Importance Sampling and Statistical Romberg method

Mohamed Ben Alaya[1], Kaouther Hajji[1], and Ahmed Kebaier*[1]

[1] Université Paris 13, Sorbonne Paris Cité, LAGA, CNRS (UMR 7539)
mba@math.univ-paris13.fr hajji@math.univ-paris13.fr kebaier@math.univ-paris13.fr

March 13, 2014

Abstract

The efficiency of Monte Carlo simulations is significantly improved when implemented with variance reduction methods. Among these methods we focus on the popular importance sampling technique based on producing a parametric transformation through a shift parameter θ. The optimal choice of θ is approximated using Robbins-Monro procedures, provided that a non explosion condition is satisfied. Otherwise, one can use either a constrained Robbins-Monro algorithm (see e.g. Arouna [2] and Lelong [17]) or a more astute procedure based on an unconstrained approach recently introduced by Lemaire and Pagès in [18]. In this article, we develop a new algorithm based on a combination of the statistical Romberg method and the importance sampling technique. The statistical Romberg method introduced by Kebaier in [12] is known for reducing efficiently the complexity compared to the classical Monte Carlo one. In the setting of discritized diffusions, we prove the almost sure convergence of the constrained and unconstrained versions of the Robbins-Monro routine, towards the optimal shift θ^* that minimizes the variance associated to the statistical Romberg method. Then, we prove a central limit theorem for the new algorithm that we called adaptive statistical Romberg method. Finally, we illustrate by numerical simulation the efficiency of our method through applications in option pricing for the Heston model.

AMS 2000 Mathematics Subject Classification. 60F05, 62F12, 65C05, 60H35.
Key Words and Phrases. Stochastic algorithm, Robbins-Monro, variance reduction, Central limit theorem, Statistical Romberg method, Euler scheme, Heston model.

1 Introduction

Monte Carlo methods have proved to be a useful tool for many of numerical computations in modern finance. These includes the pricing and hedging of complex financial products. The general problem is to estimate a real quantity $\mathbb{E}\psi(X_T)$, with $T > 0$ and $(X_t)_{0\leq t\leq T}$ is a given

*This research benefited from the support of the chair "Risques Financiers", Fondation du Risque.

diffusion, defined on $\mathcal{B} := (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, taking values in $\mathbb{R}^d$ and ψ a given function such that $\psi(X_T)$ is square integrable. Since the efficiency of the Monte Carlo simulation considerably depends on the smallness of the variance in the estimation, many variance reduction techniques were developed in the recent years. Among these methods appears the technique of importance sampling very popular for its efficiency. The working of this method is quite intuitive, if we can produce a parametric transformation such that for all $\theta \in \mathbb{R}^q$ we have

$$\mathbb{E}\psi(X_T) = \mathbb{E}g(\theta, X_T).$$

Then it is natural, to implement a Monte Carlo procedure using the optimal θ^* solution to the problem

$$\theta^* = \arg\min_{\theta\in\mathbb{R}^q} \mathbb{E}g^2(\theta, X_T),$$

since the quantity $\mathbb{E}g^2(\theta, X_T)$ denotes the main term of the limit variance in the central limit theorem associated to the Monte Carlo method. But how to compute θ^*? To solve this problem, one can use the so-called Robbins-Monro algorithm to construct recursively a sequence of random variables $(\theta_i)_{i\in\mathbb{N}}$ that approximate accurately θ^*. Convergence results of this procedure requires a quite restrictive condition known as the non explosion condition (see e.g. [4, 8, 15]) given by

$$\text{(NEC)} \qquad \mathbb{E}\left[g^2(\theta, X_T)\right] \leq C(1 + |\theta|^2), \qquad \text{for all } \theta \in \mathbb{R}^q.$$

To avoid this restrictive condition, two improved versions of this routine are proposed in the literature. The first one, based on a truncation procedure called "Projection à la Chen", is introduced by Chen in [7, 6] and investigated later by several authors (see, e.g. Andrieu, Moulines and Priouret in [1] and Lelong in [17]). The use of this procedure in the context of importance sampling is initially proposed by Arouna in [2] and investigated afterward by Lapeyre and Lelong in [16]. The second alternative, is more recent and introduced by Lemaire and Pagès in [18]. In fact, they proposed an unconstrained procedure by using extensively the regularity of the involved density and they prove the convergence of this algorithm. In what follows, these two methods will be called respectively constrained and unconstrained algorithms. In view of this, a Monte Carlo method that integrates this importance sampling recursion is recommended in practice.

The aim of this paper is to study a new algorithm based on an original combination of the statistical Romberg method and the importance sampling technique. The statistical Romberg method is known for improving the Monte Carlo efficiency when used with discretization schemes and was introduced by Kebaier in [12]. However, the main term of the limit variance in the central limit theorem associated to the statistical Romberg method is quite different from that of the crude Monte Carlo method. It turns out that the optimal θ^*, in this case, is solution to the problem

$$\theta^* = \arg\min_{\theta\in\mathbb{R}^q} \tilde{\mathbb{E}}\left(g^2(\theta, X_T) + (\nabla_x g(\theta, X_T) \cdot U_T)^2\right),$$

where $(U_t)_{t\in[0,T]}$ is a given diffusion associated to the process $(X_t)_{t\in[0,T]}$ defined on an extension $\tilde{\mathcal{B}} = (\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t\geq 0}, \tilde{\mathbb{P}})$ of the initial space $\mathcal{B}$ (see further on). Here, for $\theta \in \mathbb{R}^q$ and $x \in \mathbb{R}^d$, $\nabla_x g(\theta, x)$ denotes the gradient of the function g with respect to the second variable at the point

(θ, x). Moreover, we intend to study the discretized version of this problem. More precisely, we denote X_T^n (resp. U_T^n) the Euler scheme, with time step T/n, associated to X_T (resp. U_T) and we consider the optimal θ_n^* given by

$$\theta_n^* = \arg\min_{\theta \in \mathbb{R}^q} \tilde{\mathbb{E}}\left(g^2(\theta, X_T^n) + (\nabla_x g(\theta, X_T^n) \cdot U_T^n)^2\right).$$

The convergence of θ_n^* towards θ^* as n tends to infinity is proved in the next section. In section 3 we study the problem of estimating θ_n^* using the Robbins-Monro algorithm. More preciously, we construct recursively a sequence of random variables $(\theta_i^n)_{i,n\in\mathbb{N}}$ using either the constrained or the unconstrained procedure. The aim is to prove that

$$\lim_{i,n\to\infty} \theta_i^n = \lim_{i\to\infty}(\lim_{n\to\infty} \theta_i^n) = \lim_{n\to\infty}(\lim_{i\to\infty} \theta_i^n) = \theta^*, \ \tilde{\mathbb{P}}\text{-}a.s.$$

This assertion is slightly complicated to achieve for the unconstrained procedure. In fact, for fixed $i, n \in \mathbb{N}$, the term θ_{i+1}^n constructed with this latter procedure involves $(X_{T,i+1}^{n,(-\theta_i^n)}, U_{T,i+1}^{n,(-\theta_i^n)})$, a new pair of diffusion, with drift terms containing θ_i^n. To overcome this technical difficulty we make use of the θ-sensitivity process given by $(\frac{\partial}{\partial\theta} X_T^{n,(-\theta)}, \frac{\partial}{\partial\theta} U_T^{n,(-\theta)})$ and we obtain the announced convergence result (see Theorem 3.2 and 3.3 and Corollary 3.4). In section 4, we first introduce the new adaptive algorithm obtained by combining together the importance sampling procedure and the statistical Romberg method. Then, we prove central limit theorems for both adaptive Monte Carlo method (see Theorem 4.2 and the remark below), and adaptive statistical Romberg method (see Theorem 4.3) using the Lindeberg-Feller central limit theorem for martingale array. In Section 5 we proceed to numerical simulations to illustrate the efficiency of this new method with some applications in finance. The last section is devoted to discuss some future openings.

2 General Framework

Let $X := (X_t)_{0\le t\le T}$ be the process with values in $\mathbb{R}^d$, solution to

$$dX_t = b(X_t)dt + \sum_{j=1}^{q} \sigma_j(X_t)dW_t^j, \quad X_0 = x \in \mathbb{R}^d \tag{1}$$

where $W = (W^1, \ldots, W^q)$ is a q-dimensional Brownian motion on some given filtered probability space $\mathcal{B} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge0}, \mathbb{P})$ and $(\mathcal{F}_t)_{t\ge0}$ is the standard Brownian filtration. The functions $b : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ and $\sigma_j : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, $1 \le j \le q$, satisfy condition

$$(\mathcal{H}_{b,\sigma}) \quad \forall x, y \in \mathbb{R}^d \quad |b(x) - b(y)| + \sum_{j=1}^{q} |\sigma_j(x) - \sigma_j(y)| \le C_{b,\sigma}|x - y|, \quad \text{with } C_{b,\sigma} > 0,$$

where for $x \in \mathbb{R}^d$, $|x|^2 = x.x$ stands for the Euclidean norm associated to the inner product ".". We have also $|x|^2 = x^{tr}x$ where x^{tr} denotes the transpose of x. This ensures strong existence and uniqueness of solution of (1). In many applications, in particular for the pricing of financial securities, we are interested in the effective computation by Monte Carlo methods of the quantity $\mathbb{E}\psi(X_T)$, where ψ is a given function. From a practical point of view, we have to approximate the process X by a discretization scheme. So, let us consider the Euler continuous approximation X^n with time step $\delta = T/n$ given by

$$dX_t^n = b(X_{\eta_n(t)})dt + \sum_{j=1}^{q} \sigma_j(X_{\eta_n(t)})dW_t, \quad \eta_n(t) = [t/\delta]\delta. \tag{2}$$

It is well known that under condition $(\mathcal{H}_{b,\sigma})$ we have the almost sure convergence of X^n towards X together with the following property (see e.g. Bouleau and Lépingle [5])

$$(\mathcal{P}) \;\; \forall p \geq 1, \quad \sup_{0\leq t\leq T} |X_t|, \; \sup_{0\leq t\leq T} |X_t^n| \in L^p \quad \text{ and } \;\; \mathbb{E}\left[\sup_{0\leq t\leq T} |X_t - X_t^n|^p\right] \leq \frac{K_p(T)}{n^{p/2}},$$

where $K_p(T)$ is a positive constant depending only on b, σ, T, p and q.

The weak error is firstly studied by Talay and Tubaro in [20] and now it is well known that if ψ, b and $(\sigma_j)_{1\leq j\leq q}$ are in $\mathscr{C}_P^4$, they are four times differentiable and together with their derivatives at most polynomially growing, then we have (see Theorem 14.5.1 in Kloeden and Platen in [13])

$$\varepsilon_n := \mathbb{E}\psi(X_T^n) - \mathbb{E}\psi(X_T) = O(1/n).$$

The same result was extended in Bally and Talay in [3] for a measurable function ψ but with a non degeneracy condition of Hörmander type on the diffusion. In the context of possibly degenerate diffusions, when ψ satisfies $|\psi(x) - \psi(y)| \leq C(1 + |x|^p + |y|^p)|x - y|$ for $C > 0$, $p \geq 0$, the estimate $|\varepsilon_n| \leq \frac{c}{\sqrt{n}}$ follows easily from $(\mathcal{P})$. Moreover, Kebaier in [12] proved that in addition of assumption $(\mathcal{H}_{b,\sigma})$, if b and $(\sigma_j)_{1\leq j\leq q}$ are $\mathscr{C}^1$ and ψ satisfies condition

$$(\mathcal{H}_\psi) \quad \mathbb{P}(X_T \notin \mathcal{D}_{\dot{\psi}}) = 0, \text{ where } \mathcal{D}_{\dot{\psi}} := \{x \in \mathbb{R}^d \mid \psi \text{ is differentiable at } x\}$$

then, $\lim_{n\to\infty} \sqrt{n}\,\varepsilon_n = 0$. Conversely, under the same assumptions, he shows that the rate of convergence can be $1/n^\alpha$, for any $\alpha \in (1/2, 1]$. So, it is worth to introduce assumption

$$(\mathcal{H}_{\varepsilon_n}) \quad \text{for } \alpha \in [1/2, 1] \quad n^\alpha \varepsilon_n \to C_\psi(T, \alpha), \quad C_\psi(T, \alpha) \in \mathbb{R}.$$

In order to compute the quantity $\mathbb{E}\psi(X_T^n)$, one may use the so-called statistical Romberg method, considered by [12] and which is conceptually related to the Talay-Tubaro extrapolation. This method reduces efficiently the computational complexity of the combination of Monte Carlo method and the Euler discretization scheme. In fact, the complexity in the Monte Carlo method is equal to $n^{2\alpha+1}$ and is reduced to $n^{2\alpha+1/2}$ in the statistical Romberg method. More precisely, for two numbers of discretionary time step n and m such that $m << n$, the idea of the statistical Romberg method is to use many sample paths with a coarse time discretization step $\frac{T}{m}$ and few additional sample paths with a fine time discretization step $\frac{T}{n}$. The statistical Romberg routine approximates our initial quantity of interest $\mathbb{E}\psi(X_T)$ using two empirical means

$$\frac{1}{N_1}\sum_{i=1}^{N_1} \psi(\hat{X}_{T,i}^m) + \frac{1}{N_2}\sum_{i=1}^{N_2} \psi(X_{T,i}^n) - \psi(X_{T,i}^m).$$

The random variables of the first empirical mean are independent copies of $\psi(X_T^m)$ and the random variables in the second empirical mean are also independent copies of $\psi(X_T^n) - \psi(X_T^m)$.

The associated Brownian paths $\hat{W}$ and W are independent. Under assumptions $(\mathcal{H}_\psi)$ and $(\mathcal{H}_{\varepsilon_n})$, this method is tamed by a central limit theorem with a rate of convergence equal to n^α. More precisely, for $N_1 = n^{2\alpha}$, $N_2 = n^{2\alpha-1/2}$ and $m = \sqrt{n}$ the global error normalized by n^α converges in law to a Gaussian random variable with bias equal to $C_\psi(T,\alpha)$ and a limit variance equal to

$$\mathrm{Var}\left(\psi(X_T)\right) + \tilde{\mathrm{V}}\mathrm{ar}\left(\nabla\psi(X_T)\cdot U_T\right),$$

where U is the weak limit process of the error $\sqrt{n}(X^n - X)$ defined on $\tilde{\mathcal{B}}$ an extension of the initial space $\mathcal{B}$ (see Theorem 3.2 in Kebaier [12]). More precisely, the process U is solution to

$$dU_t = \dot{b}(X_t)U_t dt + \sum_{j=1}^{q} \dot{\sigma}_j(X_t)U_t dW_t^j - \frac{1}{\sqrt{2}}\sum_{j,\ell=1}^{q} \dot{\sigma}_j(X_t)\sigma_\ell(X_t)d\tilde{W}_t^{\ell j}, \tag{3}$$

where $\tilde{W}$ is a q^2-dimensional standard Brownian motion, defined on the extension $\tilde{\mathcal{B}}$, independent of W, and $\dot{b}$ (respectively $(\dot{\sigma}_j)_{1\le j\le q}$) is the Jacobian matrix of b (respectively $(\sigma_j)_{1\le j\le q}$).

In view to use importance sampling routine, based on the Girsanov transform, we define the family of $\mathbb{P}_\theta$, as all the equivalent probability measures with respect to $\mathbb{P}$ such that

$$L_t^\theta = \frac{d\mathbb{P}_\theta}{d\mathbb{P}}|_{\mathcal{F}_t} = \exp\left(\theta\cdot W_t - \frac{1}{2}|\theta|^2 t\right).$$

Hence, $B_t^\theta := W_t - \theta t$ is a Brownian motion under $\mathbb{P}_\theta$. This leads to

$$\mathbb{E}\psi(X_T) = \mathbb{E}_\theta\left[\psi(X_T)e^{-\theta\cdot B_T^\theta - \frac{1}{2}|\theta|^2 T}\right].$$

Let us introduce the process X_t^θ solution, under $\mathbb{P}$, to

$$dX_t^\theta = \left(b(X_t^\theta) + \sum_{j=1}^{q}\theta_j\sigma_j(X_t^\theta)\right)dt + \sum_{j=1}^{q}\sigma_j(X_t^\theta)dW_t^j, \tag{4}$$

so that the process $(B_t^\theta, X_t)_{0\le t\le T}$ under $\mathbb{P}_\theta$ has the same law as $(W_t, X_t^\theta)_{0\le t\le T}$ under $\mathbb{P}$. Henceforth, we get

$$\mathbb{E}\psi(X_T) = \mathbb{E}g(\theta, X_T^\theta, W_T), \qquad \text{with } g(\theta,x,y) = \psi(x)e^{-\theta\cdot y - \frac{1}{2}|\theta|^2 T}, \forall x\in\mathbb{R}^d \text{ and } y\in\mathbb{R}^q. \tag{5}$$

We also introduce the Euler continuous approximation $X^{n,\theta}$ of the process X^θ solution, under $\mathbb{P}$, to

$$dX_t^{n,\theta} = \left(b(X_{\eta_n(t)}^{n,\theta}) + \sum_{j=1}^{q}\theta_j\sigma_j(X_{\eta_n(t)}^{n,\theta})\right)dt + \sum_{j=1}^{q}\sigma_j(X_{\eta_n(t)}^{\theta})dW_t^j.$$

Our target now is to use the statistical Romberg method introduced above to approximate $\mathbb{E}\psi(X_T) = \mathbb{E}g(\theta, X_T^\theta, W_T)$ by

$$\frac{1}{N_1}\sum_{i=1}^{N_1} g(\theta, \hat{X}_{T,i}^{m,\theta}, \hat{W}_{T,i}) + \frac{1}{N_2}\sum_{i=1}^{N_2} g(\theta, X_{T,i}^{n,\theta}, W_{T,i}) - g(\theta, X_{T,i}^{m,\theta}, W_{T,i}).$$

Of course the Brownian paths generated by $\hat{W}$ and W have to be independent. According to Theorem 3.2 of Kebaier [12] mentioned above, we have a central limit theorem with limit variance

$$\mathrm{Var}\left(g(\theta, X_T^\theta, W_T)\right) + \tilde{\mathrm{V}}\mathrm{ar}\left(\nabla_x g(\theta, X_T^\theta, W_T)\cdot U_T^\theta\right)$$

where U^θ is the weak limit process of the error $\sqrt{n}(X^{n,\theta} - X^\theta)$ defined on the extension $\tilde{\mathcal{B}}$ and solution to

$$dU_t^\theta = \left(\dot{b}(X_t^\theta) + \sum_{j=1}^q \theta_j \dot{\sigma}_j(X_t^\theta)\right) U_t^\theta dt + \sum_{j=1}^q \dot{\sigma}_j(X_t^\theta) U_t^\theta dW_t^j - \frac{1}{\sqrt{2}} \sum_{j,\ell=1}^q \dot{\sigma}_j(X_t^\theta)\sigma_\ell(X_t^\theta) d\tilde{W}_t^{\ell j}. \quad (6)$$

Therefore, it is natural to choose the optimal θ^* minimizing the associated variance. As $\mathbb{E}g(\theta, X_T^\theta, W_T) = \mathbb{E}\psi(X_T)$ and $\tilde{\mathbb{E}}(\nabla_x g(\theta, X_T^\theta, W_T)\cdot U_T^\theta) = 0$ (see Proposition 2.1 in Kebaier [12]), it boils down to choose

$$\theta^* = \operatorname*{argmin}_{\theta\in\mathbb{R}^q} v(\theta) \quad \text{with} \quad v(\theta) := \tilde{\mathbb{E}}\left(\left[\psi(X_T^\theta)^2 + (\nabla\psi(X_T^\theta)\cdot U_T^\theta)^2\right] e^{-2\theta.W_T - |\theta|^2 T}\right). \quad (7)$$

Note that from a practical point of view the quantity $v(\theta)$ is not explicit, we use the Euler scheme to discretize (X^θ, U^θ) and we choose the associated

$$\theta_n^* := \operatorname*{argmin}_{\theta\in\mathbb{R}^q} v_n(\theta) \quad \text{with} \quad v_n(\theta) := \tilde{\mathbb{E}}\left(\left[\psi(X_T^{n,\theta})^2 + (\nabla\psi(X_T^{n,\theta})\cdot U_T^{n,\theta})^2\right] e^{-2\theta.W_T - |\theta|^2 T}\right) \quad (8)$$

where $U^{n,\theta}$ is the Euler discretization scheme of U^θ, solution to

$$dU_t^{n,\theta} = \left(\dot{b}(X_{\eta_n(t)}^{n,\theta}) + \sum_{j=1}^q \theta_j \dot{\sigma}_j(X_{\eta_n(t)}^{n,\theta})\right) U_{\eta_n(t)}^{n,\theta} dt + \sum_{j=1}^q \dot{\sigma}_j(X_{\eta_n(t)}^{n,\theta}) U_{\eta_n(t)}^{n,\theta} dW_t^j - \frac{1}{\sqrt{2}} \sum_{j,\ell=1}^q \dot{\sigma}_j(X_{\eta_n(t)}^{n,\theta})\sigma_\ell(X_{\eta_n(t)}^{n,\theta}) d\tilde{W}_t^{\ell j}. \quad (9)$$

Through the whole paper, we require $\mathbb{P}(X_T \notin \mathcal{D}_{\dot{\psi}}) = 0$ and $\mathbb{P}(X_T^n \notin \mathcal{D}_{\dot{\psi}}) = 0$, $n \in \mathbb{N}$, that make (7) and (8) well posed. Also for an integer $k \geq 1$ and $\delta \in [0,1]$, we denote by $\mathscr{C}_b^{k,\delta}$ the set of functions g in $\mathscr{C}^k$ with k^{th} order partial derivatives globally δ-Hölder and all partial derivatives up to k^{th} order bounded. In case $\delta = 0$ we simply use the usual notation $\mathscr{C}_b^k$.

The following theorem yields estimates on the L^p convergence of $U^{n,\theta}$ towards U^θ. For the reader's convenience, the proof is postponed in the Appendix.

Theorem 2.1 *Let $p \geq 1$ and $\theta \in \mathbb{R}^q$. If σ and b are in $\mathscr{C}_b^1$, then both processes $\sup_{0\leq t\leq T}|U_t^\theta|$ and $\sup_{0\leq t\leq T}|U_t^{n,\theta}|$ are in L^p. Moreover, if σ and b are in $\mathscr{C}_b^{1,1}$ then we have the almost sure convergence of $U^{n,\theta}$ towards U^θ together with the following property*

$$(\tilde{\mathcal{P}}) \;\; \forall p \geq 1, \quad \sup_{0\leq t\leq T}|U_t^\theta|, \sup_{0\leq t\leq T}|U_t^{n,\theta}| \in L^p \quad \text{and} \quad \tilde{\mathbb{E}}\left[\sup_{0\leq t\leq T}|U_t^\theta - U_t^{n,\theta}|^p\right] \leq \frac{K_p^\theta(T)}{n^{p/2}},$$

where $K_p^\theta(T)$ is a positive constant depending on b, σ, θ, T, p and q. Consequently, the above results still hold for the processes U and U^n by taking $\theta = 0$.

The existence and uniqueness of θ^* is ensured by the following result.

Proposition 2.1 *Suppose σ and b are in $\mathscr{C}_b^1$ and let ψ satisfying $\mathbb{P}(\psi(X_T) \neq 0) > 0$. If there exists $a > 1$ such that $\mathbb{E}\left[\psi^{2a}(X_T)\right]$ and $\mathbb{E}\left[|\nabla\psi(X_T)|^{2a}\right]$ are finite, then the function $\theta \mapsto v(\theta)$ is $\mathscr{C}^2$ and strictly convex with $\nabla v(\theta) = \tilde{\mathbb{E}} H(\theta, X_T, U_T, W_T)$ where*

$$H(\theta, X_T, U_T, W_T) := (\theta T - W_T)\left[\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2\right] e^{-\theta\cdot W_T + \frac{1}{2}|\theta|^2 T}. \qquad (10)$$

Moreover, there exists a unique $\theta^* \in \mathbb{R}^q$ such that $min_{\theta\in\mathbb{R}^q} v(\theta) = v(\theta^*)$.

Proof: First of all, note that according to Girsanov theorem, the process (B^θ, X, U) under $\tilde{\mathbb{P}}_\theta$ has the same law as (W, X^θ, U^θ) under $\tilde{\mathbb{P}}$. So, using a change of probability, we get

$$v(\theta) := \tilde{\mathbb{E}}\left(\left[\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2\right] e^{-\theta . W_T + \frac{1}{2}|\theta|^2 T}\right).$$

The function $\theta \mapsto \left[\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2\right] e^{-\theta . W_T + \frac{1}{2}|\theta|^2 T}$ is infinitely continuously differentiable with a first derivative equal to $H(\theta, X_T, U_T, W_T)$. Note that, for $c > 0$ we have

$$\sup_{|\theta|\le c} |H(\theta, X_T, U_T, W_T)| \le (cT + |W_T|)\left[\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2\right] e^{c|W_T| + \frac{1}{2}c^2 T}.$$

Using Hölder's inequality, it is easy to check that $\tilde{\mathbb{E}} \sup_{|\theta|\le c} |H(\theta, X_T, U_T, W_T)|$ is bounded by

$$e^{\frac{1}{2}c^2 T}\left(\|\psi^2(X_T)\|_a \|e^{c|W_T|}(cT + |W_T|)\|_{\frac{a}{a-1}} + \||\nabla\psi(X_T)|^2\|_a \||U_T|^2\|_{\frac{2a}{a-1}} \|e^{c|W_T|}(cT + |W_T|)\|_{\frac{2a}{a-1}}\right).$$

Since $\mathbb{E}\psi^{2a}(X_T)$ and $\mathbb{E}|\nabla\psi(X_T)|^{2a}$ are finite we conclude, thanks to the first assertion in the above Theorem 2.1, the boundedness of $\tilde{\mathbb{E}} \sup_{|\theta|\le c} |H(\theta, X_T, U_T, W_T)|$. According to Lebesgue's theorem we deduce that v is $\mathscr{C}^1$ in $\mathbb{R}^q$ and $\nabla v(\theta) = \tilde{\mathbb{E}} H(\theta, X_T, U_T, W_T)$. In the same way, we prove that v is of class $\mathcal{C}^2$ in $\mathbb{R}^q$. So, we have

$$\mathrm{Hess}(v(\theta)) = \tilde{\mathbb{E}}\left[\left((\theta T - W_T)(\theta T - W_T)^{tr} + T I_q\right)\left(\psi^2(X_T) + (\nabla\psi(X_T)\cdot U_T)^2\right) e^{-\theta . W_T + \frac{1}{2}|\theta|^2 T}\right].$$

Since $\mathbb{P}(\psi(X_T) \neq 0) > 0$, we get for all $u \in \mathbb{R}^q \backslash \{0\}$

$$u^{tr}\ \mathrm{Hess}(v(\theta))\ u = \tilde{\mathbb{E}}\left[T|u|^2 + (u.(\theta T - W_T))^2 (\psi^2(X_T) + (\nabla\psi(X_T)\cdot U_T)^2)\ e^{-\theta . W_T + \frac{1}{2}|\theta|^2 T}\right] > 0.$$

Hence, v is strictly convex. Consequently, to prove that the unique minimum is attained for a finite value of θ, it will be sufficient to prove that $\lim_{|\theta|\to\infty} v(\theta) = +\infty$. Recall that $v(\theta) = \tilde{\mathbb{E}}\left[(\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2) e^{-\theta . W_T + \frac{1}{2}|\theta|^2 T}\right]$. Using Fatou's lemma, we get

$$\begin{aligned} +\infty &= \tilde{\mathbb{E}}\left[\liminf_{|\theta|\to\infty} (\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2) e^{-\theta . W_T + \frac{1}{2}|\theta|^2 T}\right] \\ &\le \liminf_{|\theta|\to+\infty} \tilde{\mathbb{E}}\left[(\psi(X_T)^2 + (\nabla\psi(X_T)\cdot U_T)^2) e^{-\theta . W_T + \frac{1}{2}|\theta|^2 T}\right]. \end{aligned}$$

This completes the proof. □

The same results can be obtained for the Euler scheme X^n.

Proposition 2.2 *Suppose σ and b are in $\mathscr{C}_b^1$. Given $n \in \mathbb{N}$, let ψ satisfying $\mathbb{P}(\psi(X_T^n) \neq 0) > 0$. If there exists $a > 1$ such that $\mathbb{E}[\psi^{2a}(X_T^n)]$ and $\mathbb{E}[|\nabla\psi(X_T^n)|^{2a}]$ are finite, then the function $\theta \mapsto v_n(\theta)$ is $\mathscr{C}^2$ and strictly convex with $\nabla v_n(\theta) = \tilde{\mathbb{E}} H(\theta, X_T^n, U_T^n, W_T)$. Moreover, there exists a unique $\theta_n^* \in \mathbb{R}^q$ such that $min_{\theta\in\mathbb{R}^q} v_n(\theta) = v_n(\theta_n^*)$.*

Further, we prove the convergence of θ_n^* towards θ^* as n tends to infinity.

Theorem 2.2 *Suppose σ and b are in $\mathscr{C}_b^{1,1}$. Let ψ satisfying $\mathbb{P}(\psi(X_T) \neq 0) > 0$ and $\mathbb{P}(\psi(X_T^n) \neq 0) > 0$ for all $n \in \mathbb{N}$. If there exists $a > 1$ such that $\mathbb{E}[\psi^{2a}(X_T)]$, $\sup_{n\in\mathbb{N}} \mathbb{E}[\psi^{2a}(X_T^n)]$, $\mathbb{E}[|\nabla\psi(X_T)|^{2a}]$ and $\sup_{n\in\mathbb{N}} \mathbb{E}[|\nabla\psi(X_T^n)|^{2a}]$ are finite. Then,*

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

Proof. First of all, we will prove that $(\theta_n^*)_{n\in\mathbb{N}}$ is an $\mathbb{R}^q$-bounded sequence. By way of contradiction, let us suppose that there is a subsequence $(\theta_{n_k}^*)_{k\in\mathbb{N}}$ that diverges to infinity, $\lim_{k\to\infty} |\theta_{n_k}^*| = +\infty$. This implies that on the event $\{\psi(X_T) \neq 0\}$ we have

$$\lim_{k\to\infty} \big(\psi^2(X_T^{n_k}) + (\nabla\psi(X_T^{n_k}).U_T^n)^2\big) e^{-\theta_{n_k}^* W_T + \frac{1}{2}|\theta_{n_k}^*|^2 T} = +\infty.$$

So, by Fatou's lemma we get $\lim_{k\to\infty} v_{n_k}(\theta_{n_k}^*) = +\infty$ while

$$v_{n_k}(\theta_{n_k}^*) \leq v_{n_k}(0) \leq \sup_{n\in\mathbb{N}} \mathbb{E}\left[\psi^2(X_T^n)\right] < \infty.$$

This leads to a contradiction and we deduce that there is some $M > 0$ such that $|\theta_n^*| \leq M$ for all $n \in \mathbb{N}$. Now, it remains to prove that the set $S = \{x \in \mathbb{R}^q : \theta_{n_k}^* \to x \text{ for some subsequence } \theta_{n_k}^*\}$ is reduced to the singleton set $\{\theta^*\}$. Let us consider a subsequence $\theta_{n_k}^* \to \theta_\infty^* \in S$ as k tends to infinity. According to Proposition 2.2 above, we have

$$\nabla v_{n_k}(\theta_{n_k}^*) = \tilde{\mathbb{E}}\left[(\theta_{n_k}^* T - W_T)\big(\psi^2(X_T^{n_k}) + (\nabla\psi(X_T^{n_k}).U_T^{n_k})^2\big) e^{-\theta_{n_k}^*.W_T + \frac{1}{2}|\theta_{n_k}^*|^2 T}\right] = 0.$$

Now, let $1 < \tilde{a} < a$, using the relation $|x+y|^{\tilde{a}} \leq 2^{\tilde{a}-1}(|x|^{\tilde{a}} + |y|^{\tilde{a}})$ and applying Hölder's inequality twice with the boundedness of $\theta_{n_k}^*$ established in the first part of the proof we check easily that there exists $c_1 > 0$ depending on a, T and M such that

$$\tilde{\mathbb{E}}\left[\left|(\theta_{n_k}^* T - W_T)\big(\psi^2(X_T^{n_k}) + (\nabla\psi(X_T^{n_k}).U_T^{n_k})^2\big) e^{-\theta_{n_k}^*.W_T + \frac{1}{2}|\theta_{n_k}^*|^2 T}\right|^{\tilde{a}}\right] \leq$$
$$c_1 \left\{ \|\psi^2(X_T^{n_k})\|_a^{\tilde{a}} + \||\nabla\psi(X_T^{n_k})|^2\|_a^{\tilde{a}} \||U_T^{n_k}|^{2\tilde{a}}\|_{\frac{2a}{a-\tilde{a}}} \right\}.$$

Thanks to our assumptions $\sup_{n\in\mathbb{N}} \mathbb{E}[\psi^{2a}(X_T^n)] < \infty$ and $\sup_{n\in\mathbb{N}} \mathbb{E}[|\nabla\psi(X_T^n)|^{2a}] < \infty$ and Theorem 2.1, we get the uniform integrability. Therefore, using the almost sure convergence of $\psi^2(X_T^n)$, $\nabla\psi(X_T^n)$ and U_T^n respectively towards $\psi^2(X_T)$, $\nabla\psi(X_T)$ and U_T which is ensured by $(\mathcal{P})$, $(\tilde{\mathcal{P}})$ and $\mathbb{P}(X_T \notin \mathcal{D}_{\dot{\psi}}) = 0$, we obtain

$$\nabla v(\theta_\infty^*) = \tilde{\mathbb{E}}\left[(\theta_\infty^* T - W_T)\big(\psi^2(X_T) + (\nabla\psi(X_T).U_T)^2\big) e^{-\theta_\infty^*.W_T + \frac{1}{2}|\theta_\infty^*|^2 T}\right] = 0.$$

We complete the proof using the uniqueness of the minimum ensured by Proposition 2.1. □

3 Robbins-Monro Algorithms

The aim now is to construct for fixed n some sequences $(\theta_i^n)_{i\in\mathbb{N}}$ such that $\lim_{i\to\infty}\theta_i^n=\theta_n^*$ almost surely. It is well known that stochastic algorithms can be used to answer this issue and find an accurate approximation of $\theta_n^*=\arg\min_{\theta\in\mathbb{R}} v_n(\theta)$. Indeed, using the Robbins-Monro algorithm, we construct recursively the sequence of random variables $(\theta_i^n)_{i\in\mathbb{N}}$ in $\mathbb{R}^q$ given by

$$\theta_{i+1}^n=\theta_i^n-\gamma_{i+1}H(\theta_i^n,X_{T,i+1}^n,U_{T,i+1}^n,W_{T,i+1}),\ i\geq 0,\ \theta_0^n\in\mathbb{R}^q, \tag{11}$$

where H is given by relation (10), the gain sequence $(\gamma_i)_{i\geq 1}$ is a decreasing sequence of positive real numbers satisfying

$$\sum_{i=1}^{\infty}\gamma_i=\infty \quad\text{and}\quad \sum_{i=1}^{\infty}\gamma_i^2<\infty. \tag{12}$$

Here $(X_{T,i}^n,U_{T,i}^n)_{i\geq1}$ is a sequence of independent copies of the Euler scheme associated to (X_T^n,U_T^n) adapted to the filtration $\tilde{\mathcal{F}}_{T,i}=\sigma(W_{t,l},\tilde{W}_{t,l},l\leq i,t\leq T)$, where $(W_i,\tilde{W}_i)_{i\geq1}$ are independent copies of the pair $(W,\tilde{W})$ introduced before in Equation (3). To obtain the almost sure convergence of the above algorithm to $\theta_n^*=\arg\min_{\theta\in\mathbb{R}} v_n(\theta)$, we need to check a first condition: $\forall\theta\neq\theta_n^*$, $\langle\nabla v_n(\theta),\theta-\theta_n^*\rangle>0$, which is satisfied in our context thanks to the convexity property of v_n. Secondly we need also a sub-quadratic assumption known as the non explosion condition

$$\text{(NEC)}\qquad \tilde{\mathbb{E}}\left[|H(\theta,X_T^n,U_T^n,W_T)|^2\right]\leq C(1+|\theta|^2),\qquad \text{for all } \theta\in\mathbb{R}^q.$$

Unfortunately, this condition is not satisfied in our context and we will study two different stochastic algorithms using the Robbins-Monro procedure and avoiding the above restriction.

3.1 Constrained stochastic algorithm

The idea of the "Projection à la Chen" is to kill the classic Robbins-Monro procedure when it goes close to explosion and to restart it with a smaller step sequence. This can be described as some repeated truncations when the algorithm leaves a slowly growing compact set waiting for stabilization. Then, the algorithm behaves like the Robbins-Monro algorithm. Formally, for a fixed number of discretization time step $n\geq1$, the repeated truncations can be written in our context as follows. Let $(\mathcal{K}_i)_{i\in\mathbb{N}}$ denote an increasing sequence of compact sets satisfying $\cup_{i=0}^{\infty}\mathcal{K}_i=\mathbb{R}^d$ and $\mathcal{K}_i\subsetneq\mathring{\mathcal{K}}_{i+1},\forall i\in\mathbb{N}$. For $\theta_0^n\in\mathcal{K}_0$, $\alpha_0^n=0$ and a gain sequence $(\gamma_i)_{i\in\mathbb{N}}$ satisfying (12), we define the sequence $(\theta_i^n,\alpha_i^n)_{i\in\mathbb{N}}$ recursively by

$$\begin{cases}\textbf{if} & \theta_i^n-\gamma_{i+1}H(\theta_i^n,X_{T,i+1}^n,U_{T,i+1}^n,W_{T,i+1})\in\mathcal{K}_{\alpha_i^n},\ \textbf{then}\\ & \theta_{i+1}^n=\theta_i^n-\gamma_{i+1}H(\theta_i^n,X_{T,i+1}^n,U_{T,i+1}^n,W_{T,i+1}),\ \text{and } \alpha_{i+1}^n=\alpha_i^n\\ \textbf{else} & \theta_{i+1}^n=\theta_0^n \text{ and } \alpha_{i+1}^n=\alpha_i^n+1,\end{cases} \tag{13}$$

where the function H is given above in relation (10). For $i\in\mathbb{N}$, α_i^n represents the number of truncations of the first i iterations. In fact, as we can see, if the $(i+1)^{th}$ iteration of the Robbins-Monro is in the compact set $\mathcal{K}_{\alpha_i^n}$, then the algorithm will behave like a regular Robbins-Monro. However, if the $(i+1)^{th}$ iteration outside the compact set $\mathcal{K}_{\alpha_i^n}$, it will be reinitialized. Then, we increase the domain of projection, so we consider the new compact set $\mathcal{K}_{\alpha_i^n+1}$.

Theorem 3.1 *Suppose σ and b are in $\mathscr{C}_b^1$. Assume that for all $n \in \mathbb{N}$, $\mathbb{P}(\psi(X_T^n) \neq 0) > 0$ and there exists $a > 1$ such that $\mathbb{E}\left[\psi^{4a}(X_T^n)\right]$ and $\mathbb{E}\left[|\nabla\psi(X_T^n)|^{4a}\right]$ are finite, then the sequence $(\theta_i^n)_{i\geq 0}$ given by routine (13), satisfies*

1. *For all $n \in \mathbb{N}$, we have $\theta_i^n \underset{i\to\infty}{\longrightarrow} \theta_n^*$, almost surely where θ_n^* is given by relation (8).*
2. *Reversely, for all $i \in \mathbb{N}$, we have $\theta_i^n \underset{n\to\infty}{\longrightarrow} \theta_i$, almost surely, where the sequence $(\theta_i)_{i\geq 0}$ is obtained by replacing in routine (13), $(X_{T,i}^n, U_{T,i}^n)$ by their limit $(X_{T,i}, U_{T,i})$, $i \geq 1$.*

Proof: At the beginning, note that for $n \in \mathbb{N}$ the existence of θ_n^* is ensured by Proposition 2.2. Concerning, the first assertion, we have to check both assumptions of Theorem 3.1 in [16]. The first one given by

$$\forall \theta \neq \theta_n^*, \quad \langle \nabla v_n(\theta), \theta - \theta_n^* \rangle > 0,$$

is satisfied in our context thanks to the convexity property of v_n. So, it remains to check the second assumption given by

$$\forall c > 0, \quad \sup_{|\theta|\leq c} \tilde{\mathbb{E}}\left[|H(\theta, X_T^n, U_T^n, W_T)|^2\right] < \infty. \tag{14}$$

This assumption relaxes the usual (NEC) condition on function H used to run the Robbins-Monro algorithm. Let $c > 0$, we have

$$\sup_{|\theta|\leq c} |H(\theta, X_T^n, U_T^n, W_T)|^2 \leq 2(cT + |W_T|)^2 \left[\psi(X_T^n)^4 + (\nabla\psi(X_T^n)\cdot U_T^n)^4\right] e^{2c|W_T|+c^2T}.$$

Using several times Hölder's inequality together with property $(\tilde{\mathcal{P}})$, it is easy to check assumption (14), since $\mathbb{E}\psi^{4a}(X_T^n)$ and $\mathbb{E}|\nabla\psi(X_T^n)|^{4a}$ are finite.

The second assertion follows easily by induction on (θ_i^n, α_i^n), using that for all $i \geq 1$, the pair $(X_{T,i}^n, U_{T,i}^n)$ converges almost surely to $(X_{T,i}, U_{T,i})$ combined with the assumption $\mathbb{P}(X_T \notin \mathcal{D}_{\dot{\psi}}) = 0$. □

Now, by replacing (X_T^n, U_T^n) by their limit (X_T, U_T) in the proof of the first assertion above, we easily get the following result.

Corollary 3.1 *Suppose σ and b are in $\mathscr{C}_b^1$. Assume that $\mathbb{P}(\psi(X_T) \neq 0) > 0$ and there exists $a > 1$ such that $\mathbb{E}\left[\psi^{4a}(X_T)\right]$ and $\mathbb{E}\left[|\nabla\psi(X_T)|^{4a}\right]$ are finite, then the sequence $(\theta_i)_{i\geq 0}$ introduced in the above theorem satisfies*

$$\theta_i \underset{i\to\infty}{\longrightarrow} \theta^* \quad a.s.,$$

where θ^* is given by relation (7).

The following corollary follows immediately thanks to theorems 2.2 and 3.1 and Corollary 3.1.

Corollary 3.2 *Under assumptions of Theorem 2.2, Theorem 3.1 and Corollary 3.1, the constrained algorithm given respectively by routine (13) satisfies*

$$\lim_{i,n\to\infty} \theta_i^n = \lim_{i\to\infty}\big(\lim_{n\to\infty} \theta_i^n\big) = \lim_{n\to\infty}\big(\lim_{i\to\infty} \theta_i^n\big) = \theta^*, \quad \tilde{\mathbb{P}}\text{-}a.s.,$$

where θ^* is given by relation (7).

3.2 Unconstrained stochastic algorithm

In their recent paper [18], Lemaire and Pagès proposed a new procedure using Robbins-Monro algorithm that satisfies the classical non explosion condition (NEC). In fact, a new expression of the gradient is obtained by a third change of probability. Recall that by Proposition 2.2 we have

$$\nabla v_n(\theta) = \tilde{\mathbb{E}}\left((\theta T - W_T)\left[\psi(X_T^n)^2 + (\nabla\psi(X_T^n)\cdot U_T^n)^2\right] e^{-\theta\cdot W_T + \frac{1}{2}|\theta|^2 T}\right).$$

The aim now is to use their idea in our context. To do so, we apply Girsanov theorem, with the shift parameter $-\theta$. Let $B_t^{(-\theta)} := W_t + \theta t$ and $L_t^{(-\theta)} := \frac{d\mathbb{P}_{(-\theta)}}{d\mathbb{P}}|_{\mathcal{F}_t} = e^{-\theta . W_t - \frac{1}{2}|\theta|^2 t}$, we obtain

$$\nabla v_n(\theta) = \tilde{\mathbb{E}}_{(-\theta)}\left[(2\theta T - B_T^{(-\theta)})\left[\psi(X_T^n)^2 + (\nabla\psi(X_T^n)\cdot U_T^n)^2\right] e^{|\theta|^2 T}\right].$$

As $(B^{(-\theta)}, X^n, U^n)$ under $\tilde{\mathbb{P}}_{(-\theta)}$ has the same law as $(W, X^{n,(-\theta)}, U^{n,(-\theta)})$ under $\tilde{\mathbb{P}}$, we write

$$\nabla v_n(\theta) = \tilde{\mathbb{E}}\left[(2\theta T - W_T)\left[\psi(X_T^{n,(-\theta)})^2 + (\nabla\psi(X_T^{n,(-\theta)})\cdot U_T^{n,(-\theta)})^2\right] e^{|\theta|^2 T}\right].$$

Miming the algorithm proposed by [18], we introduce for a given $\eta > 0$, a new function

$$\tilde{H}_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T) = e^{-\eta|\theta|^2 T}(2\theta T - W_T)\left[\psi(X_T^{n,(-\theta)})^2 + (\nabla\psi(X_T^{n,(-\theta)})\cdot U_T^{n,(-\theta)})^2\right].$$

Then, we introduce for a gain sequence $(\gamma_i)_{i\in\mathbb{N}}$ satisfying (12), the algorithm

$$\theta_{i+1}^n = \theta_i^n - \gamma_{i+1}\tilde{H}_\eta(\theta_i^n, X_{T,i+1}^{n,(-\theta_i^n)}, U_{T,i+1}^{n,(-\theta_i^n)}, W_{T,i+1}), \quad \theta_0 \in \mathbb{R}^q. \tag{15}$$

This algorithm would behave like a classical Robbins-Monro one and does not suffer from the violation of (NEC). Our aim now is to establish the same results satisfied by the constrained routine (13) and given by Theorem 3.1. This is splitted into two different theorems. It is worth to note that in this context we will need to control the growth of ψ and its derivatives.

Theorem 3.2 *Suppose σ and b are in $\mathscr{C}_b^1$ and let ψ satisfying $\mathbb{P}(\psi(X_T^n) \neq 0) > 0$, for all $n \in \mathbb{N}$. In addition, assume that for $\lambda > 0$ we have*

$$|\nabla\psi(x)| \leq C_\psi(1 + |x|^\lambda) \quad \text{for all } x \in \mathcal{D}_{\dot{\psi}} \text{ and } C_\psi > 0.$$

Then, the sequence $(\theta_i^n)_{i\geq 0}$ given by routine (15), satisfies

$$\forall n \in \mathbb{N}, \quad \theta_i^n \underset{i\to\infty}{\longrightarrow} \theta_n^*, \quad a.s.$$

where θ_n^* is given by relation (8).

Proof: To prove the almost sure convergence we will use the classical Robbins-Monro theorem (see Theorem 2.2.12 page 52 in [8]). Let $n \in \mathbb{N}$, under our assumptions the existence of θ_n^* is ensured by Proposition 2.2 and we have to check first that

$$\forall\theta \neq \theta_n^* \quad \langle h_n(\theta), \theta - \theta_n^*\rangle > 0, \quad \text{where} \quad h_n(\theta) = \tilde{\mathbb{E}}\, H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T).$$

This is immediate since $h_n(\theta) = K_\eta(\theta)\nabla v_n(\theta)$ with $K_\eta > 0$ and v_n is a strictly convex function. Now it remains to prove that $\sup_{\theta\in\mathbb{R}^q} \tilde{\mathbb{E}}\left[|H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T)|^2\right] < \infty$, which guaranties the (NEC) condition. By Cauchy-Schwartz inequality we obtain

$$\tilde{\mathbb{E}}\left[|H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T)|^2\right] \le e^{-2\eta|\theta|^2 T} \left\||2\theta T - W_T|^2\right\|_2 \times \left(\left\|\psi(X_T^{n,(-\theta)})^2\right\|_2 + \left\|(\nabla\psi(X_T^{n,(-\theta)}) \cdot U_T^{n,(-\theta)})^2\right\|_2\right).$$

Using the polynomial growth assumption on $\nabla\psi$, the second and third term on the right hand side of the above inequality can be bounded respectively up to a standard positive constant by

$$1 + \left\||X_T^{n,(-\theta)}|^{2(\lambda+1)}\right\|_2 \text{ and } 1 + \left\||X_T^{n,(-\theta)}|^{4\lambda}\right\|_2 + \left\||U_T^{n,(-\theta)}|^4\right\|_2.$$

In the following proof, C will denote a positive standard constant that may change from line to line. Let $\lambda_1 = 4\lambda \vee 2(\lambda + 1)$, using the identity $(1 + x)^\rho \le C(1 + x^\rho)$ for $x \ge 0$ and $\rho \ge 1$, then we have

$$\tilde{\mathbb{E}}\left[|H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T)|^2\right] \le Ce^{-2\eta|\theta|^2 T}(1 + |\theta|^2)\left(1 + \left\||X_T^{n,(-\theta)}|^{\lambda_1}\right\|_2 + \left\||U_T^{n,(-\theta)}|^4\right\|_2\right).$$

As $(B^{(-\theta)}, X^n, U^n)$ under $\tilde{\mathbb{P}}_{(-\theta)}$ has the same law as $(W, X^{n,(-\theta)}, U^{n,(-\theta)})$ under $\tilde{\mathbb{P}}$, we write

$$\tilde{\mathbb{E}}\big|X_T^{n,(-\theta)}\big|^{2\lambda_1} = \tilde{\mathbb{E}}_{(-\theta)}\big|X_T^n\big|^{2\lambda_1} = \tilde{\mathbb{E}}\left(\big|X_T^n\big|^{2\lambda_1} e^{-\theta\cdot W_T - \frac{1}{2}|\theta|^2 T}\right)$$

and

$$\tilde{\mathbb{E}}\big|U_T^{n,(-\theta)}\big|^8 = \tilde{\mathbb{E}}_{(-\theta)}\big|U_T^n\big|^8 = \tilde{\mathbb{E}}\left(\big|U_T^n\big|^8 e^{-\theta\cdot W_T - \frac{1}{2}|\theta|^2 T}\right).$$

Now using Hölder's inequality, with $\frac{1}{r} + \frac{1}{r'} = 1$, properties $(\mathcal{P})$ and $(\tilde{\mathcal{P}})$ and $\left(\tilde{\mathbb{E}}e^{-r\theta\cdot W_T - \frac{r}{2}|\theta|^2 T}\right)^{\frac{1}{r}} = e^{\frac{r-1}{2}|\theta|^2 T}$, we obtain

$$\tilde{\mathbb{E}}\left[|H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T)|^2\right] \le C(1 + |\theta|^2)e^{-(2\eta - \frac{r-1}{4})|\theta|^2 T}.$$

Then, one sees immediately that $\sup_{\theta\in\mathbb{R}^q} \tilde{\mathbb{E}}\left[|H_\eta(\theta, X_T^{n,(-\theta)}, U_T^{n,(-\theta)}, W_T)|^2\right]$ is finite by choosing $r \in (1, 1 + 8\eta)$. This completes the proof. □

In the same way as in the constrained case, we deduce the following result if we replace (X_T^n, U_T^n) by their limit (X_T, U_T) in the above proof.

Corollary 3.3 *Suppose σ and b are in $\mathscr{C}_b^1$. Let ψ satisfying $\mathbb{P}(\psi(X_T) \neq 0) > 0$ and*

$$|\nabla\psi(x)| \le C_\psi(1 + |x|^\lambda) \quad \textit{for all } x \in \mathcal{D}_{\dot\psi} \textit{ and } C_\psi, \lambda > 0.$$

Then, the sequence $(\theta_i)_{i\ge0}$, obtained when replacing in routine (15) $(X_{T,i}^n, U_{T,i}^n)_{i\ge1}$ by their limit $(X_{T,i}, U_{T,i})_{i\ge1}$, satisfies

$$\theta_i \underset{i\to\infty}{\longrightarrow} \theta^*, \quad a.s.$$

where θ^* is given by relation (7).

The aim now is to prove that the same property *2.* in Theorem 3.1, is satisfied by the unconstrained algorithm (15). This task looks more complicated to achieve, since for a fixed $i \geq 0$ the stochastic term θ_i^n also appears in the drift part of the pair $(X_{T,i+1}^{n,(-\theta_i^n)}, U_{T,i+1}^{n,(-\theta_i^n)})$. To overcome this technical difficulty we firstly strengthen our hypothesis on the triplet (b, σ, ψ) and secondly make use of the so called θ-sensitivity process given by $(\frac{\partial}{\partial\theta} X_T^{n,(-\theta)}, \frac{\partial}{\partial\theta} U_T^{n,(-\theta)})$.

Theorem 3.3 *Let b and σ in $\mathscr{C}_b^{2,\delta}$, $\delta > 0$. Assume that ψ is $\mathscr{C}^2$ with polynomial growth as well as all its partial derivatives until order two and satisfies $\mathbb{P}(\psi(X_T) \neq 0) > 0$ and $\mathbb{P}(\psi(X_T^n) \neq 0) > 0$, for all $n \geq 1$. Then, $\forall i \in \mathbb{N}$ and $\forall p \geq 1$, there exists $C > 0$ depending only on i, p, b, σ and T such that*

$$\forall n \in \mathbb{N}^*, \quad \tilde{\mathbb{E}}|\theta_i^n - \theta_i|^{2p} \leq \frac{C}{n^p}.$$

Consequently, $\forall i \in \mathbb{N}$ $\theta_i^n \underset{n\to\infty}{\longrightarrow} \theta_i, a.s.$ where the sequence $(\theta_i)_{i\geq 0}$ is introduced in the above corollary.

Proof. We first proceed by induction on $i \in \mathbb{N}$ to prove the first assertion. The case when $i = 0$ is trivial since $\theta_0^n = \theta_0 \in \mathbb{R}^q$. We now assume the assertion holds for a fixed integer i and show that it also holds for $i + 1$. First, we write $\theta_{i+1}^n - \theta_{i+1} = \theta_i^n - \theta_i - \gamma_{i+1}(H_1 + H_2)$ where

$$H_1 := e^{-\eta|\theta_i^n|^2 T}(2\theta_i^n T - W_{T,i+1}) \\ \times \left[\psi(X_{T,i+1}^{n,(-\theta_i^n)})^2 - \psi(X_{T,i+1}^{(-\theta_i^n)})^2 + \left(\nabla\psi(X_{T,i+1}^{n,(-\theta_i^n)}).U_{T,i+1}^{n,(-\theta_i^n)}\right)^2 - \left(\nabla\psi(X_{T,i+1}^{(-\theta_i^n)}).U_{T,i+1}^{(-\theta_i^n)}\right)^2\right]$$

and

$$H_2 := e^{-\eta|\theta_i^n|^2 T}(2\theta_i^n T - W_{T,i+1}) \left[\psi(X_{T,i+1}^{(-\theta_i^n)})^2 + \left(\nabla\psi(X_{T,i+1}^{(-\theta_i^n)}).U_{T,i+1}^{(-\theta_i^n)}\right)^2\right] \\ - e^{-\eta|\theta_i|^2 T}(2\theta_i T - W_{T,i+1}) \left[\psi(X_{T,i+1}^{(-\theta_i)})^2 + \left(\nabla\psi(X_{T,i+1}^{(-\theta_i)}).U_{T,i+1}^{(-\theta_i)}\right)^2\right].$$

Hence, for all $p \geq 1$, we have

$$\tilde{\mathbb{E}}|\theta_{i+1}^n - \theta_{i+1}|^{2p} \leq 3^{2p-1}\tilde{\mathbb{E}}|\theta_i^n - \theta_i|^{2p} + 3^{2p-1}\gamma_{i+1}^{2p}(\tilde{\mathbb{E}}|H_1|^{2p} + \tilde{\mathbb{E}}|H_2|^{2p}). \tag{16}$$

Using the induction assumption we only need to control the second and third terms on the right hand side of the inequality (16) above.

Term H_1 Using that θ_i^n is $\tilde{\mathcal{F}}_{T,i}$-measurable, $W_{T,i+1} \perp\!\!\!\perp \tilde{\mathcal{F}}_{T,i}$ we write $\tilde{\mathbb{E}}|H_1|^{2p} = \tilde{\mathbb{E}}A(\theta_i^n)$ where for all $\theta \in \mathbb{R}^q$

$$A(\theta) := e^{-2p\eta|\theta|^2 T}\tilde{\mathbb{E}}\Big[|2\theta T - W_T|^{2p} \\ \times |\psi(X_T^{n,(-\theta)})^2 - \psi(X_T^{(-\theta)})^2 + (\nabla\psi(X_T^{n,(-\theta)}).U_T^{n,(-\theta)})^2 - (\nabla\psi(X_T^{(-\theta)}).U_T^{(-\theta)})^2|^{2p}\Big].$$

Since $(B^{(-\theta)}, X^n, U^n, X, U)$ under $\tilde{\mathbb{P}}_{(-\theta)}$ has the same law as $(W, X^{n,(-\theta)}, U^{n,(-\theta)}, X^{(-\theta)}, U^{(-\theta)})$ under $\tilde{\mathbb{P}}$ for all $\theta \in \mathbb{R}$, we obtain by a change of probability measure

$$A(\theta) = e^{(-2p\eta - \frac{1}{2})|\theta|^2 T} \tilde{\mathbb{E}}\Big[|\theta T - W_T|^{2p} e^{-\theta.W_T} \\ \times |\psi(X_T^n)^2 - \psi(X_T)^2 + (\nabla\psi(X_T^n).U_T^n)^2 - (\nabla\psi(X_T).U_T)^2 \,|^{2p}\Big].$$

By Hölder's inequality, we obtain $\forall r_1 \in (1, \infty)$,

$$A(\theta) \leq e^{(-2p\eta - \frac{1}{2})|\theta|^2 T} \|e^{-\theta.W_T}\|_{r_1} \||\theta T - W_T|^{2p}\|_{\frac{2r_1}{r_1-1}} \\ \times \||\psi(X_T^n)^2 - \psi(X_T)^2 + (\nabla\psi(X_T^n).U_T^n)^2 - (\nabla\psi(X_T).U_T)^2|^{2p}\|_{\frac{2r_1}{r_1-1}}.$$

As $e^{(-2p\eta - \frac{1}{2})|\theta|^2 T} \|e^{-\theta.W_T}\|_{r_1} \||\theta T - W_T|^{2p}\|_{\frac{2r_1}{r_1-1}} \leq c_1(1 + |\theta|^{2p}) e^{(\frac{r_1}{2} - 2p\eta - \frac{1}{2})|\theta|^2 T}$, with c_1 is a positive constant depending only on p, r_1 and T, one can choose $r_1 \in (1, 1 + 4p\eta)$ such that $\sup_{\theta \in \mathbb{R}^q}(1 + |\theta|^{2p}) e^{(\frac{r_1}{2} - 2p\eta - \frac{1}{2})|\theta|^2 T}$ is finite. Hence, we get the existence of a constant c_2 depending only on p, η and T such that

$$A(\theta) \leq c_2 \||\psi(X_T^n)^2 - \psi(X_T)^2 + (\nabla\psi(X_T^n).U_T^n)^2 - (\nabla\psi(X_T).U_T)^2|^{2p}\|_{\frac{2r_1}{r_1-1}}. \tag{17}$$

Since ψ is $\mathscr{C}^2$ with polynomial growth as well as all its partial derivatives until order two then the function $g(x, y) := \psi^2(x) + \nabla\psi(x).y$, $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, is $\mathscr{C}^1$ and all its partial derivatives of order one have polynomial growth. Hence, the Taylor expansion on the real-valued function g yields the existence of a point $(\bar{X}_T^n, \bar{U}_T^n)$ between (X_T^n, U_T^n) and (X_T, U_T) such that

$$g(X_T^n, U_T^n) - g(X_T, U_T) = \nabla g(\bar{X}_T^n, \bar{U}_T^n).(X_T^n - X_T, U_T^n - U_T).$$

Then by the Cauchy-Schwarz inequality, the polynomial growth of g and properties $(\mathcal{P})$ and $(\tilde{\mathcal{P}})$ we get the existence of a constant c_3 depending only on p, η, T, b, σ and ψ such that

$$\begin{aligned} \||g(X_T^n, U_T^n) - g(X_T, U_T)|^{2p}\|_{\frac{2r_1}{r_1-1}} &\leq \||\nabla g(\bar{X}_T^n, \bar{U}_T^n)|^{2p}\|_{\frac{4r_1}{r_1-1}} \||(X_T^n - X_T, U_T^n - U_T)|^{2p}\|_{\frac{4r_1}{r_1-1}} \\ &\leq \frac{c_3}{n^p}. \end{aligned} \tag{18}$$

So, (17) and (18) tell us $A(\theta) \leq \frac{c_2 c_3}{n^p}$, and we deduce the existence of a deterministic constant c_4 depending only on p, η, T, b, σ and ψ such that

$$\tilde{\mathbb{E}} H_1^{2p} \leq \frac{c_4}{n^p}. \tag{19}$$

Term H_2 Using that θ_i^n and θ_i are $\tilde{\mathcal{F}}_{T,i}$-measurable, $W_{T,i+1} \perp\!\!\!\perp \tilde{\mathcal{F}}_{T,i}$ we write $\tilde{\mathbb{E}}|H_2|^{2p} = \tilde{\mathbb{E}} B(\theta_i^n, \theta_i)$ where for all $(\theta, \theta') \in \mathbb{R}^q \times \mathbb{R}^q$

$$B(\theta, \theta') := \tilde{\mathbb{E}}|e^{-\eta|\theta|^2 T}(2\theta T - W_T) g(X_T^{(-\theta)}, U_T^{(-\theta)}) - e^{-\eta|\theta'|^2 T}(2\theta' T - W_T) g(X_T^{(-\theta')}, U_T^{(-\theta')})|^{2p}. \tag{20}$$

According to the study of θ-sensitivity of the processes $(X_t^{(-\theta)})_{t \in [0,T]}$ and $(U_t^{(-\theta)})_{t \in [0,T]}$ given in lemma 3.1 below, we have that for a time $t \in [0, T]$ the function $\theta \mapsto (X_t^{(-\theta)}, U_t^{(-\theta)})$ is almost surely $\mathscr{C}^1$. Hence, we deduce that almost surely the function $\theta \mapsto D(\theta) := e^{-\eta|\theta|^2 T}(2\theta T - W_T)g(X_T^{(-\theta)}, U_T^{(-\theta)})$ is also $\mathscr{C}^1$. This allows us to apply Taylor expansion on each component $D^{\ell'}$ of D, $\ell' \in \{1, \cdots, q\}$, and by standard evaluations we obtain a constant c_5 depending only on p and q such that

$$\begin{aligned} B(\theta,\theta') &= \tilde{\mathbb{E}}|\sum_{\ell'=1}^{q}\big(D^{\ell'}(\theta) - D^{\ell'}(\theta')\big)^2|^p = \tilde{\mathbb{E}}|\sum_{\ell'=1}^{q}\big(\sum_{\ell=1}^{q}(\theta_\ell - \theta'_\ell)\int_0^1 \frac{\partial D^{\ell'}}{\partial \theta_\ell}(t\theta' + (1-t)\theta)dt\big)^2|^p \\ &\leq c_5|\theta - \theta'|^{2p}\sum_{\ell,\ell'=1}^{q}\int_0^1 \tilde{\mathbb{E}}|\frac{\partial D^{\ell'}}{\partial \theta_\ell}(t\theta' + (1-t)\theta)|^{2p}dt. \end{aligned}$$

The term $\tilde{\mathbb{E}}|\frac{\partial D^{\ell'}}{\partial u_\ell}(u)|^{2p}$ is bounded uniformly on $u \in \mathbb{R}^q$. More precisely, we have the following result.

Lemma 3.1 *The solutions $(X_t^{(-\theta)})_{t\in[0,T]}$ and $(U_t^{(-\theta)})_{t\in[0,T]}$ of respectively Itô's stochastic differential equations (4) and (6) have modifications of $\mathscr{C}^1$ with respect to the parameter θ and their partial derivatives are L^p-bounded for all $p \geq 1$. Further, there exists a positive constant c depending only on p, q, b, σ, ψ and T such that*

$$\tilde{\mathbb{E}}|\frac{\partial D^{\ell'}}{\partial u_\ell}(u)|^{2p} \leq c \quad \forall u \in \mathbb{R}^q \text{ and } \ell, \ell' \in \{1, \cdots, q\}.$$

For the reader convenience, the proof of this lemma is postponed to the end of the current subsection. Thus, thanks to Lemma 3.1 above there is a constant c_6 depending only on p, q, b, σ, ψ and T such that $B(\theta,\theta') \leq c_6|\theta - \theta'|^{2p}$, and it follows from $\tilde{\mathbb{E}}|H_2|^{2p} = \tilde{\mathbb{E}}B(\theta_i^n, \theta_i)$ that

$$\tilde{\mathbb{E}}H_2^{2p} \leq c_6\mathbb{E}|\theta_i^n - \theta_i|^{2p}. \tag{21}$$

So, (16), (19) and (21) show that

$$\tilde{\mathbb{E}}|\theta_{i+1}^n - \theta_{i+1}|^{2p} \leq 3^{2p-1}(1 + c_6\gamma_{i+1}^{2p})\tilde{\mathbb{E}}|\theta_i^n - \theta_i|^{2p} + 3^{2p-1}\gamma_{i+1}^{2p}\frac{c_4}{n^p}.$$

Using the induction assumption for stage i, we deduce for $p \geq 1$ the existence of a positive constant C depending only on p, q, b, σ, ψ, T and i such that

$$\forall n \in \mathbb{N}^*, \quad \tilde{\mathbb{E}}|\theta_{i+1}^n - \theta_{i+1}|^{2p} \leq \frac{C}{n^p}.$$

Finally, for all $i \in \mathbb{N}$, the almost sure convergence, of θ_i^n towards θ_i as n tends to ∞ is a classical and immediate consequence of the first assertion shown above, based on the Borel-Cantelli lemma. □

The following corollary follows immediately thanks to theorems 2.2, 3.2 and 3.3 and Corollary 3.3.

Corollary 3.4 *Under assumptions of Theorem 3.3 and $\mathbb{P}(\psi(X_T) \neq 0) > 0$, the unconstrained algorithm given respectively by routine (15) satisfies*

$$\lim_{i,n\to\infty}\theta_i^n = \lim_{i\to\infty}\big(\lim_{n\to\infty}\theta_i^n\big) = \lim_{n\to\infty}\big(\lim_{i\to\infty}\theta_i^n\big) = \theta^*, \quad \tilde{\mathbb{P}}\text{-a.s.},$$

where θ^* is given by relation (7).

Our task now is to show the result given by Lemma 3.1 and used in the proof of Theorem 3.3.

Proof of Lemma 3.1 It is worth to note that all theoretical results known on the differentiation of the solution of Itô's stochastic differential equation with respect to its initial value, can be extended to any parameter. Thus, thanks to Theorem 4.6.5 in [14], our assumptions on b and σ ensures the differentiability of the processes $(X_t^{(-u)})_{0\le t\le T}$. Further, if we denote by $\partial_\ell X_t^{(-u)}$ the processes where we take the partial derivatives of all components of $(X_t^{(-u)})_{0\le t\le T}$ with respect to the ℓ^{th} variable u_l then the $\mathbb{R}^d$-valued process $(\partial_\ell X_t^{(-u)})_{0\le t\le T}$ satisfies the stochastic differential equation

$$\partial_\ell X_t^{(-u)} = \big(\dot{b}(X_t^{(-u)})\partial_\ell X_t^{(-u)} - \sigma_\ell(X_t^{(-u)})\big)dt + \sum_{j=1}^{q} \dot{\sigma}_j(X_t^{(-u)})\partial_\ell X_t^{(-u)}(dW_t^j - u_j dt). \tag{22}$$

Moreover, Corollary 4.6.7 in [14] ensures the L^p boundedness of the random variable $\partial_\ell X_t^{(-u)}$, $t\in[0,T]$ and $p\ge 1$. Concerning the process $(U_t^{(-u)})_{0\le t\le T}$, we need a more general result to study its u-sensitivity, we apply Theorem 4.6.4 in [14] to obtain its differentiability with respect to u. The process $(\partial_\ell U_t^{(-u)})_{0\le t\le T}$ is defined similarly and for $i\in\{1,\cdots,d\}$, we denote by $(\partial_\ell (U_t^{(-u)})^i)_{0\le t\le T}$ its i^{th} component satisfying the stochastic differential system

$$\begin{aligned}\partial_\ell (U_t^{(-u)})^i =\ & \big((\partial_\ell X_t^{(-u)})^{tr}\ddot{b}_i(X_t^{(-u)})U_t^{(-u)} + \dot{b}_i(X_t^{(-u)})\partial_\ell U_t^{(-u)} - \dot{\sigma}_\ell(X_t^{(-u)})U_t^{(-u)}\big)dt \\ & + \sum_{j=1}^{q}\big((\partial_\ell X_t^{(-u)})^{tr}\ddot{\sigma}_{ij}(X_t^{(-u)})U_t^{(-u)} + \dot{\sigma}_{ij}(X_t^{(-u)})\partial_\ell U_t^{(-u)}\big)(dW_t^j - u_j dt) \\ & - \frac{1}{\sqrt{2}}\sum_{j,j'=1}^{q}\big((\partial_\ell X_t^{(-u)})^{tr}\ddot{\sigma}_{ij}(X_t^{(-u)})\sigma_{j'}(X_t^{(-u)}) + \dot{\sigma}_{ij}(X_t^{(-u)})\dot{\sigma}_{j'}(X_t^{(-u)})\partial_\ell X_t^{(-u)}\big)d\tilde{W}_t^{j'j}.\end{aligned} \tag{23}$$

Moreover, the same Theorem 4.6.4 in [14] ensures that these components are also L^p bounded, for all $p\ge 1$.

Now, let us recall that the $\mathbb{R}^q$-valued function D is defined by $D(u) = e^{-\eta|u|^2T}(2uT - W_T)g(X_T^{(-u)}, U_T^{(-u)})$. For $\ell,\ell'\in\{1,\cdots,q\}$, the partial derivative of component $D^{\ell'}$ with respect to u_ℓ is given by

$$\begin{aligned}\frac{\partial D^{\ell'}}{\partial u_\ell}(u) =\ & \big(2T\delta_{\ell\ell'} - 2u_\ell T\eta(2u_{\ell'}T - W_T^{\ell'})\big)e^{-\eta|u|^2T}g(X_T^{(-u)}, U_T^{(-u)}) \\ & + e^{-\eta|u|^2T}(2u_{\ell'}T - W_T^{\ell'})\big(\nabla_x g(X_T^{(-u)}, U_T^{(-u)}).\partial_\ell X_T^{(-u)} + \nabla_y g(X_T^{(-u)}, U_T^{(-u)}).\partial_\ell U_T^{(-u)}\big).\end{aligned} \tag{24}$$

Here, for the function $g:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$, we denote by $\nabla_x g$ (resp. $\nabla_y g$) the gradient with respect to the first variable x (resp. the second variable y), and the notation $\delta_{\ell\ell'}$ stands for the Kronecker symbol. Now, let Y and Z be solution to the following stochastic differential system

$$dY_{t,\ell} = \big(\dot{b}(X_t)Y_{t,\ell} - \sigma_\ell(X_t)\big)dt + \sum_{j=1}^{q}\dot{\sigma}_j(X_t)Y_{t,\ell}dW_t^j.$$

and

$$\begin{aligned} d(Z_{t,\ell})^i &= \big((Y_{t,\ell})^{tr}\ddot{b}_i(X_t)U_t + \dot{b}_i(X_t)Z_{t,\ell} - \dot{\sigma}_\ell(X_t)U_t\big)dt \\ &\quad + \sum_{j=1}^{q}\big((Y_{t,\ell})^{tr}\ddot{\sigma}_{ij}(X_t)U_t + \dot{\sigma}_{ij}(X_t)Z_{t,\ell}\big)dW_t^j \\ &\quad -\frac{1}{\sqrt{2}}\sum_{j,j'=1}^{q}\big((Y_{t,\ell})^{tr}\ddot{\sigma}_{ij}(X_t)\sigma_{j'}(X_t) + \dot{\sigma}_{ij}(X_t)\dot{\sigma}_{j'}(X_t)Y_{t,\ell}\big)d\tilde{W}_t^{j'j}. \end{aligned}$$

These both processes can be seen as solutions of respectively (22) and (23) at point $u = 0$, consequently they are L^p bounded, $p \geq 1$. Note that (4), (6), (22) and (23) allow us to apply Girsanov theorem and deduce that $(W, X^{(-u)}, U^{(-u)}, \partial_\ell X^{(-u)}, \partial_\ell U^{(-u)})$ under $\tilde{\mathbb{P}}$ has the same law as $(B^{(-u)}, X, U, Y_{\cdot,\ell}, Z_{\cdot,\ell})$ under $\tilde{\mathbb{P}}_{(-u)}$. Hence, using relation (24) followed by a change of a probability measure, we obtain

$$\begin{aligned} \tilde{\mathbb{E}}|\frac{\partial D^{\ell'}}{\partial u_\ell}(u)|^{2p} = \tilde{\mathbb{E}}\Big[&|\big(2T\delta_{\ell\ell'} - 2u_\ell T\eta(u_{\ell'}T - W_T^{\ell'})\big)e^{-\eta|u|^2T}g(X_T, U_T) \\ &+ e^{-\eta|u|^2T}(u_{\ell'}T - W_T^{\ell'})\big(\nabla_x g(X_T, U_T).Y_{T,\ell} + \nabla_y g(X_T, U_T).Z_{T,\ell}\big)|^{2p}e^{-u.W_T - \frac{1}{2}|u|^2T}\Big]. \end{aligned}$$

Rearranging the terms in the above inequality, we get by Hölder's inequality $\forall r_1 \in (1, \infty)$

$$\begin{aligned} \tilde{\mathbb{E}}|\frac{\partial D^{\ell'}}{\partial u_\ell}(u)|^{2p} \leq\ & e^{(-2p\eta - \frac{1}{2})|u|^2T}\|e^{-u.W_T}\|_{r_1}\big\|(2T + 1 + 2|u_\ell|T\eta)^{2p}|u_{\ell'}T - W_T^{\ell'}|^{2p}\big\|_{\frac{2r_1}{r_1-1}} \\ &\times \big\|(|g(X_T, U_T)| + |\nabla_x g(X_T, U_T).Y_{T,\ell} + \nabla_y g(X_T, U_T).Z_{T,\ell}|)^{2p}\big\|_{\frac{2r_1}{r_1-1}}. \end{aligned}$$

As $e^{(-2p\eta-\frac{1}{2})|u|^2T}\|e^{-u.W_T}\|_{r_1}\big\|(2T+1+2|u_\ell|T\eta)^{2p}|u_{\ell'}T-W_T^{\ell'}|^{2p}\big\|_{\frac{2r_1}{r_1-1}} \leq c_1(1+|u|^{4p})e^{(\frac{r_1}{2}-2p\eta-\frac{1}{2})|u|^2T}$, with c_1 is a positive constant depending only on p, r_1 and T. Then, one can choose $r_1 \in (1, 1 + 4p\eta)$ such that $\sup_{u\in\mathbb{R}^q}(1 + |u|^{2p})e^{(\frac{r_1}{2}-2p\eta-\frac{1}{2})|u|^2T}$ is finite. Hence, we get the existence of a constant c_2 depending only on p, η and T such that

$$\tilde{\mathbb{E}}|\frac{\partial D^{\ell'}}{\partial u_\ell}(u)|^{2p} \leq c_2\big\|(|g(X_T, U_T)| + |\nabla_x g(X_T, U_T).Y_{T,\ell} + \nabla_y g(X_T, U_T).Z_{T,\ell}|)^{2p}\big\|_{\frac{2r_1}{r_1-1}}.$$

Since ψ is $\mathscr{C}^2$ with polynomial growth as well as all its partial derivatives until order two then the function g mapping the couple (x, y) into $\psi^2(x) + \nabla\psi(x).y$ is $\mathscr{C}^1$ and all its first partial derivatives have polynomial growth. The proof is completed, thanks to properties $(\mathcal{P})$, $(\tilde{\mathcal{P}})$ and using the L^p boundedness of Y_T and Z_T for all $p \geq 1$.

4 Central limit theorem for the adaptive procedure

In this section we prove a central limit theorem for both adaptive Monte Carlo and adaptive statistical Romberg methods. Let us recall that the adaptive importance sampling algorithm for the statistical Romberg method approximates our initial quantity of interest $\mathbb{E}\psi(X_T) = \mathbb{E}\left[\psi(X_T^\theta)e^{-\theta\cdot W_T-\frac{1}{2}|\theta|^2T}\right]$ by

$$\frac{1}{N_1}\sum_{i=1}^{N_1} g(\hat{\theta}_i^m, \hat{X}_{T,i+1}^{m,\hat{\theta}_i^m}, \hat{W}_{T,i+1}) + \frac{1}{N_2}\sum_{i=1}^{N_2}\Big(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - g(\theta_i^n, X_{T,i+1}^{m,\theta_i^n}, W_{T,i+1})\Big), \quad (25)$$

where for all $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^q$, $g(\theta,x,y) = \psi(x)e^{-\theta\cdot y-\frac{1}{2}|\theta|^2T}$. Here the paths generated by W and $\hat{W}$ are of course independent. In order to prove a central limit theorem for this algorithm, we need to study independently each of the above empirical means. This is the aim of subsections 4.2 and 4.3. We need first to recall some useful results.

4.1 Technical results

Let us recall the Central Limit Theorem for martingales array (see e.g. [8]).

Theorem 4.1 *Suppose that $(\Omega,\mathbb{F},\mathbb{P})$ is a probability space and that for each n, we have a filtration $\mathbb{F}_n = (\mathcal{F}_k^n)_{k\geq 0}$, a sequence $k_n \longrightarrow \infty$ as $n \longrightarrow \infty$ and a real square integrable vector martingale $M^n = (M_k^n)_{k\geq 0}$ which is adapted to $\mathbb{F}_n$ and has quadratic variation denoted by $(\langle M\rangle_k^n)_{k\geq 0}$. We make the following two assumptions.*

A1. There exists a deterministic symmetric positive semi-definite matrix Γ, such that

$$\langle M\rangle_{k_n}^n = \sum_{k=1}^{k_n}\mathbb{E}\left[|M_k^n - M_{k-1}^n|^2|\mathcal{F}_{k-1}^n\right] \xrightarrow[n\to\infty]{\mathbb{P}} \Gamma.$$

A2. Lindeberg's condition holds: that is, for all $\varepsilon > 0$,

$$\sum_{k=1}^{k_n}\mathbb{E}\left[|M_k^n - M_{k-1}^n|^2 1_{\{|M_k^n-M_{k-1}^n|>\varepsilon\}}|\mathcal{F}_{k-1}^n\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Then

$$M_{k_n}^n \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Gamma) \quad \text{as } n \to \infty.$$

Remark. The following assumption known as the Lyapunov condition, implies the Lindeberg's condition *A2.*,

A3. There exists a real number $a > 1$, such that

$$\sum_{k=1}^{k_n}\mathbb{E}\left[|M_k^n - M_{k-1}^n|^{2a}|\mathcal{F}_{k-1}^n\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

As a prelude to the results of this subsection, we give a double indexed version of the Toeplitz lemma that will be very helpful in the sequel.

Lemma 4.1 *Let $(a_i)_{1\leq i\leq k_n}$ a sequence of real positive numbers, where $k_n \uparrow \infty$ as n tends to infinity, and $(x_i^n)_{i\geq 1, n\geq 1}$ a double indexed sequence such that*

(i) $\lim\limits_{n\to\infty} \sum_{1\leq i\leq k_n} a_i = \infty$

(ii) $\lim\limits_{i,n\to\infty} x_i^n = \lim\limits_{i\to\infty} (\lim\limits_{n\to\infty} x_i^n) = \lim\limits_{n\to\infty} (\lim\limits_{i\to\infty} x_i^n) = x < \infty$

Then

$$\lim_{n\to+\infty} \frac{\sum_{i=1}^{k_n} a_i x_i^n}{\sum_{i=1}^{k_n} a_i} = x.$$

Proof. For all $\varepsilon > 0$, there exists $N_1(\varepsilon)$ such that for all $n \geq N_1(\varepsilon)$ and $i \geq N_1(\varepsilon)$, we have that:

$$|x_i^n - x| \leq \frac{\varepsilon}{2}.$$

As k_n goes to infinity, there exists $N_2(\varepsilon)$ such that for all $n \geq N_2(\varepsilon)$, we have $k_n \geq N_1(\varepsilon)$. Therefore, for all $n \geq \sup(N_1(\varepsilon), N_2(\varepsilon)) = N(\varepsilon)$, we can write:

$$\sum_{i=1}^{k_n} a_i |x_i^n - x| = \sum_{i=1}^{N_1(\varepsilon)-1} a_i |x_i^n - x| + \sum_{i=N_1(\varepsilon)}^{k_n} a_i |x_i^n - x|.$$

For the second term of the expression above, we have:

$$\sum_{i=N_1(\varepsilon)}^{k_n} a_i |x_i^n - x| \leq \frac{\varepsilon}{2} \sum_{i=N_1(\varepsilon)}^{k_n} a_i \leq \frac{\varepsilon}{2} \sum_{i=1}^{k_n} a_i.$$

On the other hand, by assumptions (i) and (ii) there exists $\tilde{N}(\varepsilon)$ such that for all $n \geq \tilde{N}(\varepsilon)$

$$\frac{\sup_{1\leq i\leq N_1(\varepsilon)-1} \sup_{n\geq 1} |x_i^n - x| \sum_{1\leq i\leq N_1(\varepsilon)-1} a_i}{\sum_{1\leq i\leq k_n} a_i} \leq \frac{\varepsilon}{2}.$$

Therefore, for all $n \geq \tilde{N}(\varepsilon)$

$$\left| \frac{\sum_{i=1}^{k_n} a_i x_i^n}{\sum_{i=1}^{k_n} a_i} - x \right| \leq \varepsilon.$$

This completes the proof. □

Let $\tilde{\mathcal{B}} = (\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t\geq 0}, \tilde{\mathbb{P}})$ be the extension probability space introduced in Section 2 endowed with the filtration $\tilde{\mathcal{F}}_{T,i} = \sigma(W_{t,l}, \tilde{W}_{t,l}, l \leq i, t \leq T)$ given in the very beginning of Section 3. In what follows, let $(\theta_i^n)_{i\geq 0}$, $n \in \mathbb{N}$ and $(\theta_i)_{i\geq 0}$ be a family of sequences satisfying

$$(\mathcal{H}_\theta) \quad \begin{cases} \text{For each } n \in \mathbb{N},\ (\theta_i^n)_{i\geq 0} \text{ and } (\theta_i)_{i\geq 0} \text{ are } (\tilde{\mathcal{F}}_{T,i})_{i\geq 0}\text{-adapted} \\ \lim\limits_{i\to\infty} (\lim\limits_{n\to\infty} \theta_i^n) = \lim\limits_{i\to\infty} \theta_i = \lim\limits_{n\to\infty} (\lim\limits_{i\to\infty} \theta_i^n) = \lim\limits_{n\to\infty} \theta_n^* = \theta^*, \quad \tilde{\mathbb{P}}\text{-}a.s., \end{cases}$$

with deterministic limits θ^* and θ_n^*.

4.2 The adaptive Monte Carlo method

Let us recall that the statistical Romberg algorithm (25) runs successively two independent empirical means. The first one is a crude Monte Carlo simply depending on the Euler scheme with the coarse time step T/m. However, the second empirical mean involves the functional difference between the fine Euler scheme with time step T/n and the coarse one constructed from the same Brownian path. The task now is to prove a central limit theorem for the first empirical mean.

Theorem 4.2 *Let $(\theta_i^n)_{i\geq 0}$, $n \in \mathbb{N}$ and $(\theta_i)_{i\geq 0}$ be a family of sequences satisfying $(\mathcal{H}_\theta)$. Moreover, assume that b and σ satisfy the global Lipschitz condition $(\mathcal{H}_{b,\sigma})$ and the function ψ is a real valued function satisfying assumption $(\mathcal{H}_{\varepsilon_n})$, with $\alpha \in [1/2, 1]$ and $C_\psi \in \mathbb{R}$, such that*

$$|\psi(x) - \psi(y)| \leq C(1 + |x|^p + |y|^p)|x - y|, \quad \text{for some } C, p > 0,$$

then the following convergence holds

$$n^\alpha \left(\frac{1}{n^{2\alpha}} \sum_{i=1}^{n^{2\alpha}} g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \mathbb{E}\psi(X_T) \right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(C_\psi, \sigma^2\right).$$

where $\sigma^2 := \mathbb{E}\left(\psi(X_T)^2 e^{-\theta^*.W_T + \frac{1}{2}|\theta^*|^2 T}\right) - [\mathbb{E}\psi(X_T)]^2$ and for all $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^q$, $g(\theta, x, y) = \psi(x)e^{-\theta\cdot y - \frac{1}{2}|\theta|^2 T}$. Furthermore, we have also for all $\alpha, \beta > 0$

$$n^\alpha \left(\frac{1}{n^{2\alpha}} \sum_{i=1}^{n^{2\alpha}} g(\theta_i^{n^\beta}, X_{T,i+1}^{n^\beta,\theta_i^{n^\beta}}, W_{T,i+1}) - \mathbb{E}\psi(X_T^{n^\beta}) \right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2\right).$$

Proof. At first, we rewrite the total error as follows

$$\frac{1}{n^{2\alpha}} \sum_{i=1}^{n^{2\alpha}} g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \mathbb{E}\psi(X_T) =$$
$$\frac{1}{n^{2\alpha}} \sum_{i=1}^{n^{2\alpha}} \left(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \mathbb{E}g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) \right) + \mathbb{E}\psi(X_T^n) - \mathbb{E}\psi(X_T).$$

Note that $\tilde{\mathbb{E}}g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) = \tilde{\mathbb{E}}\left(\tilde{\mathbb{E}}\left(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1})|\tilde{\mathcal{F}}_{T,i}\right)\right) = \mathbb{E}\psi(X_T^n)$. Assumption $(\mathcal{H}_{\varepsilon_n})$ ensures that $n^\alpha(\mathbb{E}\psi(X_T^n) - \mathbb{E}\psi(X_T)) \longrightarrow C_\psi$ as $n \to \infty$. Consequently, it remains to study the asymptotic behavior of the martingale array $(M_k^n)_{k\geq 1}$ given by

$$M_k^n := \frac{1}{n^\alpha} \sum_{i=1}^{k} \left(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \mathbb{E}g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) \right).$$

To do so, we split the proof into two steps devoted to apply the central limit theorem for martingales array (see Theorem 4.1 and comments their).

Step 1. We need first to study the asymptotic behavior of the quadratic variation of the martingale array $(M_k^n)_{k\geq 1}$ given by

$$\langle M\rangle_{n^{2\alpha}}^n = \frac{1}{n^{2\alpha}}\sum_{i=1}^{n^{2\alpha}}\tilde{\mathbb{E}}\left[\left(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \tilde{\mathbb{E}}g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1})\right)^2 |\tilde{\mathcal{F}}_{T,i}\right].$$

Since θ_i^n is $\tilde{\mathcal{F}}_{T,i}$-measurable and $W_{T,i+1} \perp\!\!\!\perp \tilde{\mathcal{F}}_{T,i}$, we obtain easily that

$$\langle M\rangle_{n^{2\alpha}}^n = \frac{1}{n^{2\alpha}}\sum_{i=1}^{n^{2\alpha}}\nu_n(\theta_i^n) - [\mathbb{E}\psi(X_T^n)]^2, \tag{26}$$

where for all $\theta\in\mathbb{R}^q$

$$\nu_n(\theta) := \mathbb{E}\left(\psi(X_T^{n,\theta})^2 e^{-2\theta.W_T - |\theta|^2 T}\right) = \mathbb{E}\left(\psi(X_T^n)^2 e^{-\theta\cdot W_T + \frac{1}{2}|\theta|^2 T}\right).$$

It is clear that by assumption $(\mathcal{H}_{\varepsilon_n})$, the last term on the right hand side of the relation (26) converges to $[\mathbb{E}\psi(X_T)]^2$, as n tends to infinity. Concerning the first term, we introduce $\nu(\theta) := \mathbb{E}\left(\psi(X_T)^2 e^{-\theta\cdot W_T+\frac{1}{2}|\theta|^2 T}\right)$ and we get for all $\theta\in\mathbb{R}^q$

$$|\nu_n(\theta)-\nu(\theta)| \leq \mathbb{E}\left(\left|\psi(X_T^n)^2 - \psi(X_T)^2\right| e^{-\theta\cdot W_T+\frac{1}{2}|\theta|^2 T}\right) \leq e^{\frac{3}{2}|\theta|^2 T}\|\psi(X_T^n)^2-\psi(X_T)^2\|_2.$$

Under the condition on ψ together with property $(\mathcal{P})$, there exists $C>0$ such that

$$|\nu_n(\theta)-\nu(\theta)| \leq \frac{C}{\sqrt{n}} e^{\frac{3}{2}|\theta|^2 T}, \quad \forall\theta\in\mathbb{R}^q.$$

By similar calculations, we check easily the equicontinuity of the family functions $(\nu_n)_{n\geq 1}$ and we deduce thanks to property $(\mathcal{H}_\theta)$

$$\lim_{i,n\to\infty}\nu_n(\theta_i^n) = \nu(\theta^*) \quad \tilde{\mathbb{P}}\text{-}a.s.$$

Therefore, Lemma 4.1 applies and we deduce that $\langle M\rangle_{n^{2\alpha}}^n \underset{n\to\infty}{\longrightarrow} \sigma^2$.

Step 2. We will check now the Lyapunov condition, that is assumption *A3.*, which implies the Lindeberg condition *A2.* Let $a>1$, we have

$$\begin{aligned}\sum_{i=1}^{n^{2\alpha}}\tilde{\mathbb{E}}\left[|M_i^n - M_{i-1}^n|^{2a}|\tilde{\mathcal{F}}_{T,i-1}\right] &= \frac{1}{n^{2a\alpha}}\sum_{i=1}^{n^{2\alpha}}\tilde{\mathbb{E}}\left[\left|g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - \mathbb{E}\psi(X_T^n)\right|^{2a}|\tilde{\mathcal{F}}_{T,i}\right]\\ &\leq \frac{2^{2a-1}}{n^{2a\alpha}}\sum_{i=1}^{n^{2\alpha}}\nu_{a,n}(\theta_i^n) + \frac{2^{2a-1}}{n^{2\alpha(a-1)}}[\mathbb{E}\psi(X_T^n)]^{2a}\end{aligned}$$

where for all $\theta\in\mathbb{R}^q$, $\nu_{a,n}(\theta) = \mathbb{E}\left(\psi(X_T^n)^{2a}e^{-(2a-1)\theta\cdot W_T - (a-\frac{3}{2})|\theta|^2 T}\right)$. Following the same arguments detailed in the first step, we prove that

$$\frac{1}{n^{2\alpha}}\sum_{i=1}^{n^{2\alpha}}\nu_{a,n}(\theta_i^n) \underset{n\to\infty}{\longrightarrow} \nu_a(\theta^*) \quad \tilde{\mathbb{P}}\text{-}a.s.$$

where for all $\theta\in\mathbb{R}^q$, $\nu_a(\theta) = \mathbb{E}\left(\psi(X_T)^{2a}e^{-(2a-1)\theta\cdot W_T-(a-\frac{3}{2})|\theta|^2 T}\right)$. The second assertion is easily obtained following the above proof with $\alpha,\beta>0$. This completes the proof. □

Remark. If one have in mind to reduce the variance by using an adaptive crude Monte Carlo method, it appears clear that the natural choice is

$$\theta^* = \arg\min_{\theta\in\mathbb{R}^q} \tilde{\mathbb{E}}\left(g^2(\theta, X_T)\right) \quad \text{and} \quad \theta_n^* = \arg\min_{\theta\in\mathbb{R}^q} \tilde{\mathbb{E}}\left(g^2(\theta, X_T^n)\right) \text{ for } n \geq 1.$$

Under suitable conditions on ψ, b and σ, one can of course construct sequences $(\theta_i^n)_{i\geq 0}$, $n \in \mathbb{N}$ and $(\theta_i)_{i\geq 0}$ satisfying $(\mathcal{H}_\theta)$ by either the constrained or the unconstrained Robbins-Monro algorithm.

4.3 The adaptive statistical Romberg method

As we pointed out at the beginning of the above subsection, the statistical Romberg algorithm (25) consists of two empirical means. So our task now is to study the asymptotic behavior of the second one in view to establish a central limit theorem for the method.

Theorem 4.3 *Let $(\theta_i^n)_{i\geq 0}$, $n \in \mathbb{N}$ and $(\theta_i)_{i\geq 0}$ be a family of sequences satisfying $(\mathcal{H}_\theta)$. Moreover, assume that b and σ are $\mathcal{C}^1$ functions satisfying the global Lipschitz condition $(\mathcal{H}_{b,\sigma})$ and ψ is a real valued function satisfying assumptions $(\mathcal{H}_\psi)$, $(\mathcal{H}_{\varepsilon_n})$, with constants $\alpha \in (1/2, 1]$ and $C_\psi \in \mathbb{R}$, such that*

$$|\psi(x) - \psi(y)| \leq C(1 + |x|^p + |y|^p)|x - y|, \quad \textit{for some } C, p > 0.$$

If we choose $N_1 = n^{2\alpha}$, $N_2 = n^{2\alpha-\beta}$ and $m = n^\beta$, $0 < \beta < 1$ then the statistical Romberg algorithm denoted by V_n in (25) satisfies

$$n^\alpha \left(V_n - \mathbb{E}\psi(X_T)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(C_\psi, \sigma^2 + \tilde{\sigma}^2\right) \qquad \textit{as } n \to \infty,$$

where $\sigma^2 = \mathbb{E}\left[\psi(X_T)^2 e^{-\theta^*\cdot W_T + \frac{1}{2}|\theta^*|^2 T}\right] - \left[\mathbb{E}\psi(X_T)\right]^2$, $\tilde{\sigma}^2 := \tilde{\mathbb{E}}\left[\left[\nabla\psi(X_T)\cdot U_T\right]^2 e^{-\theta^*\cdot W_T + \frac{1}{2}|\theta^*|^2 T}\right]$ and U is the process introduced from the beginning by relation (3).

Proof. First of all, note that we can rewrite the normalized total error as follows

$$n^\alpha \left(V_n - \mathbb{E}\psi(X_T)\right) := A_1^n + A_2^n$$

with $A_1^n := n^\alpha \left(V_n - \mathbb{E}\psi(X_T^{n^\beta}) - \mathbb{E}[\psi(X_T^n) - \psi(X_T^{n^\beta})]\right)$, and $A_2^n := n^\alpha \left(\mathbb{E}[\psi(X_T^n) - \psi(X_T)]\right)$. So, assumption $(\mathcal{H}_{\varepsilon_n})$ yields the convergence of the second term A_2^n towards the discretization constant C_ψ, as n tends to infinity. The first term A_1^n can be also rewritten as follows $A_1^n := A_{1,1}^n + A_{1,2}^n$, where

$$\begin{aligned}
A_{1,1}^n &:= \frac{1}{n^\alpha}\sum_{i=1}^{n^{2\alpha}}\left(g(\hat{\theta}_i^{n^\beta}, \hat{X}_{T,i+1}^{n^\beta,\hat{\theta}_i^{n^\beta}}, W_{T,i+1}) - \mathbb{E}\psi(X_T^{n^\beta})\right), \\
A_{1,2}^n &:= \frac{1}{n^{\alpha-\beta}}\sum_{i=1}^{n^{2\alpha-\beta}}\left(g(\theta_i^n, X_{T,i+1}^{n,\theta_i^n}, W_{T,i+1}) - g(\theta_i^n, X_{T,i+1}^{n^\beta,\theta_i^n}, W_{T,i+1}) - \mathbb{E}[\psi(X_T^n) - \psi(X_T^{n^\beta})]\right).
\end{aligned}$$

Using the independence between $A^n_{1,1}$ and $A^n_{1,2}$, we study separately their asymptotic behavior. Concerning the first term, the second assertion in Theorem 4.2 applies and gives the asymptotic normality of $A^n_{1,1}$,

$$A^n_{1,1} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2\right), \quad \text{as } n \to \infty. \tag{27}$$

Now, concerning the second term $A^n_{1,2}$ we introduce the martingale arrays $(M^n_k)_{k\geq 1}$

$$M^n_k := \frac{1}{n^{\alpha-\beta}} \sum_{i=1}^{k} \left(g(\theta^n_i, X^{n,\theta^n_i}_{T,i+1}, W_{T,i+1}) - g(\theta^n_i, X^{n^\beta,\theta^n_i}_{T,i+1}, W_{T,i+1}) - \mathbb{E}[\psi(X^n_T) - \psi(X^{n^\beta}_T)]\right),$$

in view to apply Theorem 4.1. To do so, we will verify both assumptions *A1.* and *A3.* in the following two steps.

• **Step 1.** The quadratic variation of M evaluated at $n^{2\alpha-\beta}$ is given by

$$\langle M\rangle^n_{n^{2\alpha-\beta}} = \frac{1}{n^{2\alpha-\beta}} \sum_{i=1}^{n^{2\alpha-\beta}} n^\beta \xi_n(\theta^n_i) - \left(n^{\frac{\beta}{2}}[\mathbb{E}\psi(X^n_T) - \mathbb{E}\psi(X^{n^\beta}_T)]\right)^2, \tag{28}$$

where $\forall\theta \in \mathbb{R}^q$, $\xi_n(\theta) := \mathbb{E}\left([\psi(X^n_T) - \psi(X^{n^\beta}_T)]^2 e^{-\theta\cdot W_T + \frac{1}{2}|\theta|^2 T}\right)$. Now, assumption $(\mathcal{H}_{\varepsilon_n})$ with $1/2 < \alpha \leq 1$ ensures that the second term on the right hand side of relation (28) vanishes as n tends to infinity. We focus now on the asymptotic behavior of $n^\beta\xi_n(\theta)$. Under assumption $(\mathcal{H}_\psi)$, we apply the Taylor expansion theorem twice to get for all $\theta \in \mathbb{R}^q$

$$n^{\frac{\beta}{2}}[\psi(X^n_T) - \psi(X^{n^\beta}_T)]e^{-\frac{1}{2}\theta\cdot W_T + \frac{1}{4}|\theta|^2 T} = n^{\frac{\beta}{2}}\nabla\psi(X_T)\cdot[X^n_T - X^{n^\beta}_T]e^{-\frac{1}{2}\theta\cdot W_T + \frac{1}{4}|\theta|^2 T} + R_n,$$

where

$$R_n := n^{\frac{\beta}{2}}(X^n_T - X_T)\varepsilon(X_T, X^n_T - X_T) - n^{\frac{\beta}{2}}(X^{n^\beta}_T - X_T)\varepsilon(X_T, X^{n^\beta}_T - X_T)$$

with $\varepsilon(X_T, X^n_T - X_T) \xrightarrow{\mathbb{P}\text{-}a.s.} 0$ and $\varepsilon(X_T, X^{n^\beta}_T - X_T) \xrightarrow{\mathbb{P}\text{-}a.s.} 0$ as $n \to \infty$, since the global Lipschitz condition $(\mathcal{H}_{b,\sigma})$ is satisfied. Further, as b and σ are $\mathcal{C}^1$ functions then according to Theorem 3.2 in [11] we have the tightness of $n^{\frac{\beta}{2}}(X^n_T - X_T)$ and $n^{\frac{\beta}{2}}(X^{n^\beta}_T - X_T)$ and we deduce the convergence in probability of the remaining term R_n to zero as n tends to infinity. Once again, by the same theorem in [11], we get for all $\theta \in \mathbb{R}^q$

$$n^{\frac{\beta}{2}}[\psi(X^n_T) - \psi(X^{n^\beta}_T)]e^{-\frac{1}{2}\theta\cdot W_T + \frac{1}{4}|\theta|^2 T} \overset{stably}{\Longrightarrow} \nabla\psi(X_T)\cdot U_T e^{-\frac{1}{2}\theta\cdot W_T + \frac{1}{4}|\theta|^2 T}. \tag{29}$$

Otherwise, $\forall\theta \in \mathbb{R}^q$ and $a' > 1$ we have by Cauchy-Schwarz inequality

$$\mathbb{E}\left|n^{\frac{\beta}{2}}[\psi(X^n_T) - \psi(X^{n^\beta}_T)]e^{-\frac{1}{2}\theta\cdot W_T + \frac{1}{4}|\theta|^2 T}\right|^{2a'} \leq n^{\beta a'}\left[\mathbb{E}\left|\psi(X^n_T) - \psi(X^{n^\beta}_T)\right|^{4a'}\right]^{\frac{1}{2}} e^{\frac{a'(2a'+1)}{2}|\theta|^2 T}.$$

Thanks to the assumption on ψ together with property $(\mathcal{P})$, we obtain

$$\sup_n \mathbb{E}\left|n^{\frac{\beta}{2}}[\psi(X^n_T) - \psi(X^{n^\beta}_T)]e^{-\frac{1}{2}\theta\cdot W_T + \frac{1}{4}|\theta|^2 T}\right|^{2a'} < \infty. \tag{30}$$

Hence, by the stable convergence obtained in (29) and the uniform integrability property given by (30) we deduce $\forall\theta \in \mathbb{R}^q$

$$\lim_{n\to\infty} n^{\beta}\xi_n(\theta) = \tilde{\mathbb{E}}\left([\nabla\psi(X_T)\cdot U_T]^2 \, e^{-\theta\cdot W_T+\frac{1}{2}|\theta|^2 T}\right) := \xi(\theta). \tag{31}$$

Using property $(\mathcal{P})$ with assumption on ψ, it is easy to check by standard evaluations the equicontinuity of the family functions $(n^{\beta}\xi_n)_{n\geq 1}$. So under assumption $(\mathcal{H}_{\theta})$, we get

$$\lim_{i,n\to\infty} n^{\beta}\xi_n(\theta_i^n) = \xi(\theta^*) \quad \tilde{\mathbb{P}}\text{-}a.s.$$

Then, Lemma 4.1 yields $\lim_{n\to\infty}\langle M\rangle^n_{n^{2\alpha-\beta}} = \xi(\theta^*)$, $\tilde{\mathbb{P}}$-$a.s.$

• **Step 2.** The second step consists on checking Lyapunov assumption *A3.* Let $a > 1$,

$$\sum_{i=1}^{n^{2\alpha-\beta}} \tilde{\mathbb{E}}\left[|M_i^n - M_{i-1}^n|^{2a}|\tilde{\mathcal{F}}_{T,i-1}\right] \leq \frac{2^{2a-1}}{n^{a(2\alpha-\beta)}}\sum_{i=1}^{n^{2\alpha-\beta}} n^{\beta a}\xi_{a,n}(\theta_i^n) + \frac{2^{2a-1}n^{\beta a}}{n^{(2\alpha-\beta)(a-1)}}|\mathbb{E}\psi(X_T^n)-\mathbb{E}\psi(X_T^{n^{\beta}})|^{2a}$$

where for all $\theta \in \mathbb{R}^q$, $\xi_{a,n}(\theta) := \mathbb{E}\left(|\psi(X_T^n) - \psi(X_T^{n^{\beta}})|^{2a}e^{-(2a-1)\theta\cdot W_T-(a-\frac{3}{2})|\theta|^2 T}\right)$. Miming the same arguments used in the first step, we prove under assumption $(\mathcal{H}_{\theta})$ using relations (29) and Lemma 4.1, that

$$\frac{1}{n^{2\alpha-\beta}}\sum_{i=1}^{n^{2\alpha-\beta}} n^{\beta a}\xi_{a,n}(\theta_i^n) \underset{n\to\infty}{\longrightarrow} \xi_a(\theta^*) := \tilde{\mathbb{E}}\left(|\nabla\psi(X_T)\cdot U_T|^{2a}e^{-(2a-1)\theta\cdot W_T-(a-\frac{3}{2})|\theta|^2 T}\right), \ \tilde{\mathbb{P}}\text{-a.s.}$$

Consequently, since $a > 1$, we conclude using assumption $(\mathcal{H}_{\varepsilon_n})$ that *A3.* holds. This gives the asymptotic normality of $A^n_{1,2,2}$ so that we have $A^n_{1,2} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\tilde{\sigma}^2)$, as $n \to \infty$. This completes the proof. □

Remark. We recall that for the adaptive statistical Romberg method the optimal choice of θ^* and θ_n^* is given respectively by relations (7) and (8). According to Corollary 3.2 (resp. Corollary 3.4), the sequences $(\theta_i^n)_{i\geq 0}$, $n \in \mathbb{N}$ and $(\theta_i)_{i\geq 0}$ obtained by the constrained Robbins-Monro algorithm (resp. the unconstrained Robbins-Monro algorithm) satisfy $(\mathcal{H}_{\theta})$ under some regularity conditions on ψ, b and σ.

Complexity analysis According to the main theorems of this section, we deduce that for a total error of order $1/n^{\alpha}$, $\alpha \in (1/2, 1]$, the minimal computational effort necessary to run the adaptive statistical Romberg algorithm is obtained for $N_1 = n^{2\alpha}$, $N_2 = n^{2\alpha-\beta}$ and $m = n^{\beta}$. This leads to a time complexity given by $C_{SR} = C \times (n^{2\alpha+\beta} + (n + n^{\beta})n^{2\alpha-\beta})$, with $C > 0$. So the time complexity reaches its minimum for the optimal choice of $\beta = 1/2$. Hence, the optimal parameters to run the method are given by $m = \sqrt{n}$, $N_1 = n^{2\alpha}$ and $N_2 = n^{2\alpha-1/2}$. Then the optimal complexity of the adaptive statistical Romberg algorithm is given by $C_{SR} \simeq C \times n^{2\alpha+\frac{1}{2}}$. However, for the same error of order $1/n^{\alpha}$, the optimal complexity of the adaptive Monte Carlo algorithm is given by $C_{MC} = C \times (N \times n) = C \times n^{2\alpha+1}$. We conclude that the adaptive statistical Romberg method is more efficient in terms of time complexity.

5 Numerical results for the Heston model

Stochastic volatility models are increasingly important in practical derivatives pricing applications. In this section we show, throughout the problem of option pricing with a stochastic volatility model, the efficiency of the importance sampling statistical Romberg method compared to the importance sampling Monte Carlo one. The popular stochastic volatility model in finance is the Heston model introduced by Heston in [10] as solution to

$$\begin{cases} dS_t = rS_t dt + \sqrt{V_t} S_t dW_t^1 \\ dV_t = \kappa(\bar{v} - V_t)dt + \sigma\sqrt{V_t}\rho dW_t^1 + \sigma\sqrt{V_t}\sqrt{1-\rho^2}dW_t^2, \end{cases}$$

where W^1 and W^2 are two independent Brownian motions. Parameters κ, σ, $\bar{v}$ and r are strictly positive constants and $|\rho| \leq 1$. In this model, κ is the rate at which V_t reverts to $\bar{v}$, $\bar{v}$ is the long run average price variance, σ is the volatility of the variance, r is the interest rate and ρ is a correlation term.

Our aim is to use the importance sampling method in order to reduce the variance when computing the price of an European call option, with strike K, under the Heston model. The payoff of the option is $\psi(S_T) = (S_T - K)_+$. Then, the price is $e^{-rT}\mathbb{E}\psi(S_T)$. After a density transformation, given by Girsanov theorem, the price will be defined by:

$$e^{-rT}\mathbb{E}\left[g(\theta, S_T^\theta)\right] = e^{-rT}\mathbb{E}\left[\psi(S_T^\theta) \quad e^{-\theta.W_T - \frac{1}{2}|\theta|^2 T}\right], \qquad \theta \in \mathbb{R}^2.$$

For more details on definitions of the function g and S_T^θ, see relation (5) and related results given in the same page. To approximate S_T^θ, we consider the step T/n and we discretize the stochastic process using the Euler scheme. For $i \in [\![0, n-1]\!]$ and $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$,

$$\begin{cases} S_{t_{i+1}}^{n,\theta} = S_{t_i}^{n,\theta}\left(1 + (r + \theta_1\sqrt{V_{t_i}^{n,\theta}})\frac{T}{n} + \sqrt{V_{t_i}^{n,\theta}\frac{T}{n}}Z_{1,i+1}\right), \\ V_{t_{i+1}}^{n,\theta} = \left|V_{t_i}^{n,\theta} + \left(\kappa(\bar{v} - V_{t_i}^{n,\theta}) + \sigma\sqrt{V_{t_i}^{n,\theta}}(\rho\theta_1 + \sqrt{1-\rho^2}\theta_2)\right)\frac{T}{n} + \sigma\sqrt{V_{t_i}^{n,\theta}\frac{T}{n}}Z_{2,i+1}\right|, \end{cases}$$

with $(Z_{1,i}, Z_{2,i})_{1\leq i\leq n}$ is a sequence of a standard Gaussian random vectors taking values in $\mathbb{R}^2$. Hence, the price of the European call option is firstly approximated by

$$e^{-rT}\mathbb{E}\left[g(\theta, S_T^{n,\theta})\right] = e^{-rT}\mathbb{E}\left[\psi(S_T^{n,\theta}) \quad e^{-\theta.W_T - \frac{1}{2}|\theta|^2 T}\right], \qquad \theta \in \mathbb{R}^2.$$

The choice of θ depends on using the classical Monte Carlo method or the statistical Romberg one. The optimal θ for the first method is given by

$$\theta_n^* = \underset{\theta\in\mathbb{R}^2}{\arg\min}\mathbb{E}\left[\psi^2(S_T^{n,\theta}) \quad e^{-2\theta.W_T - |\theta|^2 T}\right].$$

However, The optimal θ for the second one is

$$\tilde{\theta}_n^* = \underset{\theta\in\mathbb{R}^2}{\arg\min}\mathbb{E}\left[\left(\psi^2(S_T^{n,\theta}) + (\nabla\psi(S_T^{n,\theta}).U_T^{n,\theta})^2\right)e^{-2\theta.W_T - |\theta|^2 T}\right],$$

where $U^{n,\theta}$ denotes the Euler discretization scheme obtained when we replace coefficients b and σ of relation (9) by the corresponding parameters in the Heston model. Here, we have also the choice of the algorithm approximating both θ_n^* and $\tilde{\theta}_n^*$. We can use either the constrained or the unconstrained stochastic algorithms studied in section 3 above.

- **Approximation of θ_n^* by**

  i. Constrained algorithm: let $(\mathcal{K}_i)_{i\in\mathbb{N}}$ denote an increasing sequence of compact sets satisfying $\cup_{i=0}^{\infty}\mathcal{K}_i=\mathbb{R}^d$ and $\mathcal{K}_i\subsetneq\overset{\circ}{\mathcal{K}}_{i+1},\forall i\in\mathbb{N}$. For $\theta_0^n\in\mathcal{K}_0$, $\alpha_0^n=0$ and a gain sequence $(\gamma_i)_{i\in\mathbb{N}}$ satisfying (12), we define the sequence $(\theta_i^n,\alpha_i^n)_{i\in\mathbb{N}}$ recursively by

$$\begin{cases}\textbf{if} & \theta_i^n-\gamma_{i+1}H(\theta_i^n,S_{T,i+1}^n,U_{T,i+1}^n,W_{T,i+1})\in\mathcal{K}_{\alpha_i^n},\ \textbf{then}\\ & \theta_{i+1}^n=\theta_i^n-\gamma_{i+1}H(\theta_i^n,S_{T,i+1}^n,U_{T,i+1}^n,W_{T,i+1}),\ \text{and}\ \alpha_{i+1}^n=\alpha_i^n\\ \textbf{else} & \theta_{i+1}^n=\theta_0^n\ \text{and}\ \alpha_{i+1}^n=\alpha_i^n+1,\end{cases}\tag{32}$$

  where $H(\theta_i^n,S_{T,i+1}^n,W_{T,i})=(\theta_i^nT-W_{T,i+1})\ \ \psi^2(S_{T,i+1}^n)\ \ e^{-\theta_i^n.W_{T,i+1}+\frac{1}{2}|\theta_i^n|^2T}$.

  ii. Unconstrained algorithm : $\theta_{i+1}^n=\theta_i^n-\gamma_{i+1}(2\theta_i^nT-W_{T,i+1})\psi^2(S_{T,i+1}^{n,-\theta_i^n})e^{-\eta|\theta_i^n|^2T}$, with $\eta>0$.

- **Approximation of $\tilde{\theta}_n^*$ by**

  i. Constrained algorithm: we use the same routine (32) with

$$H(\theta_i^n,S_{T,i+1}^n,W_{T,i})=(\theta_i^nT-W_{T,i+1})\left(\psi^2(S_{T,i+1}^n)+(\nabla\psi(S_{T,i+1}^n).U_{T,i+1}^n)^2\right)e^{-\theta_i^n.W_{T,i+1}+\frac{1}{2}|\theta_i^n|^2T}.$$

  ii. Unconstrained algorithm: we use the routine

$$\theta_{i+1}^n=\theta_i^n-\gamma_{i+1}(2\theta_i^nT-W_{T,i+1})\left(\psi^2(S_{T,i+1}^{n,-\theta_i^n})+(\nabla\psi(S_{T,i+1}^{n,-\theta_i^n})\cdot U_{T,i+1}^n)^2\right)e^{-\eta|\theta_i^n|^2T}.$$

To compare these different routines we run a number of iterations $M=500\,000$. The parameters in the Heston model are chosen as follows: $S_0=100$, $V_0=0.01$, $K=100$, the free interest rate $r=\log(1.1)$, $\sigma=0.2$, $k=2$, $\bar{v}=0.01$, $\rho=0.5$ and maturity time $T=1$. Table 1 gives the obtained values of the two-dimensional vectors θ_n^* and $\tilde{\theta}_n^*$.

	Constrained algorithm	**Unconstrained algorithm**
θ_n^*	$(0.7906, 0.0516)$	$(0.7904, 0.0532)$
$\tilde{\theta}_n^*$	$(0.7884, 0.0587)$	$(0.7898, 0.0576)$

Table 1: *Estimation of* θ_n^* *and* $\tilde{\theta}_n^*$

In Figure 1, we test the robustness of both routines, for the computation of $\tilde{\theta}_n^*$, using the averaged algorithm "à la Ruppert & Poliak" (see e.g. [19]) known to give optimal rate for convergence. We implement this averaged algorithm using both constrained and unconstrained procedures. So, we proceed as follows,

i. first, we choose a slowly decreasing step: $\gamma_i=\gamma_0/i^\alpha$, for $\alpha\in(\frac{1}{2},1)$ and $\gamma_0>0$.

ii. Then, we compute the empirical mean of all the previous observations,

$$\bar{\theta}_{i+1}^n:=\frac{1}{i+1}\sum_{k=0}^{i}\tilde{\theta}_k^n.$$

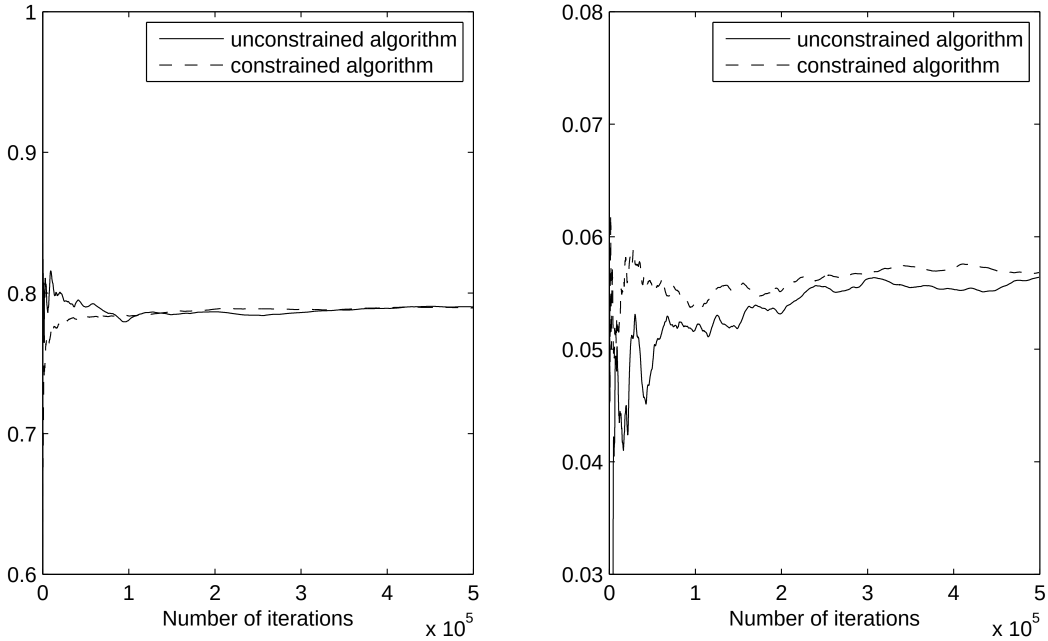

Figure 1: *Values of* $(\bar{\theta}_i^n)_{1\leq i\leq M}$ *obtained with* $n = 100$*,* $\gamma_0 = 0.01$ *and* $\alpha = 0.75$*.*

The left curve (resp. the right curve) is the representation, for $1 \leq i \leq M$, of the first component (resp. the second component) of the two-dimensional vector $\bar{\theta}_i^n$. The trajectories obtained using the constrained or the unconstrained algorithm are comparable. Consequently, since we did not notice any major difference between the two methods we have chosen to only use the constrained algorithm for approximating θ_n^* (resp. $\tilde{\theta}_n^*$) by θ_M^n (resp $\tilde{\theta}_M^n$). Our aim now, is to compare both importance sampling Monte Carlo method (denoted by MC+IS) and importance sampling statistical Romberg (denoted by SR+IS).

- MC+IS method: European call option price approximation with $N = n^2$

$$\frac{e^{-rT}}{N}\sum_{i=1}^{N} g(\theta_M^n, S_{T,i+1}^{n,\theta_M^n}) = \frac{e^{-rT}}{N}\sum_{i=1}^{N} \psi(S_{T,i+1}^{n,\theta_M^n})e^{-\theta_M^n\cdot W_{T,i+1}-\frac{1}{2}|\theta_M^n|^2T}. \tag{33}$$

- SR+IS method: European call option price approximation method with $N_1 = n^2$ and $N_2 = n^{\frac{3}{2}}$

$$\frac{e^{-rT}}{N_1}\sum_{i=1}^{N_1} g(\tilde{\theta}_M^n, \hat{S}_{T,i+1}^{\sqrt{n},\tilde{\theta}_M^n}) + \frac{e^{-rT}}{N_2}\sum_{i=1}^{N_2}\Big(g(\tilde{\theta}_M^n, S_{T,i+1}^{n,\tilde{\theta}_M^n}) - g(\tilde{\theta}_M^n, S_{T,i+1}^{\sqrt{n},\tilde{\theta}_M^n})\Big). \tag{34}$$

The first method (33) is already implemented and available in the free online version of Premia platform (`https://www.rocq.inria.fr/mathfi/Premia/index.html`) and our method (34) is now added in the latest premium version. In Table 2 (resp. Table 3), we compare for each given number of time step n, the obtained call price (resp. the sensitivity call price parameter Δ)

with the corresponding length of the 95%-confidence interval and the CPU time (per second) for both methods (33) and (34). It is worth to note that the number of time step n needed to achieve a given accuracy depends on the choice of the method.

Method	n	**Price** **Interval length**	**Confidence**	**time**
	400	9.641444	0.060094	10.38
MC+IS	900	9.661192	0.029409	91.5
	1600	9.656892	0.016538	512.29
	600	9.659409	0.057454	3.36
SR+IS	1600	9.660062	0.019933	26.79
	3600	9.65673	0.008584	194.6

Table 2: *Call Price for the Heston model*

Method	n	**Price** **Interval length**	**Confidence**	**time**
	400	0.863968	0.00721	9.39
MC+IS	900	0.863291	0.003151	91.58
	1600	0.863766	0.001774	515.31
	600	0.867441	$0,007249$	3.27
SR+IS	1600	0.864213	0.002541	27.02
	3600	0.862589	0.001095	202.2

Table 3: *Delta call price for the Heston model*

We also compare both methods (33) and (34) for a large range of time step numbers n. Then, we make a simple log-log scale plot of CPU time versus the corresponding 95%-confidence interval length. Computations are done on a PC with a 2.5 GHz Intel core i5 processor. In Figure 2 the line marked by circles denotes the MC+IS method and the line marked by squares denotes the SR+IS method. The values mentioned near the points correspond to the chosen number of steps n. Clearly, the SR+IS curve is lower than the MC+IS one, which means that the MC+IS method spends more time than the SR+IS method to achieve the same given error when computing the option price. For example for an error of 0.06, the SR+IS method reduces time by a factor of 3.33 compared to a MC+IS one. Note that, the more the imposed error is small, the better improvement is. For example for a small error 0.02, the time reduction exceeds a factor of 10.

6 Conclusion

In this paper we highlight the efficiency of the new algorithm that we propose namely the adaptive statistical Romberg method. A natural question is to produce an analogous study proving the the efficiency of importance sampling routines when used together with the so-called Multilevel Monte Carlo method. This latter method introduced by Giles in [9] reduces the complexity of the Monte Carlo Euler scheme procedure to the order of $n^2 \log n$. Proving a central limit theorem on the adaptive multilevel Monte Carlo algorithm does not seem to be immediate. In fact, this task requires a thorough study and will be the object of a forthcoming work.

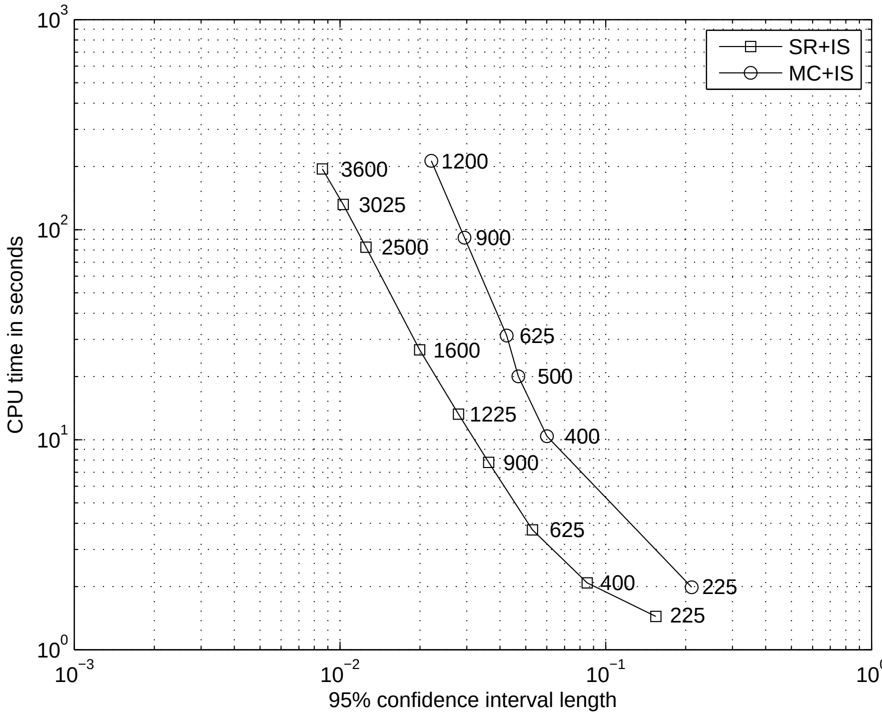

Figure 2: *CPU time versus the* 95%*-confidence interval length*

APPENDIX: PROOF OF THEOREM 2.1

Let $\theta \in \mathbb{R}^q$. First, for U^θ given by (6) we have to prove that $\sup_{0\le t\le T}|U_t^\theta|$ is in L^p. Using the integral form of the process, if F_t^1, $t \in [0,T]$, denotes the associated first term on the right-hand side of (6), then by the Hölder inequality and the boundedness of $\dot{b}$, $\dot{\sigma}_j$, $\{1,\cdots,q\}$, there is $c_1 > 0$ such that

$$\mathbb{E}\sup_{0\le s\le t}|F_s^1|^p \le \mathbb{E}\Big(\int_0^t \big(|\dot{b}(X_s^\theta)| + \sum_{j=1}^q |\theta_j||\dot{\sigma}_j(X_s^\theta)|\big)|U_s^\theta|\Big)^p ds \le c_1 \int_0^t \mathbb{E}|U_s^\theta|^p ds. \tag{35}$$

If F_t^2, $t \in [0,T]$, denotes the second term on the right-hand side of (6), then by Burkholder-Davis-Gundy's inequality there exists a constant $C_p > 0$ depending on p such that

$$\mathbb{E}\sup_{0\le s\le t}|F_s^2|^p \le q^{p-1}\sum_{j=1}^q \mathbb{E}\sup_{0\le s\le t}\Big|\int_0^s \dot{\sigma}_j(X_v^\theta)U_v^\theta dW_v^j\Big|^p \le q^{p-1}C_p \sum_{j=1}^q \mathbb{E}\Big(\int_0^t |\dot{\sigma}_j(X_s^\theta)|^2|U_s^\theta|^2 ds\Big)^{p/2}.$$

Thanks to the Hölder inequality and the boundedness of $\dot\sigma$, there is a constant $c_2 > 0$ such that

$$\mathbb{E} \sup_{0\le s\le t} |F_s^2|^p \le c_2 \int_0^t \mathbb{E}|U_s^\theta|^p ds. \tag{36}$$

Now, if F_t^3, $t \in [0,T]$, denotes the third term on the right-hand side of (6) then using the same arguments as above together with the linear growth assumption on σ and property $(\mathcal{P})$ we get the existence of $c_3 > 0$ such that

$$\mathbb{E} \sup_{0\le s\le t} |F_s^3|^p \le \frac{q^{2p-2}}{2^{p/2}} \sum_{j,\ell=1}^q \mathbb{E}\Big(\int_0^t |\dot\sigma_j(X_s^\theta)|^2 |\sigma_\ell(X_s^\theta)|^2 ds\Big)^{p/2} \le c_3. \tag{37}$$

So (35), (36), (37), and the inequality $(a+b+c)^p \le 3^{p-1}(a^p+b^p+c^p)$ tell us that there exists A and B depending on b, σ, θ, p, q and T such that

$$\mathbb{E} \sup_{0\le s\le t} |U_s^\theta|^p \le A + B\int_0^t \mathbb{E} \sup_{0\le v\le s} |U_v^\theta|^p ds.$$

Hence Gronwall's lemma yields $\mathbb{E}\sup_{0\le s\le t} |U_s^\theta|^p \le Ae^{Bt}$ for all $t\in[0,T]$ (see e.g. [5] page 269). Now, the same proof holds for $U^{n,\theta}$, where the constants obtained in the corresponding upper bound do not depend on the parameter n. Hence, we obtain the first assertion of the theorem namely $\sup_{0\le s\le T} |U_t^\theta|$ and $\sup_{0\le s\le T} |U_t^{n,\theta}|$ are in L^p, $p\ge 1$.

We now proceed to control the quantity $\mathbb{E}\sup_{0\le s\le t} |U_s^\theta - U_s^{n,\theta}|^p$ and we write

$$U_t^\theta - U_t^{n,\theta} = G_t^1 + G_t^2 + G_t^3, \quad \text{for all } t\in[0,T],$$

with G^1 is the drift term, G^2 is the sum of the stochastic integrals terms with respect to the Brownian motion W and G^3 is the sum of the stochastic integrals terms with respect to the Brownian motion $\tilde W$. Concerning the first term G^1, we write it as follows

$$\begin{aligned} G_t^1 = &\int_0^t \Big(\dot b(X_s^\theta) + \sum_{j=1}^q \theta_j \dot\sigma_j(X_s^\theta)\Big)(U_s^\theta - U_s^{n,\theta})ds + \int_0^t \Big(\dot b(X_s^\theta) + \sum_{j=1}^q \theta_j\dot\sigma_j(X_s^\theta)\Big)(U_s^{n,\theta} - U_{\eta_n(s)}^{n,\theta})ds \\ &+ \int_0^t \Big(\dot b(X_s^\theta) - \dot b(X_{\eta_n(s)}^{n,\theta}) + \sum_{j=1}^q \theta_j\big(\dot\sigma_j(X_s^\theta) - \dot\sigma_j(X_{\eta_n(s)}^{n,\theta})\big)\Big)U_{\eta_n(s)}^{n,\theta} dt. \end{aligned} \tag{38}$$

If $G_t^{1,1}$, $t\in[0,T]$, denotes the first term on the right-hand side of (38), then by the Hölder inequality and the boundedness of $\dot b$, $\dot\sigma$, there is a constant $c_4 > 0$ such that

$$\mathbb{E} \sup_{0\le s\le t} |G_s^{1,1}|^p \le c_4 \int_0^t \mathbb{E} \sup_{0\le v\le s} |U_v^\theta - U_v^{n,\theta}|^p ds. \tag{39}$$

If $G_t^{1,2}$, $t\in[0,T]$, denotes the second term on the right-hand side of (38), then by the same arguments, there is a constant $c_5 > 0$ such that

$$\mathbb{E} \sup_{0\le s\le t} |G_s^{1,2}|^p \le c_5 \int_0^T \mathbb{E}|U_t^\theta - U_{\eta_n(t)}^{n,\theta}|^p dt.$$

Noticing that

$$U_t^{n,\theta} - U_{\eta_n(t)}^{n,\theta} = \Big(\dot{b}(X_{\eta_n(t)}^{n,\theta}) + \sum_{j=1}^q \theta_j \dot{\sigma}_j(X_{\eta_n(t)}^{n,\theta})\Big) U_{\eta_n(t)}^{n,\theta}(t - \eta_n(t))$$
$$+ \sum_{j=1}^q \dot{\sigma}_j(X_{\eta_n(t)}^{n,\theta}) U_{\eta_n(t)}^{n,\theta}(W_t^j - W_{\eta_n(t)}^j) - \frac{1}{\sqrt{2}} \sum_{j,\ell=1}^q \dot{\sigma}_j(X_{\eta_n(t)}^{n,\theta}) \sigma_\ell(X_{\eta_n(t)}^{n,\theta})(\tilde{W}_t^{\ell j} - \tilde{W}_{\eta_n(t)}^{\ell j}),$$

we get thanks to the Cauchy-Schwarz inequality and the boundedness of $\dot{b}$, $\dot{\sigma}$, the existence of a constant $c_6 > 0$ such that

$$\mathbb{E}|U_t^{n,\theta} - U_{\eta_n(t)}^{n,\theta}|^p \le c_6 \Big(\mathbb{E}|U_{\eta_n(t)}^{n,\theta}|^p (t - \eta_n(t))^p$$
$$+ \sum_{j=1}^q \big(\mathbb{E}|U_{\eta_n(t)}^{n,\theta}|^{2p}\big)^{\frac{1}{2}} \big(\mathbb{E}|W_t^j - W_{\eta_n(t)}^j|^{2p}\big)^{\frac{1}{2}} + \sum_{j,\ell=1}^q \big(\mathbb{E}|\sigma_\ell(X_{\eta_n(t)}^{n,\theta})|^{2p}\big)^{\frac{1}{2}} \big(\mathbb{E}|\tilde{W}_t^{\ell j} - \tilde{W}_{\eta_n(t)}^{\ell j}|^{2p}\big)^{\frac{1}{2}} \Big).$$

Since $\mathbb{E}|W_t^j - W_{\eta_n(t)}^j|^{2p} = \mathbb{E}|\tilde{W}_t^{\ell j} - \tilde{W}_{\eta_n(t)}^{\ell j}|^{2p} = (t - \eta_n(t))^p \frac{2p!}{2^p (p)!}$ and $\sup_{0 \le s \le T} |X_t^{n,\theta}|$ and $\sup_{0 \le s \le T} |U_t^{n,\theta}|$ are in L^{2p} , we use the linear growth of σ to deduce the existence of a constant $c_7 > 0$ such that

$$\mathbb{E} \sup_{0 \le s \le t} |G_s^{1,2}|^p \le \frac{c_7}{n^{p/2}}. \tag{40}$$

If $G_t^{1,3}$, $t \in [0,T]$, denotes the third term on the right-hand side of (38), then using the Lipschitz property on $\dot{b}$, $\dot{\sigma}$, and the Cauchy-Schwarz inequality, we deduce the existence of a constant $c_8 > 0$ such that

$$\mathbb{E} \sup_{0 \le s \le t} |G_s^{1,3}|^p \le c_8 \big(\mathbb{E} \sup_{0 \le t \le T} |X_t^\theta - X_{\eta_n(t)}^{n,\theta}|^{2p}\big)^{\frac{1}{2}} \big(\mathbb{E} \sup_{0 \le t \le T} |U_{\eta_n(t)}^{n,\theta}|^{2p}\big)^{\frac{1}{2}}.$$

Now using property $(\mathcal{P})$, the proposition in page 274 of [5] and $\sup_{0 \le s \le T} |U_t^{n,\theta}| \in L^{2p}$, we deduce the existence of a constant $c_9 > 0$ such that

$$\mathbb{E} \sup_{0 \le s \le t} |G_s^{1,3}|^p \le \frac{c_9}{n^{p/2}}. \tag{41}$$

So (39), (40), (41) tell us that there exists A_1 and B_1 depending on b, σ, θ, p, q and T such that

$$\mathbb{E} \sup_{0 \le s \le t} |G_s^1|^p \le \frac{A_1}{n^{p/2}} + B_1 \int_0^t \mathbb{E} \sup_{0 \le v \le s} |U_v^\theta - U_v^{n,\theta}|^p ds. \tag{42}$$

Concerning the second term G^2, using Burkholder-Davis-Gundy's inequality followed by the Hölder's one we get the existence of a constant $c_{10} > 0$ such that

$$\mathbb{E} \sup_{0 \le s \le t} |G_s^2|^p \le c_{10} \sum_{j=1}^q \int_0^t \mathbb{E}\big|\dot{\sigma}_j(X_s^\theta) U_s^\theta - \dot{\sigma}_j(X_{\eta_n(s)}^{n,\theta}) U_{\eta_n(s)}^{n,\theta}\big|^p ds.$$

The expectation term inside the above integral is bounded up to a multiplicative constant by

$$\mathbb{E}\big|\dot{\sigma}_j(X_s^\theta)(U_s^\theta - U_s^{n,\theta})\big|^p + \mathbb{E}\big|\dot{\sigma}_j(X_s^\theta)(U_s^{n,\theta} - U_{\eta_n(s)}^{n,\theta})\big|^p + \mathbb{E}\big|\big(\dot{\sigma}_j(X_s^\theta) - \dot{\sigma}_j(X_{\eta_n(s)}^{n,\theta})\big) U_{\eta_n(s)}^{n,\theta}\big|^p.$$

The same evaluations used to get relation (42) allow us to handle separately the three terms above. Hence, we deduce in the same manner that there exists A_2 and B_2 depending on b, σ, θ, p, q and T such that

$$\mathbb{E} \sup_{0\le s\le t} |G_s^2|^p \le \frac{A_2}{n^{p/2}} + B_2 \int_0^t \mathbb{E} \sup_{0\le v\le s} |U_v^\theta - U_v^{n,\theta}|^p ds. \tag{43}$$

Concerning the third term G^3, we apply the same arguments again to get the existence of a constant $c_{11} > 0$ such that

$$\mathbb{E} \sup_{0\le s\le t} |G_s^3|^p \le c_{11} \sum_{j,\ell=1}^{q} \mathbb{E} \sup_{0\le s\le t} \left| \dot\sigma_j(X_s^\theta)\sigma_\ell(X_s^\theta) - \dot\sigma_j(X_{\eta_n(s)}^{n,\theta})\sigma_\ell(X_{\eta_n(s)}^{n,\theta}) \right|^p.$$

It follows that the expectation term in the above sum is bounded by

$$\mathbb{E} \sup_{0\le s\le t} \left| \dot\sigma_j(X_s^\theta)\sigma_\ell(X_s^\theta) - \dot\sigma_j(X_{\eta_n(s)}^{\theta})\sigma_\ell(X_{\eta_n(s)}^{\theta}) \right|^p + \mathbb{E} \sup_{0\le s\le t} \left| \dot\sigma_j(X_{\eta_n(s)}^\theta)\sigma_\ell(X_{\eta_n(s)}^\theta) - \dot\sigma_j(X_{\eta_n(s)}^{n,\theta})\sigma_\ell(X_{\eta_n(s)}^{n,\theta}) \right|^p.$$

Since σ is a Lipschitz continuous function with linear growth and $\dot\sigma$ is a Lipschitz continuous bounded function, we use again the proposition in page 274 of [5] (respectively property $(\mathcal{P})$) to get a control on the first term (respectively on the second term) of the above expression. Hence, there exists a positive constant A_3 depending on σ, θ, p, q and T such that

$$\mathbb{E} \sup_{0\le s\le t} |G_s^3|^p \le \frac{A_3}{n^{p/2}}. \tag{44}$$

Finally putting together relations (42), (43) and (44), we complete the proof by using the Gronwall lemma. □

References

[1] C. Andrieu, E. Moulines, and P. Priouret. Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.*, 44(1):283–312, 2005.

[2] B. Arouna. Adaptative Monte Carlo method, a variance reduction technique. *Monte Carlo Methods Appl.*, 10(1):1–24, 2004.

[3] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probab. Theory Related Fields*, 104(1):43–60, 1996.

[4] 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.

[5] N. Bouleau and D. Lépingle. *Numerical methods for stochastic processes.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1994. A Wiley-Interscience Publication.

[6] H. F. Chen, G. Lei, and A. J. Gao. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.*, 27(2):217–231, 1988.

[7] H. F. Chen and Y. M. Zhu. Stochastic approximation procedures with randomly varying truncations. *Sci. Sinica Ser. A*, 29(9):914–926, 1986.

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

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

[10] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

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

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

[13] P. E. Kloeden and E. Platen. Numerical methods for stochastic differential equations. In *Nonlinear dynamics and stochastic mechanics*, CRC Math. Model. Ser., pages 437–461. CRC, Boca Raton, FL, 1995.

[14] H. Kunita. *Stochastic flows and stochastic differential equations*, volume 24 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997. Reprint of the 1990 original.

[15] H. J. Kushner and 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.

[16] B. Lapeyre and J. Lelong. A framework for adaptive Monte Carlo procedures. *Monte Carlo Methods Appl.*, 17(1):77–98, 2011.

[17] J. Lelong. Almost sure convergence for randomly truncated stochastic algorithms under verifiable conditions. *Statist. Probab. Lett.*, 78(16):2632–2636, 2008.

[18] V. Lemaire and G. Pagès. Unconstrained recursive importance sampling. *Ann. Appl. Probab.*, 20(3):1029–1067, 2010.

[19] M. Pelletier. Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim.*, 39(1):49–72 (electronic), 2000.

[20] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509, 1990.