Rapport de recherche du CERMICS NR: 2006-312 Mai 2006

Central Limit Theorems for Truncating and Averaging Stochastic Algorithms: a functional approach

Jérôme Lelong

CERMICS — ENPC
6 et 8 avenue Blaise Pascal
Cité Descartes - Champs sur Marne
77455 Marne la Vallée Cedex 2

http://cermics.enpc.fr

Central Limit Theorems for Truncating and Averaging Stochastic Algorithms: a functional approach

Jérôme LELONG[*]
CERMICS, Ecole Nationale des Ponts et Chaussées, 6 et 8 av Blaise Pascal,
77455 Marne La Vallée, FRANCE.

May 15, 2006

Abstract

This article is devoted to the study of the convergence rate of stochastic algorithms using Chen's procedure throughout a functional Central Limit Theorem. We establish the convergence in the Skorokhod space of a well defined interpolation of the renormalised iterates to a stationary Ornstein Uhlenbeck process. This new result enables us to derive a CLT for the moving window averaging version of Chen's algorithm.
Key words: stochastic approximation, central limit theorem, tightness, recursive algorithms, Skorokhod, càdlàg processes, Chen's procedure, random truncations.

1 Introduction

In a stochastic context, optimisation problems often become of a great complexity. In that respect, stochastic algorithms represent an extremely valuable alternative to deterministic techniques. Nonetheless, the convergence of these algorithms is hung up to assumptions that are barely satisfied in practise, namely a sub-linear growth is required. There may also be other reasons why these algorithms cannot be used right-away. The optimisation problem may be to be solved under some constraints or may have local optima. Because of these reasons, these algorithms need to be improved for practical usage. Constrained algorithms are used to ensure that the constrains are satisfied at each step by projecting the new iterate on the constraint set.
The way to deal with fast growing functions also requires some kind of projections. The idea is to prevent the algorithm from blowing up during the first steps by forcing the iterates to remain in an increasing sequence of compact sets. This procedure, known as Chen's procedure, was first introduced in [5]. This algorithm is often needed in practise and applications in finance have recently been given in [1], where the author explains how to implement an adaptive importance sampling technique using these stochastic procedures. As the payoffs commonly involved are usually completely non-linear, the standard procedures often fail and Chen's truncation technique is used instead.
Sometimes, one uses an averaging version of Chen's procedure to smooth its numerical behaviour. Averaging algorithms have already been studied in [13], but not in combination with random truncations. The study the convergence rate of averaging algorithms can be achieved through the getting of a functional Central Limit Theorem (CLT) for the corresponding standard algorithm.

[*]E-mail: lelong@cermics.enpc.fr

Numerous results are already known on the convergence rate of unconstrained algorithms, see [8] or [7]. Duflo points out that the convergence speed of stochastic algorithms depends on the relative position of the smallest eigenvalue of the Hessian matrix at the optimum and 1/2. This remark is to be put together with Hypothesis 3. A functional CLT has also been proved for unconstrained algorithm by Bouton in [4] or Benveniste et al. in [2]. The convergence rate of constrained algorithms has been studied by Kushner et al. in [11] where he gives a functional CLT. Given that functional result, they derive a CLT for averaging constrained algorithms. The problem of multiple targets has been tackled by Pelletier in [12] where she obtained a CLT. As one can see, a lot of work has already been done around the convergence rate of stochastic algorithm. However, beyond the almost sure convergence (see [6]) no results are known for Chen's algorithm. Note that results on the convergence rate of Chen's algorithm cannot easily be deduced from the theory of classical stochastic algorithms
The purpose of this article is to prove a functional CLT for Chen's algorithm and apply it to averaging algorithms. In the first part of this work, we present the framework and the main results: namely two functional CLTs for this algorithm and a CLT for its averaging version. In the second part, we expose the proof of the functional CLTs throughout a few lemmas, whose proofs are postponed to the third part. Finally, the last section is devoted to the proof of the CLT for moving window averaging algorithms.

2 CLT for Chen's procedure

Let us consider a general problem consisting in finding the root of a continuous function $u\colon \theta \in \mathbb{R}^d \longmapsto u(\theta) \in \mathbb{R}^d$, defined as an expectation on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

$$u(\theta) = \mathbb{E}(U(\theta, Z)), \tag{1}$$

where Z is a random variable in $\mathbb{R}^m$ and U a measurable function defined on $\mathbb{R}^d \times \mathbb{R}^m$.
To solve this problem, one could use a classical stochastic algorithm. However, these procedures often show a rough numerical behaviour, especially when $\mathbb{E}(\|U(\cdot, Z)\|^2)$ grows too quickly.
One can then use alternative procedures, such as the one proposed by Chen in [5], on which we have decided to concentrate in this work. We study both the standard and averaging version of Chen's algorithm.
This technique consists in forcing the algorithm to remain in an increasing sequence of compact sets. Somehow, it prevents the algorithm from blowing up during the "first" steps.
We consider an increasing sequence of compact sets $(\mathcal{K}_j)_j$ such that $\bigcup_{j=0}^{\infty} \mathcal{K}_j = \mathbb{R}^d$. We also introduce $(Z_n)_n$ an independent and identically distributed sequence of random variables following the law of Z and $(\gamma_n)_n$ a decreasing sequence of positive real numbers. For $\theta_0 \in \mathcal{K}_0$ and $\sigma_0 = 0$, we define the sequences of random variables $(\theta_n)_n$ and $(\sigma_n)_n$. First of all, we compute the value of a candidate for θ_{n+1} denoted $\theta_{n+\frac{1}{2}}$

$$\begin{cases} & \theta_{n+\frac{1}{2}} = \theta_n - \gamma_{n+1} U(\theta_n, Z_{n+1}), \\ \text{if } \theta_{n+\frac{1}{2}} \in \mathcal{K}_{\sigma_n} & \theta_{n+1} = \theta_{n+\frac{1}{2}} \quad \text{and} \quad \sigma_{n+1} = \sigma_n, \\ \text{if } \theta_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n} & \theta_{n+1} = \theta_0 \quad \text{and} \quad \sigma_{n+1} = \sigma_n + 1. \end{cases} \tag{2}$$

Remark 1. *When $\theta_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n}$, one can set θ_{n+1} to any measurable function of $(\theta_0, \ldots, \theta_n)$ with values in a given compact set. This existence of such a compact set is definitely essential to proof the a.s. convergence of $(\theta_n)_n$.*

We introduce $\mathcal{F}_n = \sigma(Z_k; k \leq n)$ the σ-field generated by the random vectors Z_k, for $k \leq n$. Note that θ_n is $\mathcal{F}_n-$measurable since θ_0 is deterministic and U measurable. We can write $u(\theta_n) = \mathbb{E}[U(\theta_n, Z_{n+1})|\mathcal{F}_n]$.
It is often more convenient to rewrite (2) as follows

$$\theta_{n+1} = \theta_n - \gamma_{n+1}u(\theta_n) - \gamma_{n+1}\delta M_{n+1} + \gamma_{n+1}p_{n+1} \tag{3}$$

where

$$\begin{aligned} \delta M_{n+1} &= U(\theta_n, Z_{n+1}) - u(\theta_n), && (4)\\ \text{and} \quad p_{n+1} &= \begin{cases} u(\theta_n) + \delta M_{n+1} + \frac{1}{\gamma_{n+1}}(\theta_0 - \theta_n) & \text{if } \theta_{n+\frac{1}{2}} \notin \mathcal{K}_{\sigma_n}, \\ 0 & \text{otherwise.} \end{cases} && (5) \end{aligned}$$

In this paper, we only consider γ_n sequences of the type $\gamma_n = \frac{\gamma}{(n+1)^\alpha}$, with $1/2 < \alpha \leq 1$. Anyway for values of α outside of this range, there is no chance to establish a convergence result for the sequence $(\theta_n)_n$. Depending on the value of α, we obtain two slightly different results.

2.1 Hypotheses

In the following, the prime notation stands for the transpose operator. We need three different kinds of hypotheses. $(\cdot|\cdot)$ denotes the standard Euclidean scalar product on $\mathbb{R}^d$. First, we state a few hypotheses dealing with the behaviour of function u.

Hypothesis 1 (global hypothesis) $\exists!\ \theta^\star \in \mathbb{R}^d, u(\theta^\star) = 0$ *and* $\forall \theta \in \mathbb{R}^d$, $\theta \neq \theta^\star$, $(\theta - \theta^\star|u(\theta)) > 0$.

Hypothesis 2 (local hypothesis) *There exists a function* $y : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ *satisfying* $\lim_{\|x\|\to 0} \|y(x)\| = 0$ *and a symmetric definite positive matrix* A *such that*

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

Hypothesis 3 $\gamma A - \frac{I}{2}$ *is definite positive.*

We also need an hypothesis on the growth of the sequence δM_n in some $\mathbb{L}^{2+\rho}$ space and a convergence result on the covariance matrix of δM_n.

Hypothesis 4 *There exists a real number* $\rho > 0$ *such that*

$$\kappa = \sup_n \mathbb{E}\left(\|\delta M_n\|^{2+\rho}\right) < \infty.$$

Hypothesis 5 *There exists a symmetric definite positive matrix* Σ *such that*

$$\mathbb{E}\left(\delta M_n \delta M_n' | \mathcal{F}_{n-1}\right) \xrightarrow[n\to\infty]{\mathbb{P}} \Sigma.$$

Finally, we require an assumption on the geometry of the $\mathcal{K}_n$ compact sets.

Hypothesis 6 *There exists* $\eta > 0$ *such that* $\forall n \geq 0 \quad d(\theta^\star, \partial\mathcal{K}_n) > \eta$.

Thanks to Hypothesis 4, we can set $\kappa_0 = \sup_n \mathbb{E}\left(\|\delta M_n\|^2\right) < \infty$.

We define the sequence of the normalised iterates for all $n \geq 0$,

$$\Delta_n = \frac{\theta_n - \theta^\star}{\sqrt{\gamma_n}}.$$

We now introduce a sequence of interpolating times $\{t_n(u); u \geq 0, n \geq 0\}$

$$t_n(u) = \sup\left\{k \geq 0 \quad ; \quad \sum_{i=n}^{n+k} \gamma_i \leq u\right\}. \tag{6}$$

with the convention $\sup \emptyset = 0$.
We define $\Delta_n(\cdot)$ as the piecewise constant interpolation of $(\Delta_{n+p})_p$ on intervals of length $(\gamma_{n+p})_p$. More precisely,

$$\Delta_n(0) = \Delta_n \quad \text{ and } \quad \Delta_n(t) = \Delta_{n+t_n(t)+1} \quad \text{for } t \geq 0. \tag{7}$$

This means that for $t \in \left[\sum_{i=n}^{n+p} \gamma_i \ , \ \sum_{i=n}^{n+p+1} \gamma_i\right[$, $t_n(t) = p$ and $\Delta_n(t) = \Delta_{n+p+1}$. $\theta_n(\cdot)$ is defined similarly.
We also introduce $W_n(.)$

$$W_n(0) = 0 \quad \text{and} \quad W_n(t) = \sum_{i=n+1}^{n+t_n(t)+1} \sqrt{\gamma_i}\delta M_i \quad \text{for } t > 0. \tag{8}$$

Remark 2. *The processes $\Delta_n(\cdot)$ and $W_n(\cdot)$ are pure jump càdlàg processes.*

2.2 Main results

For any arbitrary $T > 0$, $\mathbb{D}([0,T])$ stands for the space of processes defined on $[0,T]$ with almost sure right-continuous paths, left-hand limits and values in $\mathbb{R}^d$. We say that a sequence of càdlàg processes X_n converges in law to X or weakly converges in $\mathbb{D}([0,T])$ ($X_n \Longrightarrow X$) in $\mathbb{D}$ if $\mathcal{L}(X_n) \to \mathcal{L}(X)$ weakly in the set of all probability measures defined on $\mathbb{D}([0,T])$. One can refer to [10] or [3] for more details on the weak convergence of càdlàg processes.
In this section, we present three Theorems. The first two state the weak convergence of the sequence of processes $(\Delta_n(\cdot))_n$ in $\mathbb{D}([0,T])$ for any T and the third one is a CLT for an averaging version of Chen's algorithm. Since the limits will be proved not to depend on T, the space $\mathbb{D}([0,T])$ will simply be denoted $\mathbb{D}$.

2.2.1 A CLT for Algorithm (3)

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

Theorem 1. *If we assume Hypotheses 1-2 and 4-6, the sequence of processes $(\Delta_n(\cdot))_n$ converges in law to a diffusion $\Delta(\cdot)$ satisfying*

$$\Delta(t) = \Delta(0)e^{-At} + \int_0^t e^{A(u-t)} dW(u),$$

where $\Delta(0)$ is a random normal variable with mean 0 and variance

$$V = \int_0^\infty e^{-Au} \Sigma e^{-Au} du$$

and W is a Wiener process w.r.t. the smallest $\sigma-$algebra that measures $(\Delta(\cdot), W(\cdot))$ with covariance matrix Σ.

A CLT for $\alpha = 1$

Theorem 2. *If we assume Hypotheses 1-6, the sequence of processes $(\Delta_n(\cdot))_n$ weakly converges in $\mathbb{D}$ to a diffusion $\Delta(\cdot)$ satisfying*

$$\Delta(t) = \Delta(0)e^{-(A-\frac{I}{2\gamma})t} + \int_0^t e^{(A-\frac{I}{2\gamma})(u-t)} dW(u),$$

where $\Delta(0)$ is a random normal variable with mean 0 and variance

$$V = \int_0^\infty e^{-(A-\frac{I}{2\gamma})u} \Sigma e^{-(A-\frac{I}{2\gamma})u} du$$

and W is a Wiener process w.r.t. the smallest $\sigma-$algebra that measures $(\Delta(\cdot), W(\cdot))$ with covariance matrix Σ.

Remark 3. *Hypothesis 3 involves the gradient of function u at the point $\theta^\star$, which is seldom available from a practical point of view but one can definitely not avoid this hypothesis. Remember that this hypothesis was already appearing in the works related to CLTs for classical stochastic algorithms as in [9] for instance. Moreover, considering the expression of V, one immediately understands the importance of this condition. If this hypothesis were not satisfied, the variance V would be non finite, which would basically mean that the rate of convergence is slower than $\sqrt{n}$.*

2.2.2 A CLT for moving window averaging algorithms

In this part, we restrict ourselves to slow decreasing step sequences, namely $\gamma_n = \frac{\gamma}{n^\alpha}$ with $\frac{1}{2} < \alpha < 1$. We assume the Hypotheses of Theorem 2. The sequence $(\theta_n)_n$ converges almost surely to $\theta^\star$ and the associated normalised sequence of processes $(\Delta_n(\cdot))_n$ weakly converges to a stationary Gaussian process in the space of càdlàg processes.
For any $t > 0$, we introduce a moving window average of the iterates

$$\hat{\theta}_n(t) = \frac{1}{t_n(t)} \sum_{i=n}^{n+t_n(t)-1} \theta_i. \tag{9}$$

One can then be tempted to centre this sequence around $\theta^\star$ and normalise it. Naturally, we define

$$\hat{\Delta}_n(t) = \frac{\hat{\theta}_n(t) - \theta^\star}{\sqrt{\gamma_n}} = \frac{1}{\sqrt{\gamma_n} t_n(t)} \sum_{i=n}^{n+t_n(t)-1} (\theta_i - \theta^\star). \tag{10}$$

Theorem 3. *Under the Hypotheses of Theorem 2, the sequence $\hat{\Delta}_n(t)$ converges in distribution to a normally distributed random variable with mean 0 and variance*

$$\hat{V} = \frac{1}{t} A^{-1} \Sigma A^{-1} + \frac{A^{-2}(e^{-At} - I)V + VA^{-2}(e^{-At} - I)}{t^2} \tag{11}$$

where V is defined in Theorem 2.

Remark 4. *Since $t_n(t) \sim \frac{t}{\gamma_n}$ (see (42)), one can replace $t_n(t)$ by its equivalent in the definitions of $(\hat{\theta}_n(t))_n$ and $(\hat{\Delta}_n(t))_n$ in Equations (9) and (10) and Theorem 3 still holds.*

3 Proofs of Theorems 1 and 2

The proofs of the Theorems announced in Section 2.2 will be achieved throughout a series of lemmas whose proofs — mainly based on tightness criteria in Skorokhod's space — are postponed to Section 3.2.

3.1 Technical lemmas

Lemma 1. *There exists $N_0 > 0$ such that if we introduce the following sequence of increasing sets*

$$A_n = \left\{ \sup_{n \geq m > N_0} \|\theta_m - \theta^\star\| < x_0 \right\}, \tag{12}$$

we have

$$\sup_{n \geq N_0} \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n} \right) < \infty. \tag{13}$$

Moreover, the sequence $(\Delta_n)_n$ is tight in $\mathbb{R}^d$.

Remark 5. *Note that $(A_n)_n$ is a decreasing sequence of measurable sets w.r.t $(\mathcal{F}_n)_n$.*

Lemma 2. *For the value of N_0 introduced in Lemma 1, we have for any $T > 0$*

$$\sup_{n \geq N_0} \mathbb{E}\left(\sup_{t \leq T} \|\Delta_n(t)\|^2 \mathbf{1}_{A_{n+t_n(T)}} \right) < \infty.$$

Moreover, the sequence $\left(\sup_{t \leq T} \|\Delta_n(t)\|^2\right)_n$ is tight in $\mathbb{R}^+$.

Lemma 3. *$(W_n(t))_{0 \leq t \leq T}$ converges in law to a process W, which is a Wiener process w.r.t. the filtration it generates with covariance matrix Σ.*

Remark 6. *The proof will show that the limit does not depend on the interval $[0, T]$. So the limit can simply be denoted W.*

Lemma 4 (Aldous' criteria)**.** *For any positive η and ε, there exists $0 < \delta < 1$ such that we have the following inequality*

$$\limsup_n \sup \left\{ \mathbb{P}\left(\|\Delta_n(\tau) - \Delta_n(S)\| \geq \eta\right); \begin{array}{c} S \text{ and } \tau \text{ stopping times in } [0,T], \\ S \leq \tau \leq (S+\delta) \wedge T \end{array} \right\} \leq \varepsilon. \tag{14}$$

Lemma 5. *$(W_n(\cdot), \Delta_n(\cdot))_n$ is tight in $\mathbb{D} \times \mathbb{D}$ and converges in law to (W, Δ) where W is a Wiener process with respect to the smallest $\sigma-$algebra that measures $(W(\cdot), \Delta(\cdot))$ with covariance matrix Σ and Δ is the stationary solution of*

$$d\Delta(t) \quad = \quad -Q\Delta(t)dt - dW(t).$$

3.2 Proofs of the Lemmas

Before proving the different Lemmas, we need a result stating the almost sure convergence of the sequence $(\theta_n)_n$. A proof of the following Proposition can be found in [5] or [7].

Proposition 1. *Under Hypotheses 1 and 4, the sequence $(\theta_n)_n$ converges a.s. to $\theta^\star$ and the sequence $(\sigma_n)_n$ is a.s. finite (i.e. for n large enough $p_n = 0$ a.s.).*

For sake of completeness, a proof of this proposition can be found in Appendix B.

3.2.1 Proof of Lemma 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.
First, we establish a recursive relation

$$\begin{aligned}\Delta_{n+1} &= \frac{\theta_{n+1}-\theta^\star}{\sqrt{\gamma_{n+1}}},\\ &= \frac{1}{\sqrt{\gamma_{n+1}}}\left(\theta_n - \theta^\star - \gamma_{n+1}u(\theta_n) - \gamma_{n+1}\delta M_{n+1} + \gamma_{n+1}p_{n+1}\right),\\ &= \sqrt{\frac{\gamma_n}{\gamma_{n+1}}}\,\Delta_n - \sqrt{\gamma_{n+1}}(u(\theta_n) + \delta M_{n+1} - p_{n+1}).\end{aligned}$$

Using Hypothesis 2, the previous equation becomes

$$\Delta_{n+1} = \left(\sqrt{\frac{\gamma_n}{\gamma_{n+1}}}I - \sqrt{\gamma_{n+1}\gamma_n}(A + y(\theta_n - \theta^\star))\right)\Delta_n - \sqrt{\gamma_{n+1}}(\delta M_{n+1} + p_{n+1}). \tag{15}$$

The following Taylor expansions hold

$$\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} = 1 + \frac{1}{2(n+1)} + \mathcal{O}\left(\frac{1}{n^2}\right) \text{ and } \sqrt{\gamma_n\gamma_{n+1}} = \gamma_n + \mathcal{O}\left(\frac{1}{n^2}\right). \tag{16}$$

We define

$$Q = A - \frac{I}{2\gamma}, \tag{17}$$

which is symmetric definite positive.
This remark enables us to simplify Equation (15) by introducing a new sequence $(\beta_n)_n$ such that for any n larger than some fixed n_0, $|\beta_n| \leq C$, where C is a positive real constant.
Equation (15) can be rewritten as

$$\begin{aligned}\Delta_{n+1} &= \Delta_n - \gamma_n Q\Delta_n - \gamma_n y(\theta_n - \theta^\star)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1}\\ &\quad + \sqrt{\gamma_{n+1}}p_{n+1} + \frac{\beta_n}{(n+1)^2}(B + y(\theta_n - \theta^\star))\Delta_n,\end{aligned} \tag{18}$$

where B is a deterministic matrix.
Let $\Delta_{n+\frac{1}{2}} = \dfrac{\theta_{n+\frac{1}{2}} - \theta^\star}{\sqrt{\gamma_{n+1}}}$.

$$\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \leq \left\|\Delta_n - \gamma_n(Q + y(\theta_n - \theta^\star))\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} + \frac{\beta_n}{(n+1)^2}(B + y(\theta_n - \theta^\star))\Delta_n\right\|^2.$$

Let us take, in the previous equality, the conditional expectation w.r.t $\mathcal{F}_n$ — denoted $\mathbb{E}_n$.

$$\mathbb{E}_n\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2\right) \leq \|\Delta_n\|^2 - 2\gamma_n{\Delta_n}'(Q + y(\theta_n - \theta^\star))\Delta_n + \mathcal{O}\left(\frac{1}{n^2}\right)\|\Delta_n\|^2 + \gamma_{n+1}\mathbb{E}_n(\|\delta M_{n+1}\|^2).$$

The definition of the A_n sets needs to be specified a little.
Since θ_n converges almost surely to $\theta^\star$,

$$\forall\varepsilon > 0,\ \forall\eta > 0, \exists N > 0 \text{ such that } \forall n \geq N\ \mathbb{P}\left(\sup_{m>n}\|\theta_m - \theta^\star\| > \eta\right) < \varepsilon. \tag{19}$$

Let $\lambda > 0$ be the smallest eigenvalue of Q. Since Q is symmetric definite positive, $\lambda > 0$. $\lim_{\|x\|\to 0} y(x) = 0$, so for $x < x_0$, $\|y(x)\| < \lambda/2$. Let $\varepsilon > 0$. Thanks to (19), there exists a rank N_0 — only depending on x_0 and ε — such that $\mathbb{P}(\sup_{m>N_0}\|\theta_m - \theta^\star\| > x_0) < \varepsilon$.

In the definition of the A_n sets (see (12)), we choose N_0 as defined above (and greater than n_0). On the set A_n, $Q + y(\theta_n - \theta^\star)$ is a definite positive matrix with smallest eigenvalue greater than $\lambda/2$. Therefore $\Delta_n{}'(Q + y(\theta_n - \theta^\star))\Delta_n > \lambda/2 \left\|\Delta_n\right\|^2$. We can assume that for $n > N_0$, $\quad \mathcal{O}(\frac{1}{n^2}) \leq \lambda/4\gamma_n$.

$$\begin{aligned}
\mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{A_n}\right) - \mathbb{E}\left(\left\|\Delta_n\right\|^2 \mathbf{1}_{A_n}\right) &\leq -\gamma_n \frac{\lambda}{4}\mathbb{E}\left(\left\|\Delta_n\right\|^2 \mathbf{1}_{A_n}\right) + c\gamma_n, \\
\mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{A_{n+1}}\right) - \mathbb{E}\left(\left\|\Delta_n\right\|^2 \mathbf{1}_{A_n}\right) &\leq -\gamma_n \frac{\lambda}{4}\mathbb{E}\left(\left\|\Delta_n\right\|^2 \mathbf{1}_{A_n}\right) + c\gamma_n, \qquad (20)
\end{aligned}$$

where c is a positive constant.
Now we would like to replace $\Delta_{n+\frac{1}{2}}$ by Δ_{n+1}.

$$\begin{aligned}
\left\|\Delta_{n+1}\right\|^2 &= \frac{\left\|\theta_0 - \theta^\star\right\|^2}{\gamma_{n+1}}\mathbf{1}_{p_{n+1}\neq 0} + \left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{p_{n+1}=0}, \\
\left\|\Delta_{n+1}\right\|^2 &\leq \left\|\Delta_{n+\frac{1}{2}}\right\|^2 + \frac{\left\|\theta_0 - \theta^\star\right\|^2}{\gamma_{n+1}}\mathbf{1}_{\theta_n - \gamma_{n+1}U(\theta_n, Z_{n+1})\notin\mathcal{K}_{\sigma_n}}.
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}_n\left\|\Delta_{n+1}\right\|^2 &\leq \mathbb{E}_n\left\|\Delta_{n+\frac{1}{2}}\right\|^2 + \frac{\left\|\theta_0 - \theta^\star\right\|^2}{\gamma_{n+1}}\mathbb{E}_n\left(\mathbf{1}_{\theta_n - \gamma_{n+1}U(\theta_n, Z_{n+1})\notin\mathcal{K}_{\sigma_n}}\right), \\
\mathbb{E}_n\left\|\Delta_{n+1}\right\|^2 \mathbf{1}_{A_n} &\leq \mathbb{E}_n\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{A_n} + \frac{\left\|\theta_0 - \theta^\star\right\|^2}{\gamma_{n+1}}\mathbf{1}_{A_n}\mathbb{E}_n\left(\mathbf{1}_{\theta_n - \gamma_{n+1}U(\theta_n, Z_{n+1})\notin\mathcal{K}_{\sigma_n}}\right), \\
\mathbb{E}\left(\left\|\Delta_{n+1}\right\|^2 \mathbf{1}_{A_{n+1}}\right) &\leq \mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2 \mathbf{1}_{A_n}\right) + \\
&\qquad \frac{\left\|\theta_0 - \theta^\star\right\|^2}{\gamma_{n+1}}\mathbb{E}\left(\mathbf{1}_{A_n}\mathbb{E}_n\left(\mathbf{1}_{\theta_n - \gamma_{n+1}U(\theta_n, Z_{n+1})\notin\mathcal{K}_{\sigma_n}}\right)\right). \qquad (21)
\end{aligned}$$

The conditional expectation on the right hand side can be rewritten

$$\begin{aligned}
\mathbb{E}_n\left(\mathbf{1}_{\theta_n - \gamma_{n+1}U(\theta_n, Z_{n+1})\notin\mathcal{K}_{\sigma_n}}\mathbf{1}_{A_n}\right) &\leq \mathbb{P}_n\left(\gamma_{n+1}\left\|U(\theta_n, Z_{n+1})\right\| \geq (\theta_n, \partial\mathcal{K}_{\sigma_n})\right)\mathbf{1}_{A_n}, \\
&\leq \frac{\gamma_{n+1}^2}{d\left(\theta_n, \partial\mathcal{K}_{\sigma_n}\right)^2}\mathbb{E}_n\left(\left\|U(\theta_n, Z_{n+1})\right\|^2\right)\mathbf{1}_{A_n}. \qquad (22)
\end{aligned}$$

Moreover, using the triangle inequality we have

$$d\left(\theta_n, \partial\mathcal{K}_{\sigma_n}\right) \geq d\left(\theta^\star, \partial\mathcal{K}_{\sigma_n}\right) - \left\|\theta_n - \theta^\star\right\|.$$

Using Hypothesis 6, $d\left(\theta^\star, \partial\mathcal{K}_{\sigma_n}\right) < \eta$ and on A_n, $\quad \left\|\theta_n - \theta^\star\right\| \leq x_0$. Hence,

$$d\left(\theta_n, \partial\mathcal{K}_{\sigma_n}\right) \geq \eta - x_0.$$

One can choose x_0 smaller than $\eta/2$ for instance, so that $(\eta - x_0)^2 > \frac{\eta^2}{4}$. Thus, Equations (22) becomes

$$\mathbb{E}\left(\mathbf{1}_{\theta_n - \gamma_{n+1}U(\theta_n, Z_{n+1})\notin\mathcal{K}_{\sigma_n}}\mathbf{1}_{A_n}\right) \leq \frac{4\gamma_{n+1}^2}{\eta^2}\mathbb{E}\left(\left\|U(\theta_n, Z_{n+1})\right\|^2 \mathbf{1}_{A_n}\right).$$

Thanks to Hypothesis 4 and the continuity of u, we get

$$\mathbb{E}\left(\left\|U(\theta_n, Z_{n+1})\right\|^2 \mathbf{1}_{A_n}\right) \leq 2\sup_n \mathbb{E}(\left\|\delta M_n\right\|^2) + 2\sup_{\left\|\theta - \theta^\star\right\| < x_0}\mathbb{E}(u(\theta)^2)$$

So, we get

$$\mathbb{E}\left(\mathbf{1}_{\theta_n-\gamma_{n+1}U(\theta_n,Z_{n+1})\notin\mathcal{K}_{\sigma_n}}\mathbf{1}_{A_n}\right) \leq c\gamma_{n+1}^2.$$

Hence, from Equation (21) we can deduce

$$\mathbb{E}\left(\|\Delta_{n+1}\|^2\,\mathbf{1}_{A_{n+1}}\right) \leq \mathbb{E}\left(\left\|\Delta_{n+\frac{1}{2}}\right\|^2\mathbf{1}_{A_n}\right) + c\gamma_n. \tag{23}$$

By combining Equations (23) and (20), we come up with

$$\mathbb{E}\left(\|\Delta_{n+1}\|^2\,\mathbf{1}_{A_{n+1}}\right) \quad\leq\quad \left(1-\gamma_n\frac{\lambda}{4}\right)\mathbb{E}\left(\|\Delta_n\|^2\,\mathbf{1}_{A_n}\right)+c\gamma_n.$$

Let $\mathcal{I}=\left\{i>N_0 : -\frac{\lambda}{4}\mathbb{E}\left(\|\Delta_i\|^2\,\mathbf{1}_{A_i}\right)+c>0\right\}$, then

$$\sup_{i\in\mathcal{I}}\mathbb{E}\left(\|\Delta_i\|^2\,\mathbf{1}_{A_i}\right)<\frac{4c}{\lambda}<\infty.$$

Otherwise for $i\notin\mathcal{I}$,

$$\mathbb{E}\left(\|\Delta_{i+1}\|^2\,\mathbf{1}_{A_{i+1}}\right)-\mathbb{E}\left(\|\Delta_i\|^2\,\mathbf{1}_{A_i}\right)\leq 0.$$

We will prove by recursion that $\forall i\geq N_0,\quad \mathbb{E}\left(\|\Delta_i\|^2\,\mathbf{1}_{A_i}\right)\leq\frac{4c}{\lambda}+\mathbb{E}\left(\|\Delta_{N_0}\|^2\,\mathbf{1}_{A_{N_0}}\right)$. It is obviously true for $i=N_0$. Let us assume that the recursion assumption holds for rank $i>N_0$. If $i+1\in\mathcal{I}$, then $\mathbb{E}\left(\|\Delta_{i+1}\|^2\,\mathbf{1}_{A_{i+1}}\right)\leq\frac{24}{\lambda}$. Otherwise if $i+1\notin\mathcal{I}$, $\mathbb{E}\left(\|\Delta_{i+1}\|^2\,\mathbf{1}_{A_{i+1}}\right)\leq\mathbb{E}\left(\|\Delta_i\|^2\,\mathbf{1}_{A_i}\right)$. So, using the hypothesis of recursion proves the result announced above. Therefore,

$$\sup_n\mathbb{E}\left(\|\Delta_n\|^2\,\mathbf{1}_{A_n}\right)<\infty.$$

In the end, this relation combined with (19) leads to the tightness of the sequence $(\Delta_n)_n$. Let $M>0$.

$$\begin{aligned}\mathbb{P}(\|\Delta_n\|>M) &\leq \mathbb{P}(\|\Delta_n\|(\mathbf{1}_{A_n}+\mathbf{1}_{A_n^c})>M),\\ &\leq \mathbb{P}(\|\Delta_n\|\,\mathbf{1}_{A_n}>M/2)+\mathbb{P}(\|\Delta_n\|\,\mathbf{1}_{A_n^c})>M/2),\\ &\leq 4/M^2\,\mathbb{E}\left(\|\Delta_n\|\,\mathbf{1}_{A_n}^2\right)+\mathbb{P}(A_n^c).\end{aligned}\tag{24}$$

There exists a value of M depending on ε such that both terms on the right hand-side of (24) are bounded above by ε. This proves the tightness of $(\Delta_n)_n$ and ends to prove Lemma 1.

Remark 7 (case $\alpha<1$)**.** *The proof is still valid if $\alpha<1$. In this case the Taylor expansions in Equation (16) become*

$$\sqrt{\frac{\gamma_n}{\gamma_{n+1}}}=1+\mathcal{O}\left(\frac{1}{n}\right)\ \textit{and}\ \sqrt{\gamma_n\gamma_{n+1}}=\gamma_n+\mathcal{O}\left(\frac{1}{n^{1+\alpha}}\right).$$

and Equation (18) is modified in the following way

$$\begin{aligned}\Delta_{n+1} \quad=\quad &\Delta_n-\gamma_nQ\Delta_n-\gamma_ny(\theta_n-\theta^\star)\Delta_n-\sqrt{\gamma_{n+1}}\delta M_{n+1}\\ &+\sqrt{\gamma_{n+1}}p_{n+1}+\frac{\beta_n}{(n+1)}(B+y(\theta_n-\theta^\star))\Delta_n,\end{aligned}$$

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

3.2.2 Proof of Lemma 2

If we go back to equation (18) and sum up this equality from n — chosen greater than N_0 — to $n+p$, we obtain

$$\begin{aligned}\Delta_{n+p} &= \Delta_n - \sum_{k=0}^{p-1} \gamma_{n+k}(\,Q \,+ y(\theta_{n+k}-\theta^\star))\Delta_{n+k} + \sqrt{\gamma_{n+k+1}}\delta M_{n+k+1} \\ &+ \sum_{k=0}^{p-1} \sqrt{\gamma_{n+k+1}} p_{n+k+1} + \frac{\beta_{n+k}}{n+k+1}\gamma_{n+k}(B + y(\theta_{n+k}-\theta^\star))\Delta_{n+k}.\end{aligned}$$

We choose $u>0$ such that $t_n(u)=p$. Since $\theta_n(\cdot)$ is piecewise constant on the subdivision defined by sequence $(\gamma_{n+p})_{p\geq 0}$, the discrete sums can be interpreted as integrals.

$$\Delta_n(u) = \Delta_n(0) - \int_0^u (Q + y(\theta_n(s)-\theta^\star))\,\Delta_n(s)ds - W_n(u) + R_n(u) + P_n(u), \tag{25}$$

where

$$\begin{aligned}P_n(u) &= \sum_{k=0}^{t_n(u)} \sqrt{\gamma_{n+k+1}} p_{n+k+1}, \\ R_n(u) &= \sum_{k=0}^{t_n(u)} \frac{\beta_{n+k}}{n+k+1}\gamma_{n+k}\,(B + y(\theta_{n+k}-\theta^\star))\Delta_{n+k}.\end{aligned}$$

Note that

$$\|R_n(u)\| \leq \frac{C}{n}\int_0^u (1 + \|y(\theta_n(s)-\theta^\star)\|)\,\|\Delta_n(s)\|\,ds.$$

Let $t>0$ and $l = n + t_n(t)$. Note that on the set A_l $P_n(u) = 0\; a.s.$ for all $u\leq t$ and

$$\|y(\theta_n(s)-\theta^\star))\Delta_n(s)\|^2 \mathbf{1}_{A_l} \leq \lambda/2\,\|\Delta_n(s)\|^2.$$

Using equation (25), we will show that $\left(\sup_{0\leq t\leq T}\|\Delta_n(t)\|^2\right)_n$ is tight in $\mathbb{R}$. Let us take the square and then the supremum over $[0,t]$ of Equation (25)

$$\begin{aligned}\sup_{u\leq t}\|\Delta_n(u)\|^2\mathbf{1}_{A_l} &\leq C'\|\Delta_n(0)\|^2\mathbf{1}_{A_l} + C'\,t\int_0^t \sup_{s\leq u}\|\Delta_n(s)\|^2\mathbf{1}_{A_l}du \\ &+ C'\sup_{u\leq t}\|W_n(u)\|^2 + C'\sup_{u\leq t}\|R_n(u)\|^2\mathbf{1}_{A_l}.\end{aligned}\tag{26}$$

The last term is bounded by $Ct\int_0^t \sup_{s\leq u}\|\Delta_n(s)\|^2\mathbf{1}_{A_l}du$.
We define $e_n(t) = \mathbb{E}\left(\sup_{u\leq t}\|\Delta_n(u)\|^2\mathbf{1}_{A_l}\right)$, then taking expectation in (26) gives

$$e_n(t) \leq Ce_n(0) + Ct\int_0^t e_n(u)du + C\,\mathbb{E}\sup_{u\leq t}\|W_n(u)\|^2. \tag{27}$$

Doob's inequality applied to $(W_n(u))_{0\leq u\leq T}$ enables us to rewrite (27)

$$e_n(t) \leq Ce_n(0) + Ct\int_0^t e_n(u)du + 4C\,\mathbb{E}\|W_n(t)\|^2.$$

Thanks to Lemma 1, $\sup_n e_n(0) < \infty$. Hence, $e_n(0)$ can be incorporated into constant C, which remains independent of n.
$\mathbb{E}\|W_n(t)\|^2 = \sum_{i=n}^{n+t_n(t)} \gamma_i \mathbb{E}(\|\delta M_i\|^2)$. So, $\sup_n \mathbb{E}\|W_n(t)\|^2$ is bounded by $\kappa_0 t$. Then, we come up with the following inequality for any $n > N_0$

$$e_n(t) \quad \leq \quad C(1+t) + CT\int_0^t e_n(u)du, \text{ for all } t \text{ in } [0,T],$$

where constant C depends neither on n nor on T.
Using Bellman-Gronwall's inequality, we obtain a key upper-bound for $e_n(t)$

$$e_n(t) \quad \leq \quad C(1+t)e^{CT^2}, \text{ for all } t \text{ in } [0,T] \text{ and } n > N_0.$$

The previous inequality can be summed up as

$$\sup_n \mathbb{E}\left(\sup_{t\leq T}\|\Delta_n(t)\|^2 \mathbf{1}_{A_{n+t_n(T)}}\right) < \infty \text{ for any } T. \tag{28}$$

Equation (28) implies that $\left(\sup_{t\leq T}\|\Delta_n(t)\|^2\right)_n$ is tight in $\mathbb{R}$.
From now on we define $\bar{e} = \sup_n \mathbb{E}\left(\sup_{t\leq T}\|\Delta_n(t)\|^2 \mathbf{1}_{A_{n+t_n(T)}}\right)$.

3.2.3 Proof of Lemma 3

To prove Lemma 3, we first prove that $(W_n(\cdot))_n$ is $\mathcal{C}-$tight and that $(\sup_{t\in[0,T]}\|W_n(t)\|)_n$ is tight in $\mathbb{R}$. These two points imply that $(W_n(\cdot))_n$ is tight in $\mathbb{D}[0,T]$ and that every converging subsequence converges in law to a continuous process. Then, we prove that any such limit is a martingale. Finally, we establish that these limit martingales have predictable quadratic variation equal to Σt. Thanks to Lévy's Theorem[1], combining these two points imply that W is a Wiener process with covariance matrix Σ.

Tightness of $(W_n(\cdot))_n$ in $\mathbb{D}[0,T]$. We have already seen that $\sup_n E\left(\|W_n(t)\|^2\right) \leq \kappa_0 t$, so the family $\{W_n(t); n \geq 1\}$ is uniformly integrable for each t in $[0,T]$.
Moreover using Doob's inequality, it comes that

$$E\left(\sup_{t\in[0,T]}\|W_n(t)\|^2\right) \leq 4\sum_{i=n}^{n+t_n(T)} \gamma_i\,\mathbb{E}\left(\|\delta M_i\|^2\right) \leq 4\kappa_0 T.$$

Thus, $\left(\sup_{t\in[0,T]}\|W_n(t)\|\right)_n$ is tight in $\mathbb{R}$.
We want to prove that the sequence of processes $(W_n(t))_{0\leq t\leq T}$ satisfies a $\mathcal{C}-$tightness criterion. It suffices to show that there exist two positive real numbers α and β such that for any (t,s) in $[0,T]^2$ the following inequality holds

$$\mathbb{E}(\|W_n(t) - W_n(s)\|^\alpha) \leq \kappa\,|t-s|^{1+\beta}.$$

[1] see [14, p. 86] for instance.

Let us choose a couple (s,t) in $[0,T]^2$ such as $s < t$ and an $\alpha > 0$. Using Burkholder-Davis-Gundy's inequality[2] we can write

$$\begin{aligned}
\mathbb{E}\left(\|W_n(t) - W_n(s)\|^\alpha\right) &= \mathbb{E}\left(\left\|W_{n+t_n(t)} - W_{n+t_n(s)}\right\|^\alpha\right), \\
&\leq \kappa\, \mathbb{E}\left\|\left[W_{n+t_n(.)} - W_{n+t_n(s)}\right]_t\right\|^{\alpha/2}, \\
&\leq \kappa\, \mathbb{E}\left\|\sum_{i=n+t_n(s)}^{n+t_n(t)} \gamma_i\, \delta M_i \delta M_i'\right\|^{\alpha/2}, \qquad (29)
\end{aligned}$$

Now, we use a well-known inequality for convex functions, assuming that $\alpha > 2$.

$$\begin{aligned}
\left\|\sum_{i=n+t_n(s)}^{n+t_n(t)} \gamma_i\, \delta M_i \delta M_i'\right\|^{\alpha/2} &\leq |t-s|^{\alpha/2} \left\|\sum_{i=n+t_n(s)}^{n+t_n(t)} \frac{\gamma_i}{t-s}\, \|\delta M_i\|^2\right\|^{\alpha/2}, \\
&\leq |t-s|^{\alpha/2-1} \sum_{i=n+t_n(s)}^{n+t_n(t)} \gamma_i\, \|\delta M_i\|^\alpha .
\end{aligned}$$

Thus, the expectation on the right hand side of (29) is bounded by

$$|t-s|^{\alpha/2-1} \sum_{i=n+t_n(s)}^{n+t_n(t)} \gamma_i\, \mathbb{E}(\|\delta M_i\|^\alpha).$$

We choose $\alpha = 2+\rho$ — ρ being defined in Hypothesis 4 — to obtain the desired inequality

$$\mathbb{E}\left(\|W_n(t) - W_n(s)\|^\alpha\right) \leq \kappa\, |t-s|^{\rho/2+1}.$$

The tightness of $(W_n(0)_n$ is given by its uniform square integrability. Thus, the sequence of processes $(W_n(\cdot))_n$ is $\mathcal{C}$-tight. Moreover, thanks to Lemma 2 $(\sup_{t\in[0,T]} \|W_n(t)\|)_n$ is tight in $\mathbb{R}$. Hence, $(W_n(\cdot))_n$ is tight in $\mathbb{D}$.

Any converging subsequence converges in law to a continuous martingale. Let $(W_n(\cdot))_n$ denote a converging subsequence with limit W. We will show that W is a continuous martingale. Since $(W_n(\cdot))_n$ is $\mathcal{C}-$tight, W is a continuous process.
For any $A > 0$, we define the continuous function f_A such that $f_A(x) = x$ if $\mathbf{1}_{\|x\|\leq A}$ and $f_A(x) = 0$ if $\mathbf{1}_{\|x\|\geq A+1}$. Therefore, f_A is a continuous bounded function. We have for all $n > 0$

$$\kappa_0 t \geq \mathbb{E}(\|W_n(t)\|^2) \geq \mathbb{E}(f_A(\|W_n(t)\|^2)).$$

Thanks to the convergence of $(W_n(\cdot))_n$, we get[3]

$$\kappa_0 t \geq \mathbb{E}(f_A(\|W(t)\|^2)).$$

$f_A(\|W(t)\|^2)$ is non decreasing w.r.t to A and positive so using the monotone convergence Theorem, we obtain

$$\kappa_0 t \geq \mathbb{E}(\|W(t)\|^2). \qquad (30)$$

[2]see [14, Theorem 48] for càdlàg martingales.

[3] In fact, we also need the continuity of $\omega \in \mathbb{D} \longmapsto \omega(t)$ on a set of measure 1 for the law W. $\omega \in \mathbb{D} \longmapsto \omega(t)$ is continuous for the topology on $\mathbb{D}$ at every point α such that α does not jump at time t. Therefore, the coordinate applications are continuous on the set of continuous paths which is of measure 1 for the law of W because W is a.s. continuous. Hence, $\omega \in \mathbb{D} \longmapsto f_A(\|\omega(t)\|)$ is continuous for the topology on $\mathbb{D}$ on a set of measure 1 for the limit law.

This proves that W is square integrable.
Let h be a continuous bounded function on $\mathbb{D}$. Since $W_n(\cdot)$ converges in law in $\mathbb{D}$ to W, we have for all $0 < s < t \leq T$

$$\mathbb{E}\left[h(W_n(u); u \leq s) f_A(W_n(t) - W_n(s))\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[h(W(u); u \leq s) f_A(W(t) - W(s))\right]. \quad (31)$$

Since W is a.s. continuous, $\lim_{A\to\infty} f_A(W(t) - W(s)) = W(t) - W(s)$. Thanks to (30), we can use the bounded convergence Theorem to show that the expectation on the right hand side of (31) tends to $\mathbb{E}\left[h(W(u); u \leq s)(W(t) - W(s))\right]$ when A goes to infinity.
Thanks to the uniform integrability of $(W_n(t))_n$ for each fixed t,

$$\sup_n \mathbb{E}\left[h(W_n(u); u \leq s)\left\{f_A(W_n(t) - W_n(s)) - (W_n(t) - W_n(s))\right\}\right] \xrightarrow[A\to\infty]{} 0.$$

So,

$$\mathbb{E}\left[h(W_n(u); u \leq s)(W_n(t) - W_n(s))\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[h(W(u); u \leq s)(W(t) - W(s))\right]. \quad (32)$$

Since for any fixed n $(W_n(t))_t$ is a martingale, $\mathbb{E}\left[h(W_n(u); u \leq s)(W_n(t) - W_n(s))\right] = 0$. Then, we come up with $\mathbb{E}\left(h(W(u); u \leq s)(W(t) - W(s))\right) = 0$, which proves that W is a martingale w.r.t. the filtration it generates.

Any limit W has predictable quadratic variation equal to Σt. Since the predictable quadratic variation process is unique, up to an evanescent set, it is sufficient to prove that $(W_t W_t' - \Sigma t)_t$ is a martingale.
As $\mathbb{E}\left\|W_n(t)\right\|^{2+\rho}$ is uniformly bounded in n, $(W_n(t)W_n(t)')_n$ is uniformly integrable for each fixed t. So using truncation functions as above, it is straightforward to prove that W is square integrable. Moreover,

$$\begin{aligned}
\mathbb{E}\left(\left\|\langle W_n, W_n\rangle_t\right\|^{1+\rho/2}\right) &\leq \mathbb{E}\left(\left(\sum_{i=n}^{n+t_n(t)} \gamma_i \mathbb{E}(\left\|\delta M_{i+1}\right\|^2 |\mathcal{F}_i)\right)^{1+\rho/2}\right), \\
&\quad \text{using a convexity inequality, we get} \\
&\leq \mathbb{E}\left(t^{\rho/2} \sum_{i=n}^{n+t_n(t)} \gamma_i \mathbb{E}(\left\|\delta M_{i+1}\right\|^{2+\rho})\right), \\
&\leq t^{1+\rho/2}\kappa.
\end{aligned}$$

So, $(\langle W_n, W_n\rangle_t)_n$ is uniformly integrable for each t.
Let h be a continuous bounded function on $\mathbb{D}$. Using Hypothesis 5, we can see that $(\langle W_n, W_n\rangle_t)$ tends in probability to Σt and thanks to the uniform integrability, the convergence also occurs in $\mathbb{L}^1$. Hence,

$$\lim_n \mathbb{E}\left[h(W_n(u); u \leq s)(\langle W_n, W_n\rangle_t - \Sigma t)\right] = 0. \quad (33)$$

Since $W_n(t)$ is martingale for any fixed n,

$$\begin{aligned}
&\mathbb{E}\left[h(W_n(u); u \leq s)(W_n(t)W_n(t)' - \langle W_n, W_n\rangle_t)\right] = \\
&\qquad\qquad \mathbb{E}\left[h(W_n(u); u \leq s)(W_n(s)W_n(s)' - \langle W_n, W_n\rangle_s)\right].
\end{aligned}$$

Once again, we use truncation functions. Since $(W_n(\cdot))_n$ converges in law in $\mathbb{D}$ to W and $(W_n(t)W_n(t)')_n$ is uniformly integrable for each t, $\mathbb{E}\left[h(W_n(u); u \leq s)(W_n(t)W_n(t)' - \Sigma t)\right]$ tends to $\mathbb{E}\left[h(W(u); u \leq s)(W(t)W(t)' - \Sigma t)\right]$.

Consequently using Equation (33), we get

$$\mathbb{E}\left[h(W(u); u \leq s)(W(t)W(t)' - \Sigma t)\right] = \mathbb{E}\left[h(W(u); u \leq s)(W(s)W(s)' - \Sigma s)\right].$$

Thus, $(W(t)W(t)' - \Sigma t)_t$ is a martingale. Since the predictable quadratic variation process is unique, up to an evanescent set, $\langle W, W\rangle_t = \Sigma t$ a.s.. Moreover W is continuous, so Lévy's characterisation of the Wiener process proves that W is a Wiener process with covariance matrix Σ.

Hence, any converging subsequence of $(W_n(\cdot))_n$ converge to a Wiener process with covariance matrix Σ, which implies that the whole converges in law to that process W.

3.2.4 Proof of Lemma 4

Let us choose some fixed positive η, ε and a corresponding δ. S and τ stands for two stopping times as introduced in Lemma 4. Let $l = n + t_n(T)$.

$$\mathbb{P}\left(\|\Delta_n(\tau) - \Delta_n(S)\| \geq 2\eta\right) \leq \mathbb{P}\left(\|\Delta_n(\tau) - \Delta_n(S)\mathbf{1}_{A_l}\| \geq \eta\right) + \mathbb{P}(A_l^c).$$

Remember that $\mathbb{P}(A_l^c) \leq \varepsilon$.

$$\begin{aligned}\mathbb{P}\left(\|\Delta_n(\tau) - \Delta_n(S)\mathbf{1}_{A_l}\| \geq \eta\right) \leq \mathbb{P}\left(\left\|\int_S^\tau \left(Q - y(\theta_n(u) - \theta^\star)\right)\Delta_n(u)\mathbf{1}_{A_l}du\right\| \geq \frac{\eta}{6}\right) \\ + \ \mathbb{P}\left(\|W_n(\tau) - W_n(S)\| \geq \frac{\eta}{6}\right) + \mathbb{P}\left(\|R_n(\tau) - R_n(S)\|\,\mathbf{1}_{A_l} \geq \frac{\eta}{6}\right). \qquad (34)\end{aligned}$$

On the set A_l, $P_n(u) = 0$ a.s. for all $u \leq T$.

The first term is handled using Markov's inequality

$$\begin{aligned}&\mathbb{P}\left(\left\|\int_S^\tau \left(Q - y(\theta_n(u) - \theta^\star)\right)\Delta_n(u)du\mathbf{1}_{A_n}\right\| \geq \frac{\eta}{6}\right)\\ &\qquad \leq \frac{c}{\eta^2}\mathbb{E}\left(\delta\int_S^{S+\delta}\|\Delta_n(u)\mathbf{1}_{A_l}\|^2 du\right),\\ &\qquad \leq \frac{c}{\eta^2}\mathbb{E}\left(\delta\int_0^T\|\Delta_n(u)\mathbf{1}_{A_l}\|^2 du\right),\\ &\qquad \leq \frac{c\delta}{\eta^2}\bar{e}\,T,\\ &\qquad \leq \frac{c}{\eta^2}K, \text{ where } c \text{ is a positive constant only depending on } T.\end{aligned}$$

The third term can be treated like the first one.

Now, we will apply Burkholder-Davis-Gundy's inequality to the stopped martingale

$(W_n(t) - W_n(t \wedge S))_t$.

$$
\begin{aligned}
\mathbb{E}\left\|W_n(\tau) - W_n(S)\right\|^{2+\rho} &\leq \mathbb{E}\left(\sum_{i=n+t_n(S)}^{n+t_n(S+\delta)} \gamma_i \left\|\delta M_i\right\|^2\right)^{1+\rho/2}, \\
&\text{using a convexity inequality, we obtain} \\
&\leq \delta^{1+\rho/2}\mathbb{E}\left(\sum_{i=n+t_n(S)}^{n+t_n(S+\delta)} \frac{\gamma_i}{\delta} \left\|\delta M_i\right\|^{2+\rho}\right), \\
&\leq \delta^{\rho/2}\left(\sum_{i=n}^{n+t_n(T)} \gamma_i \, \mathbb{E}(\left\|\delta M_i\right\|^{2+\rho})\right).
\end{aligned}
$$

Using hypothesis 4, we come up with the following upper-bound

$$
\mathbb{E}\left\|W_n(\tau) - W_n(S)\right\|^{2+\rho} \leq \delta^{\rho/2} T \sup_i \mathbb{E}(\left\|\delta M_i\right\|^{2+\rho}).
$$

Finally, we obtain a new upper bound in (34)

$$
\mathbb{P}\left(\left\|\Delta_n(\tau) - \Delta_n(S)\right\| \geq \eta\right) \leq \delta^{\rho/2}\left(\frac{C_1}{\eta^2} + \frac{C_2}{\eta^{2+\rho}}\right) + \varepsilon, \tag{35}
$$

where C_1 and C_2 are two positive constants independent of S, τ, n and η. Assuming that $\eta < 1$, Equation (35) becomes

$$
\mathbb{P}\left(\left\|\Delta_n(\tau) - \Delta_n(S)\right\| \geq \eta\right) \leq \delta^{\rho/2}\frac{C}{\eta^{2+\rho}} + \varepsilon, \tag{36}
$$

where C is a positive constant.
Choosing $\delta = (\varepsilon\eta^{2+\rho})^{1/\rho}$ shows that property (14) holds true. Since $\left(\sup_{t\in[0,T]}\left\|\Delta_n(t)\right\|\right)_n$ is tight, Equation (36) ends to prove that $(\Delta_n(\cdot))_n$ is tight in $\mathbb{D}$.

3.2.5 Proof of Lemma 5

$(W_n(\cdot))_n$ is $C-$tight and $(\Delta_n(\cdot))_n$ is tight, so it is quite straightforward[4] that the couple $(W_n(\cdot), \Delta_n(\cdot))_n$ is tight in $\mathbb{D}\times\mathbb{D}$. For a proof of the result, one can see [10, Corollary 3.33, page 317].

Thus, we can extract a converging subsequence $(W_{\phi(n)}(\cdot), \Delta_{\phi(n)}(\cdot))$ with limit $(W^\phi(\cdot), \Delta^\phi(\cdot))$. We will prove that in Equation (25), $\left(\sup_{0\leq u\leq T}\left\|R_{\phi(n)}(u)\right\|\right)_n$ and $\left(\sup_{0\leq u\leq T}\left\|P_{\phi(n)}(u)\right\|\right)_n$ tend to zero in probability.
Let $\eta > 0$ and $l = \phi(n) + t_{\phi(n)}(T)$.

$$
\mathbb{P}\left(\sup_{0\leq u\leq T}\left\|R_{\phi(n)}(u)\right\| > \eta\right) \leq \mathbb{P}\left(\frac{C}{\phi(n)}\int_0^T (1 + \left\|y(\theta_{\phi(n)}(u) - \theta^\star)\right\|)\left\|\Delta_{\phi(n)}(u)\right\| du > \eta\right).
$$

We split the probability on the right hand side on the sets A_l and A_l^c.

$$
\mathbb{P}\left(\frac{C}{\phi(n)}\int_0^T (1 + \left\|y(\theta_{\phi(n)}(u) - \theta^\star)\right\|)\left\|\Delta_{\phi(n)}(u)\right\| \mathbf{1}_{A_l^c} du > \frac{\eta}{2}\right) \leq \mathbb{P}(\mathbf{1}_{A_l^c}).
$$

[4] The pseudo continuity modulus, w', on $\mathbb{D}$ has no linearity property, but for any $\alpha, \beta \in \mathbb{D}$, any $\delta > 0$ we have $w'(\alpha+\beta, \delta) \leq w'(\alpha, \delta) + w(\beta, 2\delta)$, where w is the continuity modulus.

Now, we tackle the probability on A_l.

$$\begin{aligned}
\mathbb{P}\left(\frac{C}{\phi(n)}\int_0^T (1+\left\|y(\theta_{\phi(n)}(u)-\theta^\star)\right\|)\left\|\Delta_{\phi(n)}(u)\right\|\mathbf{1}_{A_l}du > \frac{\eta}{2}\right) \\
&\le \frac{c}{\eta^2\phi(n)^2}\mathbb{E}\left(\int_0^T (1+\left\|y(\theta_{\phi(n)}(u)-\theta^\star)\right\|)\left\|\Delta_{\phi(n)}(u)\right\|\mathbf{1}_{A_l}du\right), \\
&\le \frac{c}{\eta^2\phi(n)^2}\mathbb{E}\left(\int_0^T \left\|\Delta_{\phi(n)}(u)\right\|\mathbf{1}_{A_l}du\right), \\
&\le \frac{c}{\eta^2\phi(n)^2}T\bar{e}. \qquad (37)
\end{aligned}$$

For n large enough, the term on the right hand side of (37) can be made smaller than ε. Then, it is then clear that for n large enough

$$\mathbb{P}\left(\sup_{0\le u\le T}\left\|R_{\phi(n)}(u)\right\| > \eta\right) \le 2\varepsilon.$$

The term $\mathbb{P}\left(\sup_{0\le u\le T}\left\|P_{\phi(n)}(u)\right\| > \eta\right)$ can also be treated by splitting the probability on A_l and its complementary set. Recall that on the set A_l, $P_n(u)=0$ a.s. for all $u\le T$. Hence P_n and R_n both converge to the zero process in $\mathbb{D}$.
Remember that the integral is a continuous application from $\mathbb{D}$ into $\mathbb{R}$. More precisely for any real numbers a and b in $[0,T]$, the application $\omega\in\mathbb{D}\longmapsto\int_a^b\omega(t)dt$ is continuous.
Thanks to Lemma 3, $W^\phi(\cdot)$ is a Wiener process with covariance matrix Σ. Hence, the limit of $W_{\phi(n)}(\cdot)$ is independent of ϕ. So, letting n go to infinity in (25) enables to show that the limit $\Delta^\phi(\cdot)$ satisfies the following equation

$$\Delta^\phi(t) = \Delta^\phi(0) - \int_0^t Q\Delta^\phi(u)du - W(t), \qquad (38)$$

which is equivalent to

$$d\Delta^\phi(t) = -Q\Delta^\phi(t)dt - dW(t).$$

Equation (38) shows that the set of all possible limits of any converging subsequence of $(\Delta_n(\cdot))_n$ is a family of Ornstein Uhlenbeck processes indexed by their initial conditions. So, if we manage to prove that the set $\{\Delta^\phi(0);\phi \text{ such that } \Delta_{\phi(n)}(\cdot) \text{ converges}\}$ is reduced to a single point, we will have stated the convergence of the whole sequence $(\Delta_n(\cdot))_n$ and not only of a subsequence. Any limit $\Delta(\cdot)$ satisfies

$$\Delta(t) = e^{-Qt}\Delta(0) - \int_0^t e^{Q(u-t)}dW(u).$$

The stochastic integral converges in distribution to a random normal variable with mean 0 and covariance matrix $\int_0^\infty e^{-Qu}\Sigma e^{-Qu}du$ as t goes to infinity. So does the process Δ since the set of all possible laws for $\Delta(0)$ is tight and e^{-Qt} tends to zero when t goes to infinity. This limit happens to be the unique stationary law for the Δ process.
Now we want to prove that the set of all possible laws for $\Delta(0)$ is reduced to the stationary law described above. The way we prove it is widely inspired from [2] and [4].

Stationarity of any limit. Let $\mathsf{v}=\{\text{possible laws for } \Delta(0)\}$. v is a weakly compact set. For any $\nu\in\mathsf{v}$, let $P_\nu(t)$ denotes the law at time t of the process $\Delta(\cdot)$ of initial law ν. Let f be a continuous bounded function on $\mathbb{R}^d$ and ν_g be the stationary law described above.

Let us choose an $\varepsilon > 0$. Since $\boldsymbol{\nu}$ is weakly compact, there exists a $T > 0$ such that

$$|\langle f, P_\nu(t)\rangle - \langle f, \nu_g\rangle| \leq \varepsilon \tag{39}$$

for any $t > T$ and $\nu \in \boldsymbol{\nu}$.
We fix such a $T > 0$ and choose $\nu \in \boldsymbol{\nu}$. We can extract a converging subsequence $\left(\Delta_{\phi(n)}(\cdot)\right)_n$ such that ν is the initial laws of the limit.
We define

$$\psi(n) = \inf\left\{k \geq 0; \sum_{i=k}^{\phi(n)} \gamma_i \leq T\right\}. \tag{40}$$

For n large enough, $\psi(n) > 0$ which means that $\psi(n) + t_{\psi(n)}(T) = \phi(n)$ and ψ is an increasing function. Hence we have the equality $\Delta_{\phi(n)}(0) = \Delta_{\psi(n)}(T)$. We can extract one more subsequence such that $\Delta_{\psi(\psi'(n))}(\cdot)$ converges. If ν' denotes the initial law of the limit, we have

$$|\langle f, \nu\rangle - \langle f, \nu_g\rangle| = |\langle f, P_{\nu'}(T)\rangle - \langle f, \nu_g\rangle| \leq \varepsilon.$$

The last part of the inequality comes from (39). This proves that $\nu = \nu_g$. Henceforth, any converging subsequence of $\Delta_n(\cdot)$ converges to a stationary Ornstein Uhlenbeck process.

W is a $\mathcal{F}^{\Delta,W}-$martingale. The only remaining point to prove is that $\Delta(0)$ is independent of $\sigma(W(t); t > 0)$. This is the same as proving that W is a $\mathcal{F}^{\Delta,W}-$Wiener process, where $\mathcal{F}_t^{\Delta,W}$ is the smallest $\sigma-$algebra that measures $\{\Delta(s), W(s); s \leq t\}$.
Since we already know that W is continuous and that $\langle W, W\rangle_t = t$ a.s., it is sufficient to prove that W is a $\mathcal{F}^{\Delta,W}-$martingale.
Let h be a continuous bounded function on $\mathbb{D}$. Since $(W_n, \Delta_n) \Longrightarrow (W, \Delta)$ and $(W_n(t))_n$ is uniformly integrable for each t, it is quite obvious that

$$\mathbb{E}\left(h((\Delta_n(s), W_n(s); s \leq t)(W_n(t+\tau) - W_n(t))\right) \xrightarrow[n\to\infty]{} \mathbb{E}\left(h((\Delta(s), W(s); s \leq t))(W(t+\tau) - W(t))\right).$$

$W_n(\cdot)$ is a $\mathcal{F}_{n+t_n(\cdot)}$ martingale and $\Delta_n(\cdot)$ is measurable with respect to the shifted filtration. Hence,

$$\mathbb{E}\left(h((\Delta_n(s), W_n(s); s \leq t)(W_n(t+\tau) - W_n(t))\right) = 0.$$

Consequently,

$$\mathbb{E}\left(h((\Delta(s), W(s); s \leq t))(W(t+\tau) - W(t))\right) = 0.$$

This last equality implies that W is a $\mathcal{F}^{\Delta,W}-$Wiener process.

4 Proof of Theorem 3

One can rewrite $\hat{\Delta}_n(t)$

$$\hat{\Delta}_n(t) = \frac{1}{t_n(t)} \sum_{i=n}^{n+t_n(t)-1} \Delta_i \gamma_i \frac{1}{\sqrt{\gamma_i \gamma_n}}$$

Using equation (6), it is clear that $t \underset{n}{\sim} \gamma \int_n^{n+t_n(t)} x^{-\alpha} dx$, since $\gamma_n = \frac{\gamma}{n^\alpha}$ with $\frac{1}{2} < \alpha < 1$.
Hence,

$$t \underset{n}{\sim} \frac{\gamma}{1-\alpha} n^{1-\alpha}\left(\left(1 + \frac{t_n(t)}{n}\right)^{1-\alpha} - 1\right). \tag{41}$$

Thanks to Equation (6), it is quite easy to see that $\frac{t_n(t)}{(n+t_n(t))^\alpha} < t$. So, $\frac{t_n(t)}{n} \longrightarrow 0$. Henceforth using Taylor expansions, Equation (41) can be rewritten

$$t \underset{n}{\sim} \frac{\gamma}{1-\alpha} n^{1-\alpha}(1-\alpha)\frac{t_n(t)}{n} \underset{n}{\sim} \gamma_n t_n(t). \tag{42}$$

Moreover, for $i \in \{n, \ldots, n+t_n(t)\}$ if we write $i = n+p$, we can use the following Taylor expansions

$$\frac{1}{\sqrt{\gamma_i \gamma_n}} = \frac{1}{\gamma}\sqrt{n^\alpha\ (n+p)^\alpha} = \frac{1}{\gamma} n^\alpha \left(1 + \frac{p}{n}\right)^{\alpha/2} \underset{n}{\sim} \frac{1}{\gamma_n}$$

Finally, we obtain an equivalent for $\hat{\Delta}_n(t)$

$$\hat{\Delta}_n(t) \sim \frac{1}{t_n(t)\gamma_n} \sum_{i=n}^{n+t_n(t)-1} \Delta_i \gamma_i.$$

Thanks to (42), we get

$$\hat{\Delta}_n(t) \sim \frac{1}{t} \sum_{i=n}^{n+t_n(t)-1} \Delta_i \gamma_i = \frac{1}{t}\int_0^t \Delta_n(u) du. \tag{43}$$

Using Theorem 2, we know that $\Delta_n(\cdot)$ converges to $\Delta(\cdot)$ in the space of càdlàg processes. Since the integral over a finite time interval is a continuous function on this space, the integral in (43) converges to $\frac{1}{t}\int_0^t \Delta(u)du$.
Since $\Delta(\cdot)$ is a Gaussian process, the integral is a normally distributed random variable with mean 0 and variance $\hat{V}$.

$$\begin{aligned}\hat{V} &= \frac{1}{t^2}\operatorname{Cov}\left(\int_0^t \Delta(u)du, \int_0^t \Delta(s)ds\right), \\ &= \frac{1}{t^2}\int_0^t\int_0^t \operatorname{Cov}(\Delta(u), \Delta(s)) du\, ds,\end{aligned}$$

Let us define $\Gamma_s(\tau) = \operatorname{Cov}(\Delta(s+\tau), \Delta(s))$. Thanks to the definition of the process $\Delta(\cdot)$, it is easy to show that

$$\operatorname{Cov}(\Delta(s+\tau), \Delta(s)) = e^{-A\tau} \operatorname{Cov}(\Delta(s), \Delta(s)) \tag{44a}$$

$$\operatorname{Cov}(\Delta(s), \Delta(s+\tau)) = \operatorname{Cov}(\Delta(s), \Delta(s))\, \mathrm{e}^{-A\tau}. \tag{44b}$$

Since $\Delta(\cdot)$ is a stationary process, $\operatorname{Cov}(\Delta(s), \Delta(s)) = \operatorname{Cov}(\Delta(0), \Delta(0))$ for any $s \geq 0$. Henceforth, Equations (44a) and (44b) can be rewritten

$$\begin{aligned}\operatorname{Cov}(\Delta(s+\tau), \Delta(s)) &= e^{-A\tau} V \\ \operatorname{Cov}(\Delta(s), \Delta(s+\tau)) &= V\, \mathrm{e}^{-A\tau}.\end{aligned}$$

Let us go back to the computation of $\hat{V}$. Since Cov is a bilinear operator,

$$\begin{aligned}\hat{V} &= \frac{1}{t^2}\left(\int_0^t\int_0^t \operatorname{Cov}(\Delta(u), \Delta(s))\mathbf{1}_{u\leq v} du\, dv + \int_0^t\int_0^t \operatorname{Cov}(\Delta(u), \Delta(s))\mathbf{1}_{v\leq u} du\, dv\right), \\ &= \frac{1}{t^2}\left(\int_0^t\int_0^v V\, \mathrm{e}^{-A(v-u)}\, du\, dv + \int_0^t\int_0^u \mathrm{e}^{-A(u-v)}\, V dv\, du\right), \\ &= \frac{1}{t^2}\left\{V\left(A^{-1}t + A^{-2}[\mathrm{e}^{-At} - I]\right) + \left(A^{-1}t + A^{-2}[\mathrm{e}^{-At} - I]\right)V\right\}, \\ &= \frac{1}{t}\left(VA^{-1} + A^{-1}V\right) + \frac{1}{t^2}\left\{V\left(A^{-2}[\mathrm{e}^{-At} - I]\right) + \left(A^{-2}[\mathrm{e}^{-At} - I]\right)V\right\}. \end{aligned} \tag{46}$$

Moreover considering the definition of $V = \int_0^\infty \mathrm{e}^{-Au}\,\Sigma\,\mathrm{e}^{-Au}\ du$, it is quite easy to show that V solves the following Ricatti equation

$$AV + VA = \Sigma. \tag{47}$$

One can even prove that V is the unique solution of (47). From (47), one can deduce that $A^{-1}V + VA^{-1} = A^{-1}\Sigma A^{-1}$. Plugging this last result back into (46) gives the result announced in Theorem 3.

5 Conclusion

Acknowledgements. I wish to address my deepest gratitude to A. Alfonsi and B. Lapeyre for the fruitful discussions we had on càdlàg processes and the comments he made on a previous version of this work.

A A criterion for Sequence (3) to be a.s. compact

The following lemma gives a sufficient condition for the sequence defined by (3) to remain in a given compact set.

Lemma 6. *Let us consider the sequence $(\theta_n)_n$ defined by (3). If we assume that the series $\sum_{n>0} \gamma_n \delta M_n \mathbf{1}_{\|\theta_{n-1}-\theta^\star\|<M}$ converges and $p_n \mathbf{1}_{\|\theta_{n-1}-\theta^\star\|<M} \longrightarrow 0$ for any $M > 0$, then the sequence $(\theta_n)_n$ remains in a given compact set.*

Proof. If $\sigma_n < \infty$ a.s., then the conclusion of the Lemma is obvious. Let us suppose that $\sigma_n \longrightarrow \infty$. Since each time σ_n increases, the sequence θ_n is reinitialised to some fixed point in $\mathcal{K}_0$, the existence of a compact set $\mathcal{C}$ in which the sequence $(\theta_n)_n$ lies infinitely often is straightforward.
Let M be such that $\mathcal{C} \subset \{\theta\,;\|\theta - \theta^\star\|^2 \leq M\}$.
We can rewrite the hypotheses of the Lemma as

$$\forall \varepsilon > 0,\ \exists\, N > 0 \text{ such as } \forall\, n,\ p \geq\ N \text{ we have } \begin{cases} \left\| \sum_{k=n}^{p} \gamma_k \delta M_k \mathbf{1}_{\|\theta_{k-1}-\theta^\star\|^2 \leq M+2} \right\| < \ \varepsilon, \\ \gamma_n < \varepsilon, \\ \mathbf{1}_{\|\theta_{n-1}-\theta^\star\|^2 < M+2} \|p_n\| < \varepsilon. \end{cases}$$

Let be $\varepsilon > 0$ and a N satisfying the relation above and also $\theta_N \in \mathcal{C}$. We define

$$\theta'_n = \theta_n - \sum_{i=n+1}^{\infty} \gamma_i \delta M_i \mathbf{1}_{\|\theta_{i-1}-\theta^\star\|^2 \leq M+2}.$$

Using Equation (3), one cans show that θ'_n satisfies the following recursive relation

$$\theta'_{n+1} = \theta'_n - \gamma_{n+1}\delta M_{n+1} \mathbf{1}_{\|\theta_n - \theta^\star\|^2 > M+2} - \gamma_{n+1}(u(\theta_n) - p_{n+1}). \tag{48}$$

We will prove by recursion that the sequence $(\theta'_n)_n$ remains in $\{\theta; \|\theta - \theta^\star\|^2 \leq M+1\} = \mathcal{C}'$. The hypothesis of recursion is obviously satisfied for $n = N$ (it is sufficient to choose $\varepsilon < 1$). Assume that the hypothesis of recursion is satisfied for $N, \ldots, n$. Then, $\|\theta_n - \theta^\star\|^2 \leq M + 2$. From (48), one can deduce

$$\begin{aligned} \theta'_{n+1} &= \theta'_n - \gamma_{n+1}(u(\theta_n) - p_{n+1}), \\ \left\|\theta'_{n+1} - \theta^\star\right\|^2 &\leq \left\|\theta'_n - \theta^\star\right\|^2 - 2\gamma_{n+1}(\theta'_n - \theta^\star\ ,\ u(\theta_n)) + \gamma_{n+1} c\,\varepsilon \end{aligned}$$

where c is a positive real constant only depending on M.
If $\|\theta'_n - \theta^\star\|^2 \leq M$, thanks to the continuity of u, a proper choice of ε assures that $\gamma_{n+1} |(\theta'_n - \theta^\star \,,\, u(\theta_n))| < 1$. Hence, $\left\|\theta'_{n+1} - \theta^\star\right\|^2 \leq M + 1$.
Now, let us consider $M \leq \|\theta'_n - \theta^\star\|^2 \leq M + 1$. Thanks to the continuity of u and Hypothesis 1, $(\theta'_n - \theta^\star \,,\, u(\theta_n)) > \delta > 0$. An appropriate choice of ε guarantees that $c\varepsilon < \delta$. Thus, $\left\|\theta'_{n+1} - \theta^\star\right\|^2 \leq M + 1$.
We have proved that for $n > N$, $\|\theta'_n - \theta^\star\|^2 \leq M + 1$. Since ε can be chosen smaller that 1, we also have $\|\theta_n - \theta^\star\|^2 \leq M + 2$, for all $n > N$.
This ends to prove that the sequence $(\theta_n)_n$ remains in a compact set and consequently σ_n is a.s. finite. □

B Proof of Proposition 1

Let $q > 0$. We define $\bar{M}_n = \sum_{i=1}^n \gamma_i \delta M_i \mathbf{1}_{\|\theta_{i-1} - \theta^\star\|^2 \leq q}$. $\bar{M}_n$ is a martingale and $\mathbb{E}(M_n^2) \leq \sum_{i=1}^n \gamma_i^2 \mathbb{E}(\|\delta M_i\|^2)$. Thanks to Hypothesis 4, $\sup_n \mathbb{E}(M_n^2) < \infty$. Hence, the series $\sum_{i>0} \gamma_i \delta M_i \mathbf{1}_{\|\theta_{i-1} - \theta^\star\|^2 \leq q}$ converges a.s. for any $q > 0$. In fact, the convergence still occurs if $q = \infty$.
Let us assume that $\sigma_n \longrightarrow \infty$. This is in contradiction with the conclusion of Lemma 6, so it implies that

$$\exists\, \eta > 0,\ q > 0, \quad \forall N > 0,\ \exists n > N \quad \mathbf{1}_{\theta_n \in \mathcal{K}_q} \|p_{n+1}\| > \eta.$$

Let $\varepsilon > 0$. There exists n such that $\mathbf{1}_{\theta_n \in \mathcal{K}_q} \|p_{n+1}\| \neq 0$ and $\|\gamma_{n+1} \delta M_{n+1}\| \leq \varepsilon$.
Hence $\theta_n \in \mathcal{K}_q$ and $\theta_n - \gamma_{n+1}(u(\theta_n) + \delta M_{n+1}) \notin \mathcal{K}_{\sigma_n}$. $\|\gamma_{n+1}\delta M_{n+1}\| \leq \varepsilon$ and since u is continuous $\|\gamma_{n+1} u(\theta_n)\|$ can also be made smaller than ε. Hence a proper choice of ε assures that $\theta_n - \gamma_{n+1}(u(\theta_n) + \delta M_{n+1}) \notin \mathcal{K}_{q+1}$. Thus, $\sigma_n < q + 1$. So, σ is finite a.s..
Let us consider

$$\theta'_n = \theta_n - \sum_{i=n+1}^{\infty} \gamma_i \delta M_i.$$

Since the series $\sum_{i>0} \gamma_i \delta M_i$ converges a.s. and θ_n remains in a compact set, θ'_n also remains in a compact set $\mathcal{C}$. We set $\bar{u} = \sup_{\theta \in \mathcal{C}} \|u(\theta)\|$.

$$\theta'_{n+1} = \theta'_n - \gamma_{n+1} u(\theta'_n) + \gamma_{n+1} \varepsilon_n,$$

where $\varepsilon_n = u(\theta'_n) - u(\theta_n)$. Thanks to the continuity of u, $\|\varepsilon_n\| \longrightarrow 0$.

$$\left\|\theta'_{n+1} - \theta^\star\right\|^2 \leq \left\|\theta'_n - \theta^\star\right\|^2 - 2\gamma_{n+1}(\theta'_n - \theta^\star \,,\, u(\theta'_n)) + \gamma_{n+1}^2(\varepsilon_n^2 + \bar{u}^2) - 2\gamma_{n+1}(\theta'_n - \theta^\star \,,\, \varepsilon_n)$$

One can rewrite this inequality introducing a sequence $\varepsilon'_n \longrightarrow 0$.

$$\left\|\theta'_{n+1} - \theta^\star\right\|^2 \leq \left\|\theta'_n - \theta^\star\right\|^2 - 2\gamma_{n+1}(\theta'_n - \theta^\star \,,\, u(\theta'_n)) + \gamma_{n+1}\varepsilon'_n. \tag{49}$$

Let $\delta > 0$. We set $V_\delta = \{\theta;\ \|\theta - \theta^\star\|^2 \leq \delta\}$. if $\|\theta'_n - \theta^\star\|^2 \leq \delta$, $(\theta'_n - \theta^\star \,,\, u(\theta'_n)) > c > 0$. Hence for n large enough, Equation (49) becomes

$$\left\|\theta'_{n+1} - \theta^\star\right\|^2 \leq \left\|\theta'_n - \theta^\star\right\|^2 - \gamma_{n+1} c + \gamma_{n+1}(c + \varepsilon'_n)\mathbf{1}_{\|\theta'_n - \theta^\star\|^2 \leq \delta}.$$

Since $\sum_n \gamma_n = \infty$, each time $\|\theta'_n - \theta^\star\|^2 > \delta$ it is taken back to $\bar{B}(\theta^\star, \sqrt{\delta})$ a few iterates later. Hence, $\|\theta'_n - \theta^\star\|^2 < \delta + \gamma_{n+1}(c + \varepsilon'_n)$. Thus, $\limsup_n \|\theta'_n - \theta^\star\| \leq \delta$ for all $\delta > 0$. This proves that $\theta'_n \longrightarrow \theta^\star$. Since the series $\sum_n \gamma_{n+1} \delta M_{n+1}$ converges, we also have $\theta'_n \longrightarrow \theta^\star$.

References

[1] B. Arouna. Robbins-monro algorithms and variance reduction in finance. *The Journal of Computational Finance*, 7(2), Winter 2003/2004.

[2] A. Benveniste, M. Métivier, and P. Priouret. *Algorithmes adaptatifs et approximations stochastiques*. Masson, 1987.

[3] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

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

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

[6] H. F. Chen, G. Lei, and A. J. Gao. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.*, 27(2): 217–231, 1988. ISSN 0304-4149.

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

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

[9] P. Glynn and M. Hsieh. Confidence regions for stochastic approximation algorithms. In *2002 Winter simulation Conference*, 2002.

[10] J. Jacod and A. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag Berlin, 1987.

[11] H. Kushner and G. Yin. *Stochastic Approximation and Recursive Algorthims and Applications*. Springer-Verlag New York, second edition, 2003.

[12] M. Pelletier. Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann. Appl. Probab.*, 8(1):10–44, 1998. ISSN 1050-5164.

[13] B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.*, 30(4):838–855, 1992. ISSN 0363-0129.

[14] P. Protter. *Stochastic integration and Differential Equations*. Springer-Verlag Berlin, second edition, 2004.