The Annals of Applied Probability
2010, Vol. 20, No. 3, 1029–1067
DOI: 10.1214/09-AAP650

UNCONSTRAINED RECURSIVE IMPORTANCE SAMPLING

BY VINCENT LEMAIRE AND GILLES PAGÈS

Université Pierre et Marie Curie

We propose an *unconstrained* stochastic approximation method for finding the optimal change of measure (in an a priori parametric family) to reduce the variance of a Monte Carlo simulation. We consider different parametric families based on the Girsanov theorem and the Esscher transform (exponential-tilting). In [*Monte Carlo Methods Appl.* **10** (2004) 1–24], it described a projected Robbins–Monro procedure to select the parameter minimizing the variance in a multidimensional Gaussian framework. 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 extensively use the regularity of the density of the law without assuming smoothness of the payoff. We prove the convergence for a large class of multidimensional distributions as well as for diffusion processes.

We illustrate the efficiency of our algorithm on several pricing problems: a Basket payoff under a multidimensional NIG distribution and a barrier options in different markets.

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 through a probabilistic representation as an expectation:

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

where $X:(\Omega, \mathcal{A}, \mathbb{P}) \to (E, |\cdot|_E)$ is a random vector having values in a Banach space E and $F : E \to \mathbb{R}$ is a Borel function [such that $F(X)$ 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 a family of absolutely continuous probability

Received September 2008; revised September 2009.

AMS 2000 subject classifications. Primary 65C05, 65B99; secondary 60H35.

Key words and phrases. Stochastic algorithm, Robbins–Monro, importance sampling, Esscher transform, NIG distribution, barrier options.

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 a connected open nonempty subset of $\mathbb{R}^q$, containing 0 so that $p_0 = p$. In fact, we will assume 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

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

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 $\theta \in \mathbb{R}^q$,

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

The second equality follows from a reverse change of probability.

A typical parametrized family comes from the implementation of importance sampling by mean translation in a finite-dimensional Gaussian framework, that is,

$$X^{(\theta)} = X + \theta,\ p(x) = \frac{e^{-|x|^2/2}}{(2\pi)^{d/2}}, \qquad p_\theta(x) = p(x - \theta)$$

and

$$V(\theta) = e^{-|\theta|^2}\mathbb{E}\big[F^2(X)e^{-2\langle\theta, X\rangle}\big].$$

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 [3] (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

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

and if F satisfies $\mathbb{E}[F^2(X)\frac{p(X)}{p_\theta(X)}] < +\infty$ 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, $\operatorname{Arg\,min} V = \{\nabla V = 0\}$ *is nonempty*. Assumption $(\mathcal{H}_1)$(ii) can be localized by considering that one of the two conditions holds only on a Borel set B of $\mathbb{R}^d$ such that $\mathbb{P}_X[B \cap \{F \neq 0\}] > 0$. Also note that if $\theta \mapsto p_\theta(x)$ is *strictly* log-concave for every x in a Borel set C such that $\mathbb{P}_X[C \cap \{F \neq 0\}] > 0$, then V is strictly convex and $\operatorname{Arg\,min} V = \{\nabla V = 0\}$ is reduced to a single $\theta^* \in \mathbb{R}^q$. So is the case of importance sampling by mean translation in a finite-dimensional Gaussian space (with $B = C = \mathbb{R}^d$). 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 assume that $\operatorname{Arg\,min} V = \{\theta^*\}$ throughout this introduction section, although our main results do not need such a 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 to have any, we are naturally led to *formally* differentiate the second representation of V in (1.2) to obtain a classical representation of ∇V as

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

This representation is also known as the "likelihood ratio method" in sensitivity analysis.

Then, if we consider the function $\bar{H}_V(\theta, x)$ such that $\nabla V(\theta) = \mathbb{E}(\bar{H}_V(\theta, X))$ 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 $p(x)\lambda_d(dx)$. At a first glance, to establish the convergence of a Robbins–Monro procedure to $\theta^* = \operatorname{Arg\,min} V$ 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

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

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

[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]).

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 [3]. However, as confirmed by the numerical experiments carried out by several authors [3, 13, 15], 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 behavior 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., [2, 15]). Formally, repeated projections "à la Chen" can be written as follows:

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

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 [15], it established a central limit theorem for this version of the recursive variance reduction procedure. Some extensions to non-Gaussian frameworks 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 [13].

However, convergence occurs for this procedure after a long "stabilization phase"… provided that the sequence of compact sets has 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 [16]).

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

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

where $\widetilde{X}^{(\theta)} \sim q_\theta(x)\,dx$. The "weight" $\frac{p^2(\widetilde{X}^{(\theta)})}{p_\theta(\widetilde{X}^{(\theta)})q_\theta(\widetilde{X}^{(\theta)})}\frac{\nabla p_\theta(\widetilde{X}^{(\theta)})}{p_\theta(\widetilde{X}^{(\theta)})}$ looks complicated but the rôle of the density q_θ is to control the term $\frac{p^2(x)}{p_\theta(x)q_\theta(x)}$ by a (deterministic) quantity only depending on θ. Then, we can replace in the above Robbins–Monro procedure (AlgoRM) $\widetilde{H}_V$ by a function $H(\theta, x) = \delta(\theta)\widetilde{H}_V(\theta, x)$ where δ is a positive function used to force a sub-linear behavior of $\theta \mapsto \|H(\theta, \widetilde{X}^{(\theta)})\|_{\mathbb{L}^2}$ (note that $\{\mathbb{E}[H(\cdot, \widetilde{X}^{(\theta)})] = 0\} = \{\nabla V = 0\}$). Note that to remain within the framework of standard stochastic approximation, we have the further constraint: to represent $\widetilde{X}^{(\theta)}$ as $\widetilde{X}^{(\theta)} = \chi(\theta, \Xi)$, where Ξ is an exogeneous random vector.

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 change the measure of the process. 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 Banach space $(\mathcal{C}([0, T], \mathbb{R}^d), \|\cdot\|_\infty)$ 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

$$(E_{b,\sigma,W}) \qquad dX_t = b(t, X^t)\,dt + \sigma(t, X^t)\,dW_t, \qquad X_0 = x \in \mathbb{R}^d,$$

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 continuous functionals such that for every $t \in [0, T]$, both $b(t, \cdot)$ and $\sigma(t, \cdot)$ are Lipschitz with respect to the $\|\cdot\|_\infty$-norm on the space $\mathcal{C}([0, T], \mathbb{R}^d)$ (see [19] for more details about these path-dependent SDEs).

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

$$\mathbb{E}[F(X)] = \mathbb{E}\big[F\big(X^{(\theta)}\big)e^{-\int_0^T \langle \varphi(X^{(\theta),s})\theta(s), dW_s\rangle - 1/2\|\varphi(X^{(\theta),\cdot})\theta\|^2_{L^2_{T,q}}}\big],$$

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

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

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

The paper is organized as follows. Section 2 is devoted to the finite-dimensional setting where we recall our main tools including a slight extension of the Robbins–Monro convergence theorem in Section 2.1. The Gaussian case investigated in [3] is revisited to emphasize the new aspects of our algorithm in Section 2.2.

In Section 2, we successively investigate the translation for log-concave probability distributions and the Esscher transform. In Section 3, we introduce the functional version of our algorithm based on the Girsanov theorem to deal with the SDE framework. 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.

NOTATION: The space $\mathcal{M}(d,q)$ denotes the space of $d\times q$ real matrices.

We will denote by $S>0$ the fact that a symmetric matrix S is positive definite. $|\cdot|$ will denote the canonical Euclidean norm on $\mathbb{R}^m$ and $\langle\cdot,\cdot\rangle$ will denote the canonical inner product.

The real constant $C>0$ denotes a positive real constant that may vary from line to line.

$\|f\|_{L^2_{T,p}} := (\int_0^T f_1^2(t)+\cdots+f_p^2(t)\,dt)^{1/2}$ if $f=(f_1,\dots,f_p)$ is an $\mathbb{R}^p$-valued (class of) Borel function(s).

2. The finite-dimensional setting.

2.1. Arg min 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*,

$$\operatorname{Arg\,min} V = \{\nabla V = 0\} \neq \varnothing.$$

PROOF. By the change of probability $\frac{d\pi_\theta}{d\lambda_d}$, we have $V(\theta)=\mathbb{E}[F^2(X)\frac{p(X)}{p_\theta(X)}]$. Let x be 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.

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

- If $\lim_{|\theta|\to+\infty} p_\theta(x)=0$ for every $x\in\mathbb{R}^d$, the result follows from Fatou's lemma.
- If $\lim_{|\theta|\to+\infty}\frac{p_\theta(x)}{p^2_{\theta/2}(x)}=0$ for every $x\in\mathbb{R}^d$, we apply the reverse Hölder inequality with conjugate exponents $(\frac{1}{3},-\frac{1}{2})$ to obtain

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

(p and p_θ are probability density functions). One concludes again by Fatou's lemma. $\square$

The set $\operatorname{Arg\,min} V$, or to be precise, the random vectors taking values in $\operatorname{Arg\,min} V$ will the target(s) of our new algorithm. If V is strictly convex, for example, if

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

then $\operatorname{Arg\,min} V = \{\theta^*\}$ (of course this is only a sufficient condition). 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 $\operatorname{Arg\,min} U$ *is nonempty, it is a convex closed set* (*which coincides with* $\{\nabla U = 0\}$) *and*

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

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

At this stage, we need to recall a convergence theorem for this Robbins–Monro procedure. Since we do not wish to make a uniqueness assumption on $\operatorname{Arg\,min} V$, we propose for convenience a slight variant of the classical theorem commonly available in the literature (see, e.g., [4, 8, 14]). Its proof is rejected in the Appendix.

THEOREM 1 (Extended Robbins–Monro theorem). *Let $H : \mathbb{R}^q \times \mathbb{R}^d \to \mathbb{R}^d$ be a Borel function and let X be 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

$$\forall \theta \in \mathbb{R}^d \setminus \mathcal{T}^*, \forall \theta^* \in \mathcal{T}^* \qquad \langle \theta - \theta^*, h(\theta) \rangle > 0. \tag{2.1}$$

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

$$\sum_{n\geq 1} \gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n^2 < +\infty. \tag{2.2}$$

Suppose that

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

[*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, and let θ_0 a random vector independent of $(X_n)_{n\geq 1}$ satisfying $\mathbb{E}[|\theta_0|^2] < +\infty$, both defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then, the recursive procedure defined by

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, X_{n+1}), \qquad n \geq 0, \tag{2.3}$$

satisfies:

$$\exists \theta_\infty : (\Omega, \mathcal{A}) \to \mathcal{T}^*, \qquad \theta_\infty \in L^2(\mathbb{P}) \qquad \text{such that } \theta_n \xrightarrow{\text{a.s.}} \theta_\infty. \tag{2.4}$$

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

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: we 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 notation of the above theorem

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

so that $\mathcal{T}^* = \operatorname{Arg\,min} V$ and (2.1) is satisfied (apply Lemma 1 to V).
– *Step* 3: Specify in an appropriate way the function ρ 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 which will not be 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 [3]. It is also a kind of introduction to the infinite-dimensional diffusion setting investigated in Section 3. In the Gaussian case, the usual 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

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

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

In [3], Arouna considers the function $\bar{H}_V(\theta, x)$ defined by

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

It is clear that the sub-linear growth condition (NEC) is not satisfied even if we simplify this function by $e^{|\theta|^2/2}$ (which would not affect the zero search problem for ∇V).

A first approach: when $F(X)$ have finite moments of any order, a naive way to control directly $\|\bar{H}_V(\theta_n, X_{n+1})\|_2$ 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

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

Setting $r = 1 + \frac{1}{\varepsilon}$ and $s = 1 + \varepsilon$, for some $\varepsilon > 0$, yields

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

Then, $\bar{H}_\varepsilon(\theta, x) := e^{-(3/2+\varepsilon)|\theta|^2}\bar{H}_V(\theta, x)$ satisfies the condition (NEC) and theoretically the resulting Robbins–Monro algorithm a.s. converges and no projection nor truncation is needed. However, numerically this solution is not satisfactory because the correcting factor $e^{-(3/2+\varepsilon)|\theta|^2}$ goes to zero much too fast as θ goes to infinity: if at some iterations at the beginning of the procedure θ_n is sent "too far," then it will freeze instantly. A more robust approach needs to be developed.

A general approach: we consider the density

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

By (1.4), we have

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

with $\widetilde{X}^{(\theta)} \sim p(x+\theta)\,dx$, that is, $\widetilde{X}^{(\theta)} = X - \theta$. 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 of $\nabla V(\theta)$ up to a $e^{|\theta|^2}$ weight since $\nabla V(\theta) = e^{|\theta|^2} \times \mathbb{E}[H_V(\theta, X)]$. Note that if F is bounded, then this function H_V satisfies the condition (NEC). Otherwise, we note that thanks to this new change of variable the parameter θ now lies *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

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

then we can consider the function $H_\lambda(\theta, x) = e^{-2\lambda|\theta|}H_V(\theta, x)$ which satisfies

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

The resulting Robbins–Monro algorithm reads

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

We no longer need to tune the correcting factor like in the former approach and one verifies on simulations that it does not suffer from freezing in general, mainly because our correcting term has an exponential growth (instead of $e^{C|\theta|^2}$). 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*. In this section, we consider a random variable X with distribution $p(x)\lambda_d(x)$ satisfying

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

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

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

Importance sampling by mean translation now reads

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

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

First, we rely on (1.2) to differentiate V in the proposition below.

PROPOSITION 2. *Suppose $(\mathcal{H}_1)$ and $(\mathcal{H}^{tr}_{a,\delta})$ are satisfied and that the function F satisfies*

$$\begin{aligned} &\forall \theta \in \mathbb{R}^d \qquad \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right] < +\infty \quad \textit{and} \\ &\forall C > 0 \qquad \mathbb{E}\left[F^2(X)e^{C|X|^{a-1}}\right] < +\infty. \end{aligned} \tag{2.5}$$

Then V is finite and differentiable on $\mathbb{R}^d$ and its gradient ∇V admits the following two representations:

$$\nabla V(\theta) = \mathbb{E}\left[F^2(X)\frac{p(X)}{p^2(X-\theta)}\nabla p(X-\theta)\right] \tag{2.6}$$

$$= \mathbb{E}\left[F^2(X-\theta)\frac{p^2(X-\theta)}{p(X)p(X-2\theta)}\frac{\nabla p(X-2\theta)}{p(X-2\theta)}\right]. \tag{2.7}$$

PROOF. The formal differentiation to get (2.6) from (1.2) is obvious. So it remains to check that the Lebesgue differentiation theorem applies. Let $x \in \mathbb{R}^d$ and $\theta \in \bar{B}(0,R)$ (closed ball of radius $R > 0$). The log-concavity of p implies that

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

so that

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

Using assumption ($\mathcal{H}^{tr}_{a,\delta}$) yields,

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

owing to (2.5). To derive the representation formula (2.7) for the gradient, we proceed as follows: an elementary change of variable in (2.6) shows that

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

REMARKS. The second representation of V in (1.2), which results from a second change of probability (back to the law of X), exhibits the specificity that the parameter θ does not appear in the argument of the possibly nonsmooth function F. It makes possible the differentiation of V (since p is smooth). Our second representation formula in (2.7) results from a *third* change of probability, this time in the representation of ∇ translation $x \mapsto x-\theta$, in order to plug back the parameter θ into the function F. The motivation is that in common applications F 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 three change of probability (the first in the representation of V, the last two in the representation of ∇V) can be summed up by the following scheme:

$$X \stackrel{V}{\rightsquigarrow} X+\theta \stackrel{\nabla V}{\rightsquigarrow} X \stackrel{\nabla V}{\rightsquigarrow} X-\theta.$$

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

LEMMA 2. *If* ($\mathcal{H}^{tr}_{a,\delta}$) *holds, then there exists two real constants* A, B *such that*

$$\frac{p^2(x-\theta)}{p(x)p(x-2\theta)}\frac{|\nabla p(x-2\theta)|}{p(x-2\theta)} \le e^{2\delta|\theta|^a}(A|x|^{a-1}+A|\theta|^{a-1}+B). \tag{2.8}$$

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

$$\log\biggl(\frac{p^2(x-\theta)}{p(x)p(x-2\theta)}\biggr) = 2f(x-\theta)-\bigl(f(x)+f(x-2\theta)\bigr) \\ + \delta(|x|^a+|x-2\theta|^a-2|x-\theta|^a).$$

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

$$\forall u,v\in\mathbb{R}^d \qquad |u|^a+|v|^a\le 2\biggl(\biggl|\frac{u+v}{2}\biggr|^a+\biggl|\frac{u-v}{2}\biggr|^a\biggr)$$

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

$$\frac{p^2(x-\theta)}{p(x)p(x-2\theta)}\le e^{2\delta|\theta|^a}. \tag{2.9}$$

One concludes by ($\mathcal{H}^{tr}_{a,\delta}$)(i). $\square$

REMARK. The normal distribution satisfies ($\mathcal{H}^{tr}_{a,\delta}$) 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

$$(\mathcal{H}^{tr}_c)\qquad \begin{cases} \text{(i)} \quad \forall x,y\in\mathbb{R}^d \qquad |F(x)|\le\widetilde{F}(x) \quad \text{and} \\ \qquad\qquad\qquad\qquad\quad \widetilde{F}(x+y)\le C\bigl(1+\widetilde{F}(x)\bigr)^c\bigl(1+\widetilde{F}(y)\bigr)^c, \\ \text{(ii)} \quad \mathbb{E}\bigl[|X|^{2(a-1)}\widetilde{F}(X)^{4c}\bigr]<+\infty. \end{cases}$$

REMARK. Assumption ($\mathcal{H}^{tr}_c$) seems almost nonparametric. However, its field of application is somewhat limited by ($\mathcal{H}^{tr}_{a,\delta}$) 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 $p(x)\le Ce^{-\eta|x|^a}(|x|+1)$ for some real constant $C>0$; which in turn implies that the function F in ($\mathcal{H}^{tr}_c$) needs to satisfy $F(x)\le\widetilde{F}(x):=C'e^{\lambda|x|^b}$ for some $b\in(0,a)$ and some $\lambda>0$. (Then $c=c_b$ with $c_b=1$ if $b\in[0,1]$ and $c_b=2^{b/2}$ if $b\in(1,a)$, when $a>1$.)

THEOREM 2. *Suppose X and F satisfy* ($\mathcal{H}_1$), ($\mathcal{H}_{a,\delta}^{tr}$), (2.5) *and* ($\mathcal{H}_c^{tr}$) *for some parameters $a \in (0,2]$, $\delta > 0$ and $c > 0$, and that the step sequence $(\gamma_n)_{n\geq 1}$ satisfies the usual decreasing step assumption*

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

Then the recursive procedure defined by

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, X_{n+1}), \qquad \theta_0 \in \mathbb{R}^d, \tag{2.10}$$

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

$$H(\theta, x) := \frac{F^2(x-\theta)}{1+\widetilde{F}(-\theta)^{2c}} e^{-2\delta|\theta|^a} \frac{p^2(x-\theta)}{p(x)p(x-2\theta)} \frac{\nabla p(x-2\theta)}{p(x-2\theta)}, \tag{2.11}$$

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

PROOF. In order to apply the Robbins–Monro theorem (Theorem 1), we have to check the following fact:

– *Mean reversion*: the mean function of the procedure defined by (2.10) 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

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

where we used assumption ($\mathcal{H}_c^{tr}$) in the first line and inequality (2.8) from Lemma 2 in the second line. One derives that there exists a real constant $C > 0$ such that

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

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

Examples of distributions.

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

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

so that $(\mathcal{H}_1)$ is satisfied as well as $(\mathcal{H}^{tr}_{a,\delta})$ for $a = 2$, $\delta = \frac{1}{2}$. If $\widetilde{F}(x) = C' e^{\lambda |x|^b}$, assumption $(\mathcal{H}^{tr}_c)$ is satisfied if $(b, \lambda) \in (0, 2) \times (0, \infty) \cup \{2\} \times (0, \frac{1}{2})$. Then, the function H has a particularly simple form

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

The hyper-exponential distributions.

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

where P is a nonnegative 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 $(\mathcal{H}^{tr}_{a,\delta})$ for $a = 1 + \eta$ $[\eta \in (0, 1)]$, $\delta > 0$. If $\widetilde{F}(x) = C' e^{\lambda |x|^b}$, assumption $(\mathcal{H}^{tr}_c)$ is satisfied if $(b, \lambda) \in (0, 1) \times (0, \infty) \cup \{1\} \times (0, 1)$.

2.4. *Exponential change of measure*: *The Esscher transform* (*or exponential tilting*). Another classical approach to design an importance sampling procedure is to consider a parametrized exponential change of measure (or Esscher transform). This transformation has been consider recently for that purpose in [13] to extend the procedure with repeated projections introduced in [3]. We denote by ψ the cumulant generating function (or log-Laplace) of X, that is, $\psi(\theta) = \log \mathbb{E}[e^{\langle \theta, X\rangle}]$. We assume that $\psi(\theta) < +\infty$ for every $\theta \in \mathbb{R}^d$ (which implies that ψ is an infinitely differentiable convex function) and we define

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

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

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, that is,

$$X^{(\theta)} = g(\theta, \xi), \qquad \xi : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathcal{X}, \tag{2.12}$$

where $\mathcal{X}$ is a Borel subset of a metric space and $g : \mathbb{R}^d \times \mathcal{X}$ is an explicit Borel function.

We assume that ψ satisfies

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

By (1.2), the potential V to be minimized is $V(\theta) = \mathbb{E}[F^2(X)e^{-\langle\theta,X\rangle+\psi(\theta)}]$.

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

$$\forall\theta \in \mathbb{R}^d \qquad \mathbb{E}\big[|X|F^2(X)e^{\langle\theta,X\rangle}\big] < +\infty. \tag{2.13}$$

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

$$\nabla V(\theta) = \mathbb{E}\big[(\nabla\psi(\theta) - X)F^2(X)e^{-\langle\theta,X\rangle+\psi(\theta)}\big] \tag{2.14}$$

$$= \mathbb{E}\big[(\nabla\psi(\theta) - X^{(-\theta)})F^2(X^{(-\theta)})\big]e^{\psi(\theta)-\psi(-\theta)}, \tag{2.15}$$

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

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 of V to get (2.14) is obvious and is made rigorous owning to assumption (2.13) on F. The second expression (2.15) of the gradient uses a third change of variable

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

THEOREM 3. *We assume that ψ satisfies $(\mathcal{H}_\delta^{es})$ and F satisfies (2.13) and*

$$\forall x \in \mathbb{R}^d \qquad |F(x)| \le Ce^{\lambda/4|x|} \quad \textit{and} \quad \mathbb{E}\big[|X|^2e^{\lambda|X|}\big] < +\infty.$$

Then, the recursive procedure

$$\begin{cases} X_{n+1}^{(\theta_n)} = g(\theta_n, \xi_{n+1}), \\ \theta_{n+1} = \theta_n - \gamma_{n+1}H(\theta_n, X_{n+1}^{(-\theta_n)}), \qquad n \ge 0, \theta_0 \in \mathbb{R}^d, \end{cases}$$

where $(\xi_n)_{n\ge1}$ is an i.i.d. sequence with the same distribution as ξ in (2.12), and

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

a.s. converges toward an Arg min V*-valued (square integrable) random vector θ^*.*

PROOF. We only need to check the sub-linear growth of the function $\theta \mapsto \|H(\theta, X^{(-\theta)})\|_2$ [condition (NEC)]. We have

$$
\begin{aligned}
&\mathbb{E}\big[\big|H\big(\theta, X^{(-\theta)}\big)\big|^2\big] \\
&\quad = e^{-\lambda\sqrt{d}|\nabla\psi(-\theta)|}\mathbb{E}\big[F^4\big(X^{(-\theta)}\big)\big|\nabla\psi(\theta) - X^{(-\theta)}\big|^2\big] \\
&\quad \le Ce^{-\lambda\sqrt{d}|\nabla\psi(-\theta)|}\mathbb{E}\big[e^{\lambda|X^{(-\theta)}|}\big|\nabla\psi(\theta) - X^{(-\theta)}\big|^2\big] \\
&\quad \le Ce^{-\lambda\sqrt{d}|\nabla\psi(-\theta)|}\big(|\nabla\psi(\theta)|^2\mathbb{E}\big[e^{\lambda|X^{(-\theta)}|}\big] + \mathbb{E}\big[\big|X^{(-\theta)}\big|^2 e^{\lambda|X^{(-\theta)}|}\big]\big).
\end{aligned}
\tag{2.16}
$$

First, by the following inequality

$$
\forall x \in \mathbb{R}^d \qquad e^{\lambda|x|} \le \prod_{j=1}^{d}(e^{\lambda x_j} + e^{-\lambda x_j}) = \sum_{J\subset\{1,\dots,d\}} e^{\lambda(\sum_{j\in J} x_j - \sum_{j\in J^c} x_j)}
$$

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

$$
\begin{aligned}
\mathbb{E}\big[e^{\lambda|X^{(-\theta)}|}\big] &\le \sum_{J\subset\{1,\dots,d\}} \mathbb{E}\big[e^{\lambda\langle e_J, X^{(-\theta)}\rangle}\big] \\
&= \sum_{J\subset\{1,\dots,d\}} e^{\psi(\lambda e_J - \theta) - \psi(-\theta)}.
\end{aligned}
$$

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

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

so that

$$
\begin{aligned}
\mathbb{E}\big[e^{\lambda|X^{(-\theta)}|}\big] &\le \sum_{J\subset\{1,\dots,d\}} e^{\lambda\langle\nabla\psi(-\theta), e_J\rangle + \delta\lambda^2|e_J|^2} \\
&\le C_{d,\lambda,\delta}e^{\lambda\sqrt{d}|\nabla\psi(-\theta)|}.
\end{aligned}
\tag{2.17}
$$

Likewise, one has

$$
\begin{aligned}
&\mathbb{E}\big[\big|X^{(-\theta)}\big|^2 e^{\lambda|X^{(-\theta)}|}\big] \\
&\qquad \le \sum_{J\subset\{1,\dots,d\}} \mathbb{E}\big[\big|X^{(\lambda e_J - \theta)}\big|^2\big]e^{\psi(\lambda e_J - \theta) - \psi(-\theta)} \\
&\qquad \le C_{d,\lambda,\delta}e^{\lambda\sqrt{d}|\nabla\psi(-\theta)|} \sum_{J\subset\{1,\dots,d\}} \mathbb{E}\big[\big|X^{(\lambda e_J - \theta)}\big|^2\big].
\end{aligned}
\tag{2.18}
$$

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

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

which implies

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

The assumption ($\mathcal{H}_\delta^{es}$) 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 $J\subset\{1,\dots,d\}$,

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

Plugging this into (2.18) and using (2.17) and (2.16) yields $\mathbb{E}[|H(\theta,X^{(-\theta)})|^2]\le C(1+|\theta|^2)$. □

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)

$$(E_{b,\sigma,W})\qquad dX_t=b(t,X^t)\,dt+\sigma(t,X^t)\,dW_t,\qquad X_0=x\in\mathbb{R}^d,$$

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 [19], pages 124–130.

EXAMPLES. 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 SDEs show (see [19]) that strong existence and uniqueness of solutions for ($E_{b,\sigma,W}$) follows from the following assumption

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

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

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

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

$$\mathbb{P}[F^2(X)>0]>0 \quad \text{and} \quad F(X)\in L^2(\mathbb{P}). \tag{3.1}$$

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 $\theta\in\mathcal{L}^2_{T,p}$ by

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

a bounded Borel function, and $\theta\in L^2_{T,p}$ (represented by a Borel function) for $p\ge 1$. In what follows, we use the following notation:

$$\varphi_t(\xi):=\varphi(t,\xi^t),\qquad \Theta_t:=\Theta(t,X^t),\qquad \Theta_t^{(\theta)}:=\Theta\bigl(t,X^{(\theta),t}\bigr)$$

and

$$\Theta_t^{(-\theta)}:=\Theta\bigl(t,X^{(-\theta),t}\bigr),$$

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 $(\mathcal{H}_{b,\sigma})$ holds. The SDE $(E_{b+\sigma\Theta,\sigma,W})$ satisfies the weak existence and uniqueness assumptions and for every nonnegative 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 notation,*

$$\begin{aligned}\mathbb{E}\Bigl[G\Bigl(X,\int_0^\cdot\langle f(s,X^s),dW_s\rangle\Bigr)\Bigr]&\\ =\mathbb{E}\Bigl[G\Bigl(X^{(\theta)},\int_0^\cdot\bigl\langle f\bigl(s,X^{(\theta),s}\bigr),dW_s\bigr\rangle+\int_0^\cdot\langle f,\Theta\rangle\bigl(s,X^{(\theta),s}\bigr)\,ds\Bigr)&\\ \times e^{-\int_0^T\langle\Theta_s^{(\theta)},dW_s\rangle-1/2\|\Theta^{(\theta)}\|^2_{L^2_{T,q}}}\Bigr]&\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\Bigl[G\Bigl(X^{(\theta)},\int_0^\cdot\bigl\langle f\bigl(s,X^{(\theta),s}\bigr),dW_s\bigr\rangle\Bigr)\Bigr]&\\ =\mathbb{E}\Bigl[G\Bigl(X,\int_0^\cdot\langle f(s,X^s),dW_s\rangle-\int_0^\cdot\langle f,\Theta\rangle(s,X^s)\,ds\Bigr)&\\ \times e^{\int_0^T\langle\Theta_s,dW_s\rangle-1/2\|\Theta\|^2_{L^2_{T,q}}}\Bigr].&\end{aligned}$$

PROOF. This is a straightforward application of Theorem 1.11, page 372 (and the remark that immediately follows) in [18] 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. $\square$

REMARKS. If $\varphi(\cdot, 0)$ is continuous, $\varphi(t, \cdot)$ is Lipschitz (uniformly in t), and θ is continuous, then $X^{(\pm\theta)}$ exists as unique strong solution of $(E_{b\pm\sigma\Theta,\sigma,W})$.

The Doléans exponential $(\exp\{\int_0^t \langle \Theta_s, dW_s\rangle - 1/2\|\Theta\|^2_{L^2_{T,q}}\})_{t\in[0,T]}$ is a true martingale for any $\theta \in L^2_{T,p}$.

In fact, still following the above cited remark form [18], the above lemma holds true if we replace Θ by any progressively measurable process $\widetilde{\Theta}$ such that $\mathbb{E}[e^{1/2\int_0^T |\widetilde{\Theta}(s,\omega)|^2\,ds}] < +\infty$.

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 $\theta \in L^2_{T,p}$

$$\mathbb{E}[F(X)] = \mathbb{E}[F(X^{(\theta)})e^{-\int_0^T \langle \Theta_s^{(\theta)}, dW_s\rangle - 1/2\|\Theta^{(\theta)}\|^2_{L^2_{T,q}}}],$$

[set $G(x, y) = F(x)$]. So, finding the best variance reducer amounts to solving the minimization problem

$$\min_{\theta\in L^2_{T,p}} V(\theta) \qquad \text{with } V(\theta) := \mathbb{E}\big[F^2(X^{(\theta)})e^{-2\int_0^T \langle \Theta_s^{(\theta)}, dW_s\rangle - \|\Theta^{(\theta)}\|^2_{L^2_{T,q}}}\big].$$

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

$$V(\theta) = \mathbb{E}\big[F^2(X)e^{-\int_0^T \langle \Theta_s, dW_s\rangle + 1/2\|\Theta\|^2_{L^2_{T,q}}}\big]. \tag{3.2}$$

PROPOSITION 4. *Suppose* $\mathbb{E}F(X)^{2+\eta} < +\infty$ *for some* $\eta > 0$. *Assume the assumptions* (3.1) *and* ($\mathcal{H}_{b,\sigma}$). *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 nonatomic kernel on the event* $\{F(X) > 0\}$, *that is,*

$$\mathbb{P}\big[\{\exists\theta \in L^2_{T,p} \setminus \{0\} \text{ s.t. } \theta(s) \in \operatorname{Ker}\varphi(s, X^s)\, ds\text{-a.e. and } F^2(X) > 0\}\big] = 0, \tag{3.3}$$

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

$$\inf_{\|\theta\|_{L^2_{T,p}}=1} \int_0^T \theta(s)^*\mathbb{E}[\varphi(s, X^s)^*\varphi(s, X^s)\mathbf{1}_{\{F^2(X)>0\}}]\theta(s)\, ds > 0, \tag{3.4}$$

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

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

$$
\begin{aligned}
&\langle \nabla V(\theta), \psi \rangle_{L^2_{T,p}} \\
&\quad = \mathbb{E}\Big[F^2(X) e^{-\int_0^T \langle \Theta_s, dW_s\rangle + 1/2\|\Theta\|^2_{L^2_{T,q}}} \\
&\qquad \times \Big(\langle \Theta, \varphi(\cdot, X^{\cdot})\psi \rangle_{L^2_{T,p}} - \int_0^T \langle \varphi(s, X^s)\psi_s, dW_s\rangle \Big)\Big] \qquad (3.5)\\
&\quad = \mathbb{E}\Big[F^2\big(X^{(-\theta)}\big) e^{\|\Theta^{(-\theta)}\|^2_{L^2_{T,p}}} \\
&\qquad \times \Big(2\big\langle \Theta^{(-\theta)}, \varphi\big(X^{(-\theta),\cdot}\big)\big\rangle_{L^2_{T,p}} - \int_0^T \big\langle \varphi\big(X^{(-\theta),s}\big)\psi_s, dW_s\big\rangle \Big)\Big].
\end{aligned}
$$

REMARKS. For practical implementation, the "finite-dimensional" statement is the only result of interest since it ensures that $\operatorname{Arg\,min}_{|E} \neq \varnothing$.

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

PROOF OF PROPOSITION 4. (a) As concerns the finiteness of the function V on the whole space $L^2_{T,q}$, we rely on equality (3.2). Set $r = 1 + 2/\eta$. Owing to the Hölder inequality, if follows from

$$
\mathbb{E}\big[e^{r/2\|\Theta\|^2_{L^2_{T,q}} - r\int_0^T \langle \Theta_s, dW_s\rangle}\big] \le e^{\|\varphi\|^2_\infty \|\theta\|_{L^2_{T,p}} r(r+1)/2} < +\infty.
$$

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

$$
e^{-\int_0^T \langle \Theta_s, dW_s\rangle + 1/2\|\Theta\|^2_{L^2_{T,q}}} = \big(e^{-1/2\int_0^T \langle \Theta_s, dW_s\rangle + 1/8\|\Theta\|^2_{L^2_{T,q}}}\big)^2 e^{1/4\|\Theta\|^2_{L^2_{T,q}}}
$$

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

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

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

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

Now

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

Assumption (3.3) implies that, for every $\theta \in L^2_{T,p}$,

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

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

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

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 $\mathbb{P}[F^2(X) \geq \varepsilon] > 0$. The second claim easily follows from assumption (3.4).

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

$$\begin{aligned}&\forall\theta,\psi \in L^2_{T,p}\\ &\qquad |\Phi(\theta+\psi) - \Phi(\theta) - \langle\nabla\Phi(\theta),\psi\rangle_{L^2_{T,p}}| \\ &\qquad\leq \|\varphi\|^2_\infty \|\theta\|_{L^2_{T,p}} \|\psi\|_{L^2_{T,p}},\end{aligned} \tag{3.6}$$

where $\psi \mapsto \langle\nabla\Phi(\theta),\psi\rangle_{L^2_{T,p}} = \int_0^T \langle\varphi_s(X^s)\theta(s), \varphi_s(X^s)\psi(s)\rangle\,ds - \int_0^T \langle\varphi_s(X^s) \times \psi(s), dW_s\rangle$ is clearly a bounded random functional from $L^2_{T,p}$ into $L^r(\mathbb{P})$, with an operator norm $|||\nabla\Phi(\theta)|||_{L^2_{T,p},L^r(\mathbb{P})} \leq \|\varphi\|^2_\infty\|\theta\|_{L^2_{T,p}} + c_p\|\varphi\|_\infty$ $[c_p \in (0,+\infty)]$ (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 $e^{\Phi(\theta)}\nabla\Phi(\theta)$. This follows from standard computation based on (3.6), the elementary inequality $|e^u - 1 - u| \le \frac{1}{2}u^2(e^u + e^{-u})$ and the fact that

$$\begin{aligned}\left\| \left| \int_0^T \langle \varphi_s(X^s)\psi(s), dW_s\rangle \right| e^{\int_0^T \langle \phi(X^s)\theta(s), dW_s\rangle} \right\|_r \\ \le \left\| \int_0^T \langle \varphi_s(X^s)\psi(s), dW_s\rangle \right\|_{2r} \| e^{\int_0^T \langle \varphi_s(X^s)\theta(s), dW_s\rangle} \|_{2p} \\ \le c_p \|\varphi\|_\infty \|\theta\|_{L^2_{T,p}} \|\psi\|_{L^2_{T,p}},\end{aligned}$$

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

One concludes that $\theta \mapsto V(\theta) = \mathbb{E}[F(X)^2 e^{\Phi(\theta)}]$ 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 led to consider some nontrivial finite-dimensional subspace 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 $\theta \in E$, $\nabla V_{|E}(\theta) = \mathrm{Proj}_E(\nabla V(\theta))$ where Proj_E denotes the orthogonal projection on E, and the quadratic function $L(\theta) := \|\theta - \theta^*_E\|_{L^2_{T,p}}$ is a Lyapunov function for the problem.

Like for the finite-dimensional framework investigated in Section 2.3, our algorithm will be based on the representation (3.5) for the gradient ∇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* $(\mathcal{H}_{b,\sigma})$ *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 strong solutions of* $(E_{b,\sigma,W})$ *and* $(E_{b+\sigma\Theta,\sigma,W})$ *driven by the same Brownian motion. Then, for every* $r \ge 1$, *there exists a real constant* $C_{b,\sigma} > 0$ *such that*

$$\left\| \sup_{t\in[0,T]} |X_t - X^{(\theta)}_t| \right\|_r \le C_{b,\sigma} e^{C_{b,\sigma}T} \left\| \int_0^T |\sigma(s, X^{(\theta),s})\Theta^{(\theta)}_s|\, ds \right\|_r. \tag{3.7}$$

PROOF. The proof follows closely the lines of the proof of the strong rate of convergence of the Euler scheme (see, e.g., [5]). □

The main result of this section is the following theorem.

THEOREM 4. *Suppose that assumptions* (3.1) *and* ($\mathcal{H}_{b,\sigma}$) *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*

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

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* $(e_1,\dots,e_m)$.

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 (2.2), $(W^{(n)})_{n\geq 1}$ *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* $\xi = (\xi_t)_{t\in[0,T]}$,

$$\begin{aligned}
&\langle H_{\lambda,\eta}(\theta,\xi,W), e_i\rangle_{L^2_{T,p}} \\
&\quad = \Psi_{\lambda,\eta}(\theta,\xi) F^2(\xi) e^{\|\Theta(\cdot,\xi^\cdot)\|_{L^2_{T,q}}} \\
&\qquad \times \biggl(2\langle \Theta(\cdot,\xi^\cdot), \varphi(\cdot,\xi^\cdot)e_i\rangle_{L^2_{T,q}} - \int_0^T \langle \varphi(s,\xi^s)e_i(s), dW_s\rangle\biggr),
\end{aligned}$$

where, for any $\eta > 0$,

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

Then, the recursive sequence $(\theta_n)_{n\geq1}$ a.s. converges toward an Arg min *V-valued* (*square 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 uniformly in $t \in [0,T]$ and if θ and $\varphi(\cdot,0)$ are continuous; 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. This follows from general functional limit theorems like those established in [12].

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

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

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

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

$$\begin{aligned}\langle h(\theta), \psi\rangle_{L^2_{T,p}} &= \mathbb{E}\big[\langle H_{\lambda,\eta}(\theta, X^{(-\theta)}, W), \psi\rangle_{L^2_{T,p}}\big]\\ &= \mathbb{E}\biggl[\frac{e^{-\|\varphi\|_\infty\|\theta\|_{L^2_{T,p}}}}{1+\|\Theta^{(-\theta)}\|^{2\lambda+\eta}_{L^2_{T,q}}}\langle \nabla V_{|E}(\theta), \psi\rangle_{L^2_{T,p}}\biggr],\end{aligned}$$

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

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

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

$$\begin{aligned}&\biggl\|F(X^{(-\theta)})^2\int_0^T\langle\varphi_s(X^{(-\theta),s})e_i(s), dW_s\rangle\biggr\|_2\\ &\quad\le \|F(X^{(-\theta)})^2\|_{2+\eta'}\biggl\|\int_0^T\langle\varphi_s(X^{(-\theta),s})e_i(s), dW_s\rangle\biggr\|_{2(1+1/\eta')}\\ &\quad\le \|F(X^{(-\theta)})^2\|_{2+\eta'}\biggl\|\int_0^T|\varphi_s(X^{(-\theta),s})e_i(s)|^2ds\biggr\|_{1+1/\eta'}\\ &\quad\le \|F(X^{(-\theta)})^2\|_{2+\eta'}\|\varphi\|_\infty.\end{aligned}$$

Now

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

One shows likewise that

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

Combining these 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 $(t, x) \in [0, T] \times \mathcal{C}([0, T], \mathbb{R}^d)$,

$$\|\sigma(t, x)\| \le C_\sigma(1 + \|x\|_\infty).$$

Elementary computations based on (3.7) and Lemma 3 yield

$$\begin{aligned}
&\left\| \int_0^T |\sigma(s, X^{(\theta),s})\Theta_s^{(\theta)}| \right\|_r \\
&\qquad \le C_\sigma \|\theta\|_{L^1_{T,p}} \|\varphi\|_\infty \Bigl(1 + \Bigl\| \|X\|_\infty e^{-r/2\|\Theta/r\|^2_{L^2_{T,q}} + \int_0^T \langle \Theta_s/r, dW_s\rangle} \Bigr\|_r\Bigr) \\
&\qquad \le C_\sigma \|\theta\|_{L^1_{T,p}} \|\varphi\|_\infty \bigl(1 + e^{\|\varphi\|_\infty/(2rr')\|\theta\|^2_{L^2_{T,p}}} \| \|X\|_\infty \|_{r(1+r')}\bigr) \\
&\qquad \le C_{r,b,\sigma,\varphi}\bigl(1 + \|\theta\|_{L^2_{T,p}} e^{\|\varphi\|_\infty/(2rr')\|\theta\|^2_{L^2_{T,p}}}\bigr)
\end{aligned}$$

for every $r, r' > 0$ [assumption ($\mathcal{H}_{b,\sigma}$) implies that $\| \|X\|_\infty \|_r < +\infty$ for every $r > 0$]. Then, following the lines of the bounded case, we obtain easily the result with $\Psi_{\lambda,\eta}(\theta, \xi) = e^{-(\|\varphi\|_\infty+\eta)\|\theta\|_{L^2_{T,p}}}$. $\square$

REMARK. If the functional F is bounded ($\lambda = 0$), one can prove likewise that the algorithm without correction (i.e., $\Psi_{\lambda,\eta} = 1$) 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 [3]), 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. To be precise, one estimates $\mathbb{E}[F(X)]$ 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

$$\sqrt{N}\bigl(S_N - \mathbb{E}[F(X)]\bigr) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_*^2) \qquad \text{with } \sigma_*^2 = V(\theta^*) - \mathbb{E}[F(X)]^2.$$

This approach can be extended to the Esscher transform when we use the same innovation ξ_k [see (2.12)] 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 regular stochastic algorithm behaves as described in usual stochastic approximation theory textbooks like [4, 8, 14]. So, as soon as the set of optimal variance reducers 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 $g(\theta) := e^{-\lambda/2|\theta|^b}$ 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 $X \stackrel{d}{=} \mathcal{N}(0; 1)$ for the sake of simplicity. One can write

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

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

$$V''(\theta) = e^{|\theta|^2/2}\mathbb{E}\big[F^2(X)e^{-\theta X}\big(1 + (\theta - X)^2\big)\big].$$

Consequently, the mean functions h and h_0 related to H and H_0 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 for recursive stochastic algorithms say that if $\gamma_n = \frac{\alpha}{\beta+n}$, $\alpha, \beta > 0$ with

$$\alpha > \frac{1}{2h'(\theta^*)} = \frac{1}{2g(\theta^*)h_0'(\theta^*)}, \tag{4.1}$$

then, the Robbins–Monro algorithn related to the function H satisfies the CLT

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

where

$$\Sigma^*_\alpha = \mathrm{Var}(H(y^*, Z))\frac{\alpha^2}{2\alpha h'(y^*) - 1}. \tag{4.2}$$

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

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

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 (4.1), 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 nonexplosion of the procedure may impact the rate of convergence through 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 this specific difficulty in our numerical experiments reported below.

One classical way to obtain the optimal rate of convergence is to introduce the empirical mean of the algorithm implemented with a slowly decreasing step "à la Rupert and Poliak" (see, e.g., [17]): set $\gamma_n = \frac{c}{n^r}$, $\frac{1}{2} < r < 1$ and

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

where $(\theta_n)_{n\geq 0}$ denotes the regular Robbins–Monro algorithm defined by (2.10) starting at θ_0. Then $(\bar{\theta}_n)_{n\geq 0}$ converges toward θ^* and satisfies a CLT with the optimal asymptotic variance (4.2). See also a variant based on a gliding window developed in [15].

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

$$\theta = T(\widetilde{\theta}), \qquad \theta \in \Theta,$$

where $T:\mathbb{R}^q \to \Theta$ is a $\mathcal{C}^1$-diffeomorphism with a bounded differential (i.e., $\sup_{\widetilde{\theta}} |||DT(\widetilde{\theta})||| < +\infty$). 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 $\widetilde{H}(\widetilde{\theta}, X) := DT(\widetilde{\theta})H(T(\widetilde{\theta}), X)$ and defines recursively a procedure on $\mathbb{R}^q$ by

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

In order to establish the a.s. convergence of $\theta_n := T(\widetilde{\theta}_n)$ to $\operatorname{Arg\,min} V$, one relies on a variant of Robbins–Monro algorithm, namely a stochastic gradient approach (see [8, 14] for further details): one defines $U(\widetilde{\theta}) = V(T(\widetilde{\theta}))$ which turns out to be a Lyapunov function for the new algorithm since

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

If U satisfies $\|\widetilde{H}(\widetilde{\theta}, X)\|_2 + |\nabla U(\widetilde{\theta})| \le C(1 + U(\widetilde{\theta}))^{1/2}$ (which is a hidden constraint on the choice of T), one shows under the standard "decreasing" assumption on the step sequence that $U(\widetilde{\theta}_n) \to U_\infty \in L^1(\mathbb{P})$ and $\sum_n \gamma_{n+1}\rho(\widetilde{\theta}_n)|\nabla U(\widetilde{\theta}_n)|^2 < +\infty$. 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 approaches introduced in Section 2 to proceed recursive importance sampling. The quantity to compute is

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

where $p_{\mathrm{NIG}}(x; \alpha, \beta, \delta, \mu)$ is the normal inverse Gaussian (NIG) density of the random variable X with parameters $(\alpha, \beta, \delta, \mu)$ satisfying $\alpha > 0$, $|\beta| \le \alpha$, $\delta > 0$, $\mu \in \mathbb{R}$. Namely,

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

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

We can summarize the two variance reducer procedures presented in Section 2, the one based on translation of the mean (see Section 2.3) and the one based on the Esscher transform (see Section 2.4), by the following simplified pseudo-code:

Translation (see Section 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 Section 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' can easily be obtained by using the following classical relation satisfied by the modified Bessel function: $K_1'(x) = \frac{1}{x} K_1(x) - K_2(x)$ (see [1]).

Assumption $(\mathcal{H}^{tr}_{a,\delta})$ is satisfied with $a = 1$, and our results of Section 2.3 apply.

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

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

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

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

for every $\theta \in (-\alpha - \beta, \alpha - \beta)$. Moreover, we need $\psi(-\theta)$ to be well defined which implies $\theta \in (-\alpha + \beta, \alpha + \beta)$. To take into account these restrictions, we slightly modify the algorithm parametrization (see Section 4.3) $\theta = T(\widetilde{\theta}) := (\beta - \alpha) \frac{\widetilde{\theta}}{\sqrt{1 + \widetilde{\theta}^2}}$. We update $\widetilde{\theta} \in \mathbb{R}$ in the Robbins–Monro procedure by $\widetilde{\theta}_{n+1} = \widetilde{\theta}_n - \gamma_{n+1} \widetilde{H}_2(\widetilde{\theta}_n, X_{n+1})$ where $\widetilde{H}_2(\widetilde{\theta}, X) = T'(\widetilde{\theta}) H_2(T(\widetilde{\theta}), X)$ and $T'(\widetilde{\theta}) = \frac{\beta - \alpha}{(1 + \widetilde{\theta}^2)^{3/2}}$.

TABLE 1
Variance reduction for different strikes (one-dimensional NIG example)

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)

The payoff F is a call option with 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 Table 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*.

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 an exchange option between gas and electricity (called spark spread). We choose to model the price of the energy by the

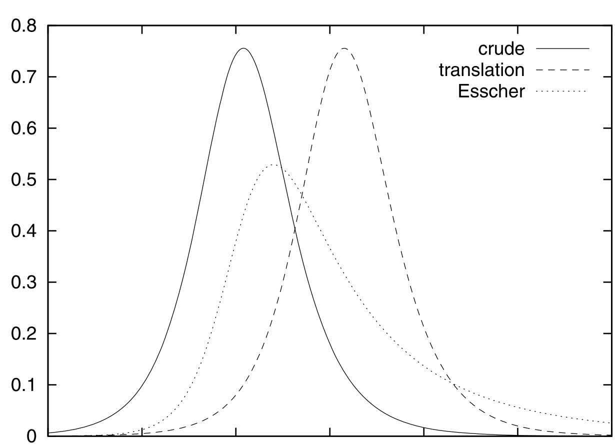


FIG. 1. *Densities of X (crude), $X + \theta$ (translation) and $X^{(\theta)}$ (Esscher) in the case $K = 1$.*

TABLE 2
Variance reduction for different strikes (spark spread example)

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	7582.0	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	4.8440	31.018
	0.6	17.955	7612.5	4.6031	34.166
	0.8	15.156	7357.3	4.4160	38.167
	1	13.027	7123.9	4.2685	42.781

exponential of a NIG distribution. A (simplified) form of the payoff is then

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

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

5.2. *Functional setting*: *Down & In Call option*. We consider a process $(X_t)_{t\geq 0}$ solution to the following diffusion equation

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad X_0 = x_0 \in \mathbb{R}.$$

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

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

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 $\min_{0\leq t\leq T} X_t$ 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 improvement is to consider the continuous Euler scheme $\bar{X}^c$ obtained by the extrapolation of the Brownian between two instants of discretization. To be precise, for every $t \in [t_k, t_{k+1}]$,

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

By preconditioning,

$$\mathbb{E}[F(X)] = \mathbb{E}\left[(\bar{X}_T - K)_+\left(1 - \prod_{k=0}^{N-1} p(\bar{X}_{t_k}, \bar{X}_{t_{k+1}})\right)\right] \tag{5.1}$$

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

$$\begin{aligned} p(x_k, x_{k+1}) &= 1 - \mathbb{P}\left[\min_{t\in[0,t_1]} W_t \leq \frac{L - x_k}{\sigma(x_k)} \middle| W_{t_1} = \frac{x_{k+1} - x_k}{\sigma(x_k)}\right] \\ &= \begin{cases} 0, & \text{if } L \geq \min(x_k, x_{k+1}), \\ 1 - e^{-2(L-x_k)(L-x_{k+1})/(\sigma^2(x_k)(t_{k+1}-t_k))}, & \\ & \text{otherwise.} \end{cases} \end{aligned} \tag{5.2}$$

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

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 $\widetilde{P}_n(t)$ defined by

$$\forall n \geq 0, \forall t \in [0, 1] \qquad \widetilde{P}_n(t) = P_n(2t - 1) \quad \text{where } P_n(t) = \frac{1}{2^n n!}\frac{d^n}{dt^n}\big((t^2 - 1)^n\big). \tag{ShLeg}$$

– the Karhunen–Loève basis defined by

$$\forall n \geq 0, \forall t \in [0, 1] \qquad e_n(t) = \sqrt{2}\sin\big((n + \tfrac{1}{2})\pi t\big). \tag{KL}$$

– the Haar basis defined by

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

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 $[b(x) = rx, \sigma(x) = \sigma x]$. We set the interest rate r to 4% and the volatility σ to

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

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

70% (which is a high volatility). The strike of the payoff F is set at $K = 115$ and the barrier level at $L = 65$. The true price $e^{-rT}\mathbb{E}[F(X)]$ of this product is 2.554. A crude Monte Carlo [with Brownian bridge interpolation, see (5.1)] gives a price of 2.596 with a variance of 230 after 500,000 trials.

The results of our algorithm, for different basis, are summarized in Table 3. 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.

Figure 2 is depicted as 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.

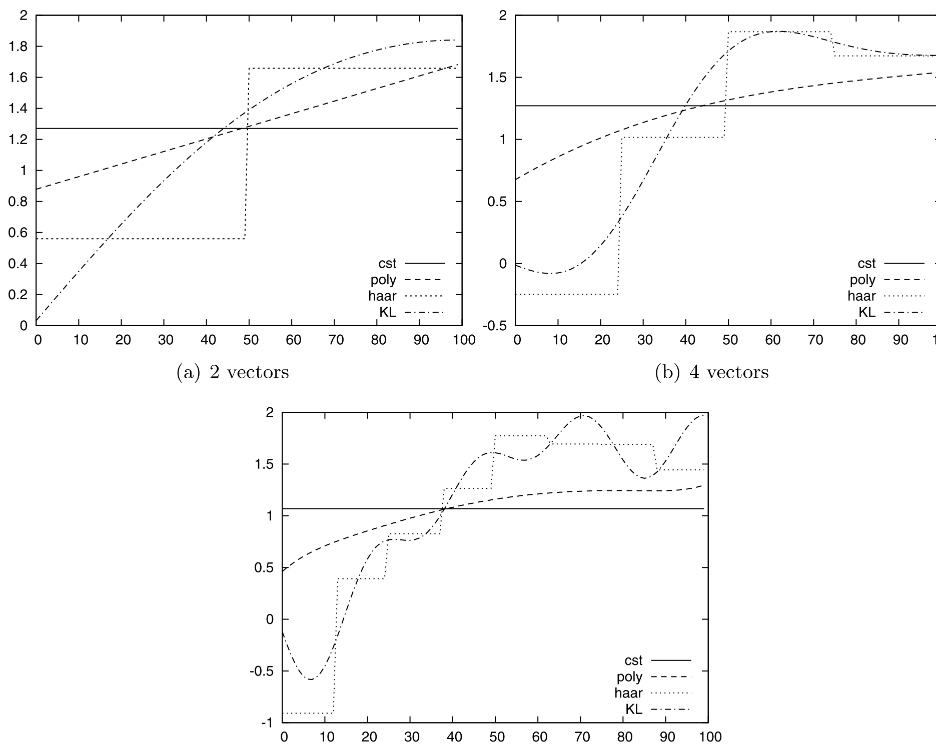
A local volatility model. To emphasize the generic feature of our algorithm, we consider the same product in a local volatility model (inspired by the CEV model) defined by

$$dX_t = rX_t\,dt + \sigma X_t^{\beta}\frac{X_t}{\sqrt{1+X_t^2}}\,dW_t \tag{5.6}$$

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. The results of our algorithm for different basis are summarized in Table 4.

Adaptive case (*nontrivial driver*). We experiment now our algorithm with a nontrivial driver φ. To be efficient, this driver must be specified according to the problem. In the case of the simple barrier option, a natural choice is to use a translation coefficient α before the underlying crosses the barrier and a different coeffi-



(a) 2 vectors (b) 4 vectors

(c) 8 vectors

FIG. 2. *Optimal θ process obtained with different basis by our algorithm using* 50,000 *trials*.

TABLE 4
Variance ratio obtained for different basis in the local volatility model (5.6)
($K = 115$, $L = 65$, *variance of the crude Monte Carlo*: 206.52)

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 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	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

cient β after. The driver is then defined for $t = t_k$ by

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

where p is defined by (5.2). Note that $\bar{p}_k = \mathbb{P}[\min_{t\in[0,t_k]} \xi_t \geq L | \xi_0, \dots, \xi_{t_k}]$ 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.

APPENDIX: PROOF OF THEOREM 1

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

PROOF OF THEOREM 1. Set $\mathcal{F}_n := \sigma(\theta_0, X_1, \dots, X_n)$, $n \geq 1$. Let $\theta^* \in \mathcal{T}^*$. Then

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

TABLE 6

Variance reduction for different strikes and barrier levels in the local volatility 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

where

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

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

$$\begin{aligned}&\mathbb{E}[|\langle \theta_n - \theta^*, H(\theta_n, X_{n+1})\rangle|] \\ &\qquad \le \tfrac{1}{2}\big(\mathbb{E}[|\theta_n - \theta^*|^2] + \mathbb{E}[|H(\theta_n, X_{n+1})|^2]\big) \\ &\qquad \le C\big(1 + \mathbb{E}[|\theta_n - \theta^*|^2]\big)\end{aligned}$$

for an appropriate real constant C (which depends on θ^*). Then, one shows by induction on n from (A.1) that $|\theta_n|$ is square integrable for every $n \ge 0$ and that ΔM_{n+1} is integrable, hence a true martingale increment. Now, one derives from assumptions (2.2) and (A.1) that

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

is a (nonnegative) super-martingale with $S_0 = |\theta_0 - \theta^*|^2 \in L^1(\mathbb{P})$. This uses the mean-reverting assumption (2.1). 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

$$
\begin{aligned}
&|\theta_n-\theta^*|^2+2\sum_{k=0}^{n-1}\gamma_{k+1}\langle\theta_k-\theta^*,h(\theta_k)\rangle\\
&\qquad\xrightarrow{\text{a.s.}}\widetilde{S}_\infty=S_\infty\prod_{n\geq 1}(1+C\gamma_n^2)\in L^1(\mathbb{P}).
\end{aligned}
\tag{A.2}
$$

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

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

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

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

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

Let $A>0$. Set

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

It follows from the a.s. finiteness of L_∞ that $\bigcup_{A>0}\Omega_A=\Omega$ a.s. Now, we consider the compact set $K_A=\mathcal{T}^*\cap\bar{B}(0,A)$. It is separable, hence there exists an everywhere dense sequence in K_A, 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

$$
\begin{aligned}
\Omega'_A:=\Bigg\{&\omega\in\Omega_A,|\theta_n(\omega)-\theta^{*,k}|^2\xrightarrow{n\to\infty}L_\infty^k(\omega),k\geq 1,\\
&\sum_{n\geq 1}\gamma_n\langle\theta_{n-1}(\omega)-\theta^*,h(\theta_{n-1}(\omega))\rangle<+\infty\Bigg\},
\end{aligned}
$$

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

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

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, $\theta_\infty(\omega)\in K_A$. Then any limiting value $\theta'_\infty(\omega)$ of the sequence $(\theta_n(\omega))_{n\geq 1}$ will satisfy

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

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 $(\theta_n(\omega))_{n\geq 0}$, that is, $\theta_n(\omega) \to \theta_\infty(\omega)$ 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 $(\theta_n - \theta^*)_{n\geq 1}$. □

REFERENCES

[1] ABRAMOWITZ, M. and STEGUN, I. A., EDS. (1992). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York. MR1225604

[2] ANDRIEU, C., MOULINES, É. and PRIOURET, P. (2005). Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim*. **44** 283–312. MR2177157

[3] AROUNA, B. (2004). Adaptative Monte Carlo method, a variance reduction technique. *Monte Carlo Methods Appl*. **10** 1–24. MR2054568

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

[5] BOULEAU, N. and LÉPINGLE, D. (1994). *Numerical Methods for Stochastic Processes*. Wiley, New York. MR1274043

[6] CHEN, H. F., LEI, G. and GAO, A. J. (1988). Convergence and robustness of the Robbins–Monro algorithm truncated at randomly varying bounds. *Stochastic Process*. *Appl*. **27** 217–231. MR931029

[7] CHEN, H. F. and ZHU, Y. M. (1986). Stochastic approximation procedures with randomly varying truncations. *Sci*. *Sinica Ser. A* **29** 914–926. MR869196

[8] DUFLO, M. (1997). *Random Iterative Models*. *Applications of Mathematics* (*New York*) **34**. Springer, Berlin. MR1485774

[9] GLASSERMAN, P. (2004). *Monte Carlo Methods in Financial Engineering*: *Stochastic Modelling and Applied Probability*. *Applications of Mathematics* (*New York*) **53**. Springer, New York. MR1999614

[10] GLASSERMAN, P., HEIDELBERGER, P. and SHAHABUDDIN, P. (1999). Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance* **9** 117–152. MR1849001

[11] GOBET, E. (2000). Weak approximation of killed diffusion using Euler schemes. *Stochastic Process*. *Appl*. **87** 167–197. MR1757112

[12] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Springer, Berlin. MR1943877

[13] KAWAI, R. (2008/09). Optimal importance sampling parameter search for Lévy processes via stochastic approximation. *SIAM J. Numer. Anal*. **47** 293–307. MR2475940

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

[15] LELONG, J. (2007). Etude Asymptotique des Algorithmes Stochastiques et Calcul des Prix des Options Parisiennes. Ph.D. thesis, Ecole nationale des ponts et chaussées, ENPC PARIS/ MARNE LA VALLEE.

[16] LEMAIRE, V. (2007). An adaptive scheme for the approximation of dissipative systems. *Stochastic Process*. *Appl*. **117** 1491–1518. MR2353037

[17] PELLETIER, M. (2000). Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim*. **39** 49–72. MR1780908

[18] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1725357

[19] ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions*, *Markov Processes*, *and Martingales*. *Cambridge Mathematical Library* **2**. Cambridge Univ. Press, Cambridge. MR1780932

UNIVERSITÉ PIERRE ET MARIE CURIE
LABORATOIRE DE PROBABILITÉS
ET MODÈLES ALÉATOIRES
UMR 7599
CASE 188, 4, PL. JUSSIEU
F-75252 PARIS CEDEX 5
FRANCE
E-MAIL: vincent.lemaire@upmc.fr
gilles.pages@upmc.fr