

Unconstrained Recursive Importance Sampling

Vincent Lemaire, Gilles Pagès

▶ **To cite this version:**

Vincent Lemaire, Gilles Pagès. Unconstrained Recursive Importance Sampling. The Annals of Applied Probability, 2010, 20 (3), pp.1029-1067. 10.1214/09-AAP650 . hal-00293466v3

HAL Id: hal-00293466
https://hal.science/hal-00293466v3

Submitted on 8 Jun 2009

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

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

Unconstrained Recursive Importance Sampling

Vincent Lemaire and Gilles Pagès *

June 8, 2009

Abstract

We propose an *unconstrained* stochastic approximation method of finding the optimal measure change (in an *a priori* parametric family) for Monte Carlo simulations. We consider different parametric families based on the Girsanov theorem and the Esscher transform (or exponential-tilting). In a multidimensional Gaussian framework, Arouna uses a projected Robbins-Monro procedure to select the parameter minimizing the variance (see [2]). In our approach, the parameter (scalar or process) is selected by a classical Robbins-Monro procedure without projection or truncation. To obtain this unconstrained algorithm we intensively use the regularity of the density of the law without assume smoothness of the payoff. We prove the convergence for a large class of multidimensional distributions and diffusion processes.

We illustrate the effectiveness of our algorithm via pricing a Basket payoff under a multidimensional NIG distribution, and pricing a barrier options in different markets.

Key words: Stochastic algorithm, Robbins-Monro, Importance sampling, Esscher transform, Girsanov, NIG distribution, Barrier options.
2000 Mathematics Subject Classification: 65C05, 65B99, 60H35.

1 Introduction

The basic problem in Numerical Probability is to *optimize* some way or another the computation by a Monte Carlo simulation of a real quantity m known by a probabilistic representation

$$m = \mathbb{E}[F(X)]$$

where $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, |\,.\,|_E)$ is a random vector having values in a Banach space E and $F : E \to \mathbb{R}$ is a Borel function (and $F(X)$ is square integrable). The space E is $\mathbb{R}^d$ but can also be a functional space of paths of a process $X = (X_t)_{t\in[0,T]}$. However, in this introduction section, we will first focus on the finite dimensional case $E = \mathbb{R}^d$.

Assume that X has an absolutely continuous distribution $\mathbb{P}_X(dx) = p(x)\lambda_d(dx)$ (λ_d denotes the Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}or(\mathbb{R}^d))$) and that $F \in L^2(\mathbb{P}_X)$ with $\mathbb{P}(F(X) \neq 0) > 0$ (otherwise the expectation is clearly 0 and the problem is meaningless). Furthermore we assume that the probability density p is *everywhere positive* on $\mathbb{R}^d$.

The paradigm of importance sampling applied to a parametrized family of distributions is the following: consider the family of absolutely continuous probability distributions $\pi_\theta(dx) := p_\theta(x)dx$, $\theta \in \Theta$, such that $p_\theta(x) > 0$, $\lambda_d(dx)$-*a.e.* . One may assume without loss of generality that Θ is an open non empty connected subset of $\mathbb{R}^q$ containing 0 so that $p_0 = p$. In fact we will assume

*Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Pierre et Marie Curie, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 5, France. E-mail:vincent.lemaire@upmc.fr, gilles.pages@upmc.fr

throughout the paper that $\Theta = \mathbb{R}^q$. Then for any $\mathbb{R}^d$-valued random variable $X^{(\theta)}$ with distribution π_θ, we have

$$\mathbb{E}[F(X)] = \mathbb{E}\left[F(X^{(\theta)})\frac{p(X^{(\theta)})}{p_\theta(X^{(\theta)})}\right]. \tag{1.1}$$

Among all these random variables having the same expectation $m = \mathbb{E}[F(X)]$, the one with the lowest variance is the one with the lowest quadratic norm: minimizing the variance amounts to finding the parameter θ^* solution (if any) to the following minimization problem

$$\min_{\theta\in\mathbb{R}^q} V(\theta)$$

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

$$V(\theta) := \mathbb{E}\left[F^2(X^{(\theta)})\frac{p^2(X^{(\theta)})}{p_\theta^2(X^{(\theta)})}\right] = \mathbb{E}\left[F^2(X)\frac{p(X)}{p_\theta(X)}\right] \le +\infty. \tag{1.2}$$

A typical situation is importance sampling by mean translation in a finite dimensional Gaussian framework *i.e.*

$$X^{(\theta)} = X + \theta, \quad p(x) = \frac{e^{-\frac{|x|^2}{2}}}{(2\pi)^{\frac{d}{2}}}, \quad p_\theta(x) = p(x-\theta) \quad \text{and} \quad V(\theta) = e^{-|\theta|^2}\mathbb{E}\Big[F^2(X)e^{-2\langle\theta,X\rangle}\Big].$$

Then the second equality in (1.2) is simply the Cameron-Martin formula. This specific framework is very important for applications, especially in Finance, and was the starting point of the new interest for recursive importance sampling procedures, mainly initiated by Arouna in [2] (see further on). In fact, as long as variance reduction is concerned, one can consider a more general framework without extra effort. As a matter of fact, if the distributions p_θ satisfy

$$\begin{cases} (i) & \forall\, x\in\mathbb{R}^d, \quad \theta\mapsto p_\theta(x) \text{ is log-concave} \\ (ii) & \forall\, x\in\mathbb{R}^d, \quad \lim_{|\theta|\to+\infty} p_\theta(x) = 0, \qquad \text{or} \qquad \forall x\in\mathbb{R}^d, \quad \lim_{|\theta|\to+\infty}\frac{p_\theta(x)}{p^2_{\theta/2}(x)} = 0, \end{cases} \tag{$\mathcal{H}_1$}$$

and F satisfies $\mathbb{E}\Big[F^2(X)\frac{p(X)}{p_\theta(X)}\Big] < +\infty$ for every $\theta\in\mathbb{R}^q$, then (see Proposition 1 below), the function V is finite, convex, goes to infinity at infinity. As a consequence $\operatorname{Argmin} V = \{\nabla V = 0\}$ *is non empty*. Assumption (ii) can be localized by by considering that one the two conditions holds only on a Borel set C of $\mathbb{R}^d$ such that $\mathbb{P}_X(C\cap\{F\neq 0\}) > 0$. If $\theta\mapsto p_\theta(x)$ is *strictly* log-concave for every x in a Borel set B such that $\mathbb{P}_X[B\cap\{F\neq 0\}] > 0$, then V is strictly convex and $\operatorname{Argmin} V = \{\nabla V = 0\}$ is reduced to a single $\theta^*\in\mathbb{R}^q$. These results follow from the second representation of V as an expectation in (1.2) which is obtained by a second change of probability (the reverse one). For notational convenience we will temporarily assume that $\operatorname{Argmin} V = \{\theta^*\}$ in this introduction section, although our main result needs no such restriction.

A classical procedure to approximate θ^* is the so-called Robbins-Monro algorithm. This is a recursive stochastic algorithm (see (AlgoRM) below) which can be seen as a stochastic counterpart of deterministic recursive zero search procedures like the Newton-Raphson one. It can be formally implemented provided the gradient of the (convex) target function V admits a representation as an expectation. Since we have no *a priori* knowledge about the regularity of F ([1]) and do not wish

[1] When F is smooth enough alternative approaches have been developed based on some large deviation estimates which provide a good approximation of θ^* by deterministic optimization methods (see [10]).

to have any, we are naturally lead to *formally* differentiate the second representation of V in (1.2) to obtain a representation of ∇V as

$$\nabla V(\theta) = \mathbb{E}\left[F^2(X)\frac{p(X)}{p_\theta(X)}\frac{\nabla_\theta\, p_\theta(X)}{p_\theta(X)}\right]. \tag{1.3}$$

Then, if we consider the function $\bar{H}_V(\theta, x)$ such that $\nabla V(\theta) = \mathbb{E}\left(\bar{H}_V(\theta, X)\right)$ naturally defined by (1.3), the derived Robbins-Monro procedure writes

$$\theta_{n+1} = \theta_n - \gamma_{n+1}\bar{H}_V(\theta_n, X_{n+1}), \tag{AlgoRM}$$

with $(\gamma_n)_{n\geq 0}$ a *step* sequence decreasing to 0 (at an appropriate rate), $(X_n)_{n\geq 0}$ a sequence of i.i.d. random variables with distribution $p(x)\lambda_d(dx)$. To establish the convergence of a Robbins-Monro procedure to $\theta^* = \operatorname{Argmin} V$ requires seemingly not so stringent assumptions. We mean by that: not so different from those needed in a deterministic framework. However, one of them turns out to be quite restrictive for our purpose: the sub-linear growth assumption in quadratic mean

$$\forall \theta \in \mathbb{R}^d, \qquad \left\|\bar{H}_V(\theta, X)\right\|_2 \leq C(1 + |\theta|). \tag{NEC}$$

which is the stochastic counterpart of the classical non-explosion condition needed in a deterministic framework. In practice, this condition is almost never satisfied in our framework due to the behaviour of the term $\frac{p(x)}{p_\theta(x)}$ as θ goes to infinity.

The origin of recursive importance sampling as briefly described above goes back to Kushner and has recently been brought back to light in a Gaussian framework by Arouna in [2]. However, as confirmed by the numerical experiments carried out by several authors ([2, 12, 14]), the regular Robbins-Monro procedure (AlgoRM) does suffer from a structural instability coming from the violation of (NEC). This phenomenon is quite similar to the behaviour of the explicit discretization schemes of an $ODE \equiv \dot{x} = h(x)$ when h has a super-linear growth at infinity. Furthermore, in a probabilistic framework no "implicit scheme" can be devised in general. Then the only way out *mutatis mutandis* is to kill the procedure when it comes close to explosion and to restart it with a smaller step sequence. Formally, this can be described as some repeated projections or truncations when the algorithm leaves a slowly growing compact set waiting for stabilization which is shown to occur *a.s.*. Then, the algorithm behaves like a regular Robbins-Monro procedure. This is the so-called "Projection à la Chen" avatar of the Robbins-Monro algorithm, introduced by Chen in [6, 7] and then investigated by several authors (see *e.g.* [1, 14]) Formally, repeated projections "à la Chen" can be written as follows:

$$\theta_{n+1} = \Pi_{K_{\sigma(n)}}\left\{\theta_n - \bar{H}_V(\theta_n, X_{n+1})\right\} \tag{AlgoP}$$

where $\Pi_{K_{\sigma(n)}}$ denotes the projection on the convex compact $K_{\sigma(n)}$ (K_p is increasing to $\mathbb{R}^d$ as $p \to \infty$). In [14] is established a a Central Limit Theorem for this version of the recursive variance reduction procedure. Some extensions to non Gaussian framework have been carried out by Arouna in his PhD thesis (with some applications to reliability) and more recently to the marginal distributions of a Lévy processes by Kawai in [12].

However, convergence occurs for this procedure after a long "stabilization phase" ... provided that the sequence of compact sets have been specified in an appropriate way. This specification turns out to be a rather sensitive phase of the "tuning" of the algorithm to be combined with that of the step sequence.

In this paper, we show that as soon as the growth of F at infinity can be explicitly controlled, it is always possible to design a regular Robbins-Monro algorithm which *a.s.* converges to a variance minimizer θ^* with no risk of explosion (and subsequently no need of repeated projections).

To this end the key is to introduce a *third* change of probability in order to control the term $\frac{p(x)}{p_\theta(x)}$. In a Gaussian framework this amounts to switching the parameter θ from the density p to the function F by a third mean translation. This of course corresponds to a new function $\bar{H}_V$ but can also be interpreted *a posteriori* as a way to introduce an *adaptive* step sequence (in the spirit of [15]).

In terms of formal importance sampling, we introduce a new positive density q_θ (everywhere positive on $\{p > 0\}$) so that the gradient writes

$$\nabla V(\theta) = \mathbb{E}\left[F^2(\widetilde{X}^{(\theta)})\frac{p^2(\widetilde{X}^{(\theta)})}{p_\theta(\widetilde{X}^{(\theta)})q_\theta(\widetilde{X}^{(\theta)})}\frac{\nabla p_\theta(\widetilde{X}^{(\theta)})}{p_\theta(\widetilde{X}^{(\theta)})}\right] = \mathbb{E}\Big[\widetilde{H}_V(\theta, \widetilde{X}^{(\theta)})\Big], \tag{1.4}$$

where $\widetilde{X}^{(\theta)} \sim q_\theta(x)dx$. The "weight" $\dfrac{p^2(\widetilde{X}^{(\theta)})}{p_\theta(\widetilde{X}^{(\theta)})q_\theta(\widetilde{X}^{(\theta)})}\dfrac{\nabla p_\theta(\widetilde{X}^{(\theta)})}{p_\theta(\widetilde{X}^{(\theta)})}$ may seem complicated but the rôle of the density q_θ is to control the critical term $\frac{p^2(x)}{p_\theta(x)q_\theta(x)}$ by a (deterministic) quantity only depending on θ. Then we can replace $\widetilde{H}_V$ by a function $H(\theta, x) = \delta(\theta)\,\widetilde{H}_V(\theta, x)$ in the above Robbins-Monro procedure (AlgoRM) where δ is a positive function used to control the behaviour of $\widetilde{H}_V(\theta, x)$ for large values of x (note that $\big\{\mathbb{E}\big[H(., \widetilde{X}^{(\theta)})\big] = 0\big\} = \{\nabla V = 0\}$).

We will first illustrate this paradigm in a finite dimensional setting with parametrized importance sampling procedures: the mean translation and the Esscher transform which coincide for Gaussian vectors on which a special emphasis will be put. Both cases correspond to a specific choice of q_θ which significantly simplifies the expression of the weight.

As a second step, we will deal with an infinite dimensional setting (path-dependent diffusion like processes) where we will rely on the Girsanov transform to play the role of mean translator. To be more precise, we want now to compute $\mathbb{E}[F(X)]$ where X is a path-dependent diffusion process and F is a functional defined on the space $\mathcal{C}([0,T], \mathbb{R}^d)$ of continuous functions defined on $[0,T]$. We consider a d-dimensional Itô process $X = (X_t)_{t\in[0,T]}$ solution of the path-dependent SDE

$$\mathrm{d}X_t = b(t, X^t)\,\mathrm{d}t + \sigma(t, X^t)\,\mathrm{d}W_t, \quad X_0 = x \in \mathbb{R}^d, \tag{$E_{b,\sigma,W}$}$$

where $W = (W_t)_{t\in[0,T]}$ is a q-dimensional standard Brownian motion, $X^t := (X_{t\wedge s})_{s\in[0,T]}$ is the stopped process at time t, $b : [0,T] \times \mathcal{C}([0,T], \mathbb{R}^d) \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathcal{C}([0,T], \mathbb{R}^d) \to \mathcal{M}(d,q)$ are Lipschitz with respect to the $\|\,.\,\|_\infty$ on the space $\mathcal{C}([0,T], \mathbb{R}^d)$ and continuous in $(t,x) \in [0,T] \times \mathcal{C}([0,T], \mathbb{R}^d)$ (see [18] for more details about these path-dependent SDE's).

Let φ be a fixed borel bounded functional on $\mathcal{C}([0,T], \mathbb{R}^d)$ with values in $\mathcal{M}(q,p)$ (where $p \geq 1$ is a free integral parameter). Then a Girsanov transform yields that for every $\theta \in L^2_{T,p} := L^2([0,T], \mathbb{R}^p)$,

$$\mathbb{E}[F(X)] = \mathbb{E}\left[F(X^{(\theta)})e^{-\int_0^T \langle \varphi(X^{(\theta),s})\theta(s), \mathrm{d}W_s\rangle - \frac{1}{2}\left\|\varphi(X^{(\theta),\cdot})\theta\right\|^2_{L^2_{T,q}}}\right]$$

where $X^{(\theta)}$ is the solution to $(E_{b+\sigma\varphi\theta,\sigma})$. The functional to be minimized is now

$$V(\theta) = \mathbb{E}\left[F(X^{(\theta)})^2 e^{-2\int_0^T \langle \varphi(X^{(\theta),s})\theta(s), \mathrm{d}W_s\rangle - \left\|\varphi(X^{(\theta),\cdot})\theta\right\|^2_{L^2_{T,q}}}\right], \qquad \theta \in L^2_{T,p}.$$

In practice we will only minimize V over a *finite dimensional subspace* of $E = \mathrm{span}\{e_1, \ldots, e_m\} \subset L^2_{T,p}$.

The paper is organized as follows. Section 2 is devoted to the finite dimensional setting where we recall the main tool including a slight extension of the Robbins-Monro theorem in the Subsection 2.1 and the gaussian case investigated in [2] is revisited to emphasize the new aspects of our algorithm in the Subsection 2.2.

In Section 2 we successively investigate the translation for log-concave distributions probability and the Esscher transform. In Section 3 we introduce a functional version of our algorithm based on the Girsanov theorem to deal the SDE. In Section 4 we provide some comments on the practical implementation and in Section 5 some numerical experiments are carried out on some option pricing problems.

NOTATIONS: • We will denote by $S > 0$ the fact that a symmetric matrix S is positive definite. $|\,.\,|$ will denote the canonical Euclidean norm on $\mathbb{R}^m$ and $\langle\,.\,,\,.\,\rangle$ will denote the canonical inner product. • The real constant $C > 0$ denotes a positive real constant that may vary from line to line. • $\|f\|_{L^2_{T,p}} := \left(\int_0^T f_1^2(t) + \cdots + f_p^2(t)dt\right)^{\frac{1}{2}}$ if $f = (f_1, \ldots, f_p)$ is an $\mathbb{R}^p$-valued (class of) Borel function(s).

2 The finite-dimensional setting

2.1 Argmin V as a target

Proposition 1 *Suppose $(\mathcal{H}_1)$ holds.*
Then the function V defined by (1.2) is convex and $\lim_{|\theta|\to+\infty} V(\theta) = +\infty$. As a consequence

$$\operatorname{Argmin} V = \{\nabla V = 0\} \neq \emptyset.$$

Proof. By the change of probability $\frac{d\pi_\theta}{d\lambda_d}$ we have $V(\theta) = \mathbb{E}\Big[F^2(X)\frac{p(X)}{p_\theta(X)}\Big]$. Let x fixed in $\mathbb{R}^d$. The function $(\theta \mapsto \log p_\theta(x))$ is concave, hence $\log(1/p_\theta(x)) = -\log p_\theta(x)$ is convex so that, owing to the Young Inequality, the function $\frac{1}{p_\theta(x)}$ is convex since it is non-negative.

To prove that V tends to infinity as $|\theta|$ goes to infinity, we consider two cases:

– If $\lim\limits_{|\theta|\to+\infty} p_\theta(x) = 0$ for every $x \in \mathbb{R}^d$, the result is trivial by Fatou's Lemma.

– If $\lim\limits_{|\theta|\to+\infty} \dfrac{p_\theta(x)}{p^2_{\theta/2}(x)} = 0$ for every $x \in \mathbb{R}^d$, we apply the reverse Hölder inequality with conjugate exponents $(\frac{1}{3}, -\frac{1}{2})$ to obtain

$$\begin{aligned} V(\theta) &\geq \mathbb{E}\left[F^{2/3}(X)\left(\frac{p^2_{\theta/2}(X)}{p(X)p_\theta(X)}\right)^{\frac{1}{3}}\right]^3 \mathbb{E}\left[\left(\frac{p(X)}{p_{\theta/2}(X)}\right)^{-1}\right]^{-2}, \\ &\geq \mathbb{E}\left[F^{2/3}(X)\left(\frac{p^2_{\theta/2}(X)}{p(X)p_\theta(X)}\right)^{\frac{1}{3}}\right]^3, \end{aligned}$$

(p and p_θ are probability density functions). One concludes again by Fatou's Lemma. ◇

The set Argmin V, or to be precise, the random vectors taking values in Argmin V will the target(s) of our new algorithm. If V is strictly convex, *e.g.* if

$$\mathbb{P}[X \in \{p_\cdot(x) \text{ strictly log-concave and } F(x) \neq 0\}] > 0,$$

then Argmin $V = \{\theta^*\}$. Nevertheless this will not be necessary owing to the combination of the two results that follow.

Lemma 1 *Let $U : \mathbb{R}^d \to \mathbb{R}_+$ be a convex differentiable function, then*

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

Furthermore, if $\operatorname{Argmin} U$ *is nonempty, it is a convex closed set (which coincide with* $\{\nabla U = 0\}$*) and*

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

A sufficient (but in no case necessary) condition for a nonnegative convex function U to attain a minimum is that $\lim_{|x|\to\infty} U(x) = +\infty$.

Now we pass to the statement of the convergence theorem on which we will rely throughout the paper. It is a slight variant of the regular Robbins-Monro procedure whose proof is rejected in an annex.

Theorem 1 *(Extended Robbins-Monro Theorem) Let $H : \mathbb{R}^q \times \mathbb{R}^d \to \mathbb{R}^d$ a Borel function and X an $\mathbb{R}^d$-valued random vector such that $\mathbb{E}[|H(\theta, X)|] < +\infty$ for every $\theta \in \mathbb{R}^d$. Then set*

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

Suppose that the function h is continuous and that $\mathcal{T}^* := \{h = 0\}$ satisfies

$$\forall\, \theta \in \mathbb{R}^d \setminus \mathcal{T}^*,\ \forall\, \theta^* \in \mathcal{T}^*, \qquad \langle \theta - \theta^*, h(\theta) \rangle > 0. \tag{2.5}$$

Let $\gamma = (\gamma_n)_{n\geq 1}$ be a sequence of gain parameters satisfying

$$\sum_{n\geq 1} \gamma_n = +\infty \qquad \text{and} \qquad \sum_{n\geq 1} \gamma_n^2 < +\infty. \tag{2.6}$$

Suppose that

$$\forall\, \theta \in \mathbb{R}^d, \qquad \mathbb{E}\big[|H(\theta, X)|^2\big] \leq C(1 + |\theta|^2) \tag{NEC}$$

(which implies $|h(\theta)|^2 \leq C(1 + |\theta|^2)$).

Let $(X_n)_{n\geq 1}$ be an i.i.d. sequence of random vectors having the distribution of X, a random vector θ_0, independent of $(X_n)_{n\geq 1}$ satisfying $\mathbb{E}\big[|\theta_0|^2\big] < +\infty$, all defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then, the recursive procedure defined by

$$\theta_n = \theta_{n-1} - \gamma_{n+1} H(\theta_n, X_{n+1}), \qquad n \geq 1 \tag{2.7}$$

satisfies:

$$\exists\, \theta_\infty : (\Omega, \mathcal{A}) \to \mathcal{T}^*,\ \theta_\infty \in L^2(\mathbb{P}), \quad \text{such that} \quad \theta_n \xrightarrow{a.s.} \theta_\infty. \tag{2.8}$$

The convergence also holds in $L^p(\mathbb{P})$, $p \in (0, 2)$.

The proof is postponed to the Appendix at the end of the paper. The natural way to apply this theorem for our purpose is the following:

- Step 1: we will show that the convex function V in (1.2) is differentiable with a gradient ∇V having a representation as an expectation formally given $\nabla V(\theta) = \mathbb{E}[\nabla_\theta v(\theta, X)]$.
- Step 2: then set $H(\theta, x) := \rho(\theta)\nabla_\theta v(\theta, x)$ where ρ is a (strictly) positive function on $\mathbb{R}^q$. As a matter of fact, with the notations of the above theorem

  $$\langle \theta - \theta^*, h(\theta) \rangle = \delta(\theta) \langle \theta - \theta^*, \nabla V(\theta) \rangle,$$

  so that $\mathcal{T}^* = \operatorname{Argmin} V$ and (2.5) is satisfied (set $U := V$ in Lemma 1).
- Step 3: Specify in an appropriate way the function δ so that the linear quadratic growth assumption is satisfied. This is the sensitive point that will lead us to modify the structure more deeply by finding a new representation of ∇V as an expectation not directly based on the local gradient $\nabla_\theta v(\theta, x)$.

2.2 A first illustration: the Gaussian case revisited

The Gaussian is the framework of [2]. It is also a kind of introduction to the infinite dimensional diffusion setting investigated in Section 3. In the Gaussian case, the natural importance sampling density is the translation of the gaussian density: $p_\theta(x) = p(x-\theta)$ for $\theta \in \mathbb{R}^d$ (*i.e.* $q = d$). We have

$$p_\theta(x) = e^{-\frac{|\theta|^2}{2} + \langle \theta, X\rangle} p(x).$$

The assumption $(\mathcal{H}_1)$ is clearly satisfied by the Gaussian density, and we assume that F satisfies $\mathbb{E}\big[F^2(X)e^{-\langle\theta,X\rangle}\big] < +\infty$ so that V is well defined.

In [2], Arouna considers the function $\bar{H}_V(\theta, x)$ defined by

$$\bar{H}_V(\theta, x) = F^2(x) e^{\frac{|\theta|^2}{2} - \langle \theta, x\rangle}(\theta - x).$$

It is clear that the condition (NEC) is not satisfied even if we simplify this function by $e^{|\theta|^2/2}$ (which does not modify the problem).

A first approach: When $F(X)$ have finite moments of any order, a naive way to control directly $\big\|\bar{H}_V(\theta_n, X_{n+1})\big\|_2$ by an explicit deterministic function of θ (in order to rescale it) is to proceed as follows: one derives from Hölder Inequality that for every couple (r, s), r, $s > 1$ of conjugate exponents

$$\big\|\bar{H}_V(\theta, X)\big\|_2 \le \big\|(\theta - X)F^2(X)\big\|_{2r} \Big\|e^{-\langle\theta,X\rangle}\Big\|_{2s} e^{\frac{|\theta|^2}{2}}.$$

Setting $r = 1 + \frac{1}{\varepsilon}$ and $s = 1 + \varepsilon$, yields

$$\big\|\bar{H}_V(\theta, X)\big\|_2 \le \Big(\big\|XF^2(X)\big\|_{2(1+\frac{1}{\varepsilon})} + \big\|F^2(X)\big\|_{2(1+\frac{1}{\varepsilon})}|\theta|\Big) e^{\left(\frac{3}{2}+\varepsilon\right)|\theta|^2}.$$

Then, $\bar{H}_\varepsilon(\theta, x) := e^{-\left(\frac{3}{2}+\varepsilon\right)|\theta|^2}\bar{H}_V(\theta, x)$ satisfies the condition (NEC) and theoretically the standard Robbins-Monro algorithm implemented with $\bar{H}_\varepsilon$ *a.s.* converges and no projection nor truncation is needed. Numerically, the solution is not satisfactory because the correcting factor $e^{-\left(\frac{3}{2}+\varepsilon\right)|\theta|^2}$ goes to zero much too fast as θ goes to infinity: if at any iteration at the beginning of the procedure θ_n is sent "too far", then it is frozen instantly. If ε is too small it will simply not prevent explosion. The tuning of ε becomes quite demanding and payoff dependent. This is in complete contradiction with our aim of *a self-controlled variance reducer*. A more robust approach needs to be developed. On the other hand this kind of behaviour suggests that we are not in the right asymptotics to control $\big\|\bar{H}_V(\theta_n, X_{n+1})\big\|_2$.

Note however that when F is bounded with a compact support, then one can set $\varepsilon = 0$ and the above approach provides an efficient answer to our problem.

A general approach: We consider the density

$$q_\theta(x) = \frac{p^2(x)}{p(x-\theta)} = p(x+\theta).$$

By (1.4), we have

$$\nabla V(\theta) = \mathbb{E}\left[F^2(\widetilde{X}^{(\theta)})\frac{\nabla p(\widetilde{X}^{(\theta)} - \theta)}{p(\widetilde{X}^{(\theta)} - \theta)}\right],$$

with $\widetilde{X}^{(\theta)} \sim p(x+\theta)dx$, *i.e.* $\widetilde{X}^{(\theta)} = X - \theta$. Since p is the Gaussian density, we have $\frac{\nabla p(x)}{p(x)} = -x$. As a consequence, the function H_V defined by

$$H_V(\theta, x) = F^2(x-\theta)(2\theta - x),$$

provides a representation $\nabla V(\theta) = \mathbb{E}[H_V(\theta, X)]$ of the gradient ∇V. As soon as F is bounded, this function satisfies the condition (NEC). Otherwise, we note that thanks to this new change of variable the parameter θ lies now *inside* the payoff function F and that the *exponential term* has disappeared from the expectation. If we have an *a priori* control on the function $F(x)$ as $|x|$ goes to infinity, say

$$\exists \lambda \in \mathbb{R}_+, \quad \forall x \in \mathbb{R}^d, \quad |F(x)| \leq c_F e^{\lambda |x|},$$

then we can consider the function $H_\lambda(\theta, x) = e^{-\lambda|\theta|} H_V(\theta, x)$ which satisfies

$$\begin{aligned} \|H_\lambda(\theta, X)\|_2 &\leq c_F^2 \left\| e^{2\lambda|X|}(2\theta - X) \right\|_2, \\ &\leq C(1 + |\theta|). \end{aligned}$$

The resulting Robbins-Monro algorithm reads

$$\theta_{n+1} = \theta_n - \gamma_{n+1} e^{-\lambda|\theta_n|} F^2(X_{n+1} - \theta_n)(2\theta_n - X_{n+1}).$$

We no longer to tune the correcting factor and one verifies on simulations that it does not suffer from freezing in general. In case of a too dissymmetric function F this may still happen but a *self-controlled* variant is proposed in Section 2.3 below to completely get rid of this effect (which cannot be compared to an explosion).

2.3 Translation of the mean: the general strongly unimodal case

We consider importance sampling by mean translation, namely we set

$$\forall x \in \mathbb{R}^d, \quad p_\theta(x) = p(x - \theta),$$

for $\theta \in \mathbb{R}^d$.

In this section we assume that

$$p \text{ is log-concave and } \lim_{|x| \to \infty} p(x) = 0,$$

so that $(\mathcal{H}_1)$ holds. Moreover, we make the following additional assumption on the probability density p

$$\exists\, a \in [1, 2] \text{ such that } \begin{cases} (i) & \frac{|\nabla p|}{p}(x) = O(|x|^{a-1}) \text{ as } |x| \to \infty \\ (ii) & \exists \delta > 0,\ \log p(x) + \delta |x|^a \text{ is convex.} \end{cases} \qquad (\mathcal{H}_a^{tr})$$

First we will use (1.2) to differentiate V since

Proposition 2 *Suppose $(\mathcal{H}_1)$ and $(\mathcal{H}_a^{tr})$ are satisfied and the function F satisfies*

$$\forall\, \theta \in \mathbb{R}^d, \quad \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right] < +\infty \quad \text{and} \quad \forall\, C > 0, \quad \mathbb{E}\left[F^2(X) e^{C|X|^{a-1}}\right] < +\infty. \qquad (2.9)$$

Then V is finite and differentiable on $\mathbb{R}^d$ with a gradient given by

$$\nabla V(\theta) = \mathbb{E}\left[F^2(X)\frac{p(X)}{p^2(X-\theta)}\nabla p(X-\theta)\right] \qquad (2.10)$$

$$= \mathbb{E}\left[F^2(X-\theta)\frac{p^2(X-\theta)}{p(X)p(X-2\theta)}\frac{\nabla p(X-2\theta)}{p(X-2\theta)}\right]. \qquad (2.11)$$

Proof. The formal differentiation to get (2.10) from (1.2) is obvious. So it remains to check the domination property for θ lying inside a compact set. Let $x \in \mathbb{R}^d$ and $\theta \in \mathbb{R}^d$. The log-concavity of p implies that

$$\log p(x) \leq \log p(x-\theta) + \frac{\langle \nabla p(x-\theta), \theta \rangle}{p(x-\theta)}$$

so that

$$0 \leq \frac{p(x)}{p(x-\theta)} \leq \exp \frac{|\nabla p(x-\theta)|}{p(x-\theta)} |\theta|.$$

Using the assumption $(\mathcal{H}_a^{tr})$ yields, for every $\theta \in B(0,R)$,

$$\begin{aligned} F^2(X) \frac{p(X)}{p^2(X-\theta)} \nabla p(X-\theta) &\leq F^2(X)(A|X|^{a-1}+B) e^{(A|X-\theta|^{a-1}+B)|\theta|} \\ &\leq C_R F^2(X) e^{C'|X|^{a-1}} =: g(X) \in L^1(\mathbb{P}) \end{aligned}$$

To derive the second expression (2.11) for the gradient, we proceed as follows: an elementary change of variable shows that

$$\begin{aligned} \nabla V(\theta) &= \int_{\mathbb{R}^d} F^2(x) \frac{p(x)}{p^2(x-\theta)} p(x) \nabla p(x-\theta) dx \\ &= \int_{\mathbb{R}^d} F^2(x-\theta) \frac{p^2(x-\theta)}{p^2(x-2\theta)} \nabla p(x-2\theta) dx \\ &= \mathbb{E}\left[F^2(X-\theta) \frac{p^2(X-\theta)}{p(X)p(X-2\theta)} \frac{\nabla p(X-2\theta)}{p(X-2\theta)} \right]. \quad \diamond \end{aligned}$$

Remark. The second change of variable (in (1.2)) has been processed to withdraw the parameter θ from the possible non smooth function F to make possible the differentiation of V (since p is smooth). The second expression (2.11) results form a *third* translation of the variable in order to plug back the parameter θ into the function F which in common applications has a known controlled growth rate at infinity.

This last statement may look strange at a first glance since θ appears in the "weight" term of the expectation that involves the probability density p. However, when $X \stackrel{d}{=} \mathcal{N}(0;1)$, this term can be controlled easily since it reduces to

$$\frac{p^2(x-\theta)}{p(x)p(x-2\theta)} \frac{\nabla p(x-2\theta)}{p(x-2\theta)} = e^{\theta^2}(2\theta - x).$$

The following lemma shows that, more generally in our strongly unimodal setting, if $(\mathcal{H}_1)$ and $(\mathcal{H}_a^{tr})$ are satisfied, this "weight" can always be controlled by a deterministic function of θ.

Lemma 2 *If $(\mathcal{H}_a^{tr})$ holds, then there exists two real constants A, B such that*

$$\frac{p^2(x-\theta)}{p(x)p(x-2\theta)} \frac{|\nabla p(x-2\theta)|}{p(x-2\theta)} \leq e^{2\delta|\theta|^a} \left(A|x|^{a-1} + A|\theta|^{a-1} + B \right). \tag{2.12}$$

Proof. Let f be the convex function defined on $\mathbb{R}^d$ by $f(x) = \log p(x) + \delta|x|^a$. Then, for every x, $\theta \in \mathbb{R}^d$,

$$\log\left(\frac{p^2(x-\theta)}{p(x)p(x-2\theta)} \right) = \log\left(\frac{f^2(x-\theta)}{f(x)f(x-2\theta)} \right) + \delta\left(|x|^a + |x-2\theta|^a - 2|x-\theta|^a \right)$$

Note that $x-\theta = \frac{1}{2}(x+(x-2\theta))$. Then, using the log-convexity of f and the elementary inequality

$$\forall\, u,\, v \in \mathbb{R}^d, \quad |u|^a + |v|^a \leq 2\left(\left|\frac{u+v}{2}\right|^a + \left|\frac{u-v}{2}\right|^a\right)$$

(valid if $a \in (0,2]$) yields

$$\frac{p^2(x-\theta)}{p(x)p(x-2\theta)} \leq e^{2\delta|\theta|^a}. \tag{2.13}$$

One concludes by the point (i) of $(\mathcal{H}_a^{tr})$. $\diamond$

Remark. Thus the normal distribution satisfies $(\mathcal{H}_a^{tr})$ with $a = 2$ and $\delta = 1/2$. Moreover, note that the last inequality in the above proof holds as an equality.

Now we are in position to derive an unconstraint (extended) Robbins-Monro algorithm to minimize the function V, provided the function F satisfies a sub-multiplicative control property, in which $c > 0$ is a real parameter and $\widetilde{F}$ a function from $\mathbb{R}^d$ to $\mathbb{R}_+$, such that, namely

$$\left\{\begin{array}{l} \forall\, x,\, y \in \mathbb{R}^d, \quad |F(x)| \leq \widetilde{F}(x) \ \text{ and } \ \widetilde{F}(x+y) \leq C(1+\widetilde{F}(x))^c(1+\widetilde{F}(y))^c \\ \\ \qquad\qquad \mathbb{E}\Big[|X|^{2(a-1)}\widetilde{F}(X)^{4c}\Big] < +\infty. \end{array}\right. \tag{$\mathcal{H}_c^{tr}$}$$

Remark. Assumption $(\mathcal{H}_c^{tr})$ seems almost non-parametric. However, its field of application is somewhat limited by $(\mathcal{H}_a^{tr})$ for the following reason: if there exists a positive real number $\eta > 0$ such that $x \mapsto \log p(x) + \eta|x|^a$ is concave, then $p(x) \leq Ce^{-\eta|x|^a}(|x|+1)$ for some real constant $C > 0$; which in turn implies that the function $\widetilde{F}$ in $(\mathcal{H}_c^{tr})$ needs to satisfy $\widetilde{F}(x) \leq C'e^{\lambda|x|^b}$ for some $b \in (0,a)$ and some $\lambda > 0$. (Then $c = c_b$ with $c_b = 1$ if $b \in [0,1]$ and $c_b = 2^{\frac{b}{2}}$ if $b \in (1,a)$, when $a > 1$).

Theorem 2 *Suppose X and F satisfy $(\mathcal{H}_1)$, $(\mathcal{H}_a^{tr})$, (2.9) and $(\mathcal{H}_c^{tr})$ for some parameters $a \in (0,2]$, $b \in (0,a)$ and $\lambda > 0$, and that the step sequence $(\gamma_n)_{n\geq 1}$ satisfies the usual decreasing step assumption*

$$\sum_{n\geq 1}\gamma_n = +\infty \qquad \textit{and} \qquad \sum_{n\geq 1}\gamma_{n+1}^2 < +\infty.$$

Then the recursive procedure defined by

$$\theta_{n+1} = \theta_n - \gamma_{n+1}H(\theta_n, X_{n+1}), \qquad \theta_0 \in \mathbb{R}^d \tag{2.14}$$

where $(X_n)_{n\geq 1}$ is an i.i.d. sequence with the same distribution as X and

$$H(\theta,x) := \frac{F^2(x-\theta)}{1+\widetilde{F}(-\theta)^{2c}}e^{-2\delta|\theta|^a}\frac{p^2(x-\theta)}{p(x)p(x-2\theta)}\frac{\nabla p(x-2\theta)}{p(x-2\theta)}, \tag{2.15}$$

a.s. converges toward an $\operatorname{Argmin} V$*-valued (square integrable) random variable θ^*.*

Proof. In order to apply Theorem 1, we have to check the following fact:

– *Mean reversion*: The mean function of the procedure defined by (2.14) reads

$$h(\theta) = \mathbb{E}[H(\theta,X)] = \frac{e^{-2\delta|\theta|^a}}{1+\widetilde{F}(-\theta)^{2c}}\nabla V(\theta)$$

so that $\mathcal{T}^* := \{h = 0\} = \{\nabla V = 0\}$ and if $\theta^* \in \mathcal{T}^*$ and $\theta \in \mathbb{R}^d \setminus \mathcal{T}^*$,

$$\langle \theta - \theta^*, h(\theta) \rangle = \frac{e^{-2\delta|\theta|^a}}{1 + \widetilde{F}(-\theta)^{2c}} \langle \nabla V(\theta), \theta - \theta^* \rangle > 0$$

for every $\theta \neq \theta^*$.

– *Linear growth of* $\theta \mapsto \|H(\theta, X)\|_2$: All our efforts in the design of the procedure are motivated by this Assumption (NEC) which prevents explosion. This condition is clearly fulfilled by H since

$$\begin{aligned}\mathbb{E}[|H(\theta, X)|]^2 &= \frac{e^{-4\delta|\theta|^a}}{(1 + \widetilde{F}(-\theta)^{2c})^2} \mathbb{E}\left[F^4(X - \theta) \left(\frac{p^2(X - \theta)}{p(X)p(X - 2\theta)} \frac{|\nabla p(X - 2\theta)|}{p(X - 2\theta)} \right)^2 \right], \\ &\leq C e^{-4\delta|\theta|^a} \mathbb{E}\Big[(1 + \widetilde{F}(X)^{2c})^2 (A(|X|^{a-1} + |\theta|^{a-1}) + B)^2 \Big],\end{aligned}$$

where we used Assumption $(\mathcal{H}_c^{tr})$ in the first line and Inequality (2.12) from Lemma 2 in the second line. One derives that there exists a real constant $C > 0$ such that

$$\mathbb{E}[|H(\theta, X)|]^2 \leq C \mathbb{E}\Big[\widetilde{F}(X)^{4c} (1 + |X|)^{2(a-1)} \Big] (1 + |\theta|^{2(a-1)}).$$

This provides the expected condition since $(\mathcal{H}_c^{tr})$ holds. $\diamond$

Examples of distributions

- The normal distribution. Its density is given on $\mathbb{R}^d$ by

  $$p(x) = (2\pi)^{-\frac{d}{2}} e^{-|x|^2/2}, \qquad x \in \mathbb{R}^d.$$

  so that $(\mathcal{H}_1)$ is satisfied as well as $(\mathcal{H}_a^{tr})$ for $a = 2$, $\delta = \frac{1}{2}$. Assumption $(\mathcal{H}_c^{tr})$ is satisfied iff $(b, \lambda) \in (0, 2) \times (0, \infty) \cup \{2\} \times (0, \frac{1}{2})$. Then the function H has a particularly simple form

  $$H(\theta, x) = e^{-\frac{\lambda}{2}|\theta|^b} F^2(x - \theta)(2\theta - x)$$

- *The hyper-exponential distributions*

  $$p(x) = C_{d,a,\sigma} e^{-\frac{|x|^a}{\sigma^a}} P(x), \quad a \in [1, 2]$$

  where P is polynomial function. This wide family includes the normal distributions, the Laplace distribution, the symmetric gamma distributions, etc.

- *The logistic distribution* Its density on the real line is given by

  $$p(x) = \frac{e^x}{(e^x + 1)^2}$$

  $(\mathcal{H}_1)$ is satisfied as well as $(\mathcal{H}_a^{tr})$ for $a = 1 + \eta$ $(\eta \in (0, 1))$, $\delta > 0$. Assumption $(\mathcal{H}_c^{tr})$ is satisfied iff $(b, \lambda) \in (0, 1) \times (0, \infty) \cup \{1\} \times (0, 1)$.

2.4 Exponential change of measure: the Esscher transform

A second classical approach is to consider an exponential change of measure (or Esscher transform). This transformation has already been consider for that purpose in [12] to extend the procedure with repeated projections introduced in [2]. We denote by ψ the cumulant generating function (or log–Laplace) of X *i.e.* $\psi(\theta) = \log \mathbb{E}\big[e^{\langle \theta, X\rangle}\big]$. We assume that $\psi(\theta) < +\infty$ for every $\theta \in \mathbb{R}^d$ (which implies that ψ is an infinitely differentiable convex function) and define

$$p_\theta(x) = e^{\langle \theta, x\rangle - \psi(\theta)} p(x), \quad x \in \mathbb{R}^d.$$

Let $X^{(\theta)}$ denote any random variable with distribution p_θ.

We assume that ψ satisfies

$$\lim_{|\theta|\to+\infty} \psi(\theta) - 2\psi\Big(\frac{\theta}{2}\Big) = +\infty \qquad \text{and} \qquad \exists\, \delta > 0, \quad \theta \mapsto \psi(\theta) - \delta|\theta|^2 \text{ is concave.} \tag{$\mathcal{H}_\delta^{es}$}$$

One must be aware that what follows makes sense as a variance reduction procedure only if the distribution of $X^{(\theta)}$ can be simulated at the same cost as X or at least at a reasonable cost *i.e.*

$$X^{(\theta)} = g(\theta, \xi), \quad \xi : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathcal{X} \tag{2.16}$$

where $\mathcal{X}$ is a Borel subset of a metric space and $g : \mathbb{R}^d \times \mathcal{X}$ is an explicit Borel function. By (1.2), the potential V to be minimized is $V(\theta) = \mathbb{E}\big[F^2(X) e^{-\langle \theta, X\rangle + \psi(\theta)}\big]$.

Proposition 3 *Suppose ψ satisfies $(\mathcal{H}_\delta^{es})$ and F satisfies*

$$\forall \theta \in \mathbb{R}^d, \quad \mathbb{E}\Big[|X| F^2(X) e^{\langle \theta, X\rangle}\Big] < +\infty. \tag{2.17}$$

Then $(\mathcal{H}_1)$ is fulfilled and the function V is differentiable on $\mathbb{R}^d$ with a gradient given by

$$\nabla V(\theta) = \mathbb{E}\Big[\big(\nabla\psi(\theta) - X\big) F^2(X) e^{-\langle \theta, X\rangle + \psi(\theta)}\Big], \tag{2.18}$$

$$= \mathbb{E}\Big[\big(\nabla\psi(\theta) - X^{(-\theta)}\big) F^2(X^{(-\theta)})\Big] e^{\psi(\theta) - \psi(-\theta)}, \tag{2.19}$$

where $\nabla\psi(\theta) = \dfrac{\mathbb{E}\big[X e^{\langle \theta, X\rangle}\big]}{\mathbb{E}\big[e^{\langle \theta, X\rangle}\big]}$.

Proof. The function ψ is clearly log-convex so that $\theta \mapsto p_\theta(x)$ is log-concave for every $x \in \mathbb{R}^d$. On the other hand, by $(\mathcal{H}_\delta^{es})$ we have $\lim \frac{p_{\theta/2}^2(x)}{p_\theta(x)} = +\infty$ for every $x \in \mathbb{R}^d$, and $(\mathcal{H}_1)$ is fulfilled.

The formal differentiation to get (2.18) is obvious and is made rigorous by applying the assumption on F. The second expression (2.19) of the gradient uses a third change of variable

$$\begin{aligned}
\nabla V(\theta) &= \int_{\mathbb{R}^d} \big(\nabla\psi(\theta) - x\big) F^2(x) e^{-\langle \theta, x\rangle + \psi(\theta)} p(x) dx, \\
&= \int_{\mathbb{R}^d} \big(\nabla\psi(\theta) - x\big) F^2(x) e^{\psi(\theta) - \psi(-\theta)} p_{-\theta}(x) dx, \\
&= \mathbb{E}\Big[\big(\nabla\psi(\theta) - X^{(-\theta)}\big) F^2(X^{(-\theta)})\Big] e^{\psi(\theta) - \psi(-\theta)}. \quad \diamondsuit
\end{aligned}$$

Theorem 3 *We assume that ψ satisfies $(\mathcal{H}_\delta^{es})$ and F satisfies* (2.17) *and*

$$\forall x \in \mathbb{R}^d, \quad |F(x)| \le C e^{\frac{\lambda}{4}|x|} \qquad \text{and} \qquad \mathbb{E}\Big[|X|^2 e^{\lambda |X|}\Big] < +\infty.$$

Then the recursive procedure

$$\left\{\begin{array}{lcl} X_{n+1}^{(\theta_n)} & = & g(\theta_n, \xi_{n+1}) \\ \theta_{n+1} & = & \theta_n - \gamma_{n+1} H(\theta_n, X_{n+1}^{(-\theta_n)}), \quad n \geq 0, \quad \theta_0 \in \mathbb{R}^d \end{array}\right.$$

where $(\xi_n)_{n\geq 1}$ *is an i.i.d. sequence with the same distribution as* ξ *in (2.16) and*

$$H(\theta, x) := e^{-\frac{\lambda}{2}\sqrt{d}|\nabla\psi(-\theta)|} F^2(x)\big(\nabla\psi(\theta) - x\big)$$

a.s. converges toward an $\operatorname{Argmin} V$*-valued (square integrable) random vector* θ^* *.*

Proof. We have to check the linear growth of the function $\theta \mapsto \big\|H(\theta, X^{(-\theta)})\big\|_2$ (condition (NEC)). We have

$$\begin{aligned} \mathbb{E}\Big[|H(\theta, X^{(-\theta)})|^2\Big] &= e^{-\lambda\sqrt{d}|\nabla\psi(-\theta)|}\mathbb{E}\Big[F^4(X^{(-\theta)})\big|\nabla\psi(\theta) - X^{(-\theta)}\big|^2\Big], \\ &\leq Ce^{-\lambda\sqrt{d}|\nabla\psi(-\theta)|}\mathbb{E}\Big[e^{\lambda|X^{(-\theta)}|}\big|\nabla\psi(\theta) - X^{(-\theta)}\big|^2\Big], \\ &\leq Ce^{-\lambda\sqrt{d}|\nabla\psi(-\theta)|}\left(|\nabla\psi(\theta)|^2\mathbb{E}\Big[e^{\lambda|X^{(-\theta)}|}\Big] + \mathbb{E}\Big[|X^{(-\theta)}|^2 e^{\lambda|X^{(-\theta)}|}\Big]\right). \end{aligned} \tag{2.20}$$

First, by the following inequality

$$\forall x \in \mathbb{R}^d, \quad e^{\lambda|x|} \leq \prod_{j=1}^d \left(e^{\lambda x_j} + e^{-\lambda x_j}\right) = \sum_{J\subset\{1,\ldots,d\}} e^{\lambda\left(\sum_{j\in J}\langle e_j, x\rangle + \sum_{j\in J^c}\langle e_j, x\rangle\right)}$$

we have $e^{\lambda|x|} \leq \displaystyle\sum_{J\subset\{1,\ldots,n\}} e^{\lambda\langle e_J, x\rangle}$ where $(e_J)_j = 1$ if $j \in J$ or -1 if $j \in J^c$. With this notation, we have

$$\mathbb{E}\Big[e^{\lambda|X^{(-\theta)}|}\Big] \leq \sum_{J\subset\{1,\ldots,d\}} \mathbb{E}\Big[e^{\lambda\langle e_J, X^{(-\theta)}\rangle}\Big] = \sum_{J\subset\{1,\ldots,d\}} e^{\psi(\lambda e_J - \theta) - \psi(-\theta)}.$$

By the concavity of $\psi - \delta\,|\,.\,|^2$, we have

$$\forall\, u,\, v \in \mathbb{R}^d, \qquad \psi(u+v) - \psi(u) \leq \langle\nabla\psi(u), v\rangle + \delta\,|v|^2$$

so that

$$\mathbb{E}\Big[e^{\lambda|X^{(-\theta)}|}\Big] \leq \sum_{J\subset\{1,\ldots,d\}} e^{\lambda\langle\nabla\psi(-\theta), e_J\rangle + \delta\lambda^2|e_J|^2} \leq C_{d,\lambda,\delta}e^{\lambda\sqrt{d}|\nabla\psi(-\theta)|}. \tag{2.21}$$

In the same way, we have

$$\begin{aligned} \mathbb{E}\Big[|X^{(-\theta)}|^2 e^{\lambda|X^{(-\theta)}|}\Big] &\leq \sum_{J\subset\{1,\ldots,d\}} \mathbb{E}\Big[|X^{(\lambda e_J - \theta)}|^2\Big]\, e^{\psi(\lambda e_J - \theta) - \psi(-\theta)}, \\ &\leq C_{d,\lambda,\delta}e^{\lambda\sqrt{d}|\nabla\psi(-\theta)|} \sum_{J\subset\{1,\ldots,d\}} \mathbb{E}\Big[|X^{(\lambda e_J - \theta)}|^2\Big]. \end{aligned} \tag{2.22}$$

Now, by differentiation of ψ it is easy to check that

$$\forall\theta \in \mathbb{R}^d, \quad \mathrm{D}^2\psi(\theta) = \frac{\int x^{\otimes 2} e^{\langle\theta, x\rangle} p(x)\,\mathrm{d}x}{e^{\psi(\theta)}} - \nabla\psi(\theta)^{\otimes 2},$$

which implies

$$\mathbb{E}\Big[|X^{(\lambda e_J-\theta)}|^2\Big] = \mathrm{Tr}\big(\mathrm{D}^2\psi(\lambda e_J-\theta)\big) + \mathrm{Tr}\big(\nabla\psi(\lambda e_J-\theta)^{\otimes 2}\big)\,.$$

The assumption $(\mathcal{H}^{es}_{\delta})$ implies that $0 \le D^2\psi(\theta) \le 2\delta\, I_d$ (for the partial order on symmetric matrices induced by nonnegative symmetric matrices) then $D^2\psi(\theta)$ is a bounded function of $\theta\in\mathbb{R}^d$ and in turn $\nabla(\theta)$ has a linear growth by the fundamental formula of calculus. Consequently, for every $J\subset(1,\ldots,d)$,

$$\mathbb{E}\Big[|X^{(\lambda e_J-\theta)}|^2\Big] \le C\,\big(1+|\theta|^2\big)\,.$$

Plugging this into (2.22) and using (2.21) and (2.20) we obtain $\mathbb{E}\big[|H(\theta,X^{(-\theta)})|^2\big] \le C\,\big(1+|\theta|^2\big)$. $\diamond$

3 Adaptive variance reduction for diffusions

3.1 Framework and preliminaries

We consider a d-dimensional Itô process $X=(X_t)_{t\in[0,T]}$ solution to the stochastic differential equation (SDE)

$$\mathrm{d}X_t = b(t,X^t)\,\mathrm{d}t + \sigma(t,X^t)\,\mathrm{d}W_t, \quad X_0=x\in\mathbb{R}^d, \tag{$E_{b,\sigma,W}$}$$

where $W=(W_t)_{t\in[0,T]}$ is a q-dimensional standard Brownian motion, $X^t:=(X_{t\wedge s})_{s\in[0,T]}$ is the stopped process at time t, $b:[0,T]\times\mathcal{C}([0,T],\mathbb{R}^d)\to\mathbb{R}^d$ and $\sigma:[0,T]\times\mathcal{C}([0,T],\mathbb{R}^d)\to\mathcal{M}(d,q)$ are measurable with respect to the canonical predictable σ-field on $[0,T]\times\mathcal{C}([0,T],\mathbb{R}^d)$. For further details we refer to [18], p. 124-130.

Thus, if $b(t,x^t)=\beta(t,x(t))$ and $\sigma(t,x^t)=\vartheta(t,x(t))$ for every $x\in\mathcal{C}([0,T],\mathbb{R}^d)$, X is a usual diffusion process with drift β and diffusion coefficient ϑ.

If $b(t,x^t)=\beta(t,x(\underline{t}))$ and $\sigma(t,x^t)=\vartheta(t,x(\underline{t}))$ for every $x\in\mathcal{C}([0,T],\mathbb{R}^d)$ where $\underline{t}:=\lfloor\frac{tn}{T}\rfloor\frac{T}{n}$, then X is *the continuous Euler scheme with step* T/n of the above diffusion with drift β and diffusion coefficient ϑ.

An easy adaptation of standard proofs for regular SDE's show (see [18]) that strong existence and uniqueness of solutions for $(E_{b,\sigma,W})$ follows from the following assumption

$$\begin{cases}(i) & b(.,0) \text{ and } \sigma(.,0) \text{ are continuous}\\ (ii) & \forall\, t\in[0,T],\ \forall\, x,\ y\in\mathcal{C}([0,T],\mathbb{R}^d),\ |b(t,y)-b(t,x)|+\|\sigma(t,y)-\sigma(t,x)\| \le C_{b,\sigma}\,\|x-y\|_{\infty}\,.\end{cases} \tag{$\mathcal{H}_{b,\sigma}$}$$

Our aim is to devise an adaptive variance reduction method inspired from Section 2 for the computation of

$$\mathbb{E}[F(X)]$$

where F is an Borel functional defined on $\mathcal{C}([0,T],\mathbb{R}^d)$ such that

$$\mathbb{P}\big[F^2(X)>0\big]>0 \quad\text{and}\quad F(X)\in L^2(\mathbb{P}). \tag{3.23}$$

In this functional setting, Girsanov Theorem will play the role of the invariance of Lebesgue measure by translation. The translation process that we consider in this section is of the form $\Theta(t,X^t)$ where Θ is defined for every $\xi\in\mathcal{C}([0,T],\mathbb{R}^d)$ and $\theta\in L^2_{T,p}$ by

$$\Theta(t,\xi) := \varphi(t,\xi^t)\,\theta_t, \qquad \text{with} \qquad \varphi:[0,T]\times\mathcal{C}([0,T],\mathbb{R}^d)\to\mathcal{M}(q,p),$$

a bounded Borel function and $\theta \in L^2_{T,p}$ (represented by a Borel function) for $p \geq 1$. In the sequel, we use the following notations $\varphi_t(\xi) := \varphi(t, \xi^t)$,

$$\Theta_t := \Theta(t, X^t), \quad \Theta_t^{(\theta)} := \Theta(t, X^{(\theta),t}), \quad \text{and} \quad \Theta_t^{(-\theta)} := \Theta(t, X^{(-\theta),t}),$$

where $X^{(\pm\theta)}$ denotes the solution to $(E_{b\pm\sigma\Theta,\sigma,W})$.

First we need the following standard abstract lemma.

Lemma 3 *Suppose $(H_{b,\sigma})$ holds.*
The SDE $(E_{b+\sigma\Theta,\sigma,W})$ satisfies the weak existence and uniqueness assumptions and for every non negative Borel functional $G : \mathcal{C}([0,T], \mathbb{R}^{d+1}) \to \mathbb{R}_+$ and $f \in \mathcal{C}([0,T], \mathbb{R}^q)$ we have, with the above notations,

$$\mathbb{E}\Big[G\Big(X, \int_0^{\cdot} \langle f(s, X^s), \mathrm{d}W_s\rangle\Big)\Big] = \mathbb{E}\Big[G\Big(X^{(\theta)}, \int_0^{\cdot} \langle f(s, X^{(\theta),s}), \mathrm{d}W_s\rangle + \int_0^{\cdot} \langle f, \Theta\rangle(s, X^{(\theta),s})\,\mathrm{d}s\Big) \times e^{-\int_0^T \langle \Theta_s^{(\theta)}, \mathrm{d}W_s\rangle - \frac{1}{2}\left\|\Theta^{(\theta)}\right\|^2_{L^2_{T,q}}}\Big],$$

and

$$\mathbb{E}\Big[G\Big(X^{(\theta)}, \int_0^{\cdot} \langle f(s, X^{(\theta),s}), \mathrm{d}W_s\rangle\Big)\Big] = \mathbb{E}\Big[G\Big(X, \int_0^{\cdot} \langle f(s, X^s), \mathrm{d}W_s\rangle - \int_0^{\cdot} \langle f, \Theta\rangle(s, X^s)\,\mathrm{d}s\Big) \times e^{\int_0^T \langle \Theta_s, \mathrm{d}W_s\rangle - \frac{1}{2}\|\Theta\|^2_{L^2_{T,q}}}\Big],$$

Proof. This is a straightforward application of Theorem 1.11, p.372 (and the remark that immediately follows) in [16] once noticed that $(t,\omega) \mapsto b(t, X^t(\omega))$, $(t,\omega) \mapsto \sigma(t, X^t(\omega))$ and $(t,\omega) \mapsto \Theta(t, X^t(\omega))$ are predictable processes with respect to the completed filtration of W. ♢

Remarks. • The Doléans exponential $\left(e^{\int_0^t \langle \Theta_s, \mathrm{d}W_s\rangle - \frac{1}{2}\|\Theta\|^2_{L^2_{T,q}}}\right)_{t\in[0,T]}$ is a true martingale for any $\theta \in L^2_{T,p}$.
• In fact, still following the above cited remark form [16], the above lemma holds true if we replace Θ by any progressively measurable process $\tilde{\Theta}$ such that $\mathbb{E}\Big[e^{\frac{1}{2}\int_0^T |\tilde{\Theta}(s,\omega)|^2\,\mathrm{d}s}\Big] < +\infty$.

It follows from the first identity in Lemma 3 that for every bounded Borel function $\varphi : [0,T] \times \mathcal{C}([0,T], \mathbb{R}^d) \to \mathcal{M}(q,p)$ and for every $\theta \in L^2_{T,p}$

$$\mathbb{E}[F(X)] = \mathbb{E}\left[F(X^{(\theta)}) e^{-\int_0^T \langle \Theta_s^{(\theta)}, \mathrm{d}W_s\rangle - \frac{1}{2}\left\|\Theta^{(\theta)}\right\|^2_{L^2_{T,q}}}\right],$$

(set $G(x,y) = F(x)$). So, finding the estimator with the lowest variance amounts to solving the minimization problem

$$\min_{\theta \in L^2_{T,p}} V(\theta) \quad \text{where} \quad V(\theta) := \mathbb{E}\Big[F^2(X^{(\theta)}) e^{-2\int_0^T \langle \Theta_s^{(\theta)}, \mathrm{d}W_s\rangle - \left\|\Theta^{(\theta)}\right\|^2_{L^2_{T,q}}}\Big],$$

Using Lemma 3 with $G(x,y) = F^2(x) e^{-2y(T) - \|\varphi(.,x^{\cdot})\theta\|^2_{L^2_{T,q}}}$ and $f = \Theta$ yields

$$V(\theta) = \mathbb{E}\Big[F^2(X) e^{-\int_0^T \langle \Theta_s, \mathrm{d}W_s\rangle + \frac{1}{2}\|\Theta\|^2_{L^2_{T,q}}}\Big]. \tag{3.24}$$

Proposition 4 *Assume* $\mathbb{E}\,F(X)^{2+\eta} < +\infty$ *for some* $\eta > 0$ *as well as Assumptions (3.23) and* $(H_{b,\sigma})$. *Then function* V *is finite on* $L^2_{T,p}$ *and* log*-convex.*

(*a*) *Assume that the bounded matrix-valued Borel function* φ *satisfies that* $\varphi(s, X^s)$ *has a non-atomic kernel on the event* $\{F(X) > 0\}$ *i.e.*

$$\mathbb{P}\big[\{\exists\, \theta \in L^2_{T,p} \setminus \{0\} \;\; s.t.\; \theta(s) \in Ker\, \varphi(s, X^s) \;\; \mathrm{d}s\text{-}a.e.\; and\; F^2(X) > 0\}\big] = 0 \tag{3.25}$$

then for every finite dimensional subspace $E \subset L^2_T$, $\lim\limits_{\|\theta\|_{L^2_{T,p}} \to +\infty, \theta \in E} V(\theta) = +\infty$. *If furthermore*

$$\inf_{\|\theta\|_{L^2_{T,p}}=1} \int_0^T \theta(s)^* \mathbb{E}\big[\varphi(s, X^s)^* \varphi(s, X^s) \mathbf{1}_{\{F^2(X)>0\}}\big]\, \theta(s)\, \mathrm{d}s > 0, \tag{3.26}$$

then $\lim\limits_{\|\theta\|_{L^2_{T,p}} \to +\infty} V(\theta) = +\infty$.

(*b*) *The function* V *is differentiable at every* $\theta \in L^2_{T,p}$ *and the differential* $\mathrm{D}V(\theta) \in L^2_{T,p}$ *is characterized on every* $\psi \in L^2_{T,p}$ *by*

$$\begin{aligned}\langle \mathrm{D}V(\theta), \psi\rangle_{L^2_{T,p}} &= \mathbb{E}\Big[F^2(X) e^{-\int_0^T \langle \Theta_s, \mathrm{d}W_s\rangle + \frac{1}{2}\|\Theta\|^2_{L^2_{T,q}}} \Big(\langle \Theta, \varphi(., X^\cdot)\psi\rangle_{L^2_{T,p}} - \int_0^T \langle \varphi(s, X^s)\psi_s, \mathrm{d}W_s\rangle\Big)\Big],\\ &= \mathbb{E}\Big[F^2(X^{(-\theta)}) e^{\|\Theta^{(-\theta)}\|^2_{L^2_{T,p}}} \Big(2\langle \Theta^{(-\theta)}, \varphi(X^{(-\theta),\cdot})\psi\rangle_{L^2_{T,p}} - \int_0^T \langle \varphi(X^{(-\theta),s})\psi_s, \mathrm{d}W_s\rangle\Big)\Big]. \end{aligned} \tag{3.27}$$

Remarks. • For practical implementation, the "finite dimensional" statement is the only result of interest since it ensures that $\mathrm{Argmin}_{|E} \neq \emptyset$.

• If $p = q$ and $\varphi = I_q$, the "infinite-dimensional" assumption is always satisfied.

Proof. (*a*) As concerns the function V, we rely on Equality (3.24). Set $r = 1 + 2/\eta$. Owing to the Hölder Inequality, showing that this function is finite on the whole space $L^2_{T,q}$ amounts to proving that

$$\mathbb{E}\Big[e^{\frac{r}{2}\|\Theta\|^2_{L^2_{T,q}} - r\int_0^T \langle \Theta_s, \mathrm{d}W_s\rangle}\Big] \leq e^{\|\varphi\|^2_\infty \|\theta\|_{L^2_{T,p}}^{r(r+1)/2}} < +\infty.$$

To show that V goes to infinity at infinity, one proceeds as follows. Using the trivial equality

$$e^{-\int_0^T \langle \Theta_s, \mathrm{d}W_s\rangle + \frac{1}{2}\|\Theta\|^2_{L^2_{T,q}}} = \left(e^{-\frac{1}{2}\int_0^T \langle \Theta_s, \mathrm{d}W_s\rangle + \frac{1}{8}\|\Theta\|^2_{L^2_{T,q}}}\right)^2 e^{\frac{1}{4}\|\Theta\|^2_{L^2_{T,q}}}$$

and the reverse Hölder inequality with conjugate exponents $(\frac{1}{3}, -\frac{1}{2})$ we obtain

$$\begin{aligned} V(\theta) &\geq \mathbb{E}\Big[F^{2/3}(X) e^{\frac{1}{12}\|\Theta\|_{L^2_{T,q}}}\Big]^3 \mathbb{E}\Big[e^{\frac{1}{2}\int_0^T \langle \Theta_s, \mathrm{d}W_s\rangle - \frac{1}{8}\|\Theta\|^2_{L^2_{T,q}}}\Big]^{-2},\\ &\geq \mathbb{E}\Big[F^{2/3}(X) e^{\frac{1}{12}\|\Theta\|^2_{L^2_{T,q}}}\Big]^3 \end{aligned}$$

by the martingale property of the Doléans exponential. Let $\varepsilon > 0$ such that $\mathbb{P}\big[F^2(X) \geq \varepsilon\big] > 0$. We have then $V(\theta) \geq \varepsilon^{1/3}\mathbb{E}\Big[\mathbf{1}_{\{F^2(X)\geq\varepsilon\}} e^{\frac{1}{12}\|\Theta\|_{L^2_{T,q}}}\Big]^3$, and by the conditional Jensen inequality

$$\begin{aligned} V(\theta) &\geq \varepsilon^{1/3}\mathbb{E}\Big[\mathbf{1}_{\{F^2(X)\geq\varepsilon\}} e^{\frac{1}{12}\mathbb{E}\left[\|\Theta\|^2_{L^2_{T,q}} \,\middle|\, F^2(X)\geq\varepsilon\right]}\Big]^3\\ &= \mathbb{E}\Big[\mathbf{1}_{\{F^2(X)\geq\varepsilon\}} e^{\frac{1}{12}\mathbb{P}\left[F^2(X)\geq\varepsilon\right]\mathbb{E}\left[\|\Theta\|^2_{L^2_{T,q}} \mathbf{1}_{\{F^2(X)\geq\varepsilon\}}\right]}\Big]^3. \end{aligned}$$

Now

$$\mathbb{E}\Big[\|\Theta\|^2_{L^2_{T,q}}\,\mathbf{1}_{\{F^2(X)\geq\varepsilon\}}\Big] = \int_0^T \theta(s)^*\mathbb{E}\big[\varphi_s(X^s)^*\varphi_s(X^s)\mathbf{1}_{\{F^2(X)\geq\varepsilon\}}\big]\,\theta(s)\,\mathrm{d}s \geq 0.$$

The assumption (3.25) implies that, for every $\theta\in L^2_{T,p}$,

$$\int_0^T \theta(s)^*\,\mathbb{E}\big[\varphi_s(X^s)^*\,\varphi_s(X^s)\mathbf{1}_{\{F^2(X)\geq\varepsilon\}}\big]\,\theta(s)\,\mathrm{d}s \geq \int_0^T \theta(s)^*\,\mathbb{E}\big[\varphi_s(X^s)^*\,\varphi_s(X^s)\mathbf{1}_{\{F^2(X)\geq 0\}}\big]\,\theta(s)\,\mathrm{d}s > 0,$$

so that if θ runs over the compact sphere of a finite dimensional subspace E of $L^2_{T,p}$

$$\inf_{\|\theta\|_{L^2_{T,p}}=1,\theta\in E}\int_0^T \theta(s)^*\,\mathbb{E}\big[\varphi_s(X^s)^*\,\varphi_s(X^s)\mathbf{1}_{\{F^2(X)\geq\varepsilon\}}\big]\,\theta(s)\,\mathrm{d}s > 0,$$

so that

$$\lim_{\|\theta\|_{L^2_{T,p}}\to\infty,\theta\in E}\mathbb{E}\Big[\|\Theta\|^2_{L^2_{T,q}}\,\mathbf{1}_{\{F^2(X)\geq\varepsilon\}}\Big] = +\infty,$$

and one concludes by Fatou's Lemma using that $\mathbb{P}\big[F^2(X)\geq\varepsilon\big] > 0$. The second claim easily follows from Assumption (3.26).

(*b*) As a first step, we show that the random functional $\Phi(\theta) := \frac{1}{2}\|\Theta\|_{L^2_{T,q}} - \int_0^T\langle\Theta(s),\mathrm{d}W_s\rangle$ from $L^2_{T,p}$ into $L^r(\mathbb{P})$ ($r\in[1,\infty)$), is differentiable. Indeed, it from the below inequality,

$$\forall\,\theta,\,\psi\in L^2_{T,p},\qquad |\Phi(\theta+\psi)-\Phi(\theta)-\langle\mathrm{D}\Phi(\theta),\psi\rangle_{L^2_{T,p}}| \leq \|\varphi\|^2_\infty\,\|\theta\|_{L^2_{T,p}}\,\|\psi\|_{L^2_{T,p}} \tag{3.28}$$

where $\psi\mapsto\langle\mathrm{D}\Phi(\theta),\psi\rangle_{L^2_{T,p}} = \int_0^T\langle\varphi_s(X^s)\theta(s),\varphi_s(X^s)\psi(s)\rangle\,\mathrm{d}s - \int_0^T\langle\varphi_s(X^s)\psi(s),\mathrm{d}W_s\rangle$ is clearly a bounded random functional from $L^2_{T,p}$ into $L^r(\mathbb{P})$, with an operator norm $|||\mathrm{D}\Phi(\theta)|||_{L^2_{T,p},L^r(\mathbb{P})} \leq \|\varphi\|^2_\infty\,\|\theta\|_{L^2_{T,p}} + c_p\,\|\varphi\|_\infty$ ($c_p\in(0,+\infty)$ (this follows from Hölder and *B.D.G.* inequalities).

Then, we derive that $\theta\mapsto e^{\Phi(\theta)}$ is differentiable form $L^2_{T,p}$ into every $L^r(\mathbb{P})$ with differential $e^{\Phi(\theta)}\,\mathrm{D}\Phi(\theta)$. This follows from standard computation based on (3.28), the elementary inequality $|e^u-1-u|\leq\frac{1}{2}u^2(e^u+e^{-u})$ and the fact that

$$\begin{aligned}\Big\|\Big|\int_0^T\langle\varphi_s(X^s)\psi(s),\mathrm{d}W_s\rangle\Big|e^{\int_0^T\langle\phi(X^s)\theta(s),\mathrm{d}W_s\rangle}\Big\|_r &\leq \Big\|\int_0^T\langle\varphi_s(X^s)\psi(s),\mathrm{d}W_s\rangle\Big\|_{2r}\Big\|e^{\int_0^T\langle\varphi_s(X^s)\theta(s),\mathrm{d}W_s\rangle}\Big\|_{2p},\\ &\leq c_p\,\|\varphi\|_\infty\,\|\theta\|_{L^2_{T,p}}\,\|\psi\|_{L^2_{T,p}},\end{aligned}$$

where we used both Hölder and *B.D.G.* inequality.

One concludes that $\theta\mapsto V(\theta)=\mathbb{E}\big[F(X)^2e^{\Phi(\theta)}\big]$ is differentiable by using the $(L^2_{T,p},L^r(\mathbb{P}))$–differentiability of $e^{\Phi(\theta)}$ with $r=1+\frac{\eta}{2}$.

The second form of the gradient is obtained by a Girsanov transform using Lemma 3. ♢

3.2 Design of the algorithm

In view of a practical implementation of the procedure we are lead to consider some non trivial finite dimensional subspaces E of $L^2_{T,p}$. The function V being strictly log-convex on E and going to infinity as $\|\theta\|_{L^2_{T,p}}$ goes to infinity, $\theta\in E$, the restriction of V on E attains a minimum θ^*_E which *de facto* becomes the target of the procedure. Furthermore, for every $\theta\in E$, $\mathrm{D}V_{|E}(\theta)=\mathrm{D}V(\theta)_{|E}$ and the quadratic function $L(\theta) := \big\|\theta-\theta^*_E\big\|_{L^2_{T,p}}$ is a Lyapunov function for the problem.

Like for the static framework investigated in Section 2.3, our algorithm will be based on the representation (3.27) for the differential $\mathrm{D}V$ of V: in this representation the variance reducer θ appears inside the functional F which makes easier a control at infinity in order to prevent from any early explosion of the procedure. However, to this end we need to control the discrepancy between X and $X^{(-\theta)}$. This is the purpose of the following Lemma.

Lemma 4 *Assume $(H_{b,\sigma})$ holds. Let φ be a bounded Borel $\mathcal{M}(q,p)$-valued function defined on $[0,T]\times\mathcal{C}([0,T],\mathbb{R}^d)$, let $\theta\in L^2_{T,p}$ and let X and $X^{(\theta)}$ denote a strong solutions of $E_{b,\sigma,W}$ and $E_{b+\sigma\Theta,\sigma,W}$ driven by the same Brownian motion. Then, for every $r\geq 1$, there exists a real constant $C_{b,\sigma}>0$ such that*

$$\left\|\sup_{t\in[0,T]}\left|X_t-X_t^{(\theta)}\right|\right\|_r\leq C_{b,\sigma}e^{C_{b,\sigma}T}\left\|\int_0^T\left|\sigma(s,X^{(\theta),s})\Theta_s^{(\theta)}\right|\mathrm{d}s\right\|_r. \tag{3.29}$$

Proof. The proof follows the lines of the proof of the strong rate of convergence of the Euler scheme (see *e.g.* [3]). $\diamond$

The main result of this section is the following theorem.

Theorem 4 *Suppose that Assumption (3.23) and $(H_{b,\sigma})$ hold.*

Let φ be a bounded Borel $\mathcal{M}(q,p)$-valued function (with $p\geq 1$) defined on $[0,T]\times\mathcal{C}([0,T],\mathbb{R}^d)$, and let F be a functional F satisfying

$$\forall\, x\in\mathcal{C}([0,T],\mathbb{R}^d),\qquad |F(x)|\leq C_{_F}(1+\|x\|_\infty^\lambda) \tag{$G_{F,\lambda}$}$$

for some positive exponent $\lambda>0$ (then $F(X)\in L^r(\mathbb{P})$ for every $r>0$). Let E be a finite dimensional subspace of $L^2_{T,p}$ spanned by an orthonormal basis $(e_1,\ldots,e_m)$.

Let $\eta>0$. We define the algorithm by

$$\theta_{n+1}=\theta_n-\gamma_{n+1}H_{\lambda,\eta}(\theta_n,X^{(-\theta_n)},W^{(n+1)})$$

where $\gamma=(\gamma_n)_{n\geq 1}$ satisfies (2.6), $(W^{(n)})_{n\geq 1}$ is a sequence of independent Brownian motions for which $X^{(-\theta_n)}=\mathcal{G}(-\theta_n,W^{(n+1)})$ is a strong solution to $(E_{b-\sigma\Theta},W^{(n+1)})$ and for every standard Brownian motion W, every $\mathcal{F}_t^W$-adapted $\mathbb{R}^p$-valued process $\xi=(\xi_t)_{t\in[0,T]}$,

$$\langle H_{\lambda,\eta}(\theta,\xi,W),e_i\rangle_{L^2_{T,p}}=\Psi_{\lambda,\eta}(\theta,\xi)F^2(\xi)e^{\|\Theta(.,\xi^\cdot)\|^2_{L^2_{T,q}}}\left(2\langle\Theta(.,\xi^\cdot),\varphi(.,\xi^\cdot)e_i\rangle_{L^2_{T,q}}-\int_0^T\langle\varphi(s,\xi^s)e_i(s),\mathrm{d}W_s\rangle\right)$$

where for $\eta>0$

$$\Psi_{\lambda,\eta}(\theta,\xi)=\begin{cases}\dfrac{e^{-\|\varphi\|_\infty\|\theta\|^2_{L^2_{T,p}}}}{1+\|\varphi(.,\xi^\cdot)\theta\|^{2\lambda+\eta}_{L^2_{T,q}}} & \text{if } \sigma \text{ is bounded,}\\ e^{-(\|\varphi\|_\infty+\eta)\|\theta\|^2_{L^2_{T,p}}} & \text{if } \sigma \text{ is unbounded.}\end{cases}$$

Then the recursive sequence $(\theta_n)_{n\geq 1}$ a.s. converges toward an $\operatorname{Argmin} V$*-valued (squared integrable) random variable θ^*.*

Remark. For a practical implementation of this algorithm, we must have *for all* Brownian motions $W^{(n+1)}$ a strong solution $X^{(-\theta_n)}$ of $(E_{b-\sigma\Theta},W^{(n+1)})$. In particular, this is the case if the driver φ is locally Lipshitz (in space) or if X is the continuous Euler scheme of a diffusion with step T/n (using the driver $\varphi(t,x^t)=f(t,x(\underline{t}))$).

Note that if φ is continuous (in space) but not necessarily locally Lipshitz, the Euler scheme converges in law to the solution of the SDE.

Proof. When the diffusion coefficient σ is bounded, it follows from Lemma 4 that, for every $r \geq 1$,

$$\left\| \sup_{t\in[0,T]} \left| X_t - X_t^{(\theta)} \right| \right\|_r \leq C_{b,\sigma,T} \left\|\varphi\right\|_\infty \left\|\theta\right\|_{L^2_{T,p}} \left\|\sigma\right\|_\infty ,$$

where $\left\|\sigma\right\|_\infty = \sup_{(t,x)\in[0,T]\times\mathcal{C}([0,T],\mathbb{R}^d)} \left\|\sigma(t,x)\right\|$.

First note that for every θ, $\psi \in E$, the mean function h of the algorithm reads

$$\langle h(\theta),\psi\rangle_{L^2_{T,p}} = \mathbb{E}\Big[\langle H_{\lambda,\eta}(\theta, X^{(-\theta)}, W),\psi\rangle_{L^2_{T,p}}\Big] = \mathbb{E}\left[\frac{e^{-\left\|\varphi\right\|_\infty \left\|\theta\right\|_{L^2_{T,p}}}}{1+\left\|\Theta^{(-\theta)}\right\|^{2\lambda+\eta}_{L^2_{T,q}}} \langle \mathrm{D}V_{|E}(\theta),\psi\rangle_{L^2_{T,p}}\right],$$

so that, for every $\theta \neq \theta^*_E$,

$$\langle h(\theta),\theta-\theta^*_E\rangle = \mathbb{E}\left[\frac{e^{-\left\|\varphi\right\|_\infty \left\|\theta\right\|_{L^2_{T,p}}}}{1+\left\|\Theta^{(-\theta)}\right\|^{2\lambda+\eta}_{L^2_{T,q}}} \langle \mathrm{D}V_{|E}(\theta),\theta-\theta^*_E\rangle_{L^2_{T,p}}\right] > 0.$$

It remains to check that for every $i \in \{1,\dots,m\}$, $\left\|H_{\lambda,\eta}(\theta, X^{(-\theta)}, W)\right\|_2 \leq C\big(1+\left\|\theta\right\|_{L^2_{T,p}}\big)$ to apply the Robbins-Zygmund Lemma which ensures the $a.s.$ convergence of the procedure (see Section 2.1). We first deal with the term $F(X^{(-\theta)})^2 \int_0^T \langle \varphi_s(X^{(-\theta),s}) e_i(s), \mathrm{d}W_s\rangle$. Let $\eta' = \frac{\eta}{2\lambda} > 0$.

$$\begin{aligned}
\left\| F(X^{(-\theta)})^2 \int_0^T \langle \varphi_s(X^{(-\theta),s}) e_i(s), \mathrm{d}W_s\rangle \right\|_2 &\leq \left\| F(X^{(-\theta)})^2 \right\|_{2+\eta'} \left\| \int_0^T \langle \varphi_s(X^{(-\theta),s}) e_i(s), \mathrm{d}W_s\rangle \right\|_{2(1+1/\eta')}, \\
&\leq \left\| F(X^{(-\theta)})^2 \right\|_{2+\eta'} \left\| \int_0^T |\varphi_s(X^{(-\theta),s}) e_i(s)|^2 \,\mathrm{d}s \right\|_{1+1/\eta'}, \\
&\leq \left\| F(X^{(-\theta)})^2 \right\|_{2+\eta'} \left\|\varphi\right\|_\infty .
\end{aligned}$$

Now

$$\begin{aligned}
\left\| F(X^{(-\theta)})^2 \right\|_{2(1+\eta')} &\leq C\left(1+\left\|\left\|X^{(-\theta)}\right\|_\infty\right\|^{2\lambda(1+\eta')}_{4\lambda(1+\eta')}\right), \\
&\leq C_{\lambda,b,\sigma,T}\left(1+\left\|\left\|X\right\|_\infty\right\|^{2\lambda(1+\eta')}_{4\lambda(1+\eta')} + \left\|\theta\right\|^{2\lambda(1+\eta')}_{L^2_{T,p}} \left\|\varphi\right\|^{2\lambda(1+\eta')}_\infty \left\|\sigma\right\|^{2\lambda(1+\eta')}_\infty\right), \\
&\leq C_{\lambda,b,\sigma,\varphi,T}\left(1+\left\|\theta\right\|^{2\lambda+\eta}_{L^2_{T,p}}\right).
\end{aligned}$$

One shows likewise that

$$\left\| F(X^{(-\theta)})^2 \right\|_2 \leq C_{\lambda,b,\sigma,\varphi,T}\left(1+\left\|\theta\right\|^{2\lambda}_{L^2_{T,p}}\right).$$

Combining theses estimates shows that $H_{\lambda,\eta}(\theta, X^{(-\theta)}, W)$ satisfies the linear growth assumption in $L^2(\mathbb{P})$.

If σ is unbounded it follows from Assumption $(\mathcal{H}_{b,\sigma})$ that, for every $(t,x) \in [0,T]\times\mathcal{C}([0,T],\mathbb{R}^d)$,

$$\left\|\sigma(t,x)\right\| \leq C_\sigma\left(1+\left\|x\right\|_\infty\right).$$

Elementary computation based on (3.29) and Lemma 3 yield

$$\begin{aligned}
\left\| \int_0^T |\sigma(s, X^{(\theta),s})\Theta^{(\theta)}_s| \right\|_r &\leq C_\sigma \left\|\theta\right\|_{L^1_{T,p}} \left\|\varphi\right\|_\infty \left(1+\left\| \left\|X\right\|_\infty e^{-\frac{r}{2}\left\|\frac{\Theta}{r}\right\|^2_{L^2_{T,q}} + \int_0^T \langle \frac{\Theta_s}{r}, \mathrm{d}W_s\rangle} \right\|_r\right), \\
&\leq C_\sigma \left\|\theta\right\|_{L^1_{T,p}} \left\|\varphi\right\|_\infty \left(1+e^{\frac{\left\|\varphi\right\|_\infty}{2rr'}\left\|\theta\right\|^2_{L^2_{T,p}}} \left\|\left\|X\right\|_\infty\right\|_{r(1+r')}\right), \\
&\leq C_{r,b,\sigma,\varphi} \left\|\theta\right\|_{L^2_{T,p}},
\end{aligned}$$

for every $r > 0$ (Assumption $(\mathcal{H}_{b,\sigma})$ implies that $\|\|X\|_\infty\|_r < +\infty$ for every $r > 0$). Following the same proof to the bounded case, we obtain easily the results with $\Psi_{\lambda,\eta}(\theta,\xi) = \frac{e^{-(\|\varphi\|_\infty+\eta)\|\theta\|_{L^2_{T,p}}}}{1+\|\varphi(.,\xi_\cdot)\theta\|^{2\lambda+\eta}_{L^2_{T,q}}}$. We conclude by noting that η is an arbitrary parameter to cancel the denominator. $\diamond$

Remark. If the functional F is bounded ($\lambda = 0$), we prove in the same way that the algorithm without correction, *i.e.* build with $\Psi_{\lambda,\eta} = 1$, a.s. converges.

4 Additional remarks

For the sake of simplicity we focus in this section on importance sampling by mean translation in a finite dimensional setting (Section 2.3) although most of the comments below can also be applied at least in the path-dependent diffusions setting.

4.1 Purely adaptive approach

As proved by Arouna (see [2]), we can consider a purely adaptive approach to reduce the variance. It consists to perform the Robbins-Monro algorithm simultaneously with the Monte Carlo approximation. More precisely, estimate $\mathbb{E}[F(X)]$ by

$$S_N = \frac{1}{N}\sum_{k=1}^{N} F(X_k + \theta_{k-1})\frac{p(X_k+\theta_{k-1})}{p(X_k)}$$

where X_k is the *same innovation* as that used in the Robbins-Monro procedure $\theta_k = \theta_{k-1} - \gamma_k H(\theta_{k-1}, X_k)$. This adaptive Monte Carlo procedure satisfies a Central Limit Theorem with the optimal asymptotic variance

$$\sqrt{N}\left(S_N - \mathbb{E}[F(X)]\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma_*^2), \quad \text{whith} \quad \sigma_*^2 = V(\theta^*) - \mathbb{E}[F(X)]^2\,.$$

This approach can be extended to the Esscher transform when we use the same innovation ξ_k (see (2.16)) for the Monte Carlo procedure (computing $X_k^{(\theta_{k-1})} = g(\theta_{k-1},\xi_k)$) and the Robbins-Monro algorithm (computing $X_k^{(-\theta_{k-1})} = g(-\theta_{k-1},\xi_k)$). Likewise in the functional setting we can combine the variance reduction procedure and the Monte Carlo simulations using the same Brownian motion.

In practice, it is not clear that this adaptive Monte Carlo is better than the naive two stage procedure: performing first Robbins-Monro with a small number of iterations (to get a rough estimate θ^*), then performing the Monte Carlo simulations with this optimized parameter.

4.2 Weak rate of convergence: Central Limit Theorem (CLT)

As concerns the rate of convergence, once again this a regular stochastic algorithm behaves as described in usual Stochastic Approximation Theory textbooks like [13], [5], [8]. So, as soon as the optimal variance reducer set is reduced to a single point θ^*, the procedure satisfies under quite standard assumptions a CLT. We will not enter into technicalities at this stage but only try to emphasize the impact of a renormalization factor $g(\theta)$ like $g(\theta) := e^{-\frac{\lambda}{2}|\theta|^b}$ or $g(\theta) := \frac{1}{1+\widetilde{F}(-\theta)^2}$ induced by the function F on the "final" rate of convergence of the algorithm toward θ^*. We will assume that $d = 1$ and that $X \overset{d}{=} \mathcal{N}(0;1)$ for the sake of simplicity. One can write

$$H(\theta,x) = g(\theta)H_0(\theta,x) \quad \text{where} \quad H_0(\theta,x) = F^2(x-\theta)(2\theta - x)$$

The function H_0 corresponds to the case of a bounded function F (then $\lambda = 0$). Under simple integration assumptions, one shows that V is twice differentiable and that

$$V''(\theta) = e^{\frac{|\theta|^2}{2}} \mathbb{E}\Big[F^2(X)e^{-\theta X}\left(1 + (\theta - X)^2\right)\Big]$$

Consequently the mean functions h and h_0 related to H and H_0 which read respectively

$$h(\theta) = g(\theta)e^{-|\theta|^2}V'(\theta) \quad \text{and} \quad h_0(x) = e^{-|\theta|^2}V'(\theta)$$

are differentiable at θ^* and

$$h'(\theta^*) = g(\theta^*)e^{-|\theta^*|^2}V'(\theta^*) \quad \text{and} \quad h_0'(\theta^*) = e^{-|\theta^*|^2}V'(\theta^*)$$

Now, general results about CLT say that if $\gamma_n = \frac{\alpha}{\beta+n}$, α, $\beta > 0$ with

$$\alpha > \frac{1}{2h'(\theta^*)} = \frac{1}{2g(\theta^*)h_0'(\theta^*)} \tag{4.30}$$

then

$$\sqrt{n}(\theta_n - \theta^*) \overset{\mathcal{L}_{stably}}{\longrightarrow} \mathcal{N}(0; \Sigma_\alpha^*)$$

where

$$\Sigma_\alpha^* = \mathrm{Var}(H(y^*, Z))\frac{\alpha^2}{2\alpha h'(y^*) - 1}. \tag{4.31}$$

The mapping $\alpha \mapsto \Sigma_\alpha$ reaches its minimum at $\alpha^* = \frac{1}{h'(\theta^*)} = \frac{1}{g(\theta^*)h_0'(\theta^*)}$ leading to the minimal asymptotic variance

$$\Sigma^* = \Sigma_{\alpha^*}^* = \frac{\mathrm{Var}(H(y^*, Z))}{h'(y^*)^2} = \frac{\mathbb{E}\big[H_0(y^*, Z)^2\big]}{h_0'(y^*)^2} = \frac{\mathbb{E}\big[F^4(X)(\theta^* - X)^2 e^{-\theta^* X}\big]}{\mathbb{E}[F^2(X)(X^2 - \theta^* X + 1)]^2}$$

by homogeneity.

So the optimal rate of convergence of the procedure is not impacted by the use of the normalizing function $g(\theta)$. However, coming back to condition (4.30), we see that this assumption on the coefficient α is more stringent since $\frac{1}{g(\theta^*)} > 1$ (in practice this factor can be rather large). Consequently, given the fact that $g(\theta^*)$ is unknown to the user, this will induce a blind choice of α biased to higher values. With the well-known consequence in practice that if α is too large the "*CLT* regime" will take place later than it would with smaller values. One solution to overcome this contradiction can be to make α depend on n and slowly decrease.

As a conclusion, the algorithm never explodes (and converges) even for strongly unbounded functions F which is a major asset compared to the version of the algorithm based on repeated projections. Nevertheless, the normalizing factor which ensures the non-explosion of the procedure may impact the rate of convergence since it has an influence on the tuning of the step sequence (which is always more or less "blind" since it depends on the target θ^*. In fact, we did not meet such difficulty in our numerical experiments reported below.

One classical way to overcome this problem can be to introduce the empirical mean of the algorithm implemented with a slowly decreasing step "à la Rupert & Poliak" (see *e.g.* [17]): Set $\gamma_n = \frac{c}{n^r}$, $\frac{1}{2} < r < 1$ and

$$\bar{\theta}_{n+1} := \frac{\theta_0 + \cdots + \theta_n}{n+1} = \bar{\theta}_n - \frac{1}{n+1}(\bar{\theta}_n - \theta_n), \quad n \geq 0$$

where $(\theta_n)_{n\geq 0}$ denotes the regular Robbins-Monro algorithm defined by (2.14) starting at θ_0. Then $(\bar{\theta}_n)_{n\geq 0}$ converges toward θ^* and satisfies a CLT with the optimal asymptotic variance (4.31). See also a variant based on a gliding window developed in [14].

4.3 Extension to more general sets of parameters

In many applications (see below with the Spark spread options with the NIG distribution) the natural set of parameters Θ is not $\mathbb{R}^q$ but an open connected subset of $\mathbb{R}^q$. Nevertheless, as illustrated below, our unconstrained approach still works provided one can proceed a diffeomorphic change of parameter by setting

$$\theta = T(\tilde{\theta}), \quad \theta \in \Theta$$

where $T : \mathbb{R}^q \to \Theta$ is a $\mathcal{C}^1$-diffeomorphism with a bounded differential (*i.e.* $\sup_{\tilde{\theta}} \|DT(\tilde{\theta})\| < +\infty$). As an illustration, let us consider the case where the state function $H(\theta, X)$ of the procedure is designed so that $h(\theta) := \mathbb{E}(H(\theta, X)) = \rho(\theta)\nabla V(\theta)$ where V is the objective function to be minimized over Θ and ρ is a bounded *positive* Borel function. Then, one replaces $H(\theta, X)$ by $\widetilde{H}(\tilde{\theta}, X) := DT(\tilde{\theta}).H(T(\tilde{\theta}), X)$ and defines recursively a procedure on $\mathbb{R}^q$ by

$$\tilde{\theta}_{n+1} = \tilde{\theta}_n - \gamma_{n+1}\widetilde{H}(\tilde{\theta}_n, X_{n+1}).$$

In order to establish the *a.s.* convergence of $\theta_n := T(\tilde{\theta}_n)$ to $\operatorname{Argmin} V$, one relies on a variant of Robbins-Monro algorithm, namely a stochastic gradient approach (see [8, 13] for further details): one defines $U(\tilde{\theta}) = V(T(\tilde{\theta}))$ which turns out to be a Lyapunov function for the new algorithm since

$$\langle \nabla U(\tilde{\theta}), \mathbb{E}(DT(\tilde{\theta})H(T(\tilde{\theta}), X))\rangle = \rho(T(\tilde{\theta}))|\nabla U(\tilde{\theta})|^2 > 0 \quad \text{on} \quad T^{-1}(\{\nabla V \neq 0\}).$$

If U satisfies $\|\widetilde{H}(\tilde{\theta}, X)\|_2 + |\nabla U(\tilde{\theta})| \leq C(1+U(\tilde{\theta}))^{\frac{1}{2}}$ (which is a hidden constraint on the choice of T), one shows under the standard "decreasing" assumption on the step sequence that $U(\tilde{\theta}_n) \to U_\infty \in L^1(\mathbb{P})$ and $\sum_n \gamma_{n+1}\rho(\tilde{\theta}_n)|\nabla U(\tilde{\theta}_n)|^2 < +\infty$. If $\lim_{\theta \to \partial\Theta} V(\theta) = +\infty$ or $\liminf_{\theta \to \partial\Theta} \rho(T(\theta))|\nabla V(\theta)|^2 > 0$, one easily derives that $\operatorname{dist}(\theta_n, \{\nabla V = 0\}) \to 0$ *a.s.* as $n \to \infty$.

5 Numerical illustrations

5.1 Multidimensional setting: the NIG distribution

First we consider a simple case to compare the two algorithms of Section 2. The quantity to compute is

$$\mathbb{E}[F(X)] = \int_{\mathbb{R}} F(x)\, p_{\mathrm{NIG}}(x; \alpha, \beta, \delta, \mu)\, \mathrm{d}x,$$

where $p_{\mathrm{NIG}}(x; \alpha, \beta, \delta, \mu)$ is the density of X a normal inverse gaussian (NIG) random variable of parameters $(\alpha, \beta, \delta, \mu)$ i.e. $\alpha > 0$, $|\beta| \leq \alpha$, $\delta > 0$, $\mu \in \mathbb{R}$,

$$p_{\mathrm{NIG}}(x; \alpha, \beta, \delta, \mu) = \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\pi\sqrt{\delta^2 + (x-\mu)^2}} e^{\delta\gamma + \beta(x-\mu)},$$

where K_1 is a modified Bessel function of the second kind and $\gamma = \sqrt{\alpha^2 - \beta^2}$.

We can summarize the two algorithms presented in section 2, more precisely the variance reduction based on translation of the density (see Subsection 2.3) and the one based on the Esscher transform (see Subsection 2.4), by the following simplified (no computation of the variance) pseudo-code:

Translation (see 2.3)

```
for n = 0 to M do
    X ~ NIG(alpha, beta, mu, delta)
    theta = theta - 1/(n+1000)*H1(theta, X)
for n = 0 to N do
    X ~ NIG(alpha, beta, mu, delta)
    mean = mean + F(X) * p(X+theta)/p(X)
```

Esscher transform (see 2.4)

```
for n = 0 to M do
    X ~ NIG(alpha, beta-theta, mu, delta)
    theta = theta - 1/(n+1000)*H2(theta, X)
for n = 0 to N do
    X ~ NIG(alpha, beta+theta, mu, delta)
    mean = mean + F(X) * exp(-theta*X)
mean = mean * exp(psi(theta))
```

– *Translation case.* We consider the function H_1 of the Robbins-Monro procedure of the first algorithm defined by

$$H_1(\theta, X) = e^{-2|\theta|} F^2(X) \frac{p'(X-2\theta)}{p(X)} \left(\frac{p(X-\theta)}{p(X-2\theta)} \right)^2,$$

where an analytic formulation of the derivative p' is easily obtained using the relation on the modified Bessel function $K_1'(x) = \frac{1}{x} K_1(x) - K_2(x)$.

The assumption $(\mathcal{H}_a^{tr})$ is satisfied with $a = 1$, and our results of Subsection 2.3 apply.

– *Esscher transform.* In the Esscher approach we consider the function H_2 defined by

$$H_2(\theta, X) = e^{-|\theta|} F^2(X) \left(\nabla\psi(\theta) - X \right).$$

Note that ψ is not well defined for every $\theta \in \mathbb{R}^d$. Indeed, the cumulant generating function of the NIG distribution is defined by

$$\psi(\theta) = \mu\theta + \delta \left(\gamma - \sqrt{\alpha^2 - (\beta+\theta)^2} \right),$$

for every $\theta \in (-\alpha-\beta, \alpha-\beta)$. Moreover, we need $\psi(-\theta)$ to be well defined *i.e.* $\theta \in (-\alpha+\beta, \alpha+\beta)$. To take account of these restrictions, we slightly modify the algorithm parametrization (see Subsection 4.3) $\theta = T(\tilde{\theta}) := (\beta - \alpha) \frac{\tilde{\theta}}{\sqrt{1+\tilde{\theta}^2}}$, and update $\tilde{\theta} \in \mathbb{R}$ in the Robbins-Monro procedure (multiply the function $H_2(T(\tilde{\theta}), X)$ by the derivative $T'(\tilde{\theta}) = \frac{\beta-\alpha}{(1+\tilde{\theta}^2)^{3/2}}$).

The payoff F is a Call option of strike K, $F(X) = 50(e^X - K)_+$. The parameters of the NIG random variable X are $\alpha = 2$, $\beta = 0.2$, $\delta = 0.8$ and $\mu = 0.04$. The variance reduction obtained for different value of K are summarized in the tabular 1. The number of iterations in the Robbins-Monro variance reduction procedure is $M = 100\,000$ and the number of Monte Carlo iterations is $N = 1\,000\,000$. Note that for each strike, the prices are computed using the same pseudo-random number generator initialized with the same *seed*.

K	mean	crude var	var. ratio. translation	(θ)	var. ratio Esscher	(θ)
0.6	42.19	8538	5.885	(0.791)	56.484	(1.322)
0.8	34.19	8388	7.525	(0.903)	39.797	(1.309)
1.0	27.66	8176	9.218	(0.982)	32.183	(1.294)
1.2	22.60	7930	10.068	(1.017)	29.232	(1.280)
1.4	18.76	7677	9.956	(1.026)	28.496	(1.268)

Table 1: Variance reduction for different strikes (one dimensional NIG example).

To complete this numerical example, Figure 1 illustrates the densities obtained after the Robbins-Monro procedure. The deformation provided by the Esscher transform is very impressive in this example. We remark that the Esscher transform modifies the parameter β which controls the asymmetric shape of the NIG distribution.

Spark spread

We consider now a exchange option between gas and electricity (called spark spread). We choose to model the price of the energy by the exponential of a NIG distribution. A simplified form of the payoff is then

$$F(X) = 50(e^{X^{elec}} - ce^{X^{gas}} - K)_+,$$

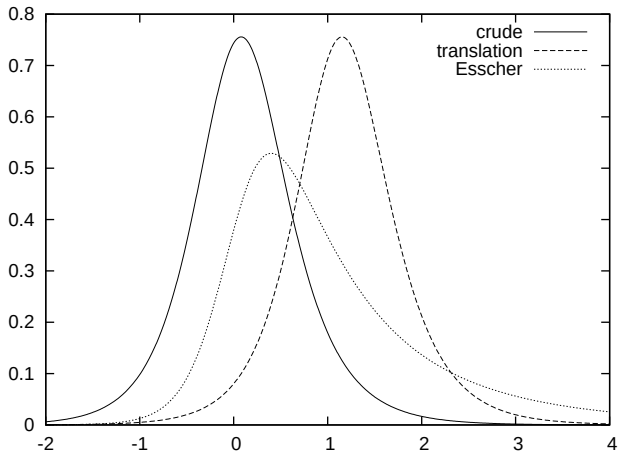


Figure 1: Densities of X (crude), $X + \theta$ (translation) and $X^{(\theta)}$ (Esscher) in the case $K = 1$.

where $X^{elec} \sim NIG(2, 0.2, 0.8, 0.04)$ and $X^{gas} \sim NIG(1.4, 0.2, 0.2, 0.04)$ are independent.

The results obtained for different strikes after 300 000 iterations of the Robbins-Monro procedure and 3 000 000 iterations of Monte Carlo, are summarized in the Table 2.

K	c	mean	crude var	var. ratio. translation	var. ratio Esscher
0.4	0.2	41.021	8540.6	5.0118	25.171
	0.4	32.719	8356.9	5.1338	27.006
	0.6	26.337	8112.2	4.9752	28.062
	0.8	21.556	7845.3	4.7569	29.964
	1	17.978	7582	4.5575	32.849
0.6	0.2	33.235	8378.4	5.2609	27.455
	0.4	26.534	8133.3	5.0604	28.669
	0.6	21.587	7862.7	4.8046	30.649
	0.8	17.931	7595.2	4.5839	33.656
	1	15.184	7344.2	4.4064	37.489
0.8	0.2	26.908	8160.1	5.1366	28.876
	0.4	21.725	7884.9	4.844	31.018
	0.6	17.955	7612.5	4.6031	34.166
	0.8	15.156	7357.3	4.416	38.167
	1	13.027	7123.9	4.2685	42.781

Table 2: Variance reduction for different strikes (spark spread example).

5.2 Functional setting: Down & In Call option

We consider a process $(X_t)_{t\geq 0}$ solution of the following diffusion

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad X_0 = x_0 \in \mathbb{R}.$$

A Down & In Call option of strike K and barrier L is a Call of strike K which is activated when the underlying X moves down and hits the barrier L. The payoff of such a European option is

defined by

$$F(X) = (X_T - K)_+ \mathbf{1}_{\left\{\min_{0\le t\le T} X_t \le L\right\}}.$$

A naive Monte Carlo approach to price this option is to consider an Euler-Maruyama scheme $\bar{X} = (\bar{X}_{t_k})_{k\in\{0,\dots,n\}}$ to discretize X and to approximate $\min_{0\le t\le T} X_t$ by $\min_{k\in\{0,\dots,n\}} \bar{X}_{t_k}$. It is well known that this approximation of the functional payoff is poor. More precisely, the weak order of convergence cannot be greater than $\frac{1}{2}$ (see [11]).

A standard approach is to consider the continuous Euler scheme $\bar{X}^c$ obtained by extrapolation of the Brownian between two instants of discretization. More precisely, for every $t \in [t_k, t_{k+1}]$,

$$\bar{X}^c_t = \bar{X}^c_{t_k} + b(\bar{X}^c_{t_k})(t - t_k) + \sigma(\bar{X}^c_{t_k})(W_t - W_{t_k}), \quad \bar{X}^c_0 = x_0 \in \mathbb{R}.$$

By preconditioning,

$$\mathbb{E}[F(X)] = \mathbb{E}\left[(\bar{X}_T - K)_+ \left(1 - \prod_{k=0}^{N-1} p(\bar{X}_{t_k}, \bar{X}_{t_{k+1}})\right)\right], \tag{5.32}$$

with $p(x_k, x_{k+1}) = \mathbb{P}\left[\min_{t\in[t_k,t_{k+1}]} \bar{X}^c_t \ge L \,\middle|\, (\bar{X}_{t_k}, \bar{X}_{t_{k+1}}) = (x_k, x_{k+1})\right]$. Now using the Girsanov Theorem and the law of the Brownian bridge (see for example [9]), we have

$$\begin{aligned} p(x_k, x_{k+1}) &= 1 - \mathbb{P}\left[\min_{t\in[0,t_1]} W_t \le \frac{L - x_k}{\sigma(x_k)} \,\middle|\, W_{t_1} = \frac{x_{k+1} - x_k}{\sigma(x_k)}\right], \\ &= \begin{cases} 0 & \text{if } L \ge \min(x_k, x_{k+1}), \\ 1 - e^{-\frac{2(L-x_k)(L-x_{k+1})}{\sigma^2(x_k)(t_{k+1}-t_k)}}, & \text{otherwise.} \end{cases} \end{aligned} \tag{5.33}$$

In the following simulations we consider an Euler scheme of step $t_k = k\frac{T}{n}$ with $n = 100$.

Deterministic case (trivial driver $\varphi \equiv 1$)

We consider three different basis of $L^2([0,1], \mathbb{R})$

– a polynomial basis composed of the shifted Legendre polynomials $\tilde{P}_n(t)$ defined by

$$\forall n \ge 0, \forall t \in [0,1], \quad \tilde{P}_n(t) = P_n(2t-1) \quad \text{where} \quad P_n(t) = \frac{1}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}t^n}\left((t^2-1)^n\right). \tag{ShLeg}$$

– the Karhunen-Loève basis defined by

$$\forall n \ge 0, \forall t \in [0,1], \quad e_n(t) = \sqrt{2}\sin\left(\left(n + \frac{1}{2}\right)\pi t\right) \tag{KL}$$

– the Haar basis defined by

$$\forall n \ge 0, \forall k = 0, \dots, 2^n - 1, \forall t \in [0,1], \quad \psi_{n,k}(t) = 2^{\frac{k}{2}}\psi(2^k t - n) \tag{Haar}$$

$$\text{where } \psi(t) = \begin{cases} 1 & \text{if } t \in [0, \frac{1}{2}) \\ -1 & \text{if } t \in [\frac{1}{2}, 1) \\ 0 & \text{otherwise} \end{cases}$$

Black&Scholes Model
First, we consider the classical Black&Scholes model. We set the interest rate r to 4% and the volatility σ to 70% (which is a high volatility). The strike of the payoff F is set at $K = 115$ and the barrier level at $L = 65$. A crude Monte Carlo (with Brownian bridge interpolation, see (5.32)) give a price of 2.596 with a variance of 230 after 500 000 trials. Note that the true price of this product is 2.554.

For different basis, the results of our algorithm are summarized in the table 5.2. In the Robbins-Monro procedure, we define the step sequence by $\gamma_n = \frac{1}{n+10x_0^2}$ and set the number of iterations at 50 000.

Basis	Dim.	Mean	CI 95%	Variance ratio
Constant	1	2.5737	±0.0230	3.4710
ShiftLegendre	2	2.5741	±0.0197	4.7225
(ShLeg)	4	2.5717	±0.0193	4.9478
	8	2.5717	±0.0193	4.9494
Karhunen-Loève	2	2.5678	±0.0164	6.8644
(KL)	4	2.5729	±0.0160	7.1851
	8	2.5705	±0.0156	7.5218
Haar	2	2.5657	±0.0192	4.9710
(Haar)	4	2.5671	±0.0163	6.9459
	8	2.5663	±0.0155	7.6574

Table 3: Variance ratio obtained for different basis in the Black&Scholes model ($K = 115$, $L = 65$, variance of the crude Monte Carlo: 230).

In figure 2 are depicted the optimal variance reducer when the optimization of V is carried out on E_m for several values of m (2, 4 and 8) in the different basis mentioned above.

A local volatility Model
To emphasize the generic feature of our algorithm we consider the same product in a local volatility model (inspired by the CEV model) defined by

$$\mathrm{d}x_t = rx_t\,\mathrm{d}t + \sigma x_t^{\beta}\frac{x_t}{\sqrt{1+x_t^2}}\,\mathrm{d}W_t, \tag{5.34}$$

with $r = 0.04$, $\sigma = 7$ and $\beta = 0.5$.

The price of the Down & In Call (strike 115, barrier 65) given by a crude Monte Carlo with Brownian interpolation after 500 000 trials is 3.194 and the variance is 206.52.

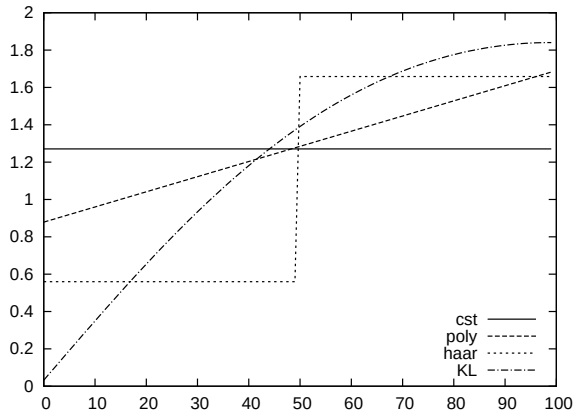
Adaptive case (non-trivial driver)

We experiment now our algorithm with a non-trivial driver φ defined for $t = t_k$ by

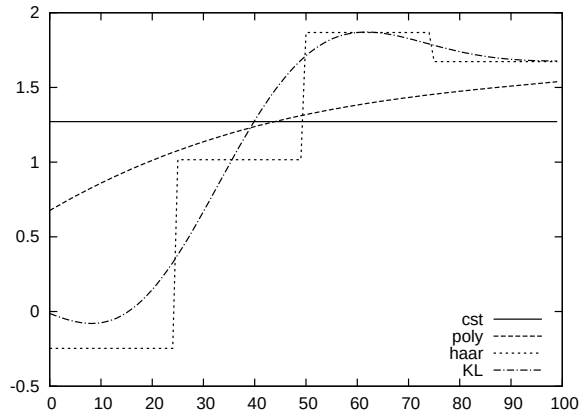
$$\varphi(t,\xi^t) = \begin{pmatrix}\bar{p}_k & 1-\bar{p}_k\end{pmatrix}, \quad \text{with} \quad \bar{p}_k = \prod_{j=0}^{k-1} p(\xi_{t_j}, \xi_{t_{j+1}}),$$

where p is defined by (5.33). Note that $\bar{p}_k = \mathbb{P}\big[\min_{t\in[0,t_k]}\xi_t \geq L \,\big|\, \xi_0,\ldots,\xi_{t_k}\big]$ so that there is no extra-computation compared to the Brownian bridge interpolation.

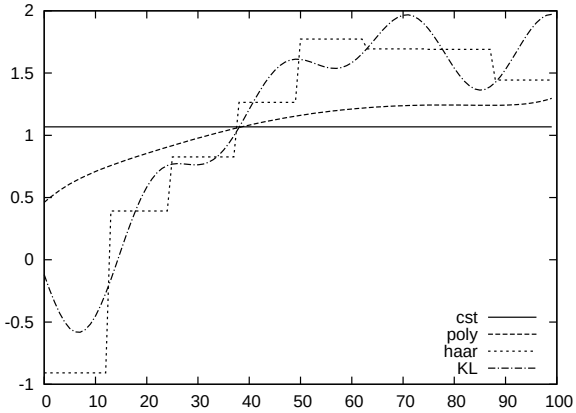
We set $p = 2$ and $E = (\mathbb{R}\mathbb{1}_{[0,T]})^2$ so that the optimal parameter $\theta_{t_k} = \alpha\bar{p}_k + \beta(1-\bar{p}_k)$ with $(\alpha,\beta) \in \mathbb{R}^2$. The results for different strikes and barrier levels are reported in Table 5 for the Black&Scholes model and in Table 6 for the local volatility model. The simulation parameters are unchanged.



(a) 2 vectors



(b) 4 vectors



(c) 8 vectors

Figure 2: Optimal θ process obtained with different basis by our algorithm using 50 000 trials.

Basis	Dim.	Mean	CI 95%	Variance ratio
Constant	1	3.1836	±0.0251	2.6297
ShiftLegendre	2	3.1830	±0.0223	3.3258
(ShLeg)	4	3.1815	±0.0215	3.5670
	8	3.1813	±0.0215	3.5659
Karhunen-Loève	2	3.1852	±0.0187	4.7254
(KL)	4	3.1862	±0.0183	4.9385
	8	3.1918	±0.0178	5.2183
Haar	2	3.1834	±0.0215	3.5699
(Haar)	4	3.1871	±0.0186	4.7896
	8	3.1864	±0.0177	5.2675

Table 4: Variance ratio obtained for different basis in the local volatility model (5.34) ($K = 115$, $L = 65$, variance of the crude Monte Carlo: 206.52).

References

[1] C. Andrieu, É. Moulines, P. Priouret (2005). Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.*, **44**(1), no. 1, 283–312 (electronic).

Strike	Barrier	Mean	CI 95%	Variance ratio	(Crude)	α	β
85	65	2.5738	±0.0115	13.49	(16.56)	-0.1752	1.6685
	75	6.0489	±0.0186	14.26	(43.39)	0.0493	1.9191
95	65	2.5704	±0.0110	14.64	(15.26)	0.0524	1.9987
	75	6.0492	±0.0190	13.67	(45.25)	0.1557	2.0560
	85	11.5970	±0.0301	12.23	(112.92)	0.4108	2.1226
105	65	2.5687	±0.0122	12.03	(18.56)	0.3888	2.1423
	75	6.0548	±0.0206	11.66	(53.08)	0.3895	2.1720
	85	11.5953	±0.0308	11.67	(118.32)	0.4524	2.1608
	95	19.2882	±0.0348	17.17	(151.04)	0.6619	1.7910
115	65	2.5706	±0.0135	9.75	(22.90)	0.5473	1.8903
	75	6.0530	±0.0211	11.16	(55.42)	0.4591	1.9371
	85	11.5976	±0.0297	12.55	(109.98)	0.4807	2.0008
	95	19.2958	±0.0347	17.21	(150.67)	0.7217	1.6380

Table 5: Variance reduction for different strikes and barrier levels in the Black&Scholes model.

Strike	Barrier	Mean	CI 95%	Variance ratio	(Crude)	α	β
85	65	3.1827	±0.0127	10.02	(20.28)	-0.3057	1.5522
	75	6.4115	±0.0190	9.96	(45.03)	-0.1428	1.7985
95	65	3.1846	±0.0124	10.65	(19.08)	-0.1141	1.9139
	75	6.4117	±0.0199	9.07	(49.42)	-0.0029	1.9814
	85	11.4478	±0.0293	8.03	(106.99)	0.1898	1.8937
105	65	3.1835	±0.0135	8.98	(22.65)	0.1487	1.9628
	75	6.4120	±0.0209	8.21	(54.59)	0.1493	2.0060
	85	11.4458	±0.0295	7.88	(108.94)	0.2503	1.8737
	95	18.6060	±0.0345	9.83	(149.07)	0.5594	1.4343
115	65	3.1817	±0.0148	7.38	(27.54)	0.3062	1.6884
	75	6.4112	±0.0209	8.18	(54.79)	0.1928	1.8119
	85	11.4470	±0.0289	8.24	(104.16)	0.2599	1.7430
	95	18.6061	±0.0346	9.79	(149.76)	0.5755	1.4313

Table 6: Variance reduction for different strikes and barrier levels in the local volatility model.

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

[3] N. Bouleau, D. Lépingle (1994). *Numerical methods for stochastic processes*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 359 pp. ISBN: 0-471-54641-0.

[4] B. Arouna, O. Bardou (2004). Efficient variance reduction for functionals of diffusions by relative entropy, technical report, CERMICS-ENPC (France).

[5] M. Benveniste, M. Métivier, P. Priouret (1990). *Adaptive algorithms and stochastic approximation*, **22**, *Applications of Mathematics*, transl. from French by S. Wilson, Springer-Verlag, Berlin.

[6] H. Chen, Y. Zhu (1986). *Stochastic Approximation Procedure with randomly varying truncations*, Scientia Sinica Series.

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

[8] M. Duflo (1997). *Iterative random models*, transl. from French, Springer-Verlag.

[9] P. Glasserman (2004). *Monte Carlo Methods in Financial Engineering*, Springer.

[10] P. Glasserman, P. Heidelberger, P. Shahabuddin (1999). Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance* **9**(2), no. 2, 117–152.

[11] E. Gobet (2000). Weak approximation of killed diffusion using Euler schemes, *Stoch. Proc. Appl.*, **87**(2), 167–197.

[12] R. Kawai (2008). Optimal importance sampling parameter search for Lévy Processes via stochastic approximation, pre-print.

[13] H.J. Kushner, G. Yin (2003). Stochastic Approximation and Recursive Algorithms and Applications, Springer.

[14] J. Lelong (2007). *Algorithmes stochastiques et Options parisiennes*, thèse de l'ENPC, 151p.

[15] V. Lemaire (2007). An adaptive scheme for the approximation of dissipative systems, *Stoch. Proc. Appl.*, **117**(10), 1491–1518.

[16] D. Revuz, M. Yor (1998). *Continuous martingales and Brownian motion*, 3^{rd} edition, Springer, Berlin, 1998 (1^{st} edition, 1990).

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

[18] L.C.G. Rogers, D. Williams (1986). *Diffusions, Markov Processes and Martingales*, 2^{nd} edition, Cambridge Mathematical Library.

6 Appendix: proof of Theorem 1

We propose below the proof of the slight extension of the regular Robbins-Monro algorithm when $\{h = 0\}$ is not reduced to a single equilibrium point. The key is still the convergence theorem for non negative super-martingales.

Proof. Set $\mathcal{F}_n := \sigma(\theta_0, Z_1, \ldots, Z_n)$, $n \geq 1$. Let $\theta^* \in \mathcal{T}^*$. Then

$$\begin{aligned} |\theta_{n+1} - \theta^*|^2 &= |\theta_n - \theta^*|^2 - 2\gamma_{n+1}\langle \theta_n - \theta^*, H(\theta_n, Z_{n+1})\rangle + \gamma_{n+1}^2 |H(\theta_n, Z_{n+1})|^2, \\ &\leq |\theta_n - \theta^*|^2 - 2\gamma_{n+1}\langle \theta_n - \theta^*, h(\theta_n)\rangle - 2\gamma_{n+1}\langle \theta_n - \theta^*, \Delta M_{n+1}\rangle + \gamma_{n+1}^2 |H(\theta_n, Z_{n+1})|^2, \end{aligned} \tag{6.35}$$

where

$$\Delta M_{n+1} = H(\theta_n, Z_{n+1}) - \mathbb{E}[H(\theta_n, Z_{n+1}) \,|\, \mathcal{F}_n] = H(\theta_n, Z_{n+1}) - h(\theta_n),$$

is an increment of (local) martingale satisfying $\mathbb{E}\big[|\Delta M_{n+1}|^2\big] \leq C(1 + \mathbb{E}\big[|\theta_n - \theta^*|^2\big])$ owing to the assumptions on H and Schwarz Inequality which also implies that

$$\mathbb{E}[|\langle \theta_n - \theta^*, H(\theta_n, Z_{n+1})\rangle|] \leq \frac{1}{2}\left(\mathbb{E}\big[|\theta_n - \theta^*|^2\big] + \mathbb{E}\big[|H(\theta_n, Z_{n+1})|^2\big]\right) \leq C(1 + \mathbb{E}\big[|\theta_n - \theta^*|^2\big]\,,$$

for an appropriate real constant C. Then, one shows by induction on n from (6.35) that $|\theta_n|$ is square integrable for every $n \geq 0$ and that ΔM_{n+1} is integrable, hence a true martingale increment. Now, one derives from the assumptions (2.6) and (6.35) that

$$S_n = \frac{|\theta_n - \theta^*|^2 + 2\sum_{k=0}^{n-1} \gamma_{k+1}\langle \theta_k - \theta^*, h(\theta_k)\rangle + C\sum_{k\geq n+1}\gamma_k^2}{\prod_{k=1}^n (1 + C\gamma_k^2)},$$

is a (non negative) super-martingale with $S_0 = |\theta_0 - \theta^*|^2 \in L^1(\mathbb{P})$. This uses the mean-reverting assumption (2.5). Hence S_n is $\mathbb{P}$-$a.s.$ converging toward an integrable r.v. S_∞. Consequently, using that $\sum_{k\geq n+1}\gamma_k^2 \to 0$, one gets

$$|\theta_n - \theta^*|^2 + 2\sum_{k=0}^{n-1} \gamma_{k+1}\langle \theta_k - \theta^*, h(\theta_k)\rangle \stackrel{a.s.}{\longrightarrow} \widetilde{S}_\infty = S_\infty \prod_{n\geq 1}(1 + C_{_L}\gamma_n^2) \in L^1(\mathbb{P}). \tag{6.36}$$

The super-martingale (S_n) being L^1-bounded, one derives likewise that $(|\theta_n - \theta^*|^2)_{n\geq 0}$ is L^1-bounded since

$$|\theta_n - \theta^*|^2 \leq \prod_{k=1}^n (1 + C_{_L}\gamma_k^2) S_n, \quad n \geq 0.$$

Now, a series with nonnegative terms which is upper bounded by an ($a.s.$) converging sequence, $a.s.$ converges in $\mathbb{R}_+$ so that

$$\sum_{n\geq 0} \gamma_{n+1}\langle \theta_n - \theta^*, h(\theta_n)\rangle < +\infty \qquad \mathbb{P}\text{-}a.s.$$

It follows from (6.36) that, $\mathbb{P}$-$a.s.$, $|\theta_n - \theta^*|^2 \stackrel{n\to\infty}{\longrightarrow} L_\infty$ which is integrable since $(|\theta_n - \theta^*|^2)_{n\geq 0}$ is L^1-bounded and consequently $a.s.$ finite.

Let $L > 0$. Set

$$\Omega_{_L} := \{\omega \in \Omega,\, \forall\, n \geq 0, |\theta_n(\omega) - \theta^*| \leq L\}\,.$$

It follows from the $a.s.$ finiteness of L_∞ that $\bigcup_{L>0} \Omega_{_L} = \Omega$ $a.s.$. Now we consider the compact set $K_{_L} = \mathcal{T}^* \cap \bar{B}(0, L)$. It is separable so there exists an everywhere dense sequence in $K_{_L}$, denoted for convenience $(\theta^{*,k})_{k\geq 1}$. The above proof shows that $\mathbb{P}$-$a.s.$, for every $k \geq 1$, $|\theta_n - \theta^{*,k}|^2 \to L_\infty^k < +\infty$ as $n \to \infty$. Then set

$$\Omega'_{_L} := \left\{ \omega \in \Omega_{_L},\ |\theta_n(\omega) - \theta^{*,k}|^2 \stackrel{n\to\infty}{\to} L_\infty^k(\omega), k \geq 1,\ \sum_{n\geq 1} \gamma_n \langle \theta_{n-1}(\omega) - \theta^*, h(\theta_{n-1}(\omega)\rangle < +\infty \right\}$$

which satisfies $\mathbb{P}(\Omega'_{_L}) = \mathbb{P}(\Omega_{_L})$. Assume $\omega \in \Omega'_{_L}$. Up to two successive extractions, there exists a subsequence $\theta_{\phi(n,\omega)}$ such that

$$\langle \theta_{\phi(n,\omega)} - \theta^*, h(\theta_{\phi(n,\omega)}(\omega))\rangle \stackrel{n\to\infty}{\longrightarrow} 0 \qquad \text{and} \qquad \theta_{\phi(n,\omega)}(\omega) \stackrel{n\to\infty}{\longrightarrow} \theta_\infty(\omega).$$

The function h being continuous $\langle \theta_\infty(\omega) - \theta^*, h(\theta_\infty(\omega))\rangle = 0$ which implies that $\theta_\infty(\omega) \in \{h = 0\}$. Hence $\theta_\infty(\omega) \in K_{_L}$. Then any limiting value $\theta'_\infty(\omega)$ of the sequence $(\theta_n(\omega))_{n\geq 1}$ will satisfy

$$\forall\, k \geq 1, \quad |\theta'_\infty(\omega) - \theta^{*,k}| = |\theta_\infty(\omega) - \theta^{*,k}| = \sqrt{L_\infty^k(\omega)}$$

which in turn implies that $\theta'_\infty(\omega) = \theta_\infty(\omega)$ by considering a subsequence $\theta^{*,k'} \to \theta_\infty(\omega)$. So, $\theta_\infty(\omega)$ is the unique limiting value of the sequence $(\theta_n(\omega))_{n\geq 0}$ *i.e.* $\theta_n(\omega) \to \theta_\infty(\omega)$ as $n \to \infty$. The fact that the resulting random vector θ_∞ is square integrable follows from Fatou's Lemma and the L^2-boundedness of the sequence $(\theta_n - \theta^*)_{n\geq 1}$. $\diamond$