

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

$$dV_t^i = \sum_{j=1}^{q'+1} \sum_{k=1}^q f_k'^{ij}(X_t) \left[V_t^k dY_t^j - \sum_{\ell=1}^{q'+1} f^{k\ell}(X_t) dZ_t^{\ell j} \right] \quad (2.5)$$

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

$$\begin{aligned} \forall (i, j) \in \llbracket 2, q'+1 \rrbracket \times \llbracket 2, q'+1 \rrbracket, Z_t^{ij} &= \frac{1}{\sqrt{2}} \sum_{1 \leq k, \ell \leq q} \int_0^t \sigma^{ik}(X_s) \sigma^{j\ell}(X_s) dB_s^{k\ell}, \\ \forall j \in \llbracket 1, q'+1 \rrbracket, Z_t^{1j} &= 0, \\ \forall i \in \llbracket 1, q'+1 \rrbracket, Z_t^{i1} &= 0, \end{aligned}$$

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 .

We will also use the following result which is due to [AK12], Theorem 4.

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

$$\sqrt{\frac{m^\ell}{(m-1)T}} (X^{m^\ell} - X^{m^{\ell-1}}) \xrightarrow{\text{stably}} V, \text{ as } \ell \rightarrow +\infty.$$

We now turn our attention to SA. There are various theorems that guarantee the *a.s.* and/or L^p convergence of SA algorithms. We provide below a general result in order to derive the *a.s.* convergence of such procedures. It is also known as *Robbins-Monro Theorem* and covers most situations (see the remark below).

Theorem 2.3 (Robbins-Monro Theorem). *Let $H : \mathbb{R}^d \times \mathbb{R}^q \rightarrow \mathbb{R}^d$ 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

$$\forall \theta \in \mathbb{R}^d, \theta \neq \theta^*, \quad \langle \theta - \theta^*, h(\theta) \rangle > 0. \quad (2.6)$$

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

$$\forall \theta \in \mathbb{R}^d, \quad \mathbb{E}|H(\theta, U)|^2 \leq C(1 + |\theta - \theta^*|^2) \quad (2.7)$$

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

$$\theta_{p+1} = \theta_p - \gamma_{p+1} H(\theta_p, U_{p+1}), \quad p \geq 0 \quad (2.8)$$

satisfies

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

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 $\lim_{|\theta| \rightarrow +\infty} L(\theta) = +\infty$ then, $\text{Argmin } L$ 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 (2.8) 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 $\text{Argmin } L$ is non-empty, then one has

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

Now, we provide a result on the weak rate of convergence of SA algorithm. In standard situations, it is well-known that a stochastic algorithm $(\theta_p)_{p \geq 1}$ converges to its target at a rate $\gamma_p^{-1/2}$. We also refer to [FM12, 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 strictly positive real parts. Assume that the function H satisfies the following regularity and growth control property*

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

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

The step sequence of the procedure (2.8) is given by $\gamma_p = \gamma(p)$, $p \geq 1$, where γ is a positive function defined on $[0, +\infty[$ decreasing to zero. We assume that γ satisfies one of the following assumptions:

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

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

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

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

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

Hence we clearly see that the optimal weak rate of convergence is achieved by choosing $\gamma_p = \gamma_0/p$ with $2\mathcal{R}e(\lambda_{min})\gamma_0 > 1$. However the main drawback 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 SA algorithm (2.8) with a slow decreasing step $\gamma = (\gamma_p)_{p \geq 1}$, where γ varies regularly with exponent $(-a)$, $a \in (1/2, 1)$ and to simultaneously compute the empirical mean $(\bar{\theta}_p)_{p \geq 1}$ of the sequence $(\theta_p)_{p \geq 0}$ by setting

$$\bar{\theta}_p = \frac{\theta_0 + \theta_1 + \dots + \theta_p}{p+1} = \bar{\theta}_{p-1} - \frac{1}{p+1} (\bar{\theta}_{p-1} - \theta_p). \quad (2.9)$$

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 at the equilibrium, that is $\Gamma(\theta^*) := \mathbb{E} [H(\theta^*, U)H(\theta^*, U)^T]$ is a positive definite deterministic matrix.

The step sequence of the procedure (2.8) is given by $\gamma_p = \gamma(p)$, $p \geq 1$, where γ varies regularly with exponent $(-a)$, $a \in (1/2, 1)$. Then, on the event $\{\theta_p \rightarrow \theta^*\}$, one has

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

2.2. Main assumptions

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

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

$$n^\rho (U^n - U) \xrightarrow{\text{stably}} V, \text{ as } n \rightarrow +\infty$$

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

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

$$m^{\ell\rho} (U^{m^\ell} - U^{m^{\ell-1}}) \xrightarrow{\text{stably}} V^m \text{ as } \ell \rightarrow +\infty$$

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

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

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

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

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

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

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

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

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

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

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

(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 $H(\theta, x) = \mathbf{1}_{\{x \leq \theta\}}$ under the additional assumption that U^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 (1.3) on a bounded convex set D (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 D , that is it holds on D instead of \mathbb{R}^d . In this case, a sufficient condition is $\inf_{\theta \in D, n \in \mathbb{N}^*} \lambda_{\min}((Dh^n(\theta) + Dh^n(\theta)^T)/2) > 0$, where $\lambda_{\min}(A)$ denotes the lowest eigenvalue of the matrix A .

We also want to point out that if 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 $(Dh^n)_{n \geq 1}$ and the convergence of $(\theta^{*,n})_{n \geq 1}$ toward θ^* , by passing to the limit $n \rightarrow +\infty$ 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 $\varepsilon \rightarrow 0$, we get $\lambda_m \geq \underline{\lambda}$.

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

$$U^n = \sum_{i=1}^{\lfloor nt \rfloor} f \left(u_n X_{\frac{i-1}{n}}, \sqrt{n} \left(X_{\frac{i}{n}} - X_{\frac{i-1}{n}} \right) \right).$$

Then under some assumptions on the function f and the coefficients b , σ , the weak and strong rate of convergence is $\rho = 1/4$, see [Jac98] for more details. Let us note that we do not know what happens when $\rho > 1/2$ which includes the case of higher order schemes for discretization schemes of SDE.

2.3. On the implicit weak error

As already observed the approximation of θ^* solution of $h(\theta) = \mathbb{E}[H(\theta, U)] = 0$ is affected by two errors: the implicit discretization error and the statistical error. Our first results concern the convergence of $\theta^{*,n}$ toward θ^* and its convergence rate as $n \rightarrow +\infty$.

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

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

Moreover, suppose 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 $\alpha \in \mathbb{R}^*$ such that

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

then, one has

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

2.4. On the optimal tradeoff between the implicit error and the statistical error

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

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

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

where

$$\Sigma^* := \int_0^\infty \exp(-s(Dh(\theta^*) - \zeta I_d))^T \mathbb{E}[H(\theta^*, U)H(\theta^*, U)^T] \exp(-s(Dh(\theta^*) - \zeta I_d)) ds \quad (2.10)$$

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 2.7, one has

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

Proof of Theorem 2.7. We decompose the error as follows:

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

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

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

and using Theorem 2.6, we also obtain

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

□

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

$$C_{SA}(\gamma) = C \times n \times \gamma^{-1}(1/n^{2\alpha}), \quad (2.11)$$

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 $2\lambda\gamma_0 > 1$, 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 $2\lambda\gamma_0 > 1$. 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 (1.3) according to the Ruppert & Polyak averaging principle.

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

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

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

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

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

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

Proof of Theorem 2.8. Similarly to the proof of Theorem 2.7 we decompose the error as follows:

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

Applying successively Theorem 2.6 and Lemma 2.4, we obtain

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

□

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

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

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

2.5. The statistical Romberg stochastic approximation method

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

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

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

$$\Theta_n^{sr} = \theta_{M_1}^{n^\beta} + \theta_{M_2}^n - \theta_{M_2}^{n^\beta}.$$

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

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

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

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

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

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

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

- γ varies regularly with exponent $(-a)$, $a \in (1/2, 1)$, that is, for any $x > 0$, $\lim_{t \rightarrow +\infty} \gamma(tx)/\gamma(t) = x^{-a}$.

- for $t \geq 1$, $\gamma(t) = \gamma_0/t$ and γ_0 satisfies $\underline{\Delta}\gamma_0 > \alpha/(2\alpha - 2\rho\beta)$.

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

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

with

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

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

$$\theta_{p+1} = \theta_p - \gamma_{p+1} H(\theta_p, (U)^{p+1}) \quad (2.12)$$

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

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

with

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

and

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

Proof of Theorem 2.9. We first write the following decomposition

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

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

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

For the first term, from Lemma 2.3 it follows

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

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

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

and use Lemma 2.5 to conclude the proof. \square

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

$$\forall \theta \in \mathbb{R}^d, \quad n^{\alpha - (\alpha - \rho\beta)\alpha} \|Dh(\theta) - Dh^{n^\beta}(\theta)\| \rightarrow 0, \quad \text{as } n \rightarrow +\infty. \quad (2.13)$$

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

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

where

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

Lemma 2.6. *Let $(\bar{\theta}_p)_{p \geq 1}$ be the empirical mean sequence associated to $(\theta_p)_{p \geq 1}$ defined by (2.12). Under the assumptions of Theorem 2.10, one has*

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

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

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

Proof of Theorem 2.10. We decompose the error as follows

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

For the first term, from Lemma 2.4 it follows that

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

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

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

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

2.6. The multi-level stochastic approximation method

As mentioned in the introduction the multi-level SA method uses $L + 1$ stochastic schemes with a sequence of bias parameter $(m^\ell)_{\ell \in \llbracket 0, L \rrbracket}$, for a fixed integer $m \geq 2$, that satisfies $m^L = n$ and estimates θ^* by computing the quantity

$$\Theta_n^{ml} = \theta_{M_0}^1 + \sum_{\ell=1}^L \left(\theta_{M_\ell}^{m^\ell} - \theta_{M_\ell}^{m^{\ell-1}} \right).$$

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

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

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

Suppose that ρ satisfies one of the following assumptions:

- if $\rho \in (0, 1/2)$, then assume that $\alpha > 2\rho$, $\lambda\gamma_0 > \alpha/(\alpha - 2\rho)$ (if $\gamma(t) = \gamma_0/t$) and

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

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

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

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

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

Then one has

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

with

$$\begin{aligned} \Sigma^* := & \int_0^\infty \left(e^{-s(Dh(\theta^*) - \zeta I_d)} \right)^T \left(\mathbb{E} [H(\theta^*, U^1)H(\theta^*, U^1)^T] \right. \\ & \left. + \tilde{\mathbb{E}} \left[\left(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V] \right) \left(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V] \right)^T \right] \right) e^{-s(Dh(\theta^*) - \zeta I_d)} ds \end{aligned}$$

Proof. We first write the following decomposition

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

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

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

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

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

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

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

$$n^\alpha \sum_{\ell=1}^L \left(\theta_{M_\ell}^{m_\ell} - \theta_{M_\ell}^{m_\ell-1} - (\theta^{*,m_\ell} - \theta^{*,m_\ell-1}) \right) \implies \mathcal{N}(0, \Theta^*), \quad n \rightarrow +\infty,$$

with

$$\Theta^* := \int_0^\infty (e^{-s(Dh(\theta^*) - \zeta I_d)})^T \tilde{\mathbb{E}} \left[\left(D_x H(\theta^*, U) V^m - \tilde{\mathbb{E}}[D_x H(\theta^*, U) V^m] \right) \left(D_x H(\theta^*, U) V^m - \tilde{\mathbb{E}}[D_x H(\theta^*, U) V^m] \right)^T \right] \times e^{-s(Dh(\theta^*) - \zeta I_d)} ds. \quad (2.14)$$

Remark 2.5. *The previous result shows that a CLT for the multi-level stochastic approximation estimator of θ^* holds if the standard weak error (and thus the implicit weak error), is of order $1/n^\alpha$ and the strong rate error is of order $1/n^\rho$ with $\alpha > \rho$ or $\alpha = 1$ and $\rho = 1/2$. Due to the non-linearity of the procedures, which leads to annoying remainder terms in the Taylor's expansions, those results do not seem to easily extend to a weak discretization error of order $1/n^\alpha$ with $\alpha < 1$ and $\rho = 1/2$ or a faster strong convergence rate $\rho > 1/2$. Moreover, for the same reason this result does not seem to extend to the empirical sequence associated to the multi-level estimator according to the Ruppert & Polyak averaging principle.*

2.7. Complexity Analysis

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

$$C_{\text{SR-SA}}(\gamma) = C \times (n^\beta \gamma^{-1}(1/n^{2\alpha}) + (n + n^\beta) \gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))) \quad (2.15)$$

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

- If we choose $\gamma(p) = \gamma_0/p$ then simple computations show that $\beta^* = 1/(1 + 2\rho)$ is the optimal choice leading to a complexity

$$C_{\text{SR-SA}}(\gamma) = C' n^{2\alpha+1/(1+2\rho)},$$

under the constraint $\Delta \gamma_0 > \alpha(1 + 2\rho)/(2\alpha(1 + 2\rho) - 2\rho)$ and $\alpha > 2\rho/(1 + 2\rho)$. Let us note that this computational cost is similar to the one achieved by the statistical Romberg Monte Carlo method for the computation of $\mathbb{E}_x[f(X_T)]$.

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

$$C_{\text{SR-SA}}(\gamma) = C' (n^{\frac{2\alpha}{a} + \beta} + n^{\frac{2\alpha}{a} - \frac{\beta}{a} + 1})$$

which is minimized for $\beta^* = a/(2\rho + a)$ leading to an optimal complexity

$$C_{\text{SR-SA}}(\gamma) = C' n^{\frac{2\alpha}{a} + \frac{a}{2\rho+a}}.$$

under the constraint $\alpha > 2\rho a/(a + 2\rho) \vee \rho$. Observe that this complexity decreases with respect to a and that it is minimal for $a \rightarrow 1$ 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 in practical implementation so that a blind choice has often to be made.

The CLT proved in Theorem 2.10 shows that for a total error of order $1/n^\alpha$, it is necessary to set $M_1 = n^{2\alpha}$, $M_2 = n^{2\alpha-2\rho\beta}$ and to simultaneously compute its empirical mean, which represents a negligible part of the total cost. Both stochastic approximation algorithm are devised with a step γ satisfying **(HS1)** with $a \in (1/2, 1)$ and $a > \frac{\alpha}{2\alpha-2\rho\beta} \vee \frac{\alpha(1-\beta)}{\alpha-\rho\beta}$. It is plain to see that $\beta^* = 1/(1+2\rho)$ is the optimal choice leading to a complexity given by

$$C_{\text{SR-RP}}(\gamma) = C \times n^{2\alpha+1/(1+2\rho)},$$

provided that $a > \frac{\alpha(1+2\rho)}{2\alpha+2\rho(2\alpha-1)}$ and $\forall \theta \in \mathbb{R}^d$, $n^{\alpha-(\alpha-\frac{\rho}{1+2\rho})a} \|Dh(\theta) - Dh^{n^{1/(1+2\rho)}}(\theta)\| \rightarrow 0$ as $n \rightarrow +\infty$ (note that when $a \rightarrow 1$ this condition is the same as in Theorem 2.9). For instance, if $\alpha = 1$ and $\rho = 1/2$, then this condition writes $a > 2/3$ and $n^{1-\frac{3}{2}a} \|Dh(\theta) - Dh^{n^{1/2}}(\theta)\| \rightarrow 0$ and a should be selected sufficiently close to 1 according to the weak discretization error of the Jacobian matrix of h . Therefore, the optimal complexity is reached for free without any condition on γ_0 thanks to the Ruppert & Polyak averaging principle. Let us also note that 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 SA algorithms.

Finally, concerning the CLT provided in Theorem 2.11 shows that in order to obtain an error of order $1/n^\alpha$, one has to set $M_0 = \gamma^{-1}(1/n^{2\alpha})$ and $M_\ell = \gamma^{-1}(m^\ell \frac{(1+2\rho)}{2} (m^{\frac{1-2\rho}{2}} - 1)/(n^{2\alpha}(n^{\frac{(1-2\rho)}{2}} - 1)))$, if $\rho \in (0, 1/2)$ or $M_0 = \gamma^{-1}(1/n^2)$ and $M_\ell = \gamma^{-1}(m^\ell \log(m)/(n^2 \log(n)(m-1)))$ if $\rho = 1/2$, $\ell = 1, \dots, L$. In both cases the complexity of the multi-level SA method is given by

$$C_{\text{ML-SA}}(\gamma) = C \times \left(\gamma^{-1}(1/n^{2\alpha}) + \sum_{\ell=1}^L M_\ell(m^\ell + m^{\ell-1}) \right). \quad (2.16)$$

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

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

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

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

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

if $\rho = 1/2$ under the constraint $\underline{\Delta}\gamma_0 > 1$. These computational costs are similar to those achieved by the multi-level Monte Carlo method for the computation of $\mathbb{E}_x[f(X_T)]$, see [Gil08b] and [AK12]. As discussed in [Gil08b], this complexity attains a minimum near $m = 7$.

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

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

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

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

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

Remark 2.6. *The value of M_0 in Theorem 2.11 seems arbitrary and is asymptotically suboptimal. Indeed choosing $M_0 = \gamma^1(1/(n^{2\alpha}n^{1-2\rho}))$ for $\rho \in (0, 1/2)$ and $M_0 = \gamma^{-1}(1/(n^2 \log(n)))$ for $\rho = 1/2$ does not change the asymptotic computational complexity and simplifies the asymptotic covariance matrix Σ^* . One easily proves that $n^\alpha(\theta_{M_0}^1 - \theta^{*,1})$ converges to 0 in probability so that Σ^* now writes*

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

3. PROOFS OF MAIN RESULTS

3.1. Proof of Theorem 2.6

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.2. Proof of Lemma 2.3

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

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

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

$$\theta_n^{n^\delta} - \theta^{*,n^\delta} = \Pi_{1,n}(\theta_0^{n^\delta} - \theta^{*,n^\delta}) + \sum_{k=1}^n \gamma_k \Pi_{k+1,n} \Delta M_k^{n^\delta} + \sum_{k=1}^n \gamma_k \Pi_{k+1,n} \left(\zeta_{k-1}^{n^\delta} + (Dh(\theta^*) - Dh^{n^\delta}(\theta^{*,n^\delta}))(\theta_{k-1}^{n^\delta} - \theta^{*,n^\delta}) \right) \quad (3.17)$$

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

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

First, since $-Dh(\theta^*)$ is a Hurwitz matrix, $\forall \lambda \in [0, \lambda_m)$, there exists $C > 0$ such that for any $k \leq n$, $\|\Pi_{k,n}\| \leq C \prod_{j=k}^n (1 - \lambda \gamma_j) \leq C \exp(-\lambda \sum_{j=k}^n \gamma_j)$. We refer to [Duf96] and [BMP90] for more details. Hence, one has for all $\eta \in (0, \lambda_m)$

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

Selecting η such that $2(\lambda_m - \eta)\gamma_0 > 2(\underline{\Delta} - \eta)\gamma_0 > 1$ 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 $\left\{ n^\alpha \sum_{k=1}^{\gamma^{-1}(1/n^{2\alpha})} \gamma_k \Pi_{k+1, \gamma^{-1}(1/n^{2\alpha})} \left(\zeta_{k-1}^{n^\delta} + (Dh(\theta^*) - Dh^{n^\delta}(\theta^{*, n^\delta}))(\theta_{k-1}^{n^\delta} - \theta^{*, n^\delta}) \right), n \geq 0 \right\}$

We focus on the last term of (3.17). Using Lemma 5.2 we get

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

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

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

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

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

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

To conclude we prove that the sequence $\{\gamma^{-1/2}(n)M_n, n \geq 0\}$, satisfies a CLT. In order to do this we apply standard results on CLT for martingale arrays. More precisely, we will apply Theorem 3.2 and Corollary 3.1, p.58 in [HH80] so that we need to prove that the conditional Lindeberg assumption is satisfied, that is $\lim_n \sum_{k=1}^n \mathbb{E} [|\gamma^{-1/2}(n) \gamma_k \Pi_{k+1, n} (h^{n^\delta}(\theta^{*, n^\delta}) - H(\theta^{*, n^\delta}, (U^{n^\delta})^{k+1}))|^p] = 0$, for some $p > 2$ and that the conditional variance $(S_n)_{n \geq 1}$ defined by

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

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

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

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

By Lemma 5.1, we have $\limsup_n \gamma^{-1+\delta/2}(n) \sum_{k=1}^n \gamma_k^{2+\delta} \|\Pi_{k+1,n}\|^{2+\delta} \leq \limsup_n \gamma^{\delta/2}(n) = 0$, so that the conditional Lindeberg condition, see [HH80], Corollary 3.1, is satisfied. Now we focus on the conditional variance. By the local uniform convergence of $(\theta \mapsto \mathbb{E}[H(\theta, U^{n^\delta})(H(\theta, U^{n^\delta}))^T])_{n \geq 0}$, the continuity of $\theta \mapsto \mathbb{E}[H(\theta, U)(H(\theta, U))^T]$ at θ^* and since $\theta^{*,n^\delta} \rightarrow \theta^*$, we have $\Gamma_n \rightarrow \Gamma^*$, so that from Lemma 5.1, it follows that

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

Hence we see that $\lim_n S_n = \lim_n \frac{1}{\gamma(n)} \sum_{k=1}^n \gamma_k^2 \Pi_{k+1,n} \Gamma^* \Pi_{k+1,n}^T$ if this latter limit exists. Let us note that Σ^* given by (2.10) is the (unique) matrix A solution to the Lyapunov equation:

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

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

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

which can be written in the following recursive form

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

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

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

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 (1.2), it follows that for $N \geq n_0$

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

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

3.3. Proof of Lemma 2.4

We freely use the notations and the intermediate results of the proof of Lemma 2.3. Using (3.17) 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

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

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 it follows

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

since by **(HS1)** one has $n\gamma_n \rightarrow 0$, $n \rightarrow +\infty$.

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

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

where we used Lemma 5.2 for the last inequality and assumption **(HS1)** with $a < 1$.

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

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

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

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

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 $\theta^{*,n^\delta} \rightarrow \theta^*$ and by **(HI)** it holds for some $R > 0$

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

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

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

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

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

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

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

since γ satisfies **(HS1)** with $a < 1$.

3.4. Proof of Lemma 2.5

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 $(\theta_p)_{p \geq 0}$ *a.s.* converges to θ^* and satisfies a CLT according to Theorem 2.4.

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

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

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

$$\begin{aligned} z_n^{n^\beta} &= \Pi_{1,n} z_0^{n^\beta} + \sum_{k=1}^n \gamma_k \Pi_{k+1,n} \Delta N_k^{n^\beta} + \sum_{k=1}^n \gamma_k \Pi_{k+1,n} \Delta R_k^{n^\beta} \\ &\quad + \sum_{k=1}^n \gamma_k \Pi_{k+1,n} \left(\zeta_{k-1}^{n^\beta} - \zeta_{k-1} + (Dh(\theta^*) - Dh^{n^\beta}(\theta^{*,n^\beta}))(\theta_{k-1}^{n^\beta} - \theta^{*,n^\beta}) \right) \end{aligned} \quad (3.18)$$

where $\Pi_{k,n} := \prod_{j=k}^n (I_d - \gamma_j Dh(\theta^*))$, with the convention that $\Pi_{n+1,n} = I_d$, and $\Delta N_k^{n^\beta} := h^{n^\beta}(\theta^*) - h(\theta^*) - (H(\theta^*, (U^{n^\beta})^{k+1}) - H(\theta^*, U^{k+1}))$, $\Delta R_k^{n^\beta} = h^{n^\beta}(\theta_k^{n^\beta}) - h^{n^\beta}(\theta^*) - (H(\theta_k^{n^\beta}, (U^{n^\beta})^{k+1}) - H(\theta^*, (U^{n^\beta})^{k+1})) + H(\theta_k, U^{k+1}) - H(\theta^*, U^{k+1}) - (h(\theta_k) - h(\theta^*))$ for $k \geq 1$. We will now investigate the asymptotic behavior of each term in the above decomposition. We will see that the second term which represents the non-linearity in the innovation variables (U^{n^β}, U) provides the announced weak rate of convergence.

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

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

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

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

Step 2: study of the sequence

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

By Lemma 5.2, one has

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

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

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

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

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

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

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

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

Regarding the third term of (3.18), namely $\sum_{k=1}^n \gamma_k \Pi_{k+1,n} \Delta R_k^{n^\beta}$, we decompose it as follows

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

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

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

From Lemma 5.1 we get $\sum_{k=1}^n \gamma_k^3 \|\Pi_{k+1,n}\|^2 = o(\gamma_n^{2\alpha/(2\alpha - 2\rho\beta)})$ and $\sum_{k=1}^n \gamma_k^2 \|\Pi_{k+1,n}\|^2 = \mathcal{O}(\gamma_n)$. Consequently, we derive $\limsup_n n^{2\alpha} A_{\gamma^{-1}(1/(n^{2\alpha - 2\rho\beta}))}^1 = 0$ and $\limsup_n n^{2\alpha} A_{\gamma^{-1}(1/(n^{2\alpha - 2\rho\beta}))}^2 = 0$. Similarly using **(HLH)** and Lemma 5.2 we derive $n^\alpha B_{\gamma^{-1}(1/(n^{2\alpha - 2\rho\beta}))} \xrightarrow{L^2(\mathbb{P})} 0$ as $n \rightarrow +\infty$ so that

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

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

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

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

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

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

Moreover simple computations lead

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

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

$$\sup_{n \geq 1} \sup_{k \in \llbracket 1, n \rrbracket} \mathbb{E} \left| n^{\rho\beta} \Delta N_k^{n^\beta} \right|^{2+\delta} < +\infty,$$

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

$$S_n := n^{2\alpha} \sum_{k=1}^{\gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \gamma_k^2 \Pi_{k+1, \gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))} \mathbb{E}_k[\Delta N_k^{n^\beta} (\Delta N_k^{n^\beta})^T] \Pi_{k+1, \gamma^{-1}(1/(n^{2\alpha-2\rho\beta}))}^T, \text{ and } V^{n^\beta} = U^{n^\beta} - U. \quad (3.19)$$

A Taylor's expansion yields

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

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

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

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

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

which combined with **(HDH)** imply

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

Hence, we have

$$\Gamma_n \rightarrow \Gamma^* := \tilde{\mathbb{E}} \left[\left(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V] \right) \left(D_x H(\theta^*, U)V - \tilde{\mathbb{E}}[D_x H(\theta^*, U)V] \right)^T \right]$$

where for $n \geq 1$ we set

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

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 Lemma 5.1, we clearly see that $\lim_n S_n = \lim_n \frac{1}{\gamma(n)} \sum_{k=1}^n \gamma_k^2 \Pi_{k+1,n} \Gamma^* \Pi_{k+1,n}^T$ if this latter limit exists. Let us note that Θ^* is the (unique) matrix A solution to the Lyapunov equation:

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

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.

3.5. Proof of Lemma 2.6

We will just prove the first assertion. The second one will readily follow. We use C to denote a constant that may change from one line to the next. Using the notations of Lemma 2.5, the sequence $(\bar{z}_p^{n^\beta})_{p \in \llbracket 0, n^{2\alpha-2\rho\beta} \rrbracket}$ can be decomposed as follows:

$$\begin{aligned} \bar{z}_{n^{2\alpha-2\rho\beta}}^{n^\beta} &= \frac{1}{n^{2\alpha-2\rho\beta} + 1} \sum_{k=0}^{n^{2\alpha-2\rho\beta}} z_k^{n^\beta} \\ &= -(Dh(\theta^*))^{-1} \frac{1}{n^{2\alpha-2\rho\beta} + 1} \left(\frac{z_{n^{2\alpha-2\rho\beta}+1}^{n^\beta} + z_0^{n^\beta}}{\gamma_{n^{2\alpha-2\rho\beta}+1}} - \frac{z_0^{n^\beta}}{\gamma_1} \right) - (Dh(\theta^*))^{-1} \frac{1}{n^{2\alpha-2\rho\beta} + 1} \sum_{k=1}^{n^{2\alpha-2\rho\beta}} \left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k+1}} \right) z_k^{n^\beta} \\ &+ (Dh(\theta^*))^{-1} \frac{1}{n^{2\alpha-2\rho\beta} + 1} \sum_{k=0}^{n^{2\alpha-2\rho\beta}} (\Delta N_{k+1}^{n^\beta} + \Delta R_{k+1}^{n^\beta}) \\ &+ (Dh(\theta^*))^{-1} \frac{1}{n^{2\alpha-2\rho\beta} + 1} \sum_{k=0}^{n^{2\alpha-2\rho\beta}} (\zeta_k^{n^\beta} - \zeta_k + (Dh(\theta^*) - Dh^{n^\beta}(\theta^*, n^\beta))(\theta_k^{n^\beta} - \theta^*, n^\beta)). \end{aligned}$$

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

Step 1: study of the sequence $\left\{ \frac{n^\alpha}{n^{2\alpha-2\rho\beta}+1} \left(\frac{z_{n^{2\alpha-2\rho\beta}+1}^{n^\beta}}{\gamma_{n^{2\alpha-2\rho\beta}+1}} - \frac{z_0^{n^\beta}}{\gamma_1} \right), n \geq 0 \right\}$:

Using Proposition 5.1 clearly yields

$$\frac{n^\alpha}{n^{2\alpha-2\rho\beta}+1} \mathbb{E} \left| \frac{z_{n^{2\alpha-2\rho\beta}+1}^{n^\beta}}{\gamma_{n^{2\alpha-2\rho\beta}+1}} - \frac{z_0^{n^\beta}}{\gamma_1} \right| \leq \frac{C}{(n^{\alpha-2\rho\beta}\gamma_{n^{2\alpha-2\rho\beta}+1})} (\mathbb{E}|\tilde{\mu}_{n^{2\alpha-2\rho\beta}}^{n^\beta}| + \mathbb{E}|\tilde{r}_{n^{2\alpha-2\rho\beta}}^{n^\beta}|) + \frac{C}{n^{\alpha-2\rho\beta}} (1 + |\theta^* - \theta^{*,n^\beta}|).$$

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

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

and

$$\frac{1}{(n^{\alpha-2\rho\beta})\gamma_{n^{2\alpha-2\rho\beta}+1}} \mathbb{E}|\tilde{r}_{n^{2\alpha-2\rho\beta}}^{n^\beta}| \leq \frac{C}{n^{\alpha-2\rho\beta}} \rightarrow 0, \text{ as } n \rightarrow +\infty.$$

From these computations we get

$$\frac{n^\alpha}{n^{2\alpha-2\rho\beta}+1} \left(\frac{z_{n^{2\alpha-2\rho\beta}+1}^{n^\beta}}{\gamma_{n^{2\alpha-2\rho\beta}+1}} - \frac{z_0^{n^\beta}}{\gamma_1} \right) \xrightarrow{L^1(\mathbb{P})} 0, \text{ as } n \rightarrow +\infty.$$

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

We use the decomposition of Proposition 5.1 to derive

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

Then taking the expectation in the previous inequality and using that $\rho < 1$ we deduce

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

For the second term, we have

$$\frac{n^\alpha}{n^{2\alpha-2\rho\beta}+1} \sum_{k=1}^{n^{2\alpha-2\rho\beta}} \left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k} \right) \mathbb{E}|\tilde{r}_k^{n^\beta}| \leq \frac{C}{n^{\alpha-2\rho\beta}} \sum_{k=1}^{n^{2\alpha-2\rho\beta}} \left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k} \right) \gamma_k \rightarrow 0,$$

since $\alpha > 2\rho\beta$ which in turn implies

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

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

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

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

since $a > \alpha/(2\alpha - 2\rho\beta)$. Now since Dh^{n^β} is Lipschitz-continuous uniformly in n we easily get

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

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

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

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

Similarly to the proof of Lemma 2.7, we decompose the sequence $\left\{ \frac{n^\alpha}{n^{2\alpha-2\rho\beta} + 1} \sum_{k=1}^{n^{2\alpha-2\rho\beta}} \Delta R_k^{n^\beta}, n \geq 1 \right\}$ as follows

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

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

$$\mathbb{E}|A_n| + \mathbb{E}|B_n| \leq \frac{1}{n^{\alpha-2\rho\beta}} \left(\sum_{k=0}^{n^{2\alpha-2\rho\beta}} \gamma_k \right)^{\frac{1}{2}} \rightarrow 0$$

since $a > \alpha/(2\alpha - 2\rho\beta) > \rho\beta/(\alpha - \rho\beta)$ so that

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

We now prove a CLT for the sequence $\left\{ \frac{n^\alpha}{n^{2\alpha-2\rho\beta} + 1} \sum_{k=0}^{n^{2\alpha-2\rho\beta}} \Delta N_k^{n^\beta}, n \geq 0 \right\}$. We first note

$$\sum_{k=0}^{n^{2\alpha-2\rho\beta}} \mathbb{E} \left| \frac{n^\alpha}{n^{2\alpha-2\rho\beta} + 1} \Delta N_k^{n^\beta} \right|^{2+\delta} \leq \sup_{n \geq 1} \sup_{k \in \llbracket 0, n \rrbracket} \mathbb{E} \left| n^{\rho\beta} \Delta N_k^{n^\beta} \right|^{2+\delta} \frac{1}{n^{\alpha\delta - \rho\beta\delta}} \rightarrow 0 \quad n \rightarrow +\infty$$

where we used assumptions **(HLH)** and **(HSR)** to derive that $\sup_{n \geq 1} \sup_{k \in \llbracket 1, n \rrbracket} \mathbb{E} \left| n^{\rho\beta} \Delta N_k^{n^\beta} \right|^{2+\delta} < +\infty$. Therefore the conditional Lindeberg condition is satisfied. Then we examine the conditional variance. Recall

that (see the the proof of Lemma 2.7) we have

$$\begin{aligned} n^{2\rho\beta} \mathbb{E}_k[\Delta N_k^{n^\beta} (\Delta N_k^{n^\beta})^T] &= n^{2\rho\beta} \mathbb{E} \left[(H(\theta^*, U^{n^\beta}) - H(\theta^*, U) - (h^{n^\beta}(\theta^*) - h(\theta^*))) \right. \\ &\quad \left. \times (H(\theta^*, U^{n^\beta}) - H(\theta^*, U) - (h^{n^\beta}(\theta^*) - h(\theta^*)))^T \right] \\ &\rightarrow \tilde{\mathbb{E}} \left[\left(D_x H(\theta^*, U) V - \tilde{\mathbb{E}}[D_x H(\theta^*, U) V] \right) \left(D_x H(\theta^*, U) V - \tilde{\mathbb{E}}[D_x H(\theta^*, U) V] \right)^T \right], \end{aligned}$$

so that if we set

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

we clearly get

$$S_n \rightarrow \tilde{\mathbb{E}} (D_x H(\theta^*, U) V) (D_x H(\theta^*, U) V)^T.$$

This completes the proof.

3.6. Proof of Lemma 2.7

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. We deal with the case $\rho \in (0, 1/2)$. The case $\rho = 1/2$ can be handled in a similar fashion.

We first write for $p \geq 0$

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

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

$$\begin{aligned} z_{M_\ell}^\ell &= \Pi_{1, M_\ell} z_0^\ell + \sum_{k=1}^{M_\ell} \gamma_k \Pi_{k+1, M_\ell} \Delta N_k^\ell + \sum_{k=1}^{M_\ell} \gamma_k \Pi_{k+1, M_\ell} \Delta R_k^\ell \\ &\quad + \sum_{k=1}^{M_\ell} \gamma_k \Pi_{k+1, M_\ell} \left(\zeta_{k-1}^\ell - \zeta_{k-1}^{\ell-1} + (Dh(\theta^*) - Dh^{m^\ell}(\theta^{*,m^\ell}))(\theta_{k-1}^{m^\ell} - \theta^{*,m^\ell}) \right. \\ &\quad \left. - (Dh(\theta^*) - Dh^{m^{\ell-1}}(\theta^{*,m^{\ell-1}}))(\theta_{k-1}^{m^{\ell-1}} - \theta^{*,m^{\ell-1}}) \right) \end{aligned} \tag{3.20}$$

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

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

Under the assumptions on the step sequence γ , for all $\eta \in (0, \lambda_m)$ we have $\|\Pi_{1, M_\ell}\| \leq \exp(-(\lambda_m - \eta) \sum_{k=1}^{M_\ell} \gamma_k) = C m^{\ell \frac{(1+2\rho)}{2} (\lambda_m - \eta) \gamma_0} / ((n^{2\alpha + \frac{1-2\rho}{2}})^{(\lambda_m - \eta) \gamma_0})$ if $\gamma(p) = \gamma_0/p$ or $\|\Pi_{1, M_\ell}\| = \mathcal{O}(\gamma(M_\ell))$ otherwise. Therefore, if $\gamma(p) = \gamma_0/p$ we select $\eta > 0$ such that $\gamma_0(\lambda_m - \eta) > \alpha/(\alpha - 2\rho)$ then one has

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

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

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

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

By Lemma 5.2, one has

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

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

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

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

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

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

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

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

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

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

Using the Cauchy-Schwarz inequality we deduce

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

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

$$n^\alpha \sum_{\ell=1}^L \left(\sum_{k=1}^{M_\ell} \gamma_k^3 \|\Pi_{k+1, M_\ell}\|^2 \right)^{1/2} \leq C n^\alpha \sum_{\ell=1}^L \gamma(M_\ell) \rightarrow 0, \quad n \rightarrow +\infty.$$

Therefore, we conclude that

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

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

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

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

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

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

Moreover, by Lemma 5.1, we have

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

Consequently we deduce

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

which in turn implies

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

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

$$S_\ell := n^{2\alpha} \sum_{k=1}^{M_\ell} \gamma_k^2 \Pi_{k+1, M_\ell} \mathbb{E}_k [\Delta N_k^\ell (\Delta N_k^\ell)^T] \Pi_{k+1, M_\ell}^T, \quad \text{and} \quad U^\ell = U^{m^\ell} - U^{m^{\ell-1}}. \quad (3.21)$$

Observe that by the very definition of M_ℓ one has

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

A Taylor's expansion yields

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

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

$$m^{\rho\ell} \left(\psi(\theta^*, U, U^{m^\ell} - U)(U^{m^\ell} - U) + \psi(\theta^*, U, U^{m^{\ell-1}} - U)(U^{m^{\ell-1}} - U) \right) \xrightarrow{\mathbb{P}} 0, \quad \ell \rightarrow +\infty.$$

Therefore using Theorem 2.1 and Lemma 2.1 yield

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

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

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

which combined with **(HDH)** imply

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

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

$$m^{2\rho\ell} \Gamma_\ell \rightarrow \Gamma^* := \tilde{\mathbb{E}} \left[\left(D_x H(\theta^*, U) V^m - \tilde{\mathbb{E}}[D_x H(\theta^*, U) V^m] \right) \left(D_x H(\theta^*, U) V^m - \tilde{\mathbb{E}}[D_x H(\theta^*, U) V^m] \right)^T \right]$$

where for $\ell \geq 1$

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

Consequently, using the following decomposition

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

with

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

which is a consequence of Lemma 5.1, we clearly see that $\frac{n^{\frac{1-2\rho}{2}} - 1}{m^\ell \frac{(1-2\rho)}{2} (m^{\frac{1-2\rho}{2}} - 1)} \lim_\ell S_\ell = \lim_{p \rightarrow +\infty} \frac{1}{\gamma(p)} \sum_{k=1}^p \gamma_k^2 \Pi_{k+1, p} \Gamma^* \Pi_{k+1, p}^\top$

if this latter limit exists. The matrix Θ^* defined by (2.14) is the (unique) matrix A solution to the Lyapunov equation:

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

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

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

4. NUMERICAL RESULTS

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

4.1. Computation of quantiles of a one dimensional diffusion process

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

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

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

$$X_t = x + \int_0^t r X_s ds + \int_0^t \sigma X_s dW_s, \quad t \in [0, T] \quad (4.22)$$

for which the quantile is explicitly known at any level l . Hence we have $\rho = 1/2$. 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 (4.22) its Euler like scheme $X^n = (X_t^n)_{t \in [0, T]}$ with time step $\Delta = T/n$. We use the following values for the parameters: $x = 100$, $r = 0.05$, $\sigma = 0.4$, $T = 1$, $l = 0.7$. The reference Black-Scholes quantile is $q_{0.7}(X_T) = 119.69$.

Remark 4.1. *Let us note that when l is close to 0 or 1 (usually less than 0.05 or more than 0.95) the convergence of the considered SA algorithm is slow and chaotic. This is mainly due to the fact that the procedure obtains few significant samples to update the estimate in this rare event situation. One solution is to combine it with a variance reduction algorithm such as an adaptive importance sampling procedure that will generate more samples in the area of interest, see e.g. [BFP09a] and [BFP09b].*

In order to illustrate the result of Theorem 2.6, we plot in Figure 1 the behaviors of $nh^n(\theta^*)$ and $n(\theta^{*,n} - \theta^*)$ for $n = 100, \dots, 500$. Actually, $h^n(\theta^*)$ is approximated by its Monte Carlo estimator and $\theta^{*,n}$ is estimated by θ_M^n , both estimators 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 2.7. The distribution of $n(\theta_{\gamma^{-1}(1/n^2)}^n - \theta^*)$, obtained with $n = 100$ and $N = 1000$ samples, is close to a normal distribution.

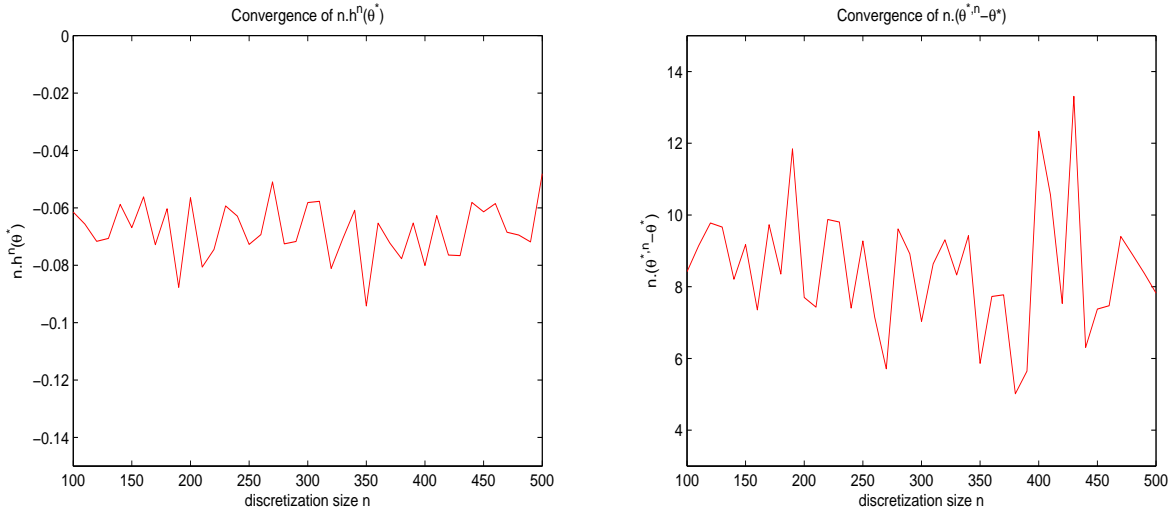


FIGURE 1. On the left: Weak discretization error $n \mapsto nh^n(\theta^*)$. On the right: Implicit discretization error $n \mapsto n(\theta^{*,n} - \theta^*)$, $n = 100, \dots, 500$.

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 (4.22) is given by

$$e^{-rT} \mathbb{E}(X_T - \theta)_+ = e^{-rT} x \phi(d_+(x, \theta, \sigma)) - e^{-rT} \theta \phi(d_-(x, \theta, \sigma)), \quad (4.23)$$

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 (4.23). The values of the parameters x, r, σ, T remain unchanged. We plot in Figure 3 the behaviors of $nh^n(\theta^*)$ and $n(\theta^{*,n} - \theta^*)$ for $n = 100, \dots, 500$. As in the previous example, $h^n(\theta^*)$ is approximated by its Monte Carlo estimator and $\theta^{*,n}$ is estimated by θ_M^n , 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 $n = 500$.

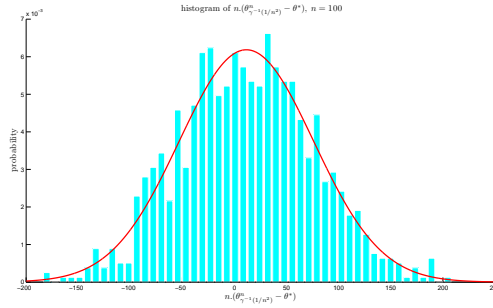


FIGURE 2. Histogram of $n(\theta_{\gamma^{-1}(1/n^2)}^n - \theta^*)$, $n = 100$, with $N = 1000$ samples.

To compare the three methods to approximate the solution to $h(\theta) = \mathbb{E}_x[H(\theta, X_T)] = 0$ with $H(\theta, x) = l - (x - \theta)_+$ in terms of computational costs, we compute the different estimators, namely $\theta_{\gamma^{-1}(1/n^2)}^n$ where $(\theta_p^n)_{p \geq 1}$ is given by (1.3), 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 (2.11), (2.15) and (2.16) 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 $\Theta_k^n = \theta_{\gamma^{-1}(1/n^2)}^n$, 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 6. 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 $\beta^* = 1/2$.

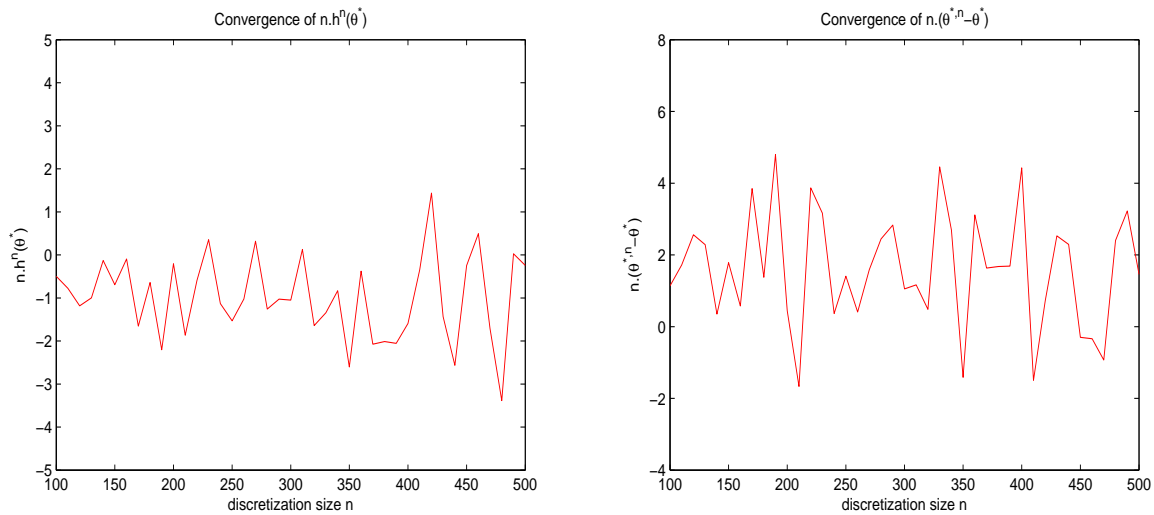


FIGURE 3. On the left: Weak discretization error $n \mapsto nh^n(\theta^*)$. On the right: Implicit discretization error $n \mapsto n(\theta^{*,n} - \theta^*)$, $n = 100, \dots, 500$.

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 $\theta_{\gamma_0 n^{3/2}}^n - \theta_{\gamma_0 n^{3/2}}^{\sqrt{n}}$ (for the statistical Romberg SA) and $\theta_{M_\ell}^{m_\ell} - \theta_{M_\ell}^{m_\ell-1}$ (for the Multi-level SA) at level ℓ are fixed to $\theta_{\gamma_0 n^2}^{\sqrt{n}}$ and to $\theta_{\gamma_0 n^2}^1 + \sum_{\ell=1}^{L-1} \theta_{M_\ell}^{m_\ell} - \theta_{M_\ell}^{m_\ell-1}$ respectively. We set $\theta_0^{n^{1/2}} = \theta_0^1 = x$ for all $k \in \{1, \dots, M\}$ to initialize the procedures. Moreover, by Lemma 5.2, the $L^1(\mathbb{P})$ -norm of an increment of a SA algorithm is of order $\sqrt{\gamma_0/p}$ since $\mathbb{E}|\theta_{p+1}^n - \theta_p^n| \leq \mathbb{E}[|\theta_{p+1}^n - \theta^{*,n}|^2]^{1/2} + \mathbb{E}[|\theta_p^n - \theta^{*,n}|^2]^{1/2} \leq C(H, \gamma)\sqrt{\gamma(p)}$. Hence, during the first iterations (say $M/100$ if M denotes the number of samples of the estimator), to ensure that the different procedures do not jump too far ahead in one step, we freeze the value of $\theta_{p+1}^{\sqrt{n}}$ (respectively $\theta_{p+1}^{m_\ell}$) and reset it to the value of the previous step as soon as $|\theta_{p+1}^{\sqrt{n}} - \theta_p^{\sqrt{n}}| \leq K/\sqrt{p}$ (respectively $|\theta_{p+1}^{m_\ell} - \theta_p^{m_\ell}| \leq K/\sqrt{p}$), for a pre-specified value of K . This is just an heuristic approach that notably prevents the algorithm from blowing up during the first steps of the procedure. We select $K = 5$ in the different procedures. Note anyway that this projection-reinitialization step does not lead to additional bias but slightly increases the complexity of each procedures. In our numerical examples, we observe that it only represents around 1-2% of the total complexity.

Now let us interpret Figure 6. 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 one of the crude SA. The difference in terms of computational cost becomes more significant as the RMSE is small, which corresponds to large values of n . The difference between the statistical romberg and the multi-level SA method is not significant for small values of n , *i.e.* for a RMSE between 1 and 0.1. For a RMSE lower than $5 \cdot 10^{-2}$, which corresponds to a number of steps n greater than about 600-700, we observe that the multi-level SA procedure becomes much more effective than both methods. For a RMSE fixed around 1 (which corresponds to $n = 100$ for the SA algorithm and Statistical Romberg SA), one divides the complexity by a factor of approximately 5 by using the statistical romberg SA. For a RMSE fixed at 10^{-1} , the computational cost gain is approximately equal to 10 by using either the statistical romberg SA algorithm or the multi-level SA one. Finally, for a RMSE fixed at $5.5 \cdot 10^{-2}$, 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 2.7, 2.9 and 2.11. The distributions of $n(\theta_{\gamma^{-1}(1/n^2)}^n - \theta^*)$, $n(\Theta_n^{sr} - \theta^*)$ and $n(\Theta_n^{ml} - \theta^*)$, obtained with $n = 4^4 = 256$ and $N = 1000$ samples, are close to a normal distribution.

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 $(-c)$, $c \in [0, 1)$, that is for any $x > 0$, $\lim_{t \rightarrow +\infty} \gamma(tx)/\gamma(t) = x^{-c}$.
- for $t \geq 1$, $\gamma(t) = \gamma_0/t$ with $b\text{Re}(\lambda_{\min})\gamma_0 > a$.

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 $\Pi_{n+1,n} = I_d$.

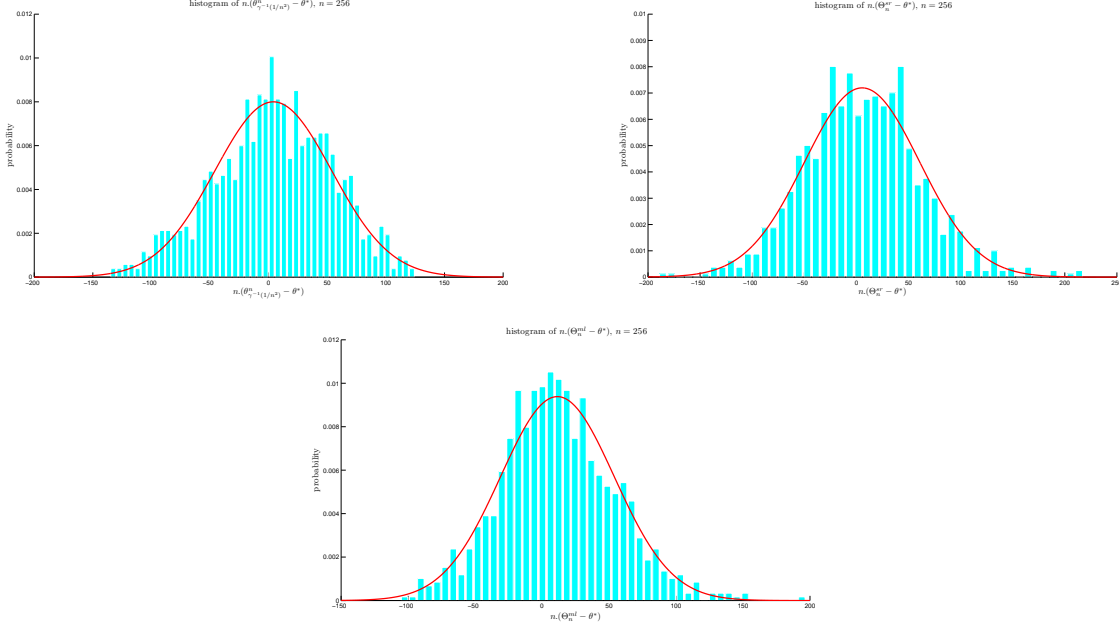


FIGURE 4. Histograms of $n(\theta_{\gamma^{-1}(1/n^2)} - \theta^*)$, $n(\Theta_n^{ST} - \theta^*)$ and $n(\Theta_n^{ml} - \theta^*)$ (from left to right), $n = 256$, with $N = 1000$ samples.

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 be written in the recursive form

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

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

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

We study now the impact of the step sequence $(\gamma_p)_{p \geq 1}$ 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 $r_p \rightarrow 0$ so that a comparison between the series and the integral yields

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

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

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

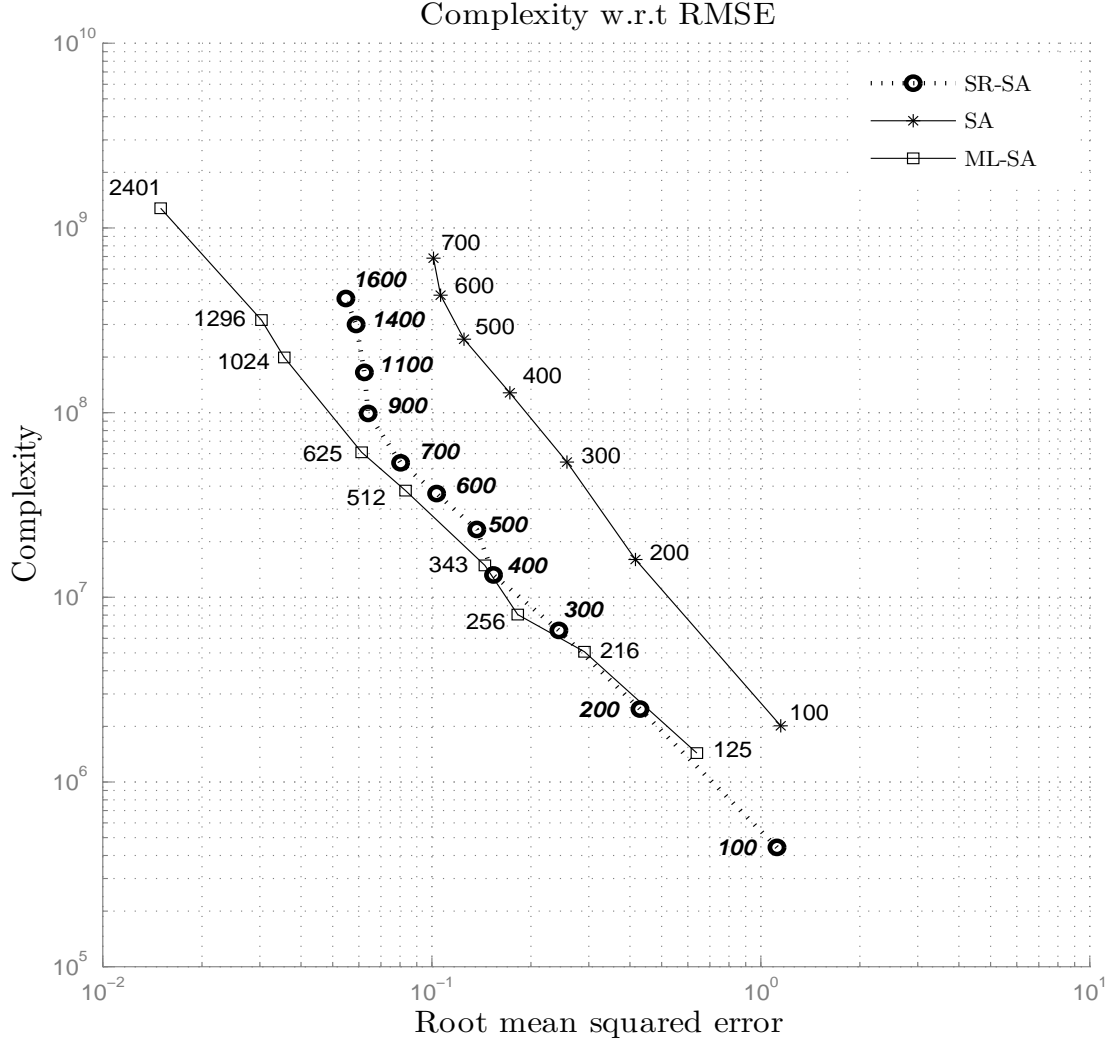


FIGURE 5. Complexity with respect to RMSE.

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

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

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

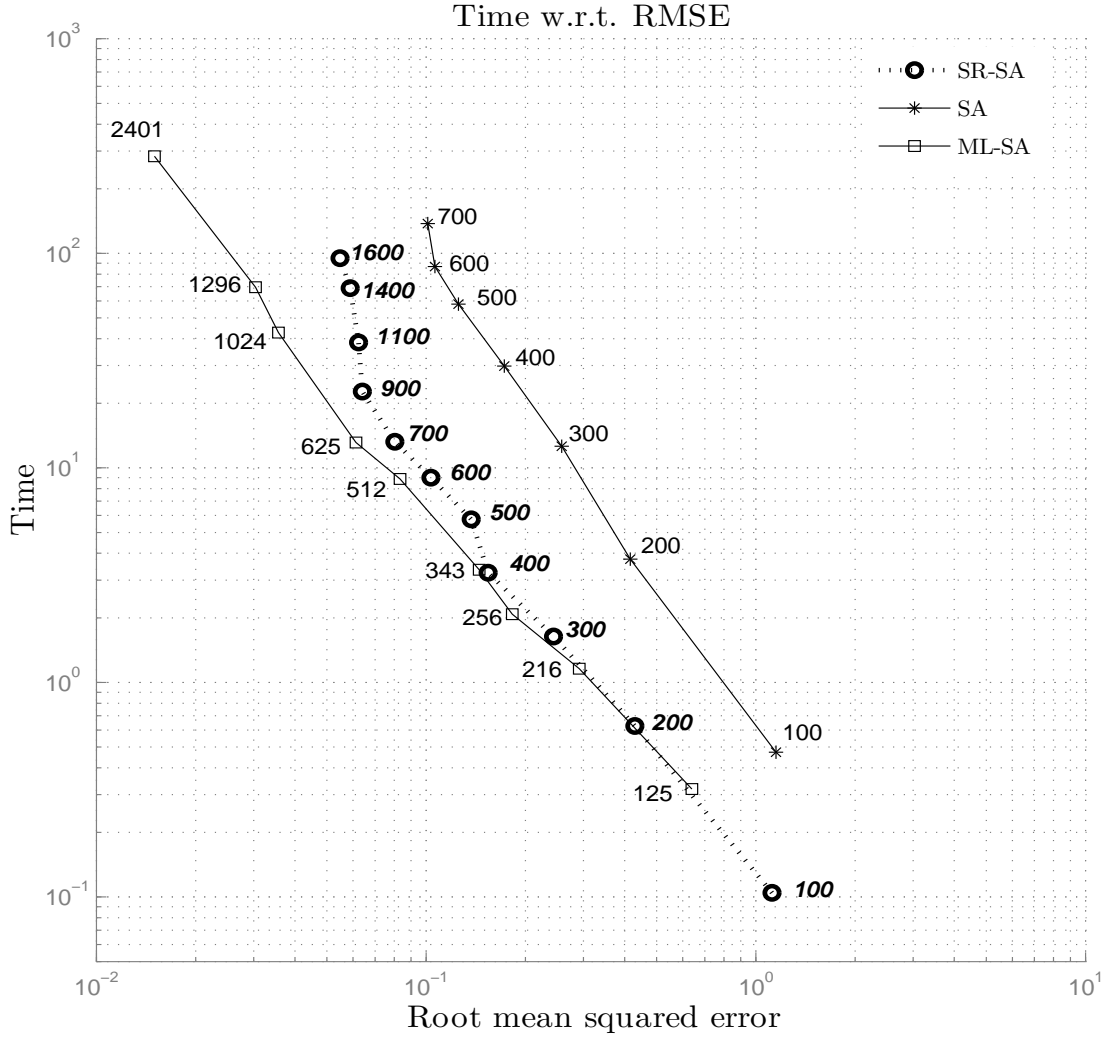


FIGURE 6. Time in second (average time for one sample) with respect to RMSE.

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

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

Hence it follows that

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

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

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

An argument similar to the previous case concludes the proof. \square

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

$$\exists \underline{\lambda} > 0, \forall n \in \mathbb{N}^*, \forall \theta \in \mathbb{R}^d, \langle \theta - \theta^{*,n}, h^n(\theta) \rangle \geq \underline{\lambda} |\theta - \theta^{*,n}|^2, \quad (5.24)$$

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

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

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

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

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

$$\begin{aligned} |\theta_{p+1}^n - \theta^{*,n}|^2 &= |\theta_p^n - \theta^{*,n}|^2 - 2\gamma_{p+1} \langle \theta_p^n - \theta^{*,n}, h^n(\theta_p^n) \rangle + 2\gamma_{p+1} \langle \theta_p^n - \theta^{*,n}, \Delta M_{p+1}^n \rangle \\ &\quad + \gamma_{p+1}^2 |H(\theta_p^n, (X_T^n)^{p+1})|^2, \end{aligned}$$

so that taking expectation in the previous equality and using assumptions (2.7) and (5.24), we easily derive

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

Now a simple induction argument yields

$$\mathbb{E}|\theta_p^n - \theta^{*,n}|^2 \leq \mathbb{E}|\theta_0^n - \theta^{*,n}|^2 \Pi_{1,p} + \sum_{k=1}^p \Pi_{k+1,p} \gamma_k^2$$

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

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

In order to prove the similar bound for the sequence $(\theta_p)_{p \geq 1}$ we first observe that since $(\theta_p)_{p \geq 1}$ converges *a.s.* to θ^* there exists a compact set K (which depends on w) such that $\theta_p \in K$, for $p \geq 0$. Then, Remark 2.3 shows that a mean reverting assumption is satisfied also for h on K with the same constant $\underline{\lambda}$. Finally we conclude using similar arguments as those used above. \square

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

$$\forall p \in \llbracket 0, n \rrbracket, \quad z_p^n = \theta_p^n - \theta^{*,n} - (\theta_p - \theta^*) = \tilde{\mu}_p^n + \tilde{r}_p^n$$

and satisfying for all $n \in \mathbb{N}$, for all $p \in \llbracket 1, n \rrbracket$

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

Proof. Using (3.18), we define the two sequences $(\tilde{\mu}_p^n)_{p \in \llbracket 0, n \rrbracket}$ and $(\tilde{r}_p^n)_{p \in \llbracket 0, n \rrbracket}$ by

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

and

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

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

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

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

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

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

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

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

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

By Lemma 5.2, we derive

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

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

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

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

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.
- [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.
- [Fri14] N. Frikha. Shortfall Risk Minimization in Discrete Time Financial Market Models. *SIAM J. Financial Math.*, 5(1):384–414, 2014.
- [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.
- [Hei01] S. Heinrich. *Multilevel Monte Carlo methods*. Springer, 2001. In Large-scale scientific computing.
- [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.
- [Jac98] J. Jacod. Rates of convergence to the local time of a diffusion. *Ann. Inst. H. Poincaré Probab. Statist.*, 34:505–544, 1998.
- [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.