A Central Limit Theorem for Robbins Monro Algorithms with Projections

LELONG Jérôme*

CERMICS, Ecole Nationale des Ponts et Chaussées, 6 et 8 av Blaise Pascal,
77455 Marne La Vallée, FRANCE.

20th September 2005

Abstract

I propose a proof of a central limit theorem for stochastic algorithms with projections. A detailed proof is given in the case of projection on a convex set. I also explain how this proof can be adapted to Chen's algorithm, in which the projection is done on an increasing sequence of compact sets.

Key words: stochastic approximation, central limit theorem, tightness, recursive algorithms.

1 Introduction

A lot of work has already been done around the rate of convergence of stochastic algorithms. Typically one normalises the iterates after centring them about the limit. Convergence in distribution to a random normal variable can then be proved for unconstrained algorithms (see [Duflo, 1997] for instance). Stronger results can be found in [Benveniste et al., 1987] and [Bouton, 1985] where functional versions of the classical central limit theorem are proposed. Their approach is based on proving some tightness criteria in Skorokhod space. The classical central limit theorem has also been adapted to constrained algorithms in [Kushner and Yin, 2003] using ordinary differential equation techniques. The case of constrained algorithms is somehow related to the problem of multiple targets. The convergence rate of algorithms with multiple targets has been studied by Pelletier in [Pelletier, 1998].

I present an alternative proof of the central limit theorem for constrained algorithms using tightness criteria, recursive relations and martingale techniques. This paper aims at proving a central limit theorem for Chen's algorithm (see [Chen and Zhu, 1986] for a convergence of the algorithm), which consists in a projection of the standard algorithm on an increasing sequence of compact sets. This new algorithm requires weaker hypotheses than the standard Robbins Monro algorithm to convergence and numerically behaves more smoothly. This kind of projected algorithm still satisfies a central limit theorem. In this last case, the same arguments hold and the proof done for the constrained algorithms can be adapted to Chen's algorithm. The required modifications are presented in the last part of this work.

These results are definitely extremely valuable to measure the quality of the convergence of the algorithm. Finally, not only do these results give a convergence speed but they also help determining the optimal step in the Robbins Monro procedure.

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

2 Robbins Monro algorithms with projection on a constraint set

2.1 Presentation

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 an underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

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

where Z is a random variable in $\mathbb{R}^m$.

Let us suppose that we want to solve this problem under some constraints on θ, which means that θ is bound to remain in a closed convex subset H of $\mathbb{R}^d$.

We define for $\theta_0 \in H$ the sequence $(\theta_n)_n$

$$\theta_{n+1} = \Pi\left(\theta_n - \gamma_{n+1} U(\theta_n, Z_{n+1})\right), \tag{2}$$

where Π stands for the Euclidean projection on H. $(Z_n)_{n\geq 0}$ is an independent and identically distributed sequence of random variables following the law of Z and $(\gamma_n)_{n\geq 0}$ a decreasing sequence of positive real numbers.
We also introduce $\mathcal{F}_n = \sigma(\theta_k, Z_k; k \leq n)$ the σ-field generated by the random vectors θ_k and Z_k for $k \leq n$, which is in fact the σ-field generated by the random vectors Z_k for $k \leq n$ since θ_0 is deterministic and U measurable. We can write $u(\theta_n) = \mathbb{E}[U(\theta_n, Z_{n+1}) | \mathcal{F}_n]$.

The following theorem gives a convergence result on the sequence $(\theta_n)_n$.

Theorem 1. *We assume that*

(H1.1) $\exists\, \theta^\star \in H, u(\theta^\star) = 0$,
$\forall \theta \in H,\ \theta \neq \theta^\star,\ \langle \theta - \theta^\star, u(\theta)\rangle > 0$.

(H1.2) $\exists K > 0,\ \forall n \geq 0\ \mathbb{E}[\|U(\theta_n, Z_{n+1})\|^2 | \mathcal{F}_n] \leq K(1 + \|\theta_n - \theta^*\|^2)$.

(H1.3) $\sum_{n=0}^{\infty} \gamma_n = \infty,\ \sum_{n=0}^{\infty} \gamma_n^2 < \infty$.

Then, the sequence $(\theta_n)_{n\geq 0}$ converges almost surely to θ^*, moreover if $\theta^* \in \mathring{H}$ there is a finite number of projections.

Proof. The proof of the convergence is based on the use of Robbins and Siegmund's lemma (see [Robbins and Siegmund, 1971]).

$$\begin{aligned}
\|\theta_{n+1} - \theta^*\|^2 &\leq \|\theta_n - \theta^* - \gamma_{n+1} U(\theta_n, Z_{n+1})\|^2, \\
&\quad \text{since } \Pi \text{ is non expansive,} \\
&\leq \|\theta_n - \theta^*\|^2 + \gamma_{n+1}^2 \|U(\theta_n, Z_{n+1})\|^2 - 2\gamma_{n+1}\langle \theta_n - \theta^*, U(\theta_n, Z_{n+1})\rangle,
\end{aligned}$$

Let us take the conditional expectation with respect to $\mathcal{F}_n$ in the previous inequality.

$$\begin{aligned}
\mathbb{E}\left(\|\theta_{n+1} - \theta^*\|^2 | \mathcal{F}_n\right) &\leq \|\theta_n - \theta^*\|^2 + \gamma_{n+1}^2 \mathbb{E}\left(\|U(\theta_n, Z_{n+1})\|^2 | \mathcal{F}_n\right) - 2\gamma_{n+1}\langle \theta_n - \theta^*, u(\theta_n)\rangle, \\
&\leq \|\theta_n - \theta^*\|^2 (1 + K\gamma_{n+1}^2) + K\gamma_{n+1}^2 - 2\gamma_{n+1}\langle \theta_n - \theta^*, u(\theta_n)\rangle, \\
&\quad \text{using (H1.2).}
\end{aligned}$$

Thanks to hypothesis (H1.1) the above scalar product is positive and the series $\sum_n \gamma_n^2$ converges. We can then apply Robbins-Siegmund's lemma which claims that $\|\theta_{n+1} - \theta^*\|^2$ and $\sum_n \gamma_{n+1}\langle \theta_n - \theta^*, u(\theta_n)\rangle$ converge almost surely.

Suppose that $\|\theta_{n+1} - \theta^*\|$ does not converge to zero, then there exist two constants m and M, $0 < m < M < \infty$, such that for n large enough, $m \leq \|\theta_{n+1} - \theta^*\| \leq M$. On the set $\{\theta; m \leq \|\theta - \theta^*\| \leq M\}$, the continuous function $\theta \longmapsto \langle \theta - \theta^*, u(\theta)\rangle$ is bounded below by $c, c > 0$. Consequently, the convergence of $\sum_n \gamma_{n+1}\langle \theta_n - \theta^*, u(\theta_n)\rangle$ would be equivalent to the one of $\sum_n \gamma_n$

which is in contradiction with hypothesis (H1.3). Hence $\theta_n \xrightarrow{a.s.} \theta^*$. This ends the proof of the convergence statement in the theorem.

Let us assume that $\theta^* \in \overset{\circ}{H}$, then there exists $\varepsilon > 0$ such that the closed ball $B(\theta^*, \varepsilon) \subset H$. Since $\theta_n \xrightarrow{a.s.} \theta^*$,

$$\exists\, n_0 \text{ such that } \forall n \geq n_0, \|\theta_n - \theta^*\| \leq \frac{\varepsilon}{2}. \tag{3}$$

If there were a non-finite number of projections, it would mean that for all n, there would exist $N > n$ such that θ_N would lie on the border of H and hence $\|\theta_N - \theta^*\|$ would be equal to ε which is in contradiction with assertion (3). Therefore, there is an almost surely finite number of projections. □

2.2 A central limit theorem for a constrained Robbins Monro algorithm

The behaviour of Robbins Monro algorithms is hung up to the choice of the step sequence $(\gamma_n)_{n\geq 0}$. The limit given by the central limit theorem is also affected by this choice. From now on, we will only consider sequences of the form $\gamma_n = \frac{\gamma}{(n+1)^\alpha}$, where $\frac{1}{2} < \alpha \leq 1$. The value $\alpha = 1$ gives a different limit and has to be treated separately.

In this section, we will assume hypotheses (H1.1), (H1.2) and (H1.3). So, the sequence $(\theta_n)_{n\geq 0}$, defined by (2), converges a.s. to θ^* thanks to theorem 1.

First, let us introduce a few notations used in the theorem and its proof. Equation (2) can be rewritten

$$\theta_{n+1} = \theta_n - \gamma_{n+1} u(\theta_n) - \gamma_{n+1}\delta M_{n+1} + \gamma_{n+1} p_{n+1}, \tag{4}$$

where

$$\begin{aligned} \delta M_{n+1} &= U(\theta_n, Z_{n+1}) - u(\theta_n), && (5)\\ \gamma_{n+1} p_{n+1} &= \Pi(\theta_n - \gamma_{n+1} U(\theta_n, Z_{n+1})) - (\theta_n - \gamma_{n+1} U(\theta_n, Z_{n+1})). && (6) \end{aligned}$$

Remark 1. One should notice that $\|p_{n+1}\| \leq \|U(\theta_n, Z_{n+1})\|$ and that almost surely for n large enough $p_n = 0$ since there is a finite number of projections (see theorem 1) .

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

$$\Delta_n = \frac{\theta_n - \theta^*}{\sqrt{\gamma_n}},$$

I will treat the case $\alpha = 1$ and the case $\frac{1}{2} < \alpha < 1$ separately, since they give two different limits and do not require exactly the same hypotheses.

2.2.1 the case $\alpha = 1$

I will give a detailed proof in this case as the case $\frac{1}{2} < \alpha < 1$ can be handled much with the same arguments.

Theorem 2. *Under hypotheses (H1.1), (H1.2), (H1.3) and the following ones:*

(H2.1) *u is C^1 in the neighbourhood of θ^*. Moreover there exists a symmetric positive definite matrix A such that*
$$u(\theta) = A(\theta - \theta^*) + o(\theta - \theta^*).$$
(H2.2) *θ^* belongs to the interior of H.*
(H2.3) *$\gamma A - \frac{I}{2}$ is definite positive.*
(H2.4) *There exists $q > 0$ and $\rho > 0$ such that* $\sup_n \mathbb{E}\left(\|\delta M_n\|^{2+\rho} \mathbf{1}_{\|\theta_n - \theta^*\| < q}\right) < \infty$.
(H2.5) *There exists $q > 0$ and a symmetric definite positive matrix Σ such that the family* $\mathbb{E}\left(U(\theta_i, Z_{i+1}) U(\theta_i, Z_{i+1})' \mathbf{1}_{\|\theta_i - \theta^*\| < q} | \mathcal{F}_i\right) \xrightarrow{\mathbb{P}} \Sigma$.

Then, the sequence $(\Delta_n)_{n\geq 0}$ converges in distribution to a random normal 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 2. Most of the time the Σ matrix introduced in (H2.5) will be $\mathbb{E}\left(U(\theta^*,Z)U(\theta^*,Z)'\right)$.

To simplify the notations, let us define the function $t_n : \mathbb{R} \longrightarrow \mathbb{N}$

$$\sum_{i=n}^{n+t_n(u)} \gamma_i \leq u < \sum_{i=n}^{n+t_n(u)+1} \gamma_i, \qquad \forall u \in \mathbb{R}. \tag{7}$$

We now define the processes $(\Delta_n(\cdot))$ as the piecewise constant interpolations of $(\Delta_p)_{p\geq n}$ on intervals of length $(\gamma_p)_{p\geq n}$:

$$\Delta_n(t) = \Delta_{n+t_n(t)} \tag{8}$$

and $W_n(.)$ by

$$W_n(t) = \sum_{i=n}^{n+t_n(t)} \sqrt{\gamma_i}\delta M_i \qquad \forall t > 0 \text{ and } W_n(0) = 0. \tag{9}$$

To prove theorem 2, we will first establish a recursive relation on the sequence $(\Delta_n)_n$. This relation will lead to the tightness of the sequence on the one hand and to an integral relation on the process $(\Delta_n(\cdot))$ on the other hand. Each term in the integral relation will be handled separately and proved to converge to zero in probability except one which will give the limit.

First, we establish a recursive expression of Δ_{n+1}.

$$\begin{aligned}
\Delta_{n+1} &= \frac{\theta_{n+1} - \theta^*}{\sqrt{\gamma_{n+1}}}, \\
&= \frac{1}{\sqrt{\gamma_{n+1}}} \left(\theta_n - \theta^* - \gamma_{n+1}u(\theta_n) - \gamma_{n+1}\delta M_{n+1} + \gamma_{n+1}p_{n+1}\right), \\
&= \sqrt{\frac{\gamma_n}{\gamma_{n+1}}}\, \Delta_n - \sqrt{\gamma_{n+1}}(u(\theta_n) + \delta M_{n+1} - p_{n+1}).
\end{aligned} \tag{10}$$

Using hypothesis (H2.1) we can introduce a function y such that $u(\theta) = A(\theta-\theta^*)+y(\theta-\theta^*)(\theta-\theta^*)$ satisfying $\lim_{\|x\|\to 0} \|y(x)\| = 0$. Then, the previous equation becomes

$$\Delta_{n+1} = \left(\sqrt{\frac{\gamma_n}{\gamma_{n+1}}}I - \sqrt{\gamma_{n+1}\gamma_n}A - \sqrt{\gamma_{n+1}\gamma_n}\, y(\theta_n - \theta^*)\right)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} + \sqrt{\gamma_{n+1}}p_{n+1}. \tag{11}$$

The following Taylor expansions hold

$$\begin{aligned}
\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} &= 1 + \frac{1}{2(n+1)} + O\left(\frac{1}{n^2}\right), \qquad (12)\\
\sqrt{\gamma_n\gamma_{n+1}} &= \gamma_n + O\left(\frac{1}{n^2}\right). \qquad (13)
\end{aligned}$$

We define $Q = A - \frac{I}{2\gamma}$ which is symmetric definite positive (see (H2.3)).

This remark enables us to simplify equation (11) by introducing a new sequence $(\beta_n)_{n\geq 0}$ such that for any n larger than some fixed n_0, $\beta_n \leq C$, where C is a positive real constant. (11) can be rewritten as

$$\begin{aligned}
\Delta_{n+1} &= \Delta_n - \gamma_n Q\Delta_n - \gamma_n y(\theta_n - \theta^*)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} \\
&\quad + \sqrt{\gamma_{n+1}}p_{n+1} + \frac{\beta_n}{n+1}\gamma_n(B + y(\theta_n - \theta^*))\Delta_n,
\end{aligned} \tag{14}$$

where B is a deterministic matrix.

We will concentrate on the tightness of the sequence $(\Delta_n)_n$. To do so, let us go back to equation (14) and let $\tilde{\theta}_n$ denote the iterate obtained before projecting. $\tilde{\Delta}_n$ is defined in the same way as Δ_n using $\tilde{\theta}_n$.

$$\tilde{\Delta}_{n+1} = \Delta_n - \gamma_n Q\Delta_n - \gamma_n y(\theta_n - \theta^*)\Delta_n - \sqrt{\gamma_{n+1}}\delta M_{n+1} + \frac{\beta_n}{n^2}(B + y(\theta_n - \theta^*))\Delta_n, \tag{15}$$

Since θ_n converges almost surely to θ^*,

$$\forall \varepsilon > 0,\ \forall \eta > 0, \exists N > 0 \text{ such that } \forall n \geq N\ \mathbb{P}\left(\sup_{m>n} \|\theta_m - \theta^*\| > \eta\right) < \varepsilon. \tag{16}$$

Let $\lambda > 0$ be the smallest eigenvalue of Q. $\lambda > 0$ since Q is symmetric definite positive. $\lim_{\|x\|\to 0} y(x) = 0$, so for $x < x_0$, $\|y(x)\| < 3\lambda/4$. One can choose x_0 smaller than constant q introduced in hypothesis (H2.5). Let $\varepsilon > 0$. Thanks to (16) there exists a rank N_0 such that $\mathbb{P}(\sup_{m>N_0} \|\theta_m - \theta^*\| > x_0) < \varepsilon$. Based on this remark we define a sequence of decreasing adapted sets $(A_n)_{n\geq N_0}$ by

$$A_n = \left\{ \sup_{n\geq m > N_0} \|\theta_m - \theta^*\| < x_0 \right\}. \tag{17}$$

Let us remark that $\forall n > N_0\ \mathbb{P}(A_n^c) \leq \varepsilon$. On the set A_n, $y(\theta_n - \theta^*)$ is bounded so we have

$$\left\|\tilde{\Delta}_{n+1}\right\|^2 = \|\Delta_n\|^2 - 2\gamma_n \Delta_n{}'(Q + y(\theta_n - \theta^*))\Delta_n + O\left(\frac{1}{n^2}\right)\|\Delta_n\|^2 - 2\sqrt{\gamma_{n+1}}\delta M_{n+1}\Delta_n + O(\gamma_n).$$

Since the projection is non-expansive, the norm of Δ_{n+1} is bound to be smaller than the one of $\tilde{\Delta}_{n+1}$. Let us take the conditional expectation, in the previous equality, with respect to $\mathcal{F}_n$, denoted $\mathbb{E}_n$.

$$\mathbb{E}_n \|\Delta_{n+1}\|^2 \leq \|\Delta_n\|^2 - 2\gamma_n \Delta_n{}'(Q + y(\theta_n - \theta^*))\Delta_n + O\left(\frac{1}{n^2}\right)\|\Delta_n\|^2 + O(\gamma_n). \tag{18}$$

On the set A_n $Q + y(\theta_n - \theta^*)$ is a definite positive matrix with smallest eigenvalue greater than $\lambda/4$. Therefore $\Delta_n{}'(Q + y(\theta_n - \theta^*))\Delta_n > \lambda/2\,\|\Delta_n\|^2$. It is important to notice that Landau's notations are used in a deterministic context, therefore we can assume that for $n > N_0 \quad O(\frac{1}{n^2}) \leq \lambda/4\gamma_n$.

$$\begin{aligned}
\mathbb{E}\left(\|\Delta_{n+1}\|^2 \mathbf{1}_{A_n}\right) - \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) &\leq -\gamma_n \frac{\lambda}{2}\mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) + c\gamma_n, \\
\mathbb{E}\left(\|\Delta_{n+1}\|^2 \mathbf{1}_{A_{n+1}}\right) - \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) &\leq -\gamma_n \frac{\lambda}{2}\mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) + c\gamma_n,
\end{aligned} \tag{19}$$

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

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

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 $\forall i \geq N_0 \quad \mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{A_i}\right) \leq \frac{2c}{\lambda} + \mathbb{E}\left(\|\Delta_{N_0}\|^2 \mathbf{1}_{A_{N_0}}\right)$. 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 $\mathbb{E}\left(\|\Delta_{i+1}\|^2 \mathbf{1}_{A_{i+1}}\right) \leq \frac{2c}{\lambda}$. Otherwise $i+1 \notin \mathcal{I}$ and hence $\mathbb{E}\left(\|\Delta_{i+1}\|^2 \mathbf{1}_{A_{i+1}}\right) \leq \mathbb{E}\left(\|\Delta_i\|^2 \mathbf{1}_{A_i}\right)$. So using the hypothesis of recursion proves the result announced above. Therefore

$$\sup_n \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) < \infty. \tag{20}$$

In the end, this relation combined with (16) will lead to the tightness of the sequence $(\Delta_n)_n$. Let $M > 0$.

$$\begin{aligned}
\mathbb{P}(\|\Delta_n\| > M) &\leq \mathbb{P}(\|\Delta_n\|(\mathbf{1}_{A_n} + \mathbf{1}_{A_n^c}) > M), \\
&\leq \mathbb{P}(\|\Delta_n\|\,\mathbf{1}_{A_n} > M/2) + \mathbb{P}(\|\Delta_n\|\,\mathbf{1}_{A_n^c}) > M/2), \\
&\leq 4/M^2\,\mathbb{E}\left(\|\Delta_n\|\,\mathbf{1}_{A_n}^2\right) + \mathbb{P}(A_n^c).
\end{aligned} \tag{21}$$

There exists a value of M depending on ε such that both terms on the right hand-side of (21) are bounded above by ε. This proves the tightness of $(\Delta_n)_n$. From now on we will assume that $n > N_0$ and large enough to ensure that $\beta_n < C$.

If we go back to equation (14) and sum up this equality from n — chosen greater than N_0 introduced above — to $n + p$ we obtain

$$\begin{aligned}\Delta_{n+p} &= \Delta_n - \sum_{k=0}^{p-1} \gamma_{n+k}(Q + y(\theta_{n+k} - \theta^*))\Delta_{n+k} + \sqrt{\gamma_{n+k}}\delta M_{n+k+1} \\ &\quad + \sum_{k=0}^{p-1} \sqrt{\gamma_{n+k}} p_{n+k+1} + \frac{\beta_{n+k}}{n+k+1}\gamma_{n+k}(B + y(\theta_{n+k} - \theta^*))\Delta_{n+k}. \end{aligned} \tag{22}$$

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

$$\Delta_n(u) = \Delta_n(0) - \int_0^u \left(Q + y(\theta_n(s) - \theta^*)\right) \Delta_n(s)ds - W_n(u) + R_n(u) + P_n(u), \tag{23}$$

where

$$P_n(u) = \sum_{k=0}^{t_n(u)} \sqrt{\gamma_{n+k}} p_{n+k+1}, \tag{24}$$

$$R_n(u) = \sum_{k=0}^{t_n(u)} \frac{\beta_{n+k}}{n+k+1}\gamma_{n+k} \left(B + y(\theta_{n+k} - \theta^*)\right)\Delta_{n+k}. \tag{25}$$

Note that

$$\|R_n(u)\| \leq \frac{C}{n}\int_0^u (1 + \|y(\theta_n(s) - \theta^*)\|) \, \|\Delta_n(s)\| \, ds. \tag{26}$$

We will show that $\lim_t \Delta_n(t)$ exists and is the random normal variable described in theorem 2. For the sake of clearness the end of the proof will be done assuming that all processes and random variables are real valued and not vector valued.
Let us go back to equation (23) and consider its equivalent differential form

$$d\Delta_n(u) = -\left(Q + y(\theta_n(s) - \theta^*)\right) \Delta_n(s)ds - dW_n(u) + dR_n(u) + dP_n(u). \tag{27}$$

We can now integrate (27) to obtain a new expression for $(\Delta_n(\cdot))$.

$$\begin{aligned}\Delta_n(t) &= e^{-Qt}\Delta_n(0) - \int_0^t e^{Q(u-t)} y(\theta_n(u) - \theta^*)\Delta_n(u)du - \int_0^t e^{Q(u-t)} dW_n(u) \\ &\quad + \int_0^t e^{Q(u-t)} dR_n(u) + \int_0^t e^{Q(u-t)} dP_n(u).\end{aligned} \tag{28}$$

The last step of the proof consists in showing that, in the previous equation, every term tends to zero in probability when letting t go to infinity except the integral with respect to $W_n(\cdot)$.
Let me remind that all the limits involved stand true for a fixed $n > N_0$.

To treat the first term in (28), one should remember that the set $\{\Delta_n(0); n \geq 0\}$ is tight. Since Q is definite positive, $e^{-Qt}\Delta_n(0)$ tends to zero in probability when t goes to infinity.

Concerning the second term, one should notice that the set $\{\Delta_n(u); u < \infty\}$ is tight. Since $\theta_n(u)$ converges almost surely to θ^* when u tends to infinity, $y(\theta_n(u) - \theta^*)$ tends to zero almost surely. These two conditions imply that $y(\theta_n(u) - \theta^*)\Delta_n(u)$ tends to zero in probability when u goes to infinity (see proposition 3). Hence proposition 2 states the convergence of the sequence of random variables $(\int_0^t e^{Q(u-t)} y(\theta_n(u) - \theta^*)\Delta_n(u)du)_t$ to zero in probability.

The fourth term can be treated using the previous result. Since $R_n(\cdot)$ is a pure jump process with a finite number of jumps on $[0,t]$, the stochastic integral can be rewritten as a discrete sum as follows

$$\int_0^t e^{Q(u-t)} dR_n(u) = \sum_{u \leq t} e^{Q(u-t)} \Delta R_n(u). \tag{29}$$

One can refer to [Rogers and Williams, 2000] to find out more on pure jump processes and stochastic integrals with respect to pure jump semi-martingales. Using (25), the previous equation can be expanded as

$$\int_0^t e^{Q(u-t)} dR_n(u) \quad = \quad \sum_{u \leq t} e^{Q(u-t)} (1 + y(\theta_n(u) - \theta^*)) \Delta_n(u) \frac{\beta_{n+t_n(u)}}{(n + t_n(u))} \gamma_{n+t_n(u)}.$$

This discrete sum behaves as the following integral

$$\int_0^t e^{Q(u-t)} (1 + y(\theta_n(u) - \theta^*)) \Delta_n(u) c_n(u) du, \tag{30}$$

where function $c_n(\cdot)$ tends to 0 towards infinity since the sequence $(\beta_n)_{n \geq 0}$ is bounded. We can reproduce what has been done while treating the second term in (28) to prove that $\int_0^t e^{Q(u-t)}(1 + y(\theta_n(u) - \theta^*))\Delta_n(u)c_n(u)du$ is tight as a sequence indexed by t. Moreover, $((1 + y(\theta_n(u) - \theta^*))\Delta_n(u))_u$ is tight and $(c_n(u))_u$ converges almost surely to zero. Therefore, using proposition 3, $((1+y(\theta_n(u)-\theta^*))\Delta_n(u)c_n(u))_u$ tends to zero in probability. Proposition 2 enables us to conclude that the fourth term in (23) tends to zero in probability when t goes to infinity.

The term due to the projection in the algorithm can be rewritten

$$\int_0^t e^{Q(u-t)} dP_n(u) = \sum_{i=0}^{t_n(t)} e^{Q(t_n^{-1}(i)-t)} \sqrt{\gamma_{i+n}}\, p_{i+n}. \tag{31}$$

This discrete sum behaves like $\int_0^t e^{Q(u-t)} \sqrt{\gamma_{n+t_n(u)}}\, p_{n+t_n(u)} du$. Since there is almost surely a finite number of projections. p_n is almost surely zero for n large enough, hence $\sqrt{\gamma_{n+t_n(u)}}\, p_{n+t_n(u)}$ converges to zero almost surely when u tends to infinity. Using proposition 1 proves that the sum in (31) tends almost surely to zero when t goes to infinity.

Looking back at (28), one realises that all the terms tend to zero in probability except the stochastic integral with respect to $W_n(\cdot)$ which converges to a random normal variable, as I am now going to prove it.
First, it is noticeable that

$$\begin{aligned} \int_0^t e^{Q(u-t)} dW_n(u) \quad &= \quad \sum_{i=n}^{n+t_n(t)} e^{Q(t_n^{-1}(i-n)-t)} \sqrt{\gamma_i} \delta M_i, & (32) \\ &= \quad \sum_{i=0}^{t_n(t)} e^{Q(t_n^{-1}(i)-t)} \sqrt{\gamma_{i+n}} \delta M_{i+n}. & (33) \end{aligned}$$

Let us fix n and define N_l^p for all $0 \leq l \leq p$ and $p > 0$

$$N_l^p = \sum_{i=0}^{l} e^{Q(t_n^{-1}(i)-t_n^{-1}(p))} \sqrt{\gamma_{i+n}} \delta M_{i+n}. \tag{34}$$

$(N_l^p)_{0 \leq l \leq p}$ is obviously a martingale with respect to $(\mathcal{F}_{n+l})_l$ and satisfies the following relation $N_p^p = \int_0^{t_n^{-1}(p)} e^{Q(u-t)} dW_n(u)$. Then we only need to prove that N_p^p converges to a random normal variable when p goes to infinity. To do so we will use a slightly modified version of the central limit theorem for martingale arrays given in [Duflo, 1997] (theorem 2.1.9).

Theorem 3 (Central Limit Theorem for martingale arrays). *Suppose that* $\{(\mathcal{F}_l^{(p)})_{0\leq l\leq p}; p>0\}$ *is a family of filtrations and* $\{(N_l^{(p)})_{0\leq l\leq p}; p>0\}$ *a square integrable martingale array with respect to the previous filtration. We also assume that:*

(H3.1) *there exists a symmetric definite positive matrix* V *such that* $\langle N\rangle_p^{(p)}$ *converges in probability to* V.
(H3.2) *There exists* $\rho>0$ *such that*

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

Then

$$N_p^{(p)}\xrightarrow{\mathcal{L}}\mathcal{N}(0,V).$$

Let us compute the angle bracket of N.

$$\langle N\rangle_p^p \quad=\quad \sum_{i=0}^{p}e^{2Q(t_n^{-1}(i)-t_n^{-1}(p))}\gamma_{i+n}\mathbb{E}\left(\delta M_{i+n}^2|\mathcal{F}_{n+i-1}\right). \tag{35}$$

First I will prove that the conditional expectation above converges in probability to Σ. Thanks to the continuity of u, $u(\theta_i)\xrightarrow[i\to\infty]{\mathbb{P}}0$.

$$\mathbb{E}\left(\delta M_{i+1}^2|\mathcal{F}_i\right)\quad=\quad\mathbb{E}\left(U(\theta_i,Z_{i+1})^2|\mathcal{F}_i\right)-u(\theta_i)^2. \tag{36}$$

Let us consider $\eta>0$

$$\begin{aligned}\mathbb{P}\left(\left|\mathbb{E}\left(U(\theta_i,Z_{i+1})^2|\mathcal{F}_i\right)-\Sigma\right|>\eta\right)\leq \mathbb{P}\left(\left|\mathbb{E}\left(U(\theta_i,Z_{i+1})^2|\mathcal{F}_i\right)-\Sigma\right|\mathbf{1}_{\|\theta_i-\theta^*\|<x_0}>\eta/2\right)+\\ \mathbb{P}\left(\left|\mathbb{E}\left(U(\theta_i,Z_{i+1})^2|\mathcal{F}_i\right)-\Sigma\right|\mathbf{1}_{\|\theta_i-\theta^*\|>x_0}>\eta/2\right).\end{aligned} \tag{37}$$

The second expectation on the right hand side is smaller than $\mathbb{P}\left(\|\theta_i-\theta^*\|>x_0\right)$ which tends to zero as i goes to infinity. $\left|\mathbb{E}\left(U(\theta_i,Z_{i+1})^2|\mathcal{F}_i\right)-\Sigma\right|\mathbf{1}_{\|\theta_i-\theta^*\|<x_0}$ is always smaller than $\left|\mathbb{E}\left(U(\theta_i,Z_{i+1})^2|\mathcal{F}_i\right)\mathbf{1}_{\|\theta_i-\theta^*\|<x_0}-\Sigma\right|$. Hence the last term of (37) tends to zero using hypothesis (H2.5). Thus, this brings to an end the proof that the sequence of conditional expectations involved in (35) converges to zero Σ in probability. Moreover, it is easy to see that $\langle N\rangle_p^p$ behaves as

$$\int_0^{t_n^{-1}(p)}e^{2Q(u-t_n^{-1}(p))}\mathbb{E}\left(\delta M_{t_n(u)+n}^2|\mathcal{F}_{n+t_n(u)-1}\right)du. \tag{38}$$

We have just seen that the conditional expectations above converge in probability to Σ. By applying proposition 2, it is clear that $\int_0^t e^{2Q(u-t)}\mathbb{E}\left(\delta M_{t_n(u)+n}^2|\mathcal{F}_{n+t_n(u)-1}\right)du$ converges in probability to $\frac{\Sigma}{2Q}$ when t goes to infinity. Moreover $t_n^{-1}(p)$ tends to infinity when p goes to infinity, hence $\langle N\rangle_p^p\xrightarrow[p\to\infty]{\mathbb{P}}\frac{\Sigma}{2Q}$.

Let ρ be the real number defined in theorem 2.

$$\sum_{l=1}^{p}\mathbb{E}\left(\left\|N_l^{(p)}-N_{l-1}^{(p)}\right\|^{2+\rho}\Big|\mathcal{F}_{l-1}^{(p)}\right)=\sum_{i=0}^{p}e^{(2+\rho)Q(t_n^{-1}(i)-t_n^{-1}(p))}\gamma_{i+n}^{1+\frac{\rho}{2}}\mathbb{E}\left(\delta M_{i+n}^{2+\rho}|\mathcal{F}_{n+i-1}\right). \tag{39}$$

This sum behaves as

$$\int_0^{t_n^{-1}(p)}e^{(2+\rho)Q(u-t_n^{-1}(p))}\gamma_{t_n(u)+n}^{\frac{\rho}{2}}\mathbb{E}\left(\delta M_{t_n(u)+n}^{2+\rho}|\mathcal{F}_{n+t_n(u)-1}\right)du. \tag{40}$$

$\gamma_{t_n(u)+n}^{\frac{\rho}{2}}$ converges to 0 when u goes to infinity and the sequence of conditional expectations converges in probability, so $\gamma_{t_n(u)+n}^{\frac{\rho}{2}}\mathbb{E}\left(\delta M_{t_n(u)+n}^{2+\rho}|\mathcal{F}_{n+t_n(u)-1}\right)$ tends to zero in probability when u goes to infinity by applying proposition 3. Proposition 2 enables to achieve the proof of the convergence of the expression defined by (40) to 0 when t goes to infinity. The hypotheses of theorem

3 are then satisfied. This precisely ends the proof that the stochastic integral $\int_0^t e^{Q(u-t)} dW_n(u)$ converges in distribution to a random normal variable with mean 0 and variance $\frac{\Sigma}{2Q}$.

Now if we go back to equation (28) considering that all the sequences are vector valued, we find out that for any fixed n

$$\Delta_n(t) \xrightarrow[t\to\infty]{\mathcal{L}} \mathcal{N}(0, V),$$

where $V = \int_0^\infty e^{Qu} \Sigma e^{Qu} du$. Moreover, the definition of the sequence of processes $(\Delta_n(\cdot))$ implies that since $\Delta_n(t)$ converges when t goes to infinity, the sequence $(\Delta_n)_n$ also converges towards the same limit when n goes to infinity.

So, the sequence $(\Delta_n)_n$ converges in distribution to a centred random normal variable with covariance matrix

$$V = \int_0^\infty e^{-Qs} \Sigma e^{-Qs} ds. \tag{41}$$

If we rewrite it using matrix A, we come up with the following expression for V.

$$V = \int_0^\infty e^{\left(\frac{I}{2\gamma} - A\right)s} \Sigma e^{\left(\frac{I}{2\gamma} - A\right)s} ds = \gamma \int_0^\infty e^{\left(\frac{I}{2} - \gamma A\right)s} \Sigma e^{\left(\frac{I}{2} - \gamma A\right)s} ds. \tag{42}$$

Hence the proof of theorem 2 is completed. I will explain how we can deduce a central limit theorem when $\frac{1}{2} < \alpha < 1$.

2.2.2 the case $\frac{1}{2} < \alpha < 1$

We use the same notations as in the case $\alpha = 1$. This time, the central limit theorem can be written as follows

Theorem 4. *Under hypotheses (H1.1), (H1.2), (H1.3), (H2.1), (H2.2), (H2.4) and (H2.5), the sequence* $(\Delta_n)_{n\geq 0}$ *converges in distribution to a random normal variable with mean* 0 *and covariance*

$$V = \int_0^\infty \exp(-At) \Sigma \exp(-At) dt.$$

Proof. The proof is almost the same as in the previous case, the only differences appear due to slightly modified expansions in equations (12) and (13) which become

$$\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} = 1 + O\left(\frac{1}{n}\right), \tag{43}$$

$$\sqrt{\gamma_n \gamma_{n+1}} = \gamma_n + O\left(\frac{1}{n^{1+\alpha}}\right). \tag{44}$$

We define $Q = A$ which is still symmetric definite positive as in the previous case. These new developments modify the recursive relation on Δ_n. Equation (14) becomes

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

The new iteration $\tilde{\Delta}_{n+1}$ obtained before any projection satisfies

$$\begin{aligned} \tilde{\Delta}_{n+1} &= \Delta_n - \gamma_n Q \Delta_n - \gamma_n y(\theta_n - \theta^*) \Delta_n - \sqrt{\gamma_{n+1}} \delta M_{n+1} \\ &\quad + \frac{\beta_n}{n+1}(B + y(\theta_n - \theta^*))\Delta_n. \end{aligned}$$

Taking the square leads to

$$\left\| \tilde{\Delta}_{n+1} \right\|^2 = \|\Delta_n\|^2 - 2\gamma_n \Delta_n{}'(Q + y(\theta_n - \theta^*))\Delta_n + O\left(\frac{1}{n}\right) \|\Delta_n\|^2 - 2\sqrt{\gamma_{n+1}} \delta M_{n+1} \Delta_n + O(\gamma_n).$$

As in the previous case, this equation will enable us to prove the tightness of the sequence $(\Delta_n)_n$, since for n large enough $O\left(\frac{1}{n}\right) < \lambda/4\gamma_n$.

$\Delta_n(u)$ satisfies relation (28) with

$$R_n(u) = \sum_{k=0}^{t_n(u)} \frac{\beta_{n+k}}{n+k+1} (1 + y(\theta_{n+k} - \theta^*))\Delta_{n+k}, \tag{45}$$

$$\|R_n(u)\| \leq \frac{C}{n^{1-\alpha}} \int_0^u (1 + y(\theta_n(s) - \theta^*))\Delta_n(s)ds. \tag{46}$$

The rest of the proof remains the same. Δ_n converges in distribution to a centred random normal variable with covariance matrix

$$V = \int_0^\infty e^{-Qs} \Sigma e^{-Qs} ds,$$

as announced in theorem 4 since $A = Q$. □

Remark 3. Practical experiments with this type of procedures quickly show that the smaller the conditional variance of Y_n is, the faster the algorithm converges. When possible, one should try to rewrite $u(\theta)$ as an expectation of a random variable with less variance. Nonetheless, it is sometimes almost impossible to reduce the variance significantly enough to ensure a satisfying convergence. A more robust algorithm, that would not take too extreme values into account, would definitely help. This is precisely what Chen offers in [Chen and Zhu, 1986].

3 Chen's projection for Robbins Monro algorithms

The basic idea consists in considering an increasing sequence $(\mathcal{K}_q)_{q\geq 0}$ of compact sets such that $\bigcup_{q=0}^{\infty} \mathcal{K}_q = \mathbb{R}^d$ and in defining a new sequence $(\bar{\theta}_n)_{n\geq 0}$ that remains in some $\mathcal{K}_q$ for a certain q. This way of forcing the sequence to remain in some compact sets was first introduced in [Chen and Zhu, 1986]. We can then define $(\bar{\theta}_n)_{n\geq 0}$

$$\bar{\theta}_{n+1} = \begin{cases} \bar{\theta}_n - \gamma_{n+1} U(\bar{\theta}_n, Z_{n+1}) & \text{if } \bar{\theta}_n - \gamma_{n+1} U(\bar{\theta}_n, Z_{n+1}) \in \mathcal{K}_{\sigma(n)}, \\ \bar{\theta}_n & \text{otherwise,} \end{cases} \tag{47}$$

where $\sigma(n)$ counts the number of projections up to step n. The following theorem, adapted from [Delyon, 1996], guarantees the convergence of this sequence to θ^*. Before stating this theorem, it might be more convenient to rewrite (47) as follows:

$$\bar{\theta}_{n+1} = \bar{\theta}_n - \gamma_{n+1} u(\bar{\theta}_n) - \gamma_{n+1} \delta M_{n+1} + \gamma_{n+1} p_{n+1} \tag{48}$$

where

$$\delta M_{n+1} = U(\bar{\theta}_n, Z_{n+1}) - u(\bar{\theta}_n), \tag{49}$$

$$\text{and } p_{n+1} = \begin{cases} U(\bar{\theta}_n, Z_{n+1}) & \text{if } \bar{\theta}_n - \gamma_{n+1} U(\bar{\theta}_n, Z_{n+1}) \notin \mathcal{K}_{\sigma(n)}, \\ 0 & \text{otherwise.} \end{cases} \tag{50}$$

3.1 A convergence result.

The following theorem states the convergence of the sequence $(\bar{\theta}_n)_n$.

Theorem 5. *Under hypotheses (H1.1), (H1.3) and*

(H5.1) *for any $q \in \mathbb{N}$, the series $\sum \gamma_n \delta M_n 1_{\|\bar{\theta}_{n-1}\| < q}$ converges,*

$(\bar{\theta}_n)_{n\geq 0}$ converges a.s. to θ^* and the projections occur a finite number of times, which means that, for any large n, $p_n = 0$.

3.2 A central limit theorem for Chen's algorithm

In this part, I present two central limit theorems which are the extensions to Chen's algorithm of the central limit theorems presented in the previous part. The notations are derived from the previous part by replacing θ_n by $\bar{\theta}_n$.

If we consider sequences $(\gamma_n)_n$ of the type $\frac{\gamma}{n+1}$, the following theorem holds

Theorem 6. *If we assume (H1.1), (H1.3), (H5.1), (H2.1), (H2.3), (H2.5) and*

(H6.1) *there exists $q > 0$ and $\rho > 0$ such that* $\sup_{\|\theta-\theta^*\|<q} \mathbb{E}\left(\|U(\theta^*, Z)\|^{2+\rho}\right) < \infty$.

(H6.2) *There exists $\eta > 0$ such that $d(\theta^*, \partial\mathcal{K}_n) > \eta\ \forall n \geq 0$.*

Then, the sequence $(\Delta_n)_{n\geq 0}$ converges in distribution to a random normal variable with mean 0 *and covariance*

$$V = \gamma \int_0^\infty \exp\left(\left(\frac{I}{2} - \gamma A\right) t\right) \Sigma \exp\left(\left(\frac{I}{2} - \gamma A\right) t\right) dt.$$

If we use sequences $\gamma_n = \frac{\gamma}{(n+1)^\alpha}$ with $1/2 < \alpha < 1$, the following theorem states the convergence of $(\Delta_n)_n$.

Theorem 7. *Under hypotheses (H1.1), (H1.3), (H2.1), (H2.4), (H2.5), (H6.1) and (H6.2), the sequence $(\Delta_n)_{n\geq 0}$ converges in distribution to a random normal variable with mean* 0 *and covariance*

$$V = \int_0^\infty \exp(-\gamma A t) \Sigma \exp(-\gamma A t) dt.$$

Before tackling the proof, let us make a few remarks on the hypotheses.

Remark 4. Hypothesis (H2.4) implies hypothesis (H5.1). Thanks to the continuity of u, it is clear that hypothesis (H6.1) implies (H2.4). Finally, hypothesis (H6.2) is the equivalent of hypothesis (H2.2).

Proof. The only difference between the classical constrained algorithm and the one using Chen's projection is that, unlike what its name suggests, Chen's projection is not non-expansive. Hence, we do not have $\|\Delta_n\|^2 \leq \left\|\tilde{\Delta}_n\right\|^2$. However, we can write $\|\Delta_{n+1}\|^2$ as

$$\begin{aligned}
\|\Delta_{n+1}\|^2 &= \|\Delta_n\|^2 \mathbf{1}_{p_{n+1}\neq 0} + \left\|\tilde{\Delta}_{n+1}\right\|^2 \mathbf{1}_{p_{n+1}=0}, \\
\|\Delta_{n+1}\|^2 &\leq \left\|\tilde{\Delta}_{n+1}\right\|^2 + \|\Delta_n\|^2 \mathbf{1}_{\bar{\theta}_n - \gamma_{n+1} U(\bar{\theta}_n, Z_{n+1}) \notin \mathcal{K}_{\sigma(n)}}.
\end{aligned}$$

Taking the conditional expectation with respect to $\mathcal{F}_n$ gives

$$\begin{aligned}
\mathbb{E}_n \|\Delta_{n+1}\|^2 &\leq \mathbb{E}_n \left\|\tilde{\Delta}_{n+1}\right\|^2 + \|\Delta_n\|^2 \mathbb{E}_n\left(\mathbf{1}_{\bar{\theta}_n - \gamma_{n+1} U(\bar{\theta}_n, Z_{n+1}) \notin \mathcal{K}_{\sigma(n)}}\right), \\
\mathbb{E}_n \|\Delta_{n+1}\|^2 \mathbf{1}_{A_n} &\leq \mathbb{E}_n \left\|\tilde{\Delta}_{n+1}\right\|^2 \mathbf{1}_{A_n} + \|\Delta_n\|^2 \mathbf{1}_{A_n} \mathbb{E}_n\left(\mathbf{1}_{\bar{\theta}_n - \gamma_{n+1} U(\bar{\theta}_n, Z_{n+1}) \notin \mathcal{K}_{\sigma(n)}}\right), \\
\mathbb{E}\left(\|\Delta_{n+1}\|^2 \mathbf{1}_{A_{n+1}}\right) &\leq \mathbb{E}\left(\left\|\tilde{\Delta}_{n+1}\right\|^2 \mathbf{1}_{A_n}\right) + \\
&\quad \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n} \mathbb{E}_n\left(\mathbf{1}_{\bar{\theta}_n - \gamma_{n+1} U(\bar{\theta}_n, Z_{n+1}) \notin \mathcal{K}_{\sigma(n)}} \mathbf{1}_{A_n}\right)\right).
\end{aligned} \tag{51}$$

The conditional expectation can be rewritten

$$\begin{aligned}
\mathbb{E}_n\left(\mathbf{1}_{\bar{\theta}_n - \gamma_{n+1} U(\bar{\theta}_n, Z_{n+1}) \notin \mathcal{K}_{\sigma(n)}} \mathbf{1}_{A_n}\right) &\leq \mathbb{P}_n\left(\gamma_{n+1}\left\|U(\bar{\theta}_n, Z_{n+1})\right\| \geq d\left(\theta_n, \partial\mathcal{K}_{\sigma(n)}\right)\right) \mathbf{1}_{A_n}, \\
&\leq \frac{\gamma_{n+1}^2}{d\left(\theta_n, \partial\mathcal{K}_{\sigma(n)}\right)^2} \mathbb{E}_n\left(\left\|U(\bar{\theta}_n, Z_{n+1})\right\|^2\right) \mathbf{1}_{A_n}.
\end{aligned} \tag{52}$$

Moreover, using the triangle inequality we have

$$d\left(\theta_n, \partial\mathcal{K}_{\sigma(n)}\right) \geq d\left(\theta^*, \partial\mathcal{K}_{\sigma(n)}\right) - \|\theta_n - \theta^*\|. \tag{53}$$

Using hypothesis (H6.2), $d\left(\theta^*, \partial\mathcal{K}_{\sigma(n)}\right) < \eta$ and on A_n, $\quad \|\theta_n - \theta^*\| \le x_0$. Hence,

$$d\left(\theta_n, \partial\mathcal{K}_{\sigma(n)}\right) \ge \eta - x_0. \tag{54}$$

We can choose x_0 smaller than $\eta/2$ for instance. Combining equation (54) and (52), we obtain

$$\mathbb{E}_n\left(\mathbf{1}_{\bar\theta_n - \gamma_{n+1}U(\bar\theta_n, Z_{n+1}) \notin \mathcal{K}_{\sigma(n)}} \mathbf{1}_{A_n}\right) \le \frac{4\gamma_{n+1}^2}{\eta^2} \mathbb{E}_n\left(\left\|U(\bar\theta_n, Z_{n+1})\right\|^2\right) \mathbf{1}_{A_n}. \tag{55}$$

Moreover, since Z_{n+1} is independent of $\mathcal{F}_n$

$$\mathbb{E}_n\left(\left\|U(\bar\theta_n, Z_{n+1})\right\|^2\right) \mathbf{1}_{A_n} \quad \le \quad \sup_\theta \mathbb{E}\left(\|U(\theta, Z)\|^2\right) \mathbf{1}_{\|\theta-\theta^*\| < x_0}. \tag{56}$$

The right-hand side of the inequality is deterministic so thanks to hypothesis (H6.1)

$$\sup_\omega \sup_n \mathbb{E}_n\left(\left\|U(\bar\theta_n, Z_{n+1})\right\|^2\right) \mathbf{1}_{A_n} < \infty. \tag{57}$$

Hence, from equation (56) we can deduce that $\sup_\omega \sup_n \mathbb{E}_n\left(\left\|U(\bar\theta_n, Z_{n+1})\right\|^2\right)$ is bounded by a constant times γ_n^2. From equation (19) we can write, in the present case,

$$\mathbb{E}\left(\left\|\tilde\Delta_{n+1}\right\|^2 \mathbf{1}_{A_n}\right) - \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) \le -\gamma_n \frac{\lambda}{2} \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) + c\gamma_n. \tag{58}$$

Combining equations (51) and (58), we get

$$\mathbb{E}\left(\|\Delta_{n+1}\|^2 \mathbf{1}_{A_{n+1}}\right) \quad \le \quad \left(1 + c'\gamma_n^2 - \gamma_n \frac{\lambda}{2}\right) \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) + c\gamma_n. \tag{59}$$

If n is large enough, $c'\gamma_n^2 \le \gamma_n \frac{\lambda}{4}$. We finally get the desired inequality

$$\mathbb{E}\left(\|\Delta_{n+1}\|^2 \mathbf{1}_{A_{n+1}}\right) \le \left(1 - \gamma_n \frac{\lambda}{4}\right) \mathbb{E}\left(\|\Delta_n\|^2 \mathbf{1}_{A_n}\right) + c\gamma_n. \tag{60}$$

The rest of the proof can be reproduced from the classical constrained algorithm. □

4 Conclusion

For fast decreasing step sequences — typically $\frac{\gamma}{n}$ — a central limit theorem holds provided that γ is large enough. This means that to converge with a speed of roughly $\sqrt{n}$, the algorithm requires a sufficient noise, which demands that γ is large enough. This type of sequence leads to the fastest convergence speed.

For slower converging step sequences — $\frac{\gamma}{n^\alpha}$ with $1/2 < \alpha < 1$ — the weight of the noise does not interfere with the existence of a central limit theorem or not. However, the convergence speed is of order $n^{\alpha/2}$. The faster the series $\sum_n \gamma_n$ diverges, the smaller the rate of convergence for Robbins Monro algorithms is.

A A convergence result for integrals

Proposition 1. *Let f be a piecewise continuous function defined on $\mathbb{R}^+$ such that $f(u) \underset{u\to\infty}{\longrightarrow} l$, where l is a real number.*
Then

$$\int_0^t e^{q(u-t)} f(u) du \underset{t\to\infty}{\longrightarrow} \frac{l}{q} \ \forall t > 0.$$

Proof. Let us fix an $\varepsilon > 0$. There exists $T_1 > 0$ such that $\forall\, u \geq T_1,\ |f(u) - l| \leq \varepsilon$. There also exists $T_2 > T_1$ such that for all $t \geq T_2$ and all $u \leq T_1$, we have $e^{q(T_1-t)} \leq \varepsilon$. Hence, for any $t \geq T_2$,

$$\int_0^t e^{q(u-t)} f(u) du = \int_0^{T_1} e^{q(u-t)} f(u) du + \int_{T_1}^t e^{q(u-t)} f(u) du. \tag{61}$$

The first integral on the right side is bounded by $\sup_{[0,T_1]} |f| \dfrac{e^{q(T_1-t)}}{q}$, which is by assumption bounded in turn by $\frac{\varepsilon}{q} \sup_{[0,\infty[} |f|$. The supremum of $|f|$ exists since f is piecewise continuous and has a limit in the neighbourhood of ∞. Let us now handle the second integral in (61).

$$\begin{aligned} \int_{T_1}^t e^{q(u-t)} \left| f(u) - l \right| du &\leq \varepsilon \int_{T_1}^t e^{q(u-t)}, & (62)\\ &\leq \varepsilon \frac{1}{q}. & (63) \end{aligned}$$

Moreover, $\int_{T_1}^t e^{q(u-t)} l = \frac{l}{q}(1 - e^{T_1 - t})$. So $\left| \int_{T_1}^t e^{q(u-t)} f(u) - \frac{l}{q} \right| \leq \varepsilon \frac{1}{q}(1 + l)$. This proves the result announced in the above lemma. □

B Integration and tightness

The following proposition is an extension of proposition 1.

Proposition 2. *Let $X(t)_{t\geq 0}$ be a piecewise constant càdlàg process. We assume that $X(t) \underset{t\to\infty}{\longrightarrow} x \in \mathbb{R}$ in probability. Let us define $Y_t = \int_0^t e^{u-t} X_u du$. Then $Y(t) \xrightarrow[t\to\infty]{\mathbb{P}} x$.*

Proof. Since the convergences in probability and in distribution to a deterministic constant are equivalent, we will prove the convergence in distribution. The first step is to prove that $(Y_t)_t$ is tight. Let $M > 0$,

$$\begin{aligned} \mathbb{P}(|Y_t| > M) &\leq \mathbb{P}\left(\int_0^t e^{u-t} \left| X_u \right| du > M \right), \\ &\leq \mathbb{P}\left(\sup_{u\in[0,t]} \left| X_u \right| > M \right). & (64) \end{aligned}$$

Since $X(t)_{t\geq 0}$ is a piecewise constant càdlàg process, the following inclusion holds

$$\left\{ \sup_{u\in[0,t]} \left| X_u \right|; t \geq 0 \right\} \subset \left\{ X_s; s \geq 0 \right\}.$$

The last set is tight since $X(\cdot)$ converges in probability. Hence, the sequence of random variables $(Y_t)_t$ is tight.

The tightness of $(Y_t)_t$ enables us to extract a converging subsequence $(Y_{t_k})_k$. The limit is denoted L.

The second step consists in proving that $L = x$ almost surely. Relying on one more extraction, the sequence $(t_k)_k$ is strictly increasing and can be chosen such that $|t_{k+1} - t_k| \geq 1$ for every k.

$\int_0^{t_k} e^{u-t_k}(X_u - L)du$ converges in distribution to zero and hence in probability. The difference between two consecutive terms of the sequence tends to zero in probability.

$$\int_0^{t_k} (1 - e^{t_k - t_{k+1}}) e^{u-t_k}(X_u - L)du + \int_{t_k}^{t_{k+1}} e^{u-t_{k+1}}(X_u - L)du \xrightarrow[t\to\infty]{\mathbb{P}} 0. \tag{65}$$

The first integral tends to zero in probability, so does the second. Now using the mean formula, the second integral can be written

$$\int_{t_k}^{t_{k+1}} e^{u-t_{k+1}}(X_u - L)du = \int_{t_k}^{t_{k+1}} e^{u-t_{k+1}}(c_k - L)du, \tag{66}$$

where

$$\inf_{u\in[t_k,t_{k+1}]} X_u \le c_k \le \sup_{u\in[t_k,t_{k+1}]} X_u. \tag{67}$$

Moreover, we had assumed that $|t_{k+1} - t_k| \ge 1$ for every k so the convergence in probability of the integral implies that $c_k - L \xrightarrow[t\to\infty]{\mathbb{P}} 0$. Since the process X is piecewise constant, its suprema are attained. So the convergence in probability to x of X_t implies the ones of $\inf_{u\in[t_k,t_{k+1}]} X_u$ and of $\sup_{u\in[t_k,t_{k+1}]} X_u$. So relation (67) enables us to state that c_k converges in probability to x. Then, the convergence of $c_k - L$ to zero in probability achieves to prove that $L = x$ almost surely. Therefore, any converging subsequence of $(Y_t)_t$ converges to zero in distribution. The set of all closure values of $\{Y_t; t \ge 0\}$ is the singleton $\{x\}$. Therefore, the whole sequence converges to x in distribution and consequently in probability. □

C Tightness and convergence in probability

Proposition 3. *Let $(X_n)_n$ be a sequence of random variables in $\mathbb{R}^d$ converging in probability to zero and $(Y_n)_n$ a tight sequence of random variables in $\mathbb{R}^d$. Then the sequence $(\langle X_n, Y_n\rangle)_n$ converges in probability to zero.*

Proof. For the sake of clearness, the proof will be done for real random variables and not vector valued ones. The same arguments would hold anyway. Since $(Y_n)_n$ is tight,

$$\forall \varepsilon > 0 \quad \exists K_\varepsilon \subset \mathbb{R}^d \text{ such that } \mathbb{P}(Y_n \notin \mathcal{K}_\varepsilon) \le \varepsilon.$$

Let $\eta > 0$ and $\varepsilon > 0$.

$$\begin{aligned}
\mathbb{P}(|X_n Y_n| > \eta) &= \mathbb{E}\left(\mathbf{1}_{|X_n Y_n| \mathbf{1}_{Y_n \in \mathcal{K}_\varepsilon} + |X_n Y_n| \mathbf{1}_{Y_n \notin \mathcal{K}_\varepsilon} > \eta}\right), \\
&\le \mathbb{E}\left(\mathbf{1}_{|X_n Y_n| \mathbf{1}_{Y_n \in \mathcal{K}_\varepsilon} > \eta/2}\right) + \mathbb{E}\left(\mathbf{1}_{|X_n Y_n| \mathbf{1}_{Y_n \notin \mathcal{K}_\varepsilon} > \eta/2}\right), \\
&\le \mathbb{P}\left(|X_n Y_n|\, \mathbf{1}_{Y_n \in \mathcal{K}_\varepsilon} > \eta/2\right) + \mathbb{P}\left(|X_n Y_n|\, \mathbf{1}_{Y_n \notin \mathcal{K}_\varepsilon} > \eta/2\right), \\
&\le \mathbb{P}\left(|X_n|\, \delta(K_\varepsilon) > \eta/2\right) + \mathbb{P}\left(\mathbf{1}_{Y_n \notin \mathcal{K}_\varepsilon} > 0\right), \\
&\le \mathbb{P}\left(|X_n|\, \delta(K_\varepsilon) > \eta/2\right) + \mathbb{P}\left(Y_n \notin \mathcal{K}_\varepsilon\right),
\end{aligned} \tag{68}$$

Since X_n tends to zero in probability, for any $n > N_\varepsilon$, the first probability in (68) is smaller than ε and by definition the second expectation is also smaller than ε. This proves the result announced in the proposition. □

References

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

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

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

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

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

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

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

[Robbins and Siegmund, 1971] Robbins, H. and Siegmund, D. (1971). A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing methods in statistics (Proc. Sympos., Ohio State Univ., Columbus, Ohio, 1971)*, pages 233–257. Academic Press, New York.

[Rogers and Williams, 2000] Rogers, L. and Williams, D. (2000). *Diffusions, Markov Processes and Martingales*, volume 2 (Itô Calculus). Cambridge University Press.

Contents

1 Introduction **1**

2 Robbins Monro algorithms with projection on a constraint set **2**
- 2.1 Presentation . . . 2
- 2.2 A central limit theorem for a constrained Robbins Monro algorithm . . . 3
  - 2.2.1 the case $\alpha = 1$ . . . 3
  - 2.2.2 the case $\frac{1}{2} < \alpha < 1$ . . . 9

3 Chen's projection for Robbins Monro algorithms **10**
- 3.1 A convergence result. . . . 10
- 3.2 A central limit theorem for Chen's algorithm . . . 11

4 Conclusion **12**

A A convergence result for integrals **13**

B Integration and tightness **13**

C Tightness and convergence in probability **14**

References **15**