HAL
open science

Asymptotic normality of randomly truncated stochastic algorithms

Jérôme Lelong

To cite this version:

Jérôme Lelong. Asymptotic normality of randomly truncated stochastic algorithms. ESAIM: Probability and Statistics, 2013, 17, pp.105-119. 10.1051/ps/2011110 . hal-00464380v2

HAL Id: hal-00464380

https://hal.science/hal-00464380v2

Submitted on 20 May 2011

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

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

Asymptotic normality of randomly truncated stochastic algorithms

Jérôme Lelong*

May 20, 2011

Abstract

We study the convergence rate of randomly truncated stochastic algorithms, which consist in the truncation of the standard Robbins-Monro procedure on an increasing sequence of compact sets. Such a truncation is often required in practice to ensure convergence when standard algorithms fail because the expected-value function grows too fast. In this work, we give a self contained proof of a central limit theorem for this algorithm under local assumptions on the expected-value function, which are fairly easy to check in practice.

Key words: stochastic approximation, central limit theorem, randomly truncated stochastic algorithms, martingale arrays.

1 Introduction

The use of stochastic algorithms is widespread for solving stochastic optimization problems. These algorithms are extremely valuable for a practical use and particularly well suited to localize the zero of a function u. Such algorithms go back to the pioneering work of Robbins and Monro [1951], who considered the sequence

$$X_{n+1} = X_n - \gamma_{n+1} u(X_n) - \gamma_{n+1} \delta M_{n+1} \tag{1}$$

to estimate the zero of the function u. The sequence $(\gamma_n)_n$ classically denotes the gain or step sequence of the algorithm and $(\delta M_n)_n$ depicts a random measurement error. Nevertheless, the assumptions required to ensure the convergence — basically, a sub-linear growth of u on average — are barely satisfied in practice, which dramatically reduces the range of applications. Chen and Zhu [1986] proposed a modified algorithm to deal with fast growing functions. Their new algorithm can be summed up as

$$X_{n+1} = X_n - \gamma_{n+1} u(X_n) - \gamma_{n+1} \delta M_{n+1} + \gamma_{n+1} p_{n+1} \tag{2}$$

where $(p_n)_n$ is a truncation term ensuring that the sequence $(X_n)_n$ cannot jump too far ahead in one step.

In this paper, we are concerned with the rate of convergence of Equation (2). Numerous results are known for the sequence defined by Equation (1), which is known to converge at

*Laboratoire Jean Kuntzmann, Université de Grenoble et CNRS, BP 53, 38041 Grenoble Cédex 9, FRANCE, e-mail : jerome.lelong@imag.fr

the rate $\sqrt{\gamma_n}$ when γ_n is of the form $\frac{\gamma}{n^\alpha}$ with $1/2 < \alpha \leq 1$ (see Delyon [1996], Duflo [1997] or Buche and Kushner [2001] for instance). Let $x^\star$ denote the unique root of u, $u(x^\star) = 0$. When $\gamma_n = \frac{\gamma}{n}$ and $\nabla u(x)$ is of the form λI, Duflo [1997] showed that the convergence rate depends on the relative position of λ and $\frac{\gamma}{2}$. A functional central limit theorem for this algorithm was proved by Bouton [1985] and Benveniste et al. [1990]. The convergence rate of constrained algorithms was studied by Kushner and Yin [2003]. The problem of multiple targets was tackled by Pelletier [1998] who proved a Central Limit Theorem. However, very few results are known about the convergence rate of the algorithm devised by Chen and Zhu [1986]. Chen [2002] briefly studied the convergence rate under global hypotheses on the noise sequence $(\delta M_n)_n$. Here, we aim at giving a clarified, self-contained and elementary proof of this result under local assumptions (see Section 2.3 for a detailed comparison of the two results). Besides giving a clarified and self-contained proof of the central limit theorem for randomly truncated algorithm, the improvement brought by our work is the use of the local condition $\sup_n \mathbb{E}[|\delta M_n|^{2+\rho}\, \mathbf{1}_{\{|X_{n-1}-x^\star|\leq\eta\}}] < \infty$ with some $\rho > 0$ and $\eta > 0$ replacing the global condition $\sup_n \mathbb{E}[|\delta M_n|^{2+\rho}] < \infty$.

First, we define the general framework and explain the algorithm developed by Chen and Zhu [1986]. Our main results are stated in Theorems 2.1 and 2.2 (see page 4) depending on the decreasing speed of the sequence $(\gamma_n)_n$. In Sections 2.3 and 2.4, we discuss the improvements brought by our new results and we give a concrete example to show the benefits of using local assumptions. Section 3 is devoted to the proof of the main results.

Notations

- If z is a complex number, we denote by $\mathrm{Re}(z)$ (resp. $\mathrm{Im}(z)$) its real (resp. imaginary) part.
- The prime notation denotes the transpose operator, eg. A' stands for the transpose of A.
- For any vector x, $|x|$ denotes its Eucildean norm and $\cdot$ the associated scalar product.
- For any square matrix A, $||A||$ denotes the matrix norm associated to $|\cdot|$, ie. $||A|| = \sup_{|x|\leq 1} |Ax|$. For any matrix A, $||A||$ is equal to the square root of the largest eigenvalue of $A'A$.
- If A is a square matrix, we denote by $\mathrm{Sp}(A)$ the eigenvalues of A.
- We say that a square matrix A is repulsive if all the eigenvalues of A have positive real parts.

2 A CLT for randomly truncated stochastic algorithms

It is quite common to look for the root of a continuous function $u\colon x \in \mathbb{R}^d \longmapsto u(x) \in \mathbb{R}^d$, which is not easily tractable. We assume that we can only access u up to a measurement error embodied in the following by the sequence $(\delta M_n)_n$ and that the norm $|u(x)|^2$ grows faster than $|x|^2$ such that the standard Robbins-Monro algorithm (see Equation (1)) quickly fails. Instead, we consider the alternative procedure introduced by Chen and Zhu [1986]. This

technique consists in forcing the algorithm to remain in an increasing sequence of compact sets $(K_j)_j$ such that

$$\bigcup_{j=0}^{\infty} K_j = \mathbb{R}^d \quad \text{and} \quad \forall j,\ K_j \subsetneq \operatorname{int}(K_{j+1}).$$

It prevents the algorithm from blowing up during the first iterates. Let $(\gamma_n)_n$ be a decreasing sequence of positive real numbers satisfying $\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$. For $X_0 \in \mathbb{R}^d$ and $\sigma_0 = 0$, we define the sequences of random variables $(X_n)_n$ and $(\sigma_n)_n$ by

$$\begin{cases} & X_{n+\frac{1}{2}} = X_n - \gamma_{n+1} u(X_n) - \gamma_{n+1}\delta M_{n+1}, \\ \text{if } X_{n+\frac{1}{2}} \in K_{\sigma_n} & X_{n+1} = X_{n+\frac{1}{2}} \quad \text{and} \quad \sigma_{n+1} = \sigma_n, \\ \text{if } X_{n+\frac{1}{2}} \notin K_{\sigma_n} & X_{n+1} = X_0 \quad \text{and} \quad \sigma_{n+1} = \sigma_n + 1. \end{cases} \tag{3}$$

Let $\mathcal{F}_n$ denote the $\sigma-$algebra generated by $(\delta M_k, k \leq n)$, $\mathcal{F}_n = \sigma(\delta M_k, k \leq n)$. We assume that $(\delta M_n)_n$ is a sequence of martingale increments, i.e. $\mathbb{E}(\delta M_{n+1}|\mathcal{F}_n) = 0$.

Remark 1. $X_{n+\frac{1}{2}}$ is actually drawn from the dynamics of the Robbins-Monro algorithm (see Equation (1)). If the standard algorithm wants to jump too far ahead it is reset to a fixed value. When $X_{n+\frac{1}{2}} \notin K_{\sigma_n}$, one can set X_{n+1} to any measurable function of $(X_0, \ldots, X_n)$ with values in a given compact set. The existence of such a compact set is crucial to prove the a.s. convergence of $(X_n)_n$.

It is more convenient to rewrite Equation (3) as follows

$$X_{n+1} = X_n - \gamma_{n+1} u(X_n) - \gamma_{n+1}\delta M_{n+1} + \gamma_{n+1} p_{n+1} \tag{4}$$

where

$$p_{n+1} = \left(u(X_n) + \delta M_{n+1} + \frac{1}{\gamma_{n+1}}(X_0 - X_n) \right) \mathbf{1}_{\left\{ X_{n+\frac{1}{2}} \notin K_{\sigma_n} \right\}}.$$

In this paper, we only consider gain sequences of the type $\gamma_n = \frac{\gamma}{(n+1)^\alpha}$, with $1/2 < \alpha \leq 1$. If $\alpha = 1$, we obtain a slightly different limit. For values of α outside this range, the almost sure convergence is not even guarantied.

2.1 Hypotheses

In the following, the prime notation stands for the transpose operator. We introduce the following hypotheses.

(A1) *i.* $\exists x^\star \in \mathbb{R}^d$ s.t. $u(x^\star) = 0$ and $\forall x \in \mathbb{R}^d$, $x \neq x^\star$, $(x - x^\star) \cdot u(x) > 0$.

ii. There exist a function $y : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ satisfying $\lim_{|x|\to 0} \|y(x)\| = 0$ and a repulsive matrix A such that

$$u(x) = A(x - x^\star) + y(x - x^\star)(x - x^\star).$$

(A2) For any $q > 0$, the series $\sum_n \gamma_{n+1}\delta M_{n+1}\mathbf{1}_{\{|X_n - x^\star| \leq q\}}$ converges almost surely.

(A3) *i.* There exist two real numbers $\rho > 0$ and $\eta > 0$ such that

$$\kappa = \sup_n \mathbb{E}\left(|\delta M_n|^{2+\rho} \mathbf{1}_{\{|X_{n-1} - x^\star| \leq \eta\}} \right) < \infty.$$

ii. There exists a symmetric positive definite matrix Σ such that

$$\mathbb{E}\left(\delta M_n \delta M_n' \big| \mathcal{F}_{n-1}\right) \mathbf{1}_{\{|X_{n-1}-x^\star|\leq\eta\}} \xrightarrow[n\to\infty]{\mathbb{P}} \Sigma.$$

(A4) There exists $\mu > 0$ such that $\forall n \geq 0,\ d(x^\star, \partial K_n) \geq \mu$.

Remark 2. *Comments on the assumptions.*

1. Hypothesis (**A1-i**) is satisfied as soon as u can be interpreted as the gradient of a strictly convex function. The Hypothesis (**A1-ii**) is equivalent to saying that u is differentiable at $x^\star$.

2. Hypothesis (**A2**) ensures that $X_n \longrightarrow x^\star$ a.s. and σ_n is almost surely finite, see Lelong [2009] for a proof of this result.

3. Hypothesis (**A3-i**) corresponds to some local uniform integrability condition and reminds of Lindeberg's condition. (**A3-ii**) guaranties the convergence of the angle bracket of the martingale of interest.

4. Hypothesis (**A4**) is only required for technical reasons but one does not need to be concerned with it in practical applications. It reminds of the case of constrained stochastic algorithms for which the CLT can only be proved for non saturated constraints.

The idea of using local assumptions, ie. assumptions to be checked in the neighbourhood of $x^\star$ was already used in the results of Duflo [1996] but unfortunately her proof cannot be easily adapted as being in a neighbourhood of $x^\star$ does not ensure that there is no projection anymore.

2.2 Main results

For $n \geq 0$, we define the renormalized and centered error

$$\Delta_n = \frac{X_n - x^\star}{\sqrt{\gamma_n}}.$$

A CLT for $1/2 < \alpha < 1$

Theorem 2.1. *If we assume Hypotheses (**A1**) to (**A4**), the sequence $(\Delta_n)_n$ converges in distribution to a normal random variable with mean 0 and covariance*

$$V = \int_0^\infty \exp(-At)\Sigma \exp(-A't)dt.$$

A CLT for $\alpha = 1$

Theorem 2.2. *We assume Hypotheses (**A1**) to (**A4**) and*

(A5) *$\gamma A - \frac{1}{2}I$ is repulsive*

Then, the sequence $(\Delta_n)_n$ *converges in distribution to a normal random variable with mean* 0 *and covariance*

$$V = \gamma \int_0^\infty \exp\left(\left(\frac{I}{2} - \gamma A\right) t\right) \Sigma \exp\left(\left(\frac{I}{2} - \gamma A'\right) t\right) dt.$$

Remark 3. Hypothesis (**A5**) involves the gradient of function u at the point $x^\star$, which is seldom tractable from a practical point of view but one can definitely not avoid it. The condition $\mathrm{Re}(\mathrm{Sp}(\gamma A - \frac{1}{2}I)) > 0$ is the border of two different convergence regimes as already noted by Duflo [1997] for the Robbins-Monro algorithm.

2.3 Discussion around the assumptions of Theorem 2.2

Theorem 2.2 is actually an extension of Chen [2002, Theorem 3.3.1]. The main improvements brought by our new result concern the conditions imposed on the noise term. Our Assumption (**A3**) is weaker than the one imposed by Chen [2002, A3.3.3] since we only assume local conditions on the noise terms; namely, unlike Chen, we only need to monitor the behavior of $(\delta M_n)_n$ in a small neighborhood of the optimum $x^\star$ (see Assumptions (**A3-i**) and (**A3-ii**)). Moreover, we only assume the local convergence in probability of the angle bracket of the martingale of interest built with $(\delta M_n)_n$ whereas Chen [2002, Equation (3.3.22)] requires the almost sure convergence which may be a little harder to prove in practical applications.

Concerning Assumption (**A1-ii**), it essentially means that u must be differentiable at $x^\star$. This is to be compared to the Hölder continuity property of the remainder of the first order expansion of u at $x^\star$ required by Chen [2002, A3.3.4], which is not so obvious to check in practice. Our goal in this work was not only to state a theorem with weaker assumptions but also to present a self contained and elementary proof of a central limit theorem for truncated stochastic algorithms. In particular, Lemma 3.1 provides a smart way of handling the truncation terms.

2.4 Example of applications

In this part, we present an application of Algorithm (3) to adaptive variance reduction and show the improvement brought by the localisation of the assumptions. The adaptive Monte–Carlo framework recalled hereafter was investigated by Arouna [2004] and later revisited by Lemaire and Pagès [2010], Lapeyre and Lelong [2011].

Let G be a $d-$dimensional standard normal random vector. For any measurable function $h : \mathbb{R}^d \longrightarrow \mathbb{R}$ such that $\mathbb{E}(|h(G)|) < \infty$, one has for all $\theta \in \mathbb{R}^d$

$$\mathbb{E}\left(h(G)\right) = \mathbb{E}\left(e^{-\theta\cdot G - \frac{|\theta|^2}{2}} h(G+\theta)\right). \tag{5}$$

Assume we want to compute $\mathbb{E}(f(G))$ using Monte–Carlo simulations for a measurable function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ such that $\mathbb{P}(f(G) \neq 0) > 0$ and there exists $\varepsilon > 0$, s.t. $\mathbb{E}(|f(G)|^{4+\varepsilon}) < \infty$.

By applying equality (5) to $h = f$ and $h(x) = f^2(x)\,\mathrm{e}^{-\theta\cdot x + \frac{|\theta|^2}{2}}$, one obtains that the expectation and the variance of the random variable $f(G+\theta)\,\mathrm{e}^{-\theta\cdot G\, - \frac{|\theta|^2}{2}}$ are respectively equal to $\mathbb{E}(f(G))$ and $v(\theta) - \mathbb{E}^2(f(G))$ where

$$v(\theta) = \mathbb{E}\left(f^2(G)\,\mathrm{e}^{-\theta\cdot G + \frac{|\theta|^2}{2}}\right).$$

The function v can be proved to be of infinitely continuously differentiable and strongly convex. Moreover,

$$\nabla v(\theta) = \mathbb{E}\left(f^2(G)\,\mathrm{e}^{-\theta\cdot G+\frac{|\theta|^2}{2}}(\theta - G)\right).$$

Hence, v has a unique minimizer characterised by $\nabla v(\theta^\star) = 0$.

As the computation of the expectation $\mathbb{E}(f(G))$ will be carried out using simulations, it is advised to make the most of the free parameter θ in Equation (5) by choosing $\theta = \theta^\star$. Obviously, this heavily relies on the ability to compute $\theta^\star$ numerically.

Note that ∇v can be written as an expectation of a function $U : \mathbb{R}^d \times \mathbb{R}^d \longmapsto \mathbb{R}^d$ defined by $U(\theta, x) = f^2(x)\,\mathrm{e}^{-\theta\cdot x+\frac{|\theta|^2}{2}}(\theta - x)$. If we consider Algorithm (3) for our problem, we get

$$\begin{cases} & \theta_{n+\frac{1}{2}} = \theta_n - \gamma_{n+1}v(\theta_n) - \gamma_{n+1}\delta M_{n+1}, \\ \text{if } \theta_{n+\frac{1}{2}} \in K_{\sigma_n} & \theta_{n+1} = \theta_{n+\frac{1}{2}} \quad \text{and} \quad \sigma_{n+1} = \sigma_n, \\ \text{if } \theta_{n+\frac{1}{2}} \notin K_{\sigma_n} & \theta_{n+1} = \theta_0 \quad \text{and} \quad \sigma_{n+1} = \sigma_n + 1 \end{cases}$$

with $\delta M_{n+1} = U(\theta_n, G_{n+1}) - \nabla v(\theta_n)$ where $(G_n)_n$ is a sequence of .i.i.d. standard normal random vectors.

Remark 4. Note that using Hölder's inequality and the assumptions on f, one can show that for any $\rho < \varepsilon/2$ and $A > 0$, $\mathbb{E}(\sup_{|\theta|<A} |U(\theta, G)|^{2+\rho}) < \infty$ Let us have a look at the assumptions of Theorems 2.1 and 2.2.

- Assumption (**A1**) is satisfied as v is strongly convex and the matrix A appearing in the assumption is symmetric positive definite.
- Let $q > 0$. The sequence $Y_n = \sum_{i=0}^n \gamma_{i+1}\delta M_{i+1}\mathbf{1}_{\{|\theta_i-\theta^\star|\le q\}}$ is a martingale with angle bracket $\langle Y\rangle_n = \sum_{i=0}^n \gamma_{i+1}^2\mathbb{E}(\delta M_{i+1}\delta M'_{i+1}|\mathcal{F}_i)\mathbf{1}_{\{|\theta_i-\theta^\star|\le q\}}$. As $\sum \gamma_i^2 < \infty$, by Remark 4, $\mathbb{E}(\langle Y\rangle_\infty) < \infty$ which implies, using the strong law of large numbers for square integrable martingales, that Y_n converges almost surely.
- Let $\eta > 0$ and $\rho < \varepsilon/2$. C denotes a positive constant.

  $$\mathbb{E}(|\delta M_n|^{2+\rho}\,\mathbf{1}_{\{|\theta_{n-1}-\theta^\star|\le\eta\}}|\mathcal{F}_{n-1}) \le C\left(\mathbb{E}(\sup_{|\theta-\theta^\star|\le\eta} |U(\theta,G)|^{2+\rho}) + \sup_{|\theta-\theta^\star|\le\eta} |\nabla v(\theta)|^{2+\rho}\right)$$

  The second term on the r.h.s is bounded as ∇v is continuous and the first term on the r.h.s. is bounded by applying Remark 4. The indicator of the set $\{|\theta_{n-1} - \theta^\star| \le \eta\}$ is definitely essential to prove the boundedness.
- Assumptions (**A1**) and (**A2**) imply that θ_n converges almost surely to $\theta^\star$.

  $$\mathbb{E}(\delta M_{i+1}\delta M'_{i+1}|\mathcal{F}_i) = \mathbb{E}(U(\theta, G)U(\theta, G))_{|\theta=\theta_i} - \nabla v(\theta_i)\nabla v(\theta_i)'.$$

  The continuity of ∇v ensures the convergence of $\nabla v(\theta_i)\nabla v(\theta_i)'$ to $\nabla v(\theta^\star)\nabla v(\theta^\star)'$. Moreover thanks to Remark 4, the expectation $\mathbb{E}(U(\theta, G)U(\theta, G)')$ is continuous w.r.t. θ. Hence, the matrix Σ of Assumption (**Aii**) is given by $\mathrm{Var}(U(\theta^\star, G))$ which is therefore positive definite.

As we have just seen it, having local assumptions makes them pretty easy to check in practical situations.

3 Proofs of Theorems 2.1 and 2.2

In this section, we prove the Theorems presented in Section 2.2 through a series of three lemmas. The proofs of these lemmas are postponed to Section 3.2.

3.1 Technical lemmas

For any fixed $n > 0$, we introduce $s_{n,k} = \sum_{i=1}^{k} \gamma_{n+i}$ for $k > 0$ and we set $s_{n,0} = 0$. $(s_{n,k})_{k\geq 0}$ can be interpreted as a discretisation grid of $[0,\infty)$ because $\lim_{k\to\infty} s_{n,k} = \infty$.

Theorems 2.1 and 2.2 are based on the following three lemmas.

Lemma 3.1. *Let $\varepsilon > 0$ and $\eta > 0$ as in Hypothesis (**A3**). There exists $N_0 > 0$, such that if we define for $n \geq N_0$*

$$A_n = \left\{ \sup_{n\geq m\geq N_0} |X_m - x^\star| \leq \eta \right\},$$

then

$$\mathbb{P}(A_n) \geq 1-\varepsilon \quad \forall n \geq N_0 \quad \textit{and} \quad \sup_{n\geq N_0} \mathbb{E}\left(|\Delta_n|^2 \mathbf{1}_{\{A_n\}} \right) < \infty.$$

Lemma 3.2. *For any integers $t > 0$ and $n > 0$*

$$\Delta_{n+t} = e^{-s_{n,t}Q}\Delta_n - \sum_{k=0}^{t-1} \mathrm{e}^{Q(s_{n,k}-s_{n,t})} \sqrt{\gamma_{n+k+1}}\delta M_{n+k+1} - \sum_{k=0}^{t-1} \mathrm{e}^{Q(s_{n,k}-s_{n,t})} \gamma_{n+k} R_{n+k}, \quad (6)$$

where

- *if $\alpha = 1$,*

$$\begin{cases} Q &= A - \frac{1}{2\gamma} I \\ R_m &= -y(X_m - x^\star)\Delta_m + \frac{1}{\sqrt{\gamma_{m+1}}} p_{m+1} \\ & \quad +\gamma_m(a_m I + b_m(A + y(X_m - x^\star)) + \mathcal{O}(\gamma_m))\Delta_m, \end{cases} \quad (7)$$

- *if $1/2 < \alpha < 1$,*

$$\begin{cases} Q &= A \\ R_m &= y(X_m - x^\star))\Delta_m - \frac{1}{\sqrt{\gamma_{m+1}}} p_{m+1} \\ & \quad -\frac{1}{m\gamma_m}(a_m I + b_m \gamma_n (A + y(X_m - x^\star)))\Delta_m + \mathcal{O}(\gamma_m)\Delta_m \end{cases} \quad (8)$$

  with $(a_n)_n$ and $(b_n)_n$ two real valued and bounded sequences.

Moreover, the last term in (6) tends to zero in probability.

Lemma 3.3. *In Equation (6), the sequence $(\sum_{k=0}^{t-1} \mathrm{e}^{Q(s_{n,k}-s_{n,t})} \sqrt{\gamma_{n+k+1}}\delta M_{n+k+1})_t$ converges in distribution to $\mathcal{N}(0,V)$ for any fixed n when t goes to infinity, where $V = \int_0^\infty \mathrm{e}^{-Qu}\,\Sigma\,\mathrm{e}^{-Q'u}\,du$.*

Proof of Theorems 2.1 and 2.2. Let us consider Equation (6) for a fixed $n > N_0$, where N_0 is defined in Lemma 3.1. Because the matrix Q is definite positive and Δ_n is almost surely finite, $e^{-s_{n,t}Q}\Delta_n$ tends to zero almost surely when t goes to infinity. Thanks to Lemma 3.2, the last term in Equation (6) tends to zero in probability when t goes to infinity.

Combining these two convergences in probability to zero with Lemma 3.3 yields the convergence in distribution of $(\Delta_{n+t})_t$ to a normal random variable with mean 0 and variance V when t goes to infinity, where V is defined in Lemma 3.3. Plugging the value of the matrix Q (see Equations (7) and (8)) in the expression of V yields the result. ■

Note that the proof for the classical Robbins Monro algorithm is much simpler since we do not need to introduce the A_n sets, which are only used here to handle the truncation terms.

3.2 Proofs of the lemmas

3.2.1 Proof of Lemma 3.1

We only do the proof in the case $\alpha = 1$, as in the other case, it is sufficient to slightly modify a few Taylor expansions and the same results still hold. From Equation (4), we have the following recursive relation

$$\Delta_{n+1} = \frac{X_{n+1} - x^\star}{\sqrt{\gamma_{n+1}}} = \sqrt{\frac{\gamma_n}{\gamma_{n+1}}}\,\Delta_n - \sqrt{\gamma_{n+1}}(u(X_n) + \delta M_{n+1} - p_{n+1}).$$

Using Hypothesis (**A1-ii**), the previous equation becomes

$$\Delta_{n+1} = \left(\sqrt{\frac{\gamma_n}{\gamma_{n+1}}}I - \sqrt{\gamma_{n+1}\gamma_n}(A + y(X_n - x^\star))\right)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} + \sqrt{\gamma_{n+1}}p_{n+1}. \quad (9)$$

The following Taylor expansions hold

$$\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} = 1 + \frac{\gamma_n}{2\gamma} + \mathcal{O}\left(\gamma_n^2\right) \text{ and } \sqrt{\gamma_n\gamma_{n+1}} = \gamma_n + \mathcal{O}\left(\gamma_n^2\right). \quad (10)$$

There exist two real valued and bounded sequences $(a_n)_n$ and $(b_n)_n$ such that

$$\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} = 1 + \frac{\gamma_n}{2\gamma} + \gamma_n^2 a_n \text{ and } \sqrt{\gamma_n\gamma_{n+1}} = \gamma_n + \gamma_n^2 b_n.$$

This enables us to simplify Equation (9)

$$\begin{aligned}\Delta_{n+1} =& \Delta_n - \gamma_n Q\Delta_n - \gamma_n y(X_n - x^\star)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} \\ &+ \sqrt{\gamma_{n+1}}p_{n+1} + \gamma_n^2(a_n I + b_n(A + y(X_n - x^\star)))\Delta_n, \end{aligned} \quad (11)$$

where $Q = A - \frac{I}{2\gamma}$. Let $\Delta_{n+\frac{1}{2}} = \dfrac{X_{n+\frac{1}{2}} - x^\star}{\sqrt{\gamma_{n+1}}}$, where $X_{n+\frac{1}{2}}$, defined by Equation (3), is the value of the new iterate obtained before truncation.

$$\begin{aligned}\left|\Delta_{n+\frac{1}{2}}\right|^2 =& |\Delta_n - \gamma_n Q\Delta_n - \gamma_n y(X_n - x^\star)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} \\ &+\gamma_n^2(a_n I + b_n(A + y(X_n - x^\star)I))\Delta_n\big|^2 \\ \leq& |\Delta_n - \gamma_n Q\Delta_n - \gamma_n y(X_n - x^\star)\Delta_n|^2 + \gamma_{n+1}|\delta M_{n+1}|^2 \\ &+ \gamma_n^4|(a_n I + b_n(A + y(X_n - x^\star)I))\Delta_n|^2 \\ &+ 2\gamma_n(\Delta_n - \gamma_n Q\Delta_n - \gamma_n y(X_n - x^\star)\Delta_n)'\delta M_{n+1} \\ &+ 2\gamma_n^4((a_n I + b_n(A + y(X_n - x^\star)I))\Delta_n)'(\Delta_n - \gamma_n Q\Delta_n - \gamma_n y(X_n - x^\star)\Delta_n) \\ &+ 2\gamma_n^{5/2}\delta M'_{n+1}(a_n I + b_n(A + y(X_n - x^\star)I))\Delta M_n.\end{aligned}$$

If we take the conditional expectation with respect to $\mathcal{F}_n$ — denoted $\mathbb{E}_n$ — in the previous equality, we find

$$\begin{aligned}\mathbb{E}_n\left|\Delta_{n+\frac{1}{2}}\right|^2 \leq& |\Delta_n - \gamma_n Q\Delta_n - \gamma_n y(X_n - x^\star)\Delta_n|^2 + \gamma_{n+1}\mathbb{E}_n|\delta M_{n+1}|^2 \\ &+ \gamma_n^4 \left\|a_n I + b_n(A + y(X_n - x^\star)I)\right\|^2 |\Delta_n|^2 \\ &+ 2\gamma_n^4 \left\|(a_n I + b_n(A + y(X_n - x^\star)I))\right\|(1 + \gamma_n \left\|Q + y(X_n - x^\star)\right\|)|\Delta_n|^2.\end{aligned}$$

$$\begin{aligned}\mathbb{E}_n\left(\left|\Delta_{n+\frac{1}{2}}\right|^2\right) \leq& \Delta_n'(I - 2\gamma_n Q)\Delta_n + \gamma_n \left\|y(X_n - x^\star)\right\| |\Delta_n|^2 + \gamma_{n+1}\mathbb{E}_n|\delta M_{n+1}|^2 \\ &+ \mathcal{O}\left(\gamma_n^2\right)(1 + |y(X_n - x^\star)|)\,|\Delta_n|^2. \end{aligned} \tag{12}$$

Note that in the previous equation the quantity $\mathcal{O}\left(\gamma_n^2\right)$ is non random.

Let $\varepsilon > 0$. Since $(X_n)_n$ converges almost surely to $x^\star$, there exists a rank N_0 such that

$$\mathbb{P}(\sup_{m>N_0} |X_m - x^\star| > \eta) < \varepsilon.$$

Hence, $\mathbb{P}(A_n) \geq 1 - \varepsilon$ for all $n > N_0$.

Let $\lambda = \min\{\mathrm{Re}(l); l \in \mathrm{Sp}(Q)\} > 0$ and $0 < \zeta < \lambda$. We can assume that for $n > N_0$, $\gamma_n \leq 1/\zeta$, then it ensues from Proposition A.1 that $\|I - \gamma_n Q\| \leq 1 - \gamma_n\zeta$. Moreover, since $\lim_{|x|\to 0}\|y(x)\| = 0$, there exists $\eta > 0$ such that for all $|x| < \eta$, $\|y(x)\| < \zeta/2$. We assume that this value of η satisfies Hypothesis (**A3**). On the set A_n,

$$\Delta_n'(I - 2\gamma_n Q)\Delta_n + \gamma_n \left\|y(X_n - x^\star)\right\| |\Delta_n|^2 \leq ((1 - \gamma_n\zeta) + \gamma_n\zeta/2)\,|\Delta_n|^2 \leq (1 - \gamma_n\zeta)\,|\Delta_n|^2.$$

Hence, we can deduce from Equation (12) that

$$\begin{aligned}\mathbb{E}\left(\left|\Delta_{n+\frac{1}{2}}\right|^2 \mathbf{1}_{\{A_n\}}\right) - \mathbb{E}\left(|\Delta_n|^2 \mathbf{1}_{\{A_n\}}\right) \leq& -\gamma_n\zeta\mathbb{E}\left(|\Delta_n|^2 \mathbf{1}_{\{A_n\}}\right) + \gamma_n\kappa \\ &+ \mathcal{O}(\gamma_n^2)(1 + \frac{1}{2}\zeta)\mathbb{E}\left(|\Delta_n|^2 \mathbf{1}_{\{A_n\}}\right).\end{aligned}$$

We can assume that for $n > N_0$, $\quad |\mathcal{O}(\gamma_n^2)(1 + \zeta/2)| \leq \gamma_n\zeta/2$. Hence we get, for $n \geq N_0$,

$$\mathbb{E}\left(\left|\Delta_{n+\frac{1}{2}}\right|^2 \mathbf{1}_{\{A_n\}}\right) - \mathbb{E}\left(|\Delta_n|^2 \mathbf{1}_{\{A_n\}}\right) \leq -\gamma_n\frac{\zeta}{2}\mathbb{E}\left(|\Delta_n|^2 \mathbf{1}_{\{A_n\}}\right) + \gamma_n\kappa$$

Since $A_{n+1} \subset A_n$,

$$\mathbb{E}\left(\left|\Delta_{n+\frac{1}{2}}\right|^2 \mathbf{1}_{\{A_{n+1}\}}\right) - \mathbb{E}\left(|\Delta_n|^2 \mathbf{1}_{\{A_n\}}\right) \leq -\gamma_n \frac{\zeta}{2}\mathbb{E}\left(|\Delta_n|^2 \mathbf{1}_{\{A_n\}}\right) + \kappa\gamma_n, \tag{13}$$

Now, we would like to replace $\Delta_{n+\frac{1}{2}}$ by Δ_{n+1} in Equation (13).

$$\begin{aligned} |\Delta_{n+1}|^2 &= \frac{|X_0 - x^\star|^2}{\gamma_{n+1}} \mathbf{1}_{\{p_{n+1}\neq 0\}} + \left|\Delta_{n+\frac{1}{2}}\right|^2 \mathbf{1}_{\{p_{n+1}=0\}}, \\ |\Delta_{n+1}|^2 &\leq \left|\Delta_{n+\frac{1}{2}}\right|^2 + \frac{|X_0 - x^\star|^2}{\gamma_{n+1}} \mathbf{1}_{\left\{X_{n+\frac{1}{2}} \notin K_{\sigma_n}\right\}}. \end{aligned}$$

Taking the conditional expectation w.r.t. $\mathcal{F}_n$ on the set A_n gives

$$\begin{aligned} \mathbb{E}_n |\Delta_{n+1}|^2 &\leq \mathbb{E}_n \left|\Delta_{n+\frac{1}{2}}\right|^2 + \frac{|X_0 - x^\star|^2}{\gamma_{n+1}} \mathbb{P}\left(X_{n+\frac{1}{2}} \notin K_{\sigma_n} | \mathcal{F}_n\right), \\ \mathbb{E}_n |\Delta_{n+1}|^2 \mathbf{1}_{\{A_n\}} &\leq \mathbb{E}_n \left|\Delta_{n+\frac{1}{2}}\right|^2 \mathbf{1}_{\{A_n\}} + \frac{|X_0 - x^\star|^2}{\gamma_{n+1}} \mathbf{1}_{\{A_n\}} \mathbb{P}\left(X_{n+\frac{1}{2}} \notin K_{\sigma_n} | \mathcal{F}_n\right), \\ \mathbb{E}\left(|\Delta_{n+1}|^2 \mathbf{1}_{\{A_{n+1}\}}\right) &\leq \mathbb{E}\left(\left|\Delta_{n+\frac{1}{2}}\right|^2 \mathbf{1}_{\{A_n\}}\right) + \frac{|X_0 - x^\star|^2}{\gamma_{n+1}} \mathbb{P}\left(A_n \cap \{X_{n+\frac{1}{2}} \notin K_{\sigma_n}\}\right). \end{aligned} \tag{14}$$

The probability on the right hand side can be rewritten

$$\mathbb{P}\left(A_n \cap \{X_{n+\frac{1}{2}} \notin K_{\sigma_n}\}\right) = \mathbb{E}\left(\mathbf{1}_{\{\gamma_{n+1}|u(X_n)+\delta M_{n+1}| \geq d(X_n, \partial K_{\sigma_n})\}} \mathbf{1}_{\{A_n\}}\right)$$

Moreover using the triangle inequality, we have $d(X_n, \partial K_{\sigma_n}) \geq d(x^\star, \partial K_{\sigma_n}) - |X_n - x^\star|$. Due to Hypothesis (**A4**), $d(x^\star, \partial K_{\sigma_n}) \geq \mu$ and on A_n, $|X_n - x^\star| \leq \eta$. Hence, $d(X_n, \partial K_{\sigma_n}) \geq \mu - \eta$. One can choose $\eta < \mu/2$ for instance, so that $d(X_n, \partial K_{\sigma_n}) > \frac{\mu}{2}$.

$$\begin{aligned} \mathbb{P}\left(A_n \cap \{X_{n+\frac{1}{2}} \notin K_{\sigma_n}\}\right) &\leq \mathbb{E}\left(\mathbb{E}_n\left(\mathbf{1}_{\left\{\gamma_{n+1}|u(X_n)+\delta M_{n+1}| \geq \frac{\mu}{2}\right\}}\right) \mathbf{1}_{\{A_n\}}\right), \\ &\leq \frac{8\gamma_{n+1}^2}{\mu^2} \mathbb{E}\left(|u(X_n)|^2 \mathbf{1}_{\{A_n\}} + |\delta M_{n+1}|^2) \mathbf{1}_{\{A_n\}}\right). \end{aligned} \tag{15}$$

Thanks to Hypothesis (**A3**) and the continuity of u, the expectation on the r.h.s of (15) is bounded by a constant $\overline{c} > 0$ independent of n. So, we get

$$\mathbb{P}\left(X_{n+\frac{1}{2}} \notin K_{\sigma_n}, A_n\right) \leq \overline{c}\gamma_{n+1}^2.$$

Hence, from Equation (14) we can deduce

$$\mathbb{E}\left(|\Delta_{n+1}|^2 \mathbf{1}_{\{A_{n+1}\}}\right) \leq \mathbb{E}\left(\left|\Delta_{n+\frac{1}{2}}\right|^2 \mathbf{1}_{\{A_n\}}\right) + \overline{c}\gamma_n. \tag{16}$$

By combining Equations (16) and (13), we come up with

$$\mathbb{E}\left(|\Delta_{n+1}|^2 \mathbf{1}_{\{A_{n+1}\}}\right) \leq \left(1 - \gamma_n \frac{\zeta}{2}\right) \mathbb{E}\left(|\Delta_n|^2 \mathbf{1}_{\{A_n\}}\right) + c\gamma_n,$$

where $c = \overline{c} + \kappa$.

Let $\mathcal{I} = \left\{ i \geq N_0 : -\frac{\zeta}{2}\mathbb{E}\left(|\Delta_i|^2 \mathbf{1}_{\{A_i\}}\right) + c > 0 \right\}$, then

$$\sup_{i\in\mathcal{I}} \mathbb{E}\left(|\Delta_i|^2 \mathbf{1}_{\{A_i\}}\right) < \frac{2c}{\zeta} < \infty.$$

Note that we can always assume that $2c/\zeta \geq \mathbb{E}\left(|\Delta_{N_0}|^2 \mathbf{1}_{\{A_{N_0}\}}\right)$, such that the set $\mathcal{I}$ is non empty. Assume $i \notin \mathcal{I}$, let $i_0 = \sup\{k < i \ : \ k \in \mathcal{I}\}$.

$$\mathbb{E}\left(|\Delta_i|^2 \mathbf{1}_{\{A_i\}}\right) - \mathbb{E}\left(|\Delta_{i_0}|^2 \mathbf{1}_{\{A_{i_0}\}}\right) \leq \sum_{k=i_0}^{i-1} \gamma_k \left(c - \frac{\zeta}{2}\mathbb{E}\left(|\Delta_k|^2 \mathbf{1}_{\{A_k\}}\right)\right)$$

Since all the terms for $k = i_0 + 1, \ldots, i-1$ are negative and $i_0 \in \mathcal{I}$, we find

$$\mathbb{E}\left(|\Delta_i|^2 \mathbf{1}_{\{A_i\}}\right) \leq \gamma_{i_0} c + \frac{2c}{\zeta}.$$

Finally, we come with the following upper bound.

$$\sup_{n\geq N_0} \mathbb{E}\left(|\Delta_n|^2 \mathbf{1}_{\{A_n\}}\right) < \infty.$$

Remark 5 (case $1/2 < \alpha < 1$)**.** This proof is still valid for $\alpha < 1$ if we replace the Taylor expansions of Equation (10) by

$$\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} = 1 + \frac{a_n}{n} \text{ and } \sqrt{\gamma_n\gamma_{n+1}} = \gamma_n + \frac{\gamma_n b_n}{n}.$$

Then, Equation (11) becomes

$$\begin{aligned}\Delta_{n+1} =& \Delta_n - \gamma_n Q\Delta_n - \gamma_n y(X_n - x^\star)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} \\ &+ \sqrt{\gamma_{n+1}} p_{n+1} + \frac{1}{n}(a_n I + b_n\gamma_n(A + y(X_n - x^\star))\Delta_n,\end{aligned}$$

with $Q = A$ this time, which is still positive definite.

3.2.2 Proof of Lemma 3.2

Let us go back to Equation (11). For any $n > N_0$ and $k > 0$, we can write

$$\Delta_{n+k} = \Delta_{n+k-1} - \gamma_{n+k-1}Q\Delta_{n+k-1} - \sqrt{\gamma_{n+k}}\delta M_{n+k} + \gamma_{n+k-1}\overline{R}_{n+k-1}$$

where

$$\overline{R}_m = -y(X_m - x^\star)\Delta_m + \frac{1}{\sqrt{\gamma_{m+1}}}p_{m+1} + \gamma_m(a_m I + b_m(A + y(X_m - x^\star)))\Delta_m,$$

We can actually notice that the previous equation pretty much looks like a discrete time ODE. Based on this remark, it is natural to multiply the previous equation by $\mathrm{e}^{s_{n,k}Q}$ to find

$$\mathrm{e}^{s_{n,k}Q}\,\Delta_{n+k} - (\mathrm{e}^{s_{n,k}Q} - \mathrm{e}^{s_{n,k}Q}\,\gamma_{n+k-1}Q)\Delta_{n+k-1} = -\,\mathrm{e}^{s_{n,k}Q}\,\sqrt{\gamma_{n+k}}\delta M_{n+k} + \gamma_{n+k-1}\,\mathrm{e}^{s_{n,k}Q}\,\overline{R}_{n+k-1}$$

Note that $\mathrm{e}^{s_{n,k}Q} - \mathrm{e}^{s_{n,k}Q}\,\gamma_{n+k-1}Q = \mathrm{e}^{s_{n,k-1}Q}(1+\mathcal{O}(\gamma_{n+k-1}^2))$. Hence, we come up with the following equation

$$\mathrm{e}^{s_{n,k}Q}\,\Delta_{n+k} - \mathrm{e}^{s_{n,k-1}Q}\,\Delta_{n+k-1} = -\,\mathrm{e}^{s_{n,k}Q}\,\sqrt{\gamma_{n+k}}\delta M_{n+k} + \gamma_{n+k-1}\,\mathrm{e}^{s_{n,k}Q}\,R_{n+k-1}$$

where

$$R_m = -y(X_m - x^\star)\Delta_m + \frac{1}{\sqrt{\gamma_{m+1}}}p_{m+1} + \gamma_m(a_m I + b_m(A + y(X_m - x^\star)) + \mathcal{O}(1))\Delta_m, \quad (17)$$

When summing the previous equalities for $k = 1, \ldots, t-1$ for any integer $t > 0$, we get

$$\Delta_{n+t} = \mathrm{e}^{-s_{n,t}Q}\,\Delta_n - \sum_{k=0}^{t-1}\mathrm{e}^{(s_{n,k}-s_{n,t})Q}\,\sqrt{\gamma_{n+k+1}}\delta M_{n+k+1} - \sum_{k=0}^{t-1}\mathrm{e}^{(s_{n,k}-s_{n,t})Q}\,\gamma_{n+k}R_{n+k},$$

Let us a have a closer look at the different terms of Equation (17)

- $\lim_m y(X_m - x^\star)\Delta_m \mathbf{1}_{\{|X_m - x^\star|>\eta\}} = 0$ a.s. thanks to the a.s. convergence of $(X_m)_m$ and using Lemma 3.1, the sequence $(y(X_m - x^\star)\Delta_m \mathbf{1}_{\{|X_m - x^\star|\le\eta\}})_m$ is uniformly integrable and tends to zero in probability because $\lim_m y(X_m - x^\star) = 0$ a.s.
- p_m is almost surely equal to 0 for m large enough thanks to Remark 2, so $\frac{1}{\sqrt{\gamma_m}}p_m = 0$ a.s. for m large enough.
- $\gamma_m(a_m I + b_m(A + y(X_m - x^\star)) + \mathcal{O}(1))\Delta_m \mathbf{1}_{\{|X_m - x^\star|>\eta\}} \longrightarrow 0$ almost surely because for m large enough the indicator equals 0. The sequence $\gamma_m(a_m I + b_m(A + y(X_m - x^\star)) + \mathcal{O}(1))\Delta_m \mathbf{1}_{\{|X_m - x^\star|\le\eta\}}$ is uniformly integrable by Lemma 3.1 and tends to zero in probability because $\gamma_m \longrightarrow 0$.

Hence, R_m can be split in two terms : one tending to zero almost surely and an other one which is uniformly integrable and tends to zero in probability. Then, we can apply Propositions A.2 and A.3 to prove the convergence in probability of $(\sum_{k=0}^{t-1}\mathrm{e}^{(s_{n,k}-s_{n,t})Q}\,\gamma_{n+k}R_{n+k})_t$. This last point ends the proof of Lemma 3.2.

3.2.3 Proof of Lemma 3.3

To prove Lemma 3.3, we need a result on the rate of convergence of martingale arrays. First, note that for $\frac{1}{\sqrt{\gamma_n}}\,\delta M_n \mathbf{1}_{\{|X_{n-1}-x^\star|>\eta\}}$ tends to 0 a.s. when n goes to infinity because $\mathbf{1}_{\{|X_{n-1}-x^\star|>\eta\}} = 0$ for n large enough. Then, it ensues from Proposition A.2 that $\sum_{k=1}^{t}\mathrm{e}^{Q(s_{n,k}-s_{n,t})}\,\sqrt{\gamma_{n+k-1}}\delta M_{n+k}\mathbf{1}_{\{|X_{n+k-1}-x^\star|>\eta\}}$ converges to zero in probability when t goes to infinity. Henceforth, it is sufficient to prove a localized version of Lemma 3.3 by considering $\sum_{k=0}^{t}\mathrm{e}^{Q(s_{n,k}-s_{n,t})}\,\sqrt{\gamma_{n+k}}\delta M_{n+k}\mathbf{1}_{\{|X_{n+k-1}-x^\star|\le\eta\}}$.

We will use the following Central Limit Theorem for martingale arrays adapted from Duflo [1997, Theorem 2.1.9].

Theorem 3.4. *Suppose that $\{(\mathcal{F}_l^t)_{0\le l\le t}; t>0\}$ is a family of filtrations and $\{(N_l^t)_{0\le l\le t}; t>0\}$ a square integrable martingale array with respect to the previous filtration. Assume that :*

(A6) *there exists a symmetric positive definite matrix Γ such that $\langle N\rangle_t^t \xrightarrow[t\to\infty]{\mathbb{P}} \Gamma$.*

(A7) *There exists $\rho > 0$ such that*

$$\sum_{l=1}^{t} \mathbb{E}\left(\left|N_l^t - N_{l-1}^t\right|^{2+\rho} \left|\mathcal{F}_{l-1}^t\right.\right) \xrightarrow[t\to\infty]{\mathbb{P}} 0.$$

Then,

$$N_t^t \xrightarrow[t\to\infty]{\mathcal{L}} \mathcal{N}(0,\Gamma).$$

Using this theorem, we can now prove Lemma 3.3.

Proof of Lemma 3.3. Let us define N_l^t for all $0 \le l \le t$ and $t > 0$

$$N_l^t = \sum_{k=1}^{l} \mathrm{e}^{(s_{n,k}-s_{n,t})Q} \sqrt{\gamma_{n+k}}\delta M_{n+k} \mathbf{1}_{\{|X_{n+k-1}-x^\star|\le\eta\}}.$$

We should have written $s_{n,k-1}$ instead of $s_{n,k}$, but as $\mathrm{e}^{(s_{n,k-1}-s_{n,t})Q} = \mathrm{e}^{(s_{n,k}-s_{n,t})Q}\,\mathrm{e}^{-\gamma_{n+k}Q}$ and $\mathrm{e}^{-\gamma_{n+k}Q}$ converges to the identity matrix, we can make this little change without altering the rigor of the proof and this way, N_l^t naturally fits in the framework of Propositions A.2 and A.4.

$(N_l^t)_{0\le l\le p}$ is obviously a martingale with respect to $(\mathcal{F}_{n+l})_l$. Let us compute its angle bracket

$$\langle N\rangle_t^t = \sum_{k=1}^{t} \mathrm{e}^{(s_{n,k}-s_{n,t})Q}\, \gamma_{n+k}\mathbb{E}\left(\delta M_{n+k}\delta M'_{n+k}\mathbf{1}_{\{|X_{n+k-1}-x^\star|\le\eta\}}|\mathcal{F}_{n+k-1}\right) \mathrm{e}^{(s_{n,k}-s_{n,t})Q'} \quad (18)$$

Thanks to Hypotheses (**A3**), the conditional expectation in (18) is uniformly integrable and converges in probability to Σ when k goes to infinity. Applying Proposition A.4 proves the convergence in probability of $\langle N\rangle_t^t$ to $\int_0^\infty \mathrm{e}^{-Q'u}\,\Sigma\,\mathrm{e}^{-Qu}\,du$. Let ρ be the real number defined in Theorem 2.1.

$$\sum_{l=1}^{t} \mathbb{E}\left(\left|N_l^t - N_{l-1}^t\right|^{2+\rho}\right) = \sum_{k=1}^{t} \left\|\mathrm{e}^{(s_{n,k}-s_{n,t})Q}\right\|^{2+\rho} \gamma_{n+k}^{1+\frac{\rho}{2}}\mathbb{E}\left(\left|\delta M_{n+k}\right|^{2+\rho}\mathbf{1}_{\{|X_{n+k-1}-x^\star|\le\eta\}}\right). \quad (19)$$

$\gamma_{n+k}^{\frac{\rho}{2}}$ converges to 0 when k goes to infinity and the sequence of expectations is bounded using Hypothesis (**A3**), so $\gamma_{n+k}^{\frac{\rho}{2}}\mathbb{E}\left(\left|\delta M_{n+k}\right|^{2+\rho}\mathbf{1}_{\{|X_{n+k-1}-x^\star|\le\eta\}}\right)$ tends to zero when k goes to infinity. Proposition A.2 proves that the l.h.s. of Equation (19) tends to 0 when t goes to infinity. Hence, $\sum_{l=1}^{t}\mathbb{E}\left(\left|N_l^t - N_{l-1}^p\right|^{2+\rho}\left|\mathcal{F}_{l-1}^t\right.\right)$ tends to zero in $\mathbb{L}^1$, and consequently in probability. Then, the Hypotheses of Theorem 3.4 are satisfied.
Finally, we have proved that

$$\sum_{k=0}^{t} \mathrm{e}^{Q(s_{n,k}-s_{n,t})} \sqrt{\gamma_{n+k}}\delta M_{n+k} \xrightarrow[t\to\infty]{law} \mathcal{N}\left(0, \int_0^\infty \mathrm{e}^{-Qu}\,\Sigma\,\mathrm{e}^{-Q'u}\,du\right).$$

■

4 Conclusion

In this work, we have proved a Central Limit Theorem with rate $\sqrt{\gamma_n}$ for randomly truncated stochastic algorithms under local assumptions. We have also tried to clarify the proof of the convergence rate of randomly truncated stochastic algorithms under assumptions which can be easily verified in practice. The improvement brought by this new set of assumptions is that all they should only be checked in a neighbourhood of the target value $x^\star$, which means that in the case where $u(x) = \mathbb{E}(U(x, Z))$ the assumptions can be reformulated in terms of some local regularity properties of U.

A Some elementary results

Proposition A.1. *Let Q be a square matrix such that all its eigenvalues have positive real parts. Let $\overline{\lambda} > 0 = \min\{\mathrm{Re}(\mu); \mu \in Sp(Q)\}$. Then, for all $0 < \lambda < \overline{\lambda}$, there exists $0 < \overline{\gamma} < 1/\lambda$, such that for all $\gamma \leq \overline{\gamma}$, $\|I - \gamma Q\| \leq (1 - \gamma\lambda)$.*

Proof. We consider the Jordan decomposition J of Q such that $Q = PJP^{-1}$. $\|I - \gamma Q\| \leq \|I - \gamma J\|$. Let $M = \sup\{\mathrm{Im}(\mu); \mu \in Sp(Q)\}$ and $\overline{\gamma}_1 = \max\{\mathrm{Re}(\mu); \mu \in Sp(Q)\}$. If $\gamma < \max(\overline{\gamma}_1, 1/\overline{\lambda})$, the diagonal terms of $I - \gamma J$ are bounded by $\sqrt{(1 - \gamma\overline{\lambda})^2 + (\gamma M)^2} = (1 - \gamma\overline{\lambda})(1 + O(\gamma^2))$. The conclusion follows easily. ■

Proposition A.2. *Let $(Y_n)_n$ be a sequence of random vectors of $\mathbb{R}^d$ converging almost surely to a non random vector $x \in \mathbb{R}^d$. For any fixed integer $n > 0$ and repulsive matrix $Q \in \mathbb{R}^{d\times d}$, we define, for all integers $t \geq 0$, $Z_t = \sum_{k=0}^{t} \mathrm{e}^{Q(s_{n,k} - s_{n,t})} \gamma_{n+k} Y_{n+k}$. Then, $\lim_t Z_t = \int_0^\infty \mathrm{e}^{-Qu}\, x\, du$ almost surely.*

Proof. It is clear that $\lim_{t\to\infty} \int_0^{s_{n,t}} \mathrm{e}^{-Qu}\, du\ x = \int_0^\infty \mathrm{e}^{-Qu}\, du\ x$. Hence, it is sufficient to consider

$$\left| Z_t - \int_0^{s_{n,t}} \mathrm{e}^{-Qu}\, du\, x \right| \leq \sum_{k=0}^{t} \gamma_{n+k} \left\| \mathrm{e}^{Q(s_{n,k} - s_{n,t})} \right\| |Y_{n+k} - x| + \left\| \sum_{k=0}^{t} \gamma_{n+k}\, \mathrm{e}^{Q(s_{n,k} - s_{n,t})} - \int_0^{s_{n,t}} \mathrm{e}^{-Qu}\, du \right\| |x|. \tag{20}$$

Let $\underline{q} = \min\{\mathrm{Re}(\lambda); \lambda \in \mathrm{Sp}(Q)\} > 0$ and $\overline{q} = \max\{\mathrm{Re}(\lambda); \lambda \in \mathrm{Sp}(Q)\} > 0$.

Step 1 : We will prove that the first term in Equation (20) tends to 0 almost surely.

$$\sum_{k=0}^{t} \gamma_{n+k} \left\| \mathrm{e}^{Q(s_{n,k} - s_{n,t})} \right\| |Y_{n+k} - x| \leq \sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{\underline{q}(s_{n,k} - s_{n,t})} |Y_{n+k} - x|\, du \leq \int_0^{s_{n,t}} \mathrm{e}^{\underline{q}(u - s_{n,t})}\, \mathrm{e}^{\underline{q}\gamma_{n+t_n(u)}} |Y_{n+t_n(u)} - x|\, du \tag{21}$$

where for any real number $u > 0$, $t_n(u)$ is the largest integer k such that $s_{n,k-1} \leq u < s_{n,k}$. Note that $\lim_{u\to+\infty} t_n(u) = +\infty$. $\lim_{u\to+\infty} \mathrm{e}^{\underline{q}\gamma_{n+t_n(u)}} |Y_{n+t_n(u)} - x| = 0$ a.s., hence it is obvious that the term on the r.h.s of Equation (21) tend to 0 almost surely.

Step 2 : We will now prove that the second term in Equation (20) tends to 0. We use the convention $s_{n,-1} = 0$ and recall that $s_{n,k} = s_{n,k-1} + \gamma_{n+k}$. Note that $\int_0^{s_{n,t}} \mathrm{e}^{-Qu}\, du = \int_0^{s_{n,t}} \mathrm{e}^{Q(u-s_{n,t})}\, du$, hence the following inequality holds

$$\begin{aligned}\left\|\sum_{k=0}^{t} \gamma_{n+k}\, \mathrm{e}^{Q(s_{n,k}-s_{n,t})} - \int_0^{s_{n,t}} \mathrm{e}^{-Qu}\, du\right\| &\leq \sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \left\|\mathrm{e}^{Q(s_{n,k}-s_{n,t})} - \mathrm{e}^{Q(u-s_{n,t})}\right\| du \\ &\leq \sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \left\|\mathrm{e}^{Q(u-s_{n,t})}\right\| \left\|\mathrm{e}^{Q(s_{n,k}-u)} - I\right\| du. \\ &\leq \sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{\underline{q}(u-s_{n,t})}(\mathrm{e}^{\overline{q}\gamma_{n+k}} - 1) du. \end{aligned} \tag{22}$$

Let $\varepsilon > 0$, there exits $T_1 > 0$ such that for all $t \geq T_1$, $(\mathrm{e}^{\overline{q}\gamma_{n+t}} - 1) \leq \varepsilon$, hence for all $t > T_1$,

$$\begin{aligned}\sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{\underline{q}(u-s_{n,t})}(\mathrm{e}^{\overline{q}\gamma_{n+k}} - 1) du &\leq \sum_{k=0}^{T_1} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{\underline{q}(u-s_{n,t})}(\mathrm{e}^{\overline{q}} - 1) du + \varepsilon \sum_{k=T_1+1}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{\underline{q}(u-s_{n,t})}\, du, \\ &\leq \int_0^{s_{n,T_1}} \mathrm{e}^{\underline{q}(u-s_{n,t})}(\mathrm{e}^{\overline{q}} - 1) du + \varepsilon \int_{s_{n,T_1}}^{s_{n,t}} \mathrm{e}^{\underline{q}(u-s_{n,t})}\, du, \\ &\leq (\mathrm{e}^{\underline{q}(s_{n,T_1}-s_{n,t})} - \mathrm{e}^{-\underline{q}s_{n,t}}) \frac{\mathrm{e}^{\overline{q}} - 1}{\underline{q}} + \varepsilon \frac{1}{\underline{q}}.\end{aligned}$$

There exists $T_2 > T_1$ such that for all $t > T_2$, $(\mathrm{e}^{\underline{q}(s_{n,T_1}-s_{n,t})} - \mathrm{e}^{-\underline{q}s_{n,t}})(\mathrm{e}^{\overline{q}} - 1) \leq \varepsilon$, hence for all $t > T_2$,

$$\sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{\underline{q}(u-s_{n,t})}(\mathrm{e}^{\overline{q}\gamma_{n+k}} - 1) du \leq \frac{2\varepsilon}{\underline{q}}.$$

This ends to prove that the second term in Equation (20) tends to 0 when t goes to infinity. ■

Proposition A.3. *Let $(Y_n)_n$ be a sequence of random vectors of $\mathbb{R}^d$ uniformly integrable and converging in probability to a non random vector $x \in \mathbb{R}^d$. For any fixed integer $n > 0$ and repulsive matrix $Q \in \mathbb{R}^{d\times d}$, we define, for all integers $t \geq 0$, $Z_t = \sum_{k=0}^{t} \mathrm{e}^{Q(s_{n,k}-s_{n,t})}\, \gamma_{n+k} Y_{n+k}$. Then, $\lim_t Z_t = \int_0^{\infty} \mathrm{e}^{-Qu}\, x\, du$ in probability.*

Proof. We recall the decomposition given by Equation (20)

$$\begin{aligned}\left|Z_t - \int_0^{s_{n,t}} \mathrm{e}^{-Qu}\, du\, x\right| &\leq \sum_{k=0}^{t} \gamma_{n+k} \left\|\mathrm{e}^{Q(s_{n,k}-s_{n,t})}\right\| |Y_{n+k} - x| \\ &\quad + \left\|\sum_{k=0}^{t} \gamma_{n+k}\, \mathrm{e}^{Q(s_{n,k}-s_{n,t})} - \int_0^{s_{n,t}} \mathrm{e}^{-Qu}\, du\right\| |x|.\end{aligned}$$

The last term in the above equation has already been proved to tend to 0 in the proof of Proposition A.2 Step 2. So, we only need to prove that $\lim_{u\to+\infty} \sum_{k=0}^{t} \gamma_{n+k} \left\|\mathrm{e}^{Q(s_{n,k}-s_{n,t})}\right\| |Y_{n+k} - x| = 0$ in probability.

Let $\underline{q} = \min\{\mathrm{Re}(\lambda); \lambda \in \mathrm{Sp}(Q)\} > 0$ and $\overline{q} = \max\{\mathrm{Re}(\lambda); \lambda \in \mathrm{Sp}(Q)\} > 0$.

$$\begin{aligned}\sum_{k=0}^{t} \gamma_{n+k} \left\| \mathrm{e}^{Q(s_{n,k}-s_{n,t})} \right\| |Y_{n+k} - x| &\leq \sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{\underline{q}(s_{n,k}-s_{n,t})} |Y_{n+k} - x| \, du \\ &\leq \int_0^{s_{n,t}} \mathrm{e}^{\underline{q}(u-s_{n,t})} \, \mathrm{e}^{\underline{q}\gamma_{n+t_n(u)}} |Y_{n+t_n(u)} - x| \, du\end{aligned}$$

where for any real number $u > 0$, $t_n(u)$ is the largest integer k such that $s_{n,k-1} \leq u < s_{n,k}$. Let $\overline{Y}_k = \gamma_k |Y_k - x|$. The sequence $(\overline{Y}_k)_k$ tends to zero in probability, is uniformly integrable and positive.

$$\mathbb{E}\left(\int_0^{s_{n,t}} \mathrm{e}^{\underline{q}(u-s_{n,t})} \, \overline{Y}_{n+t_n(u)} \, du \right) = \int_0^{s_{n,t}} \mathrm{e}^{\underline{q}(u-s_{n,t})} \, \mathbb{E}(\overline{Y}_{n+t_n(u)}) du. \tag{23}$$

Since $(\overline{Y}_k)_k$ is uniformly integrable and converges to 0 in probability, $\lim_{u\to+\infty} \mathbb{E}(\overline{Y}_{n+t_n(u)}) = 0$, hence the term on the r.h.s of Equation (23) tends to 0 when t goes to infinity. This proves that $\lim_{u\to+\infty} \sum_{k=0}^{t} \gamma_{n+k} \left\| \mathrm{e}^{Q(s_{n,k}-s_{n,t})} \right\| |Y_{n+k} - x| = 0$ in $\mathbb{L}^1$ and in probability. ■

Proposition A.4. *Let $(Y_n)_n \in \mathbb{R}^{d\times d}$ be a sequence of random square matrices $\mathbb{R}^d$ uniformely integrable and converging in probability to a non random matrix $X \in \mathbb{R}^{d\times d}$. For any fixed integer $n > 0$ and repulsive matrix $Q \in \mathbb{R}^{d\times d}$, we define, for all integers $t \geq 0$, $Z_t = \sum_{k=0}^{t} \mathrm{e}^{Q(s_{n,k}-s_{n,t})} \gamma_{n+k} Y_{n+k} \mathrm{e}^{Q'(s_{n,k}-s_{n,t})}$. Then, $\lim_t Z_t = \int_0^\infty \mathrm{e}^{-Qu} X \mathrm{e}^{-Q'u} \, du$ in probability.*

Proof. The proof is very similar to the previous but for sake of completeness we make it. Let $\underline{q} = \min\{\mathrm{Re}(\lambda); \lambda \in \mathrm{Sp}(Q)\} > 0$ and $\overline{q} = \max\{\mathrm{Re}(\lambda); \lambda \in \mathrm{Sp}(Q)\} > 0$.
We use a decomposition similar to Equation (20)

$$\begin{aligned}\left| Z_t - \int_0^{s_{n,t}} \mathrm{e}^{-Qu} X \mathrm{e}^{-Q'u} \, du \right| &\leq \sum_{k=0}^{t} \gamma_{n+k} \left\| \mathrm{e}^{Q(s_{n,k}-s_{n,t})} (Y_{n+k} - X) \, \mathrm{e}^{Q'(s_{n,k}-s_{n,t})} \right\| \\ &\quad + \left\| \sum_{k=0}^{t} \gamma_{n+k} \, \mathrm{e}^{Q(s_{n,k}-s_{n,t})} X \, \mathrm{e}^{Q'(s_{n,k}-s_{n,t})} - \int_0^{s_{n,t}} \mathrm{e}^{-Qu} X \mathrm{e}^{-Q'u} \, du \right\| \\ &\leq \sum_{k=0}^{t} \gamma_{n+k} \, \mathrm{e}^{2\underline{q}(s_{n,k}-s_{n,t})} \|Y_{n+k} - X\| \\ &\quad + \left\| \sum_{k=0}^{t} \gamma_{n+k} \, \mathrm{e}^{Q(s_{n,k}-s_{n,t})} X \, \mathrm{e}^{Q'(s_{n,k}-s_{n,t})} - \int_0^{s_{n,t}} \mathrm{e}^{-Qu} X \mathrm{e}^{-Q'u} \, du \right\|\end{aligned}$$

The convergence in probability to zero of the first term on the r.h.s of the above equation is

ensured by Proposition A.3. We rewrite the second term as follows

$$
\begin{aligned}
&\left\| \sum_{k=0}^{t} \gamma_{n+k}\, \mathrm{e}^{Q(s_{n,k}-s_{n,t})}\, X\, \mathrm{e}^{Q'(s_{n,k}-s_{n,t})} - \int_0^{s_{n,t}} \mathrm{e}^{-Qu}\, X\, \mathrm{e}^{-Q'u}\, du \right\| \\
&= \left\| \sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{Q(s_{n,k}-s_{n,t})}\, X\, \mathrm{e}^{Q'(s_{n,k}-s_{n,t})} - \mathrm{e}^{Q(u-s_{n,t})}\, X\, \mathrm{e}^{Q'(u-s_{n,t})}\, du \right\| \\
&= \left\| \sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{Q(u-s_{n,t})} (\mathrm{e}^{Q(s_{n,k}-u)}\, X\, \mathrm{e}^{Q'(s_{n,k}-u)} - X)\, \mathrm{e}^{Q'(u-s_{n,t})}\, du \right\| \\
&= \left\| \sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{Q(u-s_{n,t})} (\mathrm{e}^{Q(s_{n,k}-u)} - I) X\, \mathrm{e}^{Q'(s_{n,k}-u)} + X(\mathrm{e}^{Q'(s_{n,k}-u)} - I)\, \mathrm{e}^{Q'(u-s_{n,t})}\, du \right\| \\
&\leq \sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{2\underline{q}(u-s_{n,t})} \left\{ (\mathrm{e}^{\overline{q}(s_{n,k}-u)} - 1)\, \mathrm{e}^{\overline{q}(s_{n,k}-u)} + (\mathrm{e}^{\overline{q}(s_{n,k}-u)} - 1) \right\} du\, \|X\| \\
&\leq \sum_{k=0}^{t} \int_{s_{n,k-1}}^{s_{n,k}} \mathrm{e}^{2\underline{q}(u-s_{n,t})} (\mathrm{e}^{\overline{q}(s_{n,k}-u)} - 1)(\mathrm{e}^{\overline{q}(s_{n,k}-u)} + 1) du\, \|X\|
\end{aligned}
$$

The proof is ended by closely following the same reasoning as to prove that the r.h.s of Equation (22) tends to 0. ■

References

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

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

C. Bouton. *Approximation Gaussienne d'algorithmes stochastiques à dynamique Markovienne*. PhD thesis, Université Pierre et Marie Curie - Paris 6, 1985.

R. Buche and H. J. Kushner. Rate of convergence for constrained stochastic approximation algorithms. *SIAM J. Control Optim.*, 40(4):1011–1041 (electronic), 2001.

H. Chen and Y. Zhu. *Stochastic Approximation Procedure with randomly varying truncations*. Scientia Sinica Series, 1986.

H.-F. Chen. *Stochastic approximation and its applications*, volume 64 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 2002.

B. Delyon. General results on the convergence of stochastic algorithms. *IEEE Transactions on Automatic Control*, 41(9):1245–1255, 1996.

M. Duflo. *Algorithmes stochastiques (Mathématiques et Applications)*. Springer, 1996.

M. Duflo. *Random Iterative Models*. Springer-Verlag Berlin and New York, 1997.

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

B. Lapeyre and J. Lelong. A framework for adaptive Monte–Carlo procedures. *Monte Carlo Methods Appl.*, 2011.

J. Lelong. Almost sure convergence of randomly truncated stochastic agorithms under verifiable conditions. *Statistics & Probability Letters*, 78, 2009.

V. Lemaire and G. Pagès. Unconstrained Recursive Importance Sampling. *Ann. Appl. Probab.*, 20(3):1029–1067, 2010.

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

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