CENTRAL LIMIT THEOREMS FOR STOCHASTIC APPROXIMATION WITH CONTROLLED MARKOV CHAIN DYNAMICS

Gersende Fort[1]

Abstract. This paper provides a Central Limit Theorem (CLT) for a process $\{\theta_n, n \geq 0\}$ satisfying a stochastic approximation (SA) equation of the form $\theta_{n+1} = \theta_n + \gamma_{n+1} H(\theta_n, X_{n+1})$; a CLT for the associated average sequence is also established. The originality of this paper is to address the case of controlled Markov chain dynamics $\{X_n, n \geq 0\}$ and the case of multiple targets. The framework also accomodates (randomly) truncated SA algorithms. Sufficient conditions for CLT's hold are provided as well as comments on how these conditions extend previous works (such as independent and identically distributed dynamics, the Robbins−Monro dynamic or the single target case). The paper gives a special emphasis on how these conditions hold for SA with controlled Markov chain dynamics and multiple targets; it is proved that this paper improves on existing works.

1. Introduction

Stochastic Approximation (SA) algorithms were introduced for finding roots of an unknown function h (for recent surveys on SA, see *e.g.* [6,9,20,22,28]). SA defines iteratively a sequence $\{\theta_n, n \geq 0\}$ by the update rule

$$\theta_{n+1} = \theta_n + \gamma_{n+1} \Xi_{n+1}, \tag{1.1}$$

where $\{\gamma_n, n \geq 1\}$ is a sequence of deterministic step-size and Ξ_{n+1} is a random variable (r.v.) standing for a noisy measurement of the unknown quantity $h(\theta_n)$.

Our aim is to establish the rate of convergence of the sequence $\{\theta_n, n \geq 0\}$ to a limiting point $\theta_\star$ in the following framework.

Let $\Theta \subseteq \mathbb{R}^d$; the sequence $\{\theta_n, n \geq 0\}$ is a Θ-valued random sequence defined on the filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \{\mathcal{F}_n, n \geq 0\})$ and given by

$$\theta_{n+1} = \theta_n + \gamma_{n+1} \left(h(\theta_n) + e_{n+1} + r_{n+1} \right), \qquad \theta_0 \in \Theta;$$

where $h : \Theta \to \mathbb{R}^d$ is a measurable function, $\{e_n, n \geq 1\}$ is a $\mathcal{F}_n$-adapted $\mathbb{P}$-martingale increment sequence and $\{r_n, n \geq 1\}$ is a vanishing $\mathcal{F}_n$-adapted random sequence. Such a general description covers many SA algorithms: as discussed below (see Sect. 2.1), it covers the case when Ξ_{n+1} is of the form $H(\theta_n, X_{n+1})$ where $\{X_n, n \geq 1\}$

---

[1] LTCI, CNRS & TELECOM ParisTech, 46 rue Barrault, 75634 Paris cedex 13, France. gersende.fort@telecom-paristech.fr

are independent and identically distributed (i.i.d.) r.v. such that (s.t.) $\mathbb{E}\left[H(\theta, X)\right] = h(\theta)$; and the more general case when $\{X_n, n \geq 1\}$ is an adapted (non stationary) Markov chain with transition kernel driven by the current value of the SA sequence $\{\theta_n, n \geq 0\}$. It also covers the case of fixed truncated and randomly truncated SA algorithms *i.e.* situations when given a (possibly random) sequence of subsets $\{\mathcal{K}_n, n \geq 0\}$ of Θ, the update rule is given by

$$\theta_{n+1} = \begin{cases} \theta_n + \gamma_{n+1}\Xi_{n+1} , & \text{if } \theta_n + \gamma_{n+1}\Xi_{n+1} \in \mathcal{K}_{n+1}. \\ \theta_0 & \text{otherwise.} \end{cases} \tag{1.2}$$

Such a truncated algorithm is used for example to solve optimization problem on a constraint set Θ (in this case, $\mathcal{K}_n = \Theta$ for any n), or to ensure stability of the random sequence $\{\theta_n, n \geq 0\}$ in situations where the location of the sought-for root is unknown (in this case, $\mathcal{K}_n$ is an increasing sequence of sets, see [9, 10], Chap. 2).

Our second aim is to extend the previous results to the case of multiple targets: we provide asymptotic convergence rates of $\{\theta_n, n \geq 0\}$ to a point $\theta_\star$ given the event $\{\lim_q \theta_q = \theta_\star\}$ for some $\theta_\star$ in the interior of Θ. Note that this paper is devoted to convergence rates so that sufficient conditions for the convergence is out of the scope of the paper; for convergence, the interested reader can refer to [2–4, 6, 9, 13].

The originality of this paper consists in deriving rates of convergence in a new framework characterized by (i) general assumptions on the noisy measurement Ξ_{n+1} of $h(\theta_n)$ which weaken the conditions in the literature and (ii) the multiple targets problem. In Section 2.2, our framework will be carefully compared to the literature.

We derive sufficient conditions on the step-size sequence $\{\gamma_n, n \geq 1\}$, on the random sequences $\{e_n, r_n, n \geq 1\}$ and on the limiting point $\theta_\star$ so that $\gamma_n^{-1/2}(\theta_n - \theta_\star)$ converges in distribution under the conditional probability $\mathbb{P}(\cdot | \lim_q \theta_q = \theta_\star)$. The limiting distribution is a (mixture of) centered Gaussian distribution(s) and this distribution is explicitly characterized. We also address the rate of convergence of the associated averaged process $\{\bar{\theta}_n, n \geq 0\}$ defined by

$$\bar{\theta}_n \stackrel{\text{def}}{=} \frac{1}{n+1} \sum_{k=0}^{n} \theta_k. \tag{1.3}$$

We prove that this averaged sequence reaches the optimal rate and the optimal variance (in a sense discussed below); such a result was already established in the literature in a more restrictive framework.

The paper is organized as follows. Section 2 (resp. Sect. 3) is devoted to the SA sequence $\{\theta_n, n \geq 0\}$ (resp. the averaged SA sequence $\{\bar{\theta}_n, n \geq 0\}$). We successively introduce the assumptions, comment these conditions, compare our framework to the literature and state a Central Limit Theorem (CLT). In Section 4, our results are applied to a randomly truncated SA algorithm with controlled Markov chain dynamics; since our conditions are quite weak, we are able to obtain better convergence rates than the rates obtained in Delyon [11]. All the proofs are postponed in Section 5.

2. A CENTRAL LIMIT THEOREM FOR STOCHASTIC APPROXIMATION

2.1. Assumptions

Let $\Theta \subseteq \mathbb{R}^d$. We consider the $\mathbb{R}^d$-valued sequence satisfying for $n \geq 0$,

$$\theta_{n+1} = \theta_n + \gamma_{n+1}h(\theta_n) + \gamma_{n+1}e_{n+1} + \gamma_{n+1}r_{n+1} , \qquad \theta_0 \in \Theta; \tag{2.1}$$

and we establish a Central Limit Theorem along sequences $\{\theta_n, n \geq 0\}$ converging to some point $\theta_\star \in \Theta$ which is a root of the function h. We assume the following conditions on the attractive target $\theta_\star$.

C1 (a) $\theta_\star$ is in the interior of Θ and $h(\theta_\star) = 0$.
　　(b) The mean field $h : \Theta \to \mathbb{R}^d$ is measurable and twice continuously differentiable in a neighborhood of $\theta_\star$.
　　(c) The gradient $\nabla h(\theta_\star)$ is a Hurwitz matrix. Denote by $-L$, $L > 0$, the largest real part of its eigenvalues.

Let $\{e_n, n \geq 1\}$ be a $\mathbb{R}^d$-valued random variables defined on the filtered space $(\Omega, \mathcal{A}, \mathbb{P}, \{\mathcal{F}_n, n \geq 0\})$. We will denote by $|\cdot|$ the Euclidean norm on $\mathbb{R}^d$; and by x^T the transpose of a matrix x. By convention, vectors are column-vectors. For a set A, $\mathbb{1}_A$ is the indicator function. It is assumed

C2 (a) $\{e_n, \ n \geq 1\}$ is a $\mathcal{F}_n$-adapted $\mathbb{P}$-martingale-increment sequence *i.e.* $\mathbb{E}[e_n|\mathcal{F}_{n-1}] = 0$ $\mathbb{P}$-almost surely.

(b) For any $m \geq 1$, there exists a sequence of measurable sets $\{\mathcal{A}_{m,k}, k \geq 0\}$ such that $\mathcal{A}_{m,k} \in \mathcal{F}_k$ and there exists $\tau > 0$ such that

$$\sup_{k \geq 0} \mathbb{E}\left[|e_{k+1}|^{2+\tau} \mathbb{1}_{\mathcal{A}_{m,k}}\right] < \infty.$$

In addition, for any $m \geq 1$, $\lim_k \mathbb{1}_{\mathcal{A}_{m,k}} \mathbb{1}_{\lim_q \theta_q = \theta_\star} = \mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star}$ and the limiting set satisfies $\lim_m \mathbb{P}(\mathcal{A}_m | \lim_q \theta_q = \theta_\star) = 1$.

(c) $\mathbb{E}\left[e_{k+1} e_{k+1}^T | \mathcal{F}_k\right] = U_\star + D_k^{(1)} + D_k^{(2)}$ where $U_\star$ is a symmetric positive definite matrix and

$$\begin{cases} D_k^{(1)} \xrightarrow{\text{a.s.}} 0, & \text{on the set } \{\lim_q \theta_q = \theta_\star\} \\ \lim_n \gamma_n \ \mathbb{E}\left[\left|\sum_{k=1}^n D_k^{(2)}\right| \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m}\right] = 0; \end{cases} \tag{2.2}$$

the sequence $\{\mathcal{A}_m, m \geq 1\}$ is defined in C2b.

We will show (see Rem. 5.3 in Sect. 5) that the condition on the r.v. $\{D_k^{(2)}, k \geq 1\}$ can be replaced with: $D_k^{(2)} = D_k^{(2,a)} + D_k^{(2,b)}$

$$\lim_n \gamma_n \ \mathbb{E}\left[\left|\sum_{k=1}^n D_k^{(2,a)} \mathbb{1}_{\mathcal{A}_{m,k}} \mathbb{1}_{A_k}\right| + \left|\sum_{k=1}^n D_k^{(2,b)}\right| \mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star}\right] = 0, \quad \forall m \geq 1, \tag{2.3}$$

where $\{A_k, k \geq 1\}$ is any $\mathcal{F}_k$-adapted sequence of sets satisfying $\lim_k \mathbb{1}_{A_k} = \mathbb{1}_{\lim_q \theta_q = \theta_\star}$; and $\mathcal{A}_{m,k}$ is given by C2b.

For a sequence of $\mathbb{R}^d$-valued r.v. $\{Z_n, n \geq 0\}$, we write $Z_n = O_{w.p.1.}(1)$ if $\sup_n |Z_n| < \infty$ w.p.1; and $Z_n = o_{L^p}(1)$ if $\lim_n \mathbb{E}[|Z_n|^p] = 0$. Let $\{r_n, n \geq 1\}$ be a $\mathbb{R}^d$-valued random variables defined on the filtered space $(\Omega, \mathcal{A}, \mathbb{P}, \{\mathcal{F}_n, n \geq 0\})$.

C3 r_n is $\mathcal{F}_n$-adapted. $r_n = r_n^{(1)} + r_n^{(2)}$ with, for any $m \geq 1$,

$$\begin{cases} \gamma_n^{-1/2} r_n^{(1)} \ \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} = O_{w.p.1}(1) o_{L^1}(1), \\ \sqrt{\gamma_n} \sum_{k=1}^n r_k^{(2)} \ \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} = O_{w.p.1}(1) o_{L^1}(1). \end{cases}$$

The sequence $\{\mathcal{A}_m, m \geq 1\}$ is defined in C2b. The last assumption is on the step-size sequence.

C4 One of the following conditions is satisfied:

(a) $\sum_k \gamma_k = +\infty$, $\sum_k \gamma_k^2 < \infty$ and $\log(\gamma_{k-1}/\gamma_k) = o(\gamma_k)$.

(b) $\sum_k \gamma_k = +\infty$, $\sum_k \gamma_k^2 < \infty$ and there exists $\gamma_\star > 1/(2L)$ such that $\log(\gamma_{k-1}/\gamma_k) \sim \gamma_k/\gamma_\star$.

2.2. Comments on the assumptions

The framework described by (2.1) and the conditions C1 to C4 is general enough to cover many scenarios studied in the literature and to address new ones.

For SA algorithms (1.1) with $\Xi_{n+1} = H(\theta_n, X_{n+1})$, $\{X_n, n \geq 1\}$ i.i.d. r.v. (and independent of θ_0) such that $h(\theta) = \mathbb{E}[H(\theta, X)]$, equation (2.1) is satisfied with

$$e_{n+1} = H(\theta_n, X_{n+1}) - h(\theta_n), \qquad r_{n+1} = 0;$$

and $\mathbb{E}[e_{n+1}|\mathcal{F}_n] = 0$. Our framework also addresses the case when $\{X_n, n \geq 1\}$ is a $\mathcal{F}_n$-adapted controlled Markov chain *i.e.* when there exists a family of transition kernels $\{Q_\theta, \theta \in \Theta\}$ such that

$$\mathbb{P}(X_{n+1} \in \cdot | \mathcal{F}_n) = Q_{\theta_n}(X_n, \cdot),$$

each kernel possessing an invariant probability distribution π_θ and $h(\theta) = \int H(\theta, x) \pi_\theta(\mathrm{d}x)$ − hereafter, these algorithms will be called "SA with controlled Markov chain dynamics". Introduce the solution $\widehat{H}_\theta$ of the Poisson equation $H(\theta, \cdot) - h(\theta) = \widehat{H}_\theta - Q_\theta \widehat{H}_\theta$ (see *e.g.* [18], Chap. 8 or [24], Chap. 17), and set

$$e_{n+1} = \widehat{H}_{\theta_n}(X_{n+1}) - Q_{\theta_n} \widehat{H}_{\theta_n}(X_n), \qquad r_{n+1} = Q_{\theta_n} \widehat{H}_{\theta_n}(X_n) - Q_{\theta_n} \widehat{H}_{\theta_n}(X_{n+1});$$

then $\mathbb{E}[e_{n+1}|\mathcal{F}_n] = 0$ $\mathbb{P}$-almost surely. We will provide in Section 4 sufficient conditions on the transition kernels Q_θ so that these sequences $\{e_n, r_n, n \geq 1\}$ exist and satisfy the conditions C2 and C3. Note that the i.i.d. case is a special case of the controlled Markov chain framework (set $Q_\theta = \pi_\theta = \pi$ for any θ); and the so-called Robbins−Monro case corresponds to $Q_\theta = \pi_\theta$ for any θ.

Truncated SA algorithms (1.2) can be written as

$$\theta_{n+1} = \theta_n + \gamma_{n+1} \Xi_{n+1} + (\theta_0 - \theta_n - \gamma_{n+1} \Xi_{n+1}) \mathbb{1}_{\theta_n + \gamma_{n+1} \Xi_{n+1} \notin \mathcal{K}_{n+1}};$$

in most (if not any) proof of convergence of this sequence to limiting points in the interior of Θ, the first step consists in proving that $\mathbb{P}$-almost-surely, the number of truncations is finite (see *e.g.* Andrieu *et al.* [2], Thm. 1). Therefore, the term $(\theta_0 - \theta_n - \gamma_{n+1} \Xi_{n+1}) \mathbb{1}_{\theta_n + \gamma_{n+1} \Xi_{n+1} \notin \mathcal{K}_{n+1}}$ is null for any large n on the set $\{\lim_q \theta_q = \theta_\star\}$ thus showing that it is part of $\gamma_{n+1} r_{n+1}^{(1)}$ in the expansion (2.1).

The condition C1 considers a limiting target $\theta_\star$ which is assumed to be stable and such that the linear term in the Taylor's expansion of h at $\theta_\star$ does not vanish (see condition C1c). Results for the case of vanishing linear term can be found in Chen ([9], Sect. 3.2). When h is a gradient function so that the SA algorithm is a stochastic gradient procedure, the condition C1a assumes that $\theta_\star$ is a root of the gradient. Therefore, our assumptions do not cover the case of constrained optimization problem with solutions on the boundaries of the constraint set Θ. For rates of convergence for these constrained SA algorithms, see *e.g.* Buche and Kushner [8].

The conditions C2 and C3 are designed to address the case of multiple targets, a framework which improves on many published results. It is usually assumed in the literature that there is an unique limiting target (see *e.g.* Fabian [14], Kushner and Huang [21], Bouton [7], Buche and Kushner [8], Chen [9] Chap. 3, and Lelong [23]). While we are interested in proving a Central Limit Theorem given the tail event $\{\lim_q \theta_q = \theta_\star\}$, it is assumed in C2a that the r.v. e_{n+1} in the expansion (2.1) is a martingale increment with respect to (w.r.t.) the probability $\mathbb{P}$. As discussed above, such an expansion is easily verified. Note that since the event $\{\lim_q \theta_q = \theta_\star\}$ is in the tail σ-field $\sigma(\bigvee_n \mathcal{F}_n)$, it is not true that $\{e_n, n \geq 1\}$ are martingale-increments w.r.t. the probability $\mathbb{P}(\cdot | \lim_q \theta_q = \theta_\star)$. Therefore, our framework is not a special case of the single target framework.

The main use of C2 is to prove that the sequence $\{e_n, n \geq 1\}$ satisfies a CLT under the conditional distribution $\mathbb{P}(\cdot | \lim_q \theta_q = \theta_\star)$. We could weaken some of the assumptions, for example by relaxing the $2+\tau$-moment condition C2b which is a way to easily check the Lindeberg condition for martingale difference array. Nevertheless, our goal is not only to state a theorem with weaker assumptions but also to provide easy-to-check conditions.

When there exists $\tau > 0$ such that $\sup_{k \geq 1} \mathbb{E}[|e_k|^{2+\tau}] < \infty$, C2b is satisfied with $\mathcal{A}_m = \mathcal{A}_{m,k} = \Omega$. When there exist $\tau, \delta > 0$ such that

$$\sup_{k \geq 0} \mathbb{E}[|e_{k+1}|^{2+\tau} \mathbb{1}_{|\theta_k - \theta_\star| \leq \delta}] < \infty, \tag{2.4}$$

then C2b is satisfied with $\mathcal{A}_{m,k} = \bigcap_{m \leq j \leq k} \{|\theta_j - \theta_\star| \leq \delta\}$ and $\mathcal{A}_m = \bigcap_{j \geq m} \{|\theta_j - \theta_\star| \leq \delta\}$. In most contributions, rates of convergence are derived under the condition (2.4) (see *e.g.* the recent works by Pelletier [25] and Lelong [23]). This framework is too restrictive to address the case of SA with controlled Markov chain dynamics when the ergodic properties of the transition kernels $\{Q_\theta, \theta \in \Theta\}$ are not uniform in θ. Our assumption C2b is designed to address this framework as it will be shown in Section 4.

C2c is an assumption on the conditional variance of the martingale-increment term e_n, which is more general than what is usually assumed. In Zhu [29], Pelletier [25], Chen [9] and Lelong [23] (resp. in Delyon [11]), a CLT is proved under the assumption that $\mathbb{E}[e_{k+1} e_{k+1}^T | \mathcal{F}_k] = U_\star + D_k^{(1)}$ (resp. $\mathbb{E}[e_{k+1} e_{k+1}^T | \mathcal{F}_k] = U_\star + D_k^{(2)}$) where $D_k^{(1)}, D_k^{(2)}$ satisfy (2.2) and $U_\star$ is a deterministic symmetric positive definite matrix. The improvement is in the

combination $D_k^{(1)} + D_k^{(2)}$. The introduction of the term $D_k^{(2)}$ is a strong improvement since it covers the case of SA with controlled Markov chain dynamic: observe indeed that in this case $\mathbb{E}\left[e_{k+1}e_{k+1}^T|\mathcal{F}_k\right]$ is a function of (X_k, θ_k) and it is really unlikely that this term converges almost-surely to a (random) variable along the set $\{\lim_q \theta_q = \theta_\star\}$. Allowing an additional term $D_k^{(2)}$ such that the sum $\sum_{k=1}^n D_k^{(2)}$ converges in some sense to zero introduces more flexibility (see Sect. 4 for more details). We will also show in Section 4 how our framework improves on Delyon [11]. Examples of SA algorithm where C2c holds with resp. Robbins−Monro and controlled Markov chain dynamics can be found resp. in Bianchi *et al.* [5] and Fort *et al.* [15].

Kushner and Huang [21] establish a CLT (as a consequence of some weak convergence of a suitable continuous interpolation of the sequence $(\theta_n - \theta_\star)/\sqrt{\gamma_n}$) for the algorithm (1.1) when Ξ_{n+1} is of the form $H(\theta_n, X_{n+1})$. Their proof relies on a linearization of the algorithm (1.1) but with no explicit introduction of a martingale-increment noise term; nevertheless, their conditions require that some remainder terms vanish almost-surely, as well as some kind of stationary assumption on the sequence $(X_n)_n$ (see [21], Assumption A.2 which is really restrictive).

Examples of sequences satisfying the condition C4 are the polynomial ones. The step size $\gamma_n \sim \gamma_\star n^{-a}$ for $a \in (1/2, 1)$ satisfies C4a. The step size $\gamma_n \sim \gamma_\star/n$ satisfies C4b; note that the condition on $(\gamma_\star, L)$ is well known in the literature (see *e.g.* Chen [9], assumption A3.1.4).

2.3. Main result

Theorem 2.1. *Choose $\theta_0 \in \Theta$ and consider the sequence $\{\theta_n, n \geq 0\}$ given by (2.1). Assume* C1, C2, C3 *and* C4. *Let V be the positive definite matrix satisfying*

$$\begin{cases} V\nabla h(\theta_\star)^T + \nabla h(\theta_\star)V = -U_\star , & \text{in case } C4a , \\ V(\text{Id} + 2\gamma_\star \nabla h(\theta_\star)^T) + (\text{Id} + 2\gamma_\star \nabla h(\theta_\star))V = -2\gamma_\star U_\star , & \text{in case } C4b. \end{cases}$$

Under the conditional probability $\mathbb{P}\left(\cdot|\lim_q \theta_q = \theta_\star\right)$, $\{\gamma_n^{-1/2} \ (\theta_n - \theta_\star), n \geq 1\}$ converges in distribution to a centered multidimensional Gaussian distribution with covariance matrix V.

Given matrices A, E, existence of a solution to the equation $VA + A^T V = -E$ is solved by the Lyapunov theorem (see *e.g.* Horn and Johnson [19], Thm. 2.2.1). When A is a (negative) stable real matrix and E is positive definite, then there exists an unique positive definite matrix V satisfying the Lyapunov equation $VA + A^T V = -E$ (see *e.g.* Horn and Johnson [19], Thm. 2.2.3.).

Sketch of the proof of Theorem 2.1. The proof of Theorem 2.1 is detailed in Section 5. The key ingredient is the Central Limit Theorem for martingale arrays. As commented in Section 2.2, e_n is not a martingale-increment w.r.t. the conditional probability $\mathbb{P}(\cdot|\lim_q \theta_q = \theta_\star)$. To overcome this technical difficulty, we use that

$$e_{n+1} = e_{n+1}\mathbb{1}_{A_n} + e_{n+1}\left(1 - \mathbb{1}_{A_n}\right) \tag{2.5}$$

where $\{A_n, n \geq 1\}$ is a $\mathcal{F}_n$-adapted sequence of sets converging to $\{\lim_q \theta_q = \theta_\star\}$ (such a sequence always exists, see Lem. 5.6). Along the event $\{\lim_q \theta_q = \theta_\star\}$, the second term in the right hand side (rhs) of (2.5) is null for any n larger than some almost-surely finite random time.

We write $\theta_n - \theta_\star = \mu_n + \rho_n$, where μ_n satisfies the equation

$$\mu_{n+1} = (\text{Id} + \gamma_{n+1}\nabla h(\theta_\star))\mu_n + \gamma_{n+1}e_{n+1}; \qquad \mu_0 = 0.$$

Id denotes the $d \times d$ identity matrix. Roughly speaking, the sequence $\{\mu_n, n \geq 0\}$ captures the linear approximation of $h(\theta_n)$ and the martingale-increment noise sequence $\{e_n, n \geq 1\}$.

We prove that $\gamma_n^{-1/2} \rho_n \mathbb{1}_{\lim_q \theta_q = \theta_\star}$ converges to zero in probability so that $\{\mu_n, n \geq 0\}$ is the leading term. We then establish that for any $t \in \mathbb{R}^d$,

$$\lim_n \mathbb{E}\left[\mathbb{1}_{\lim_q \theta_q = \theta_\star} \ \exp\left(i\gamma_n^{-1/2} t^T \mu_n\right)\right] = \mathbb{E}\left[\mathbb{1}_{\lim_q \theta_q = \theta_\star} \ \exp\left(-\frac{1}{2}t^T V t\right)\right].$$

3. A CENTRAL LIMIT THEOREM FOR ITERATE AVERAGING

Theorem 2.1 shows that the rate of convergence of the sequence $\{\theta_n, n \geq 0\}$ to $\theta_\star$ is $O(n^{a/2})$ when $\gamma_n \sim \gamma_\star/n^a$ for some $a \in (1/2, 1]$. The maximal rate is reached by choosing $\gamma_n \sim \gamma_\star/n$, for some $\gamma_\star$ satisfying the conditions C4b. The main drawback with such a choice of the step-size sequence $\{\gamma_n, n \geq 1\}$ is that in practice, $-L$ *i.e.* the largest real part of the eigenvalues of $\nabla h(\theta_\star)$ is unknown so that the condition C4b is difficult to check.

The second comment is on the limiting covariance matrix when the rate is maximal (*i.e.* in the case $\gamma_n \sim \gamma_\star/n$). For any non-singular matrix Γ, we could define the algorithm

$$\tilde{\theta}_{n+1} = \tilde{\theta}_n + \gamma_{n+1}\Gamma h(\tilde{\theta}_n) + \gamma_{n+1}\Gamma e_{n+1} + \gamma_{n+1}\Gamma r_{n+1} , \qquad \tilde{\theta}_0 \in \Theta.$$

This equation is of the form (2.1) with a mean field $\tilde{h} = \Gamma h$ and noises $\{e_n, r_n, n \geq 1\}$ replaced with $\{\Gamma e_n, \Gamma r_n, n \geq 1\}$. Then, Theorem 2.1 gives sufficient conditions so that a CLT for the sequence $\{\tilde{\theta}_n, n \geq 0\}$ holds: the matrix V is replaced with $\tilde{V} = \tilde{V}(\Gamma)$ satisfying

$$\tilde{V}(\mathrm{Id} + 2\gamma_\star \nabla h(\theta_\star)^T \Gamma^T) + (\mathrm{Id} + 2\gamma_\star \nabla h(\theta_\star)\Gamma)\tilde{V} = -2\gamma_\star \Gamma U_\star \Gamma^T.$$

A natural question is the "optimal" choice of the gain matrix Γ, defined as the matrix $\Gamma_\star$ such that for any $\lambda \in \mathbb{R}^d$, $\lambda^T \tilde{V}(\Gamma)\lambda \geq \lambda^T \tilde{V}(\Gamma_\star)\lambda$. Following the same lines as in Benveniste *et al.* ([4], Prop. 4, Chap. 3, Part I), it can be proved that $\Gamma_\star = -\gamma_\star^{-1}\nabla h(\theta_\star)^{-1}$ and in this case,

$$\tilde{V}(\Gamma_\star) = \gamma_\star^{-1}\nabla h(\theta_\star)^{-1}U_\star \nabla h(\theta_\star)^{-T}.$$

Theorem 3.2 below shows that by considering the averaged sequence $\{\bar{\theta}_n, n \geq 0\}$, the optimal rate of convergence (*i.e.* the rate $\sqrt{n}$) and the optimal asymptotic covariance matrix (optimal in the sense discussed above) can be reached whatever the sequence $\{\gamma_n, n \geq 1\}$ satisfying C4a used in the basic SA sequence (2.1). Therefore, such an optimality can be obtained even when $\nabla h(\theta_\star)$ is unknown. Note also that on a practical point of view, slow decreasing step-size γ_n are better (see *e.g.* Spall [28], Sect. 4.4.) and this simple averaging procedure improves the rate of convergence of the estimate of $\theta_\star$.

These properties of the averaged sequence were simultaneously established by Ruppert [27] and Polyak and Juditsky [26] under more restrictive conditions than those stated below.

3.1. Assumptions

AVER1

(a) $\{e_n, \ n \geq 1\}$ is a $\mathcal{F}_n$-adapted $\mathbb{P}$-martingale-increment sequence.

(b) There exists a sequence $\{\mathcal{A}_m, m \geq 1\}$ such that $\lim_m \mathbb{P}(\mathcal{A}_m | \lim_q \theta_q = \theta_\star) = 1$, and for any $m \geq 1$,

$$\sup_k \mathbb{E}\left[|e_k|^2 \mathbb{1}_{\mathcal{A}_{m,k-1}}\right] < \infty,$$

where $\mathcal{A}_{m,k-1} \in \mathcal{F}_{k-1}$ and $\lim_k \mathbb{1}_{\mathcal{A}_{m,k}} = \mathbb{1}_{\mathcal{A}_m}$ almost-surely on the set $\{\lim_q \theta_q = \theta_\star\}$.

(c) Let

$$\mathcal{E}_{n+1} = \frac{1}{\sqrt{n+1}} \sum_{k=0}^{n} e_{k+1}.$$

There exists a positive definite deterministic matrix $U_\star$ such that for any $t \in \mathbb{R}^d$,

$$\lim_n \mathbb{E}\left[\mathbb{1}_{\lim_q \theta_q = \theta_\star} \ \exp(it^T \mathcal{E}_{n+1})\right] = \mathbb{E}\left[\mathbb{1}_{\lim_q \theta_q = \theta_\star} \ \exp(-\frac{1}{2} \ t^T U_\star t)\right].$$

We prove in Lemma 5.5 that when $\lim_n n\gamma_n > 0$, assumption C2 implies AVER1. Note also that since $\lim_m \mathbb{P}(\mathcal{A}_m | \lim_q \theta_q = \theta_\star) = 1$, AVER1c is equivalent to the condition: for any $m \geq 1$,

$$\lim_n \mathbb{E}\left[\mathbb{1}_{\lim_q \theta_q = \theta_\star} \exp(it^T \mathcal{E}_{n+1}) \mathbb{1}_{\mathcal{A}_m}\right] = \mathbb{E}\left[\mathbb{1}_{\lim_q \theta_q = \theta_\star} \exp(-\frac{1}{2} t^T U_\star t) \mathbb{1}_{\mathcal{A}_m}\right].$$

For a sequence of $\mathbb{R}^d$-valued r.v. $\{Z_n, n \geq 0\}$, we write $Z_n = O_{L^p}(1)$ if $\sup_n \mathbb{E}[|Z_n|^p] < \infty$.

AVER2 r_n is $\mathcal{F}_n$-adapted. $r_n = r_n^{(1)} + r_n^{(2)}$ with for any $m \geq 1$,
    (a) $\gamma_n^{-1/2} r_n^{(1)} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} = O_{w.p.1}(1) O_{L^2}(1)$.
    (b) $\sqrt{\gamma_n} \sum_{k=1}^n r_k^{(2)} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} = O_{w.p.1}(1) O_{L^2}(1)$.
    (c) $n^{-1/2} \sum_{k=0}^n r_{k+1} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \xrightarrow{\mathbb{P}} 0$.
    The sequence $\{\mathcal{A}_m, m \geq 1\}$ is defined in AVER1b.

Note that AVER2c is equivalent to $n^{-1/2} \sum_{k=0}^n r_{k+1} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} \xrightarrow{\mathbb{P}} 0$ for any $m \geq 1$.

AVER3 $\lim_n n\gamma_n = +\infty$ and

$$\lim_n \frac{1}{\sqrt{n}} \sum_{k=1}^n \gamma_k^{-1/2} \left|1 - \frac{\gamma_k}{\gamma_{k+1}}\right| = 0, \qquad \lim_n \frac{1}{\sqrt{n}} \sum_{k=1}^n \gamma_k = 0.$$

The step size $\gamma_n \sim \gamma_\star n^{-a}$ for $a \in (1/2, 1)$ satisfies AVER3 but the step size $\gamma_n \sim \gamma_\star/n$ does not. Observe that if the sequence $\{\gamma_n, n \geq 0\}$ is non-increasing (or ultimately non-increasing) then (see *e.g.* Delyon [11], proof of Thm. 26)

$$\lim_n n\gamma_n = +\infty \Longrightarrow \lim_n \frac{1}{\sqrt{n}} \sum_{k=1}^n \gamma_k^{-1/2} \left|1 - \frac{\gamma_k}{\gamma_{k+1}}\right| = 0.$$

3.2. Main results

We show that the above conditions allow a control of the L^2-moment of the errors $\{\theta_n - \theta_\star, n \geq 0\}$. This result is a cornerstone for the proof of Theorem 3.2. The proof is given in Section 5.

Proposition 3.1. *Assume* C1, C4, AVER1a-b *and* AVER2a-b. *Then, for any* $m \geq 1$

$$\gamma_n^{-1} \|\theta_n - \theta_\star\|^2 \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} = O_{w.p.1}(1) \, O_{L^1}(1).$$

Theorem 3.2. *Choose* $\theta_0 \in \Theta$ *and consider the averaged sequence given by* (1.3). *Assume* C1, C4a, AVER1, AVER2 *and* AVER3. *Then for any* $t \in \mathbb{R}^d$,

$$\lim_n \mathbb{E}\left[\mathbb{1}_{\lim_q \theta_q = \theta_\star} \exp\left(i\sqrt{n}\, t^T (\bar{\theta}_n - \theta_\star)\right)\right] = \mathbb{P}\left(\lim_q \theta_q = \theta_\star\right) \exp\left(-\frac{1}{2} t^T \nabla h(\theta_\star)^{-1} U_\star (\nabla h(\theta_\star)^{-1})^T t\right).$$

Sketch of the proof of Theorem 3.2. The proof is detailed in Section 5. Since $\lim_m \mathbb{P}(\mathcal{A}_m | \lim_q \theta_q = \theta_\star) = 1$, we only have to prove that for any $m \geq 1$ and $t \in \mathbb{R}^d$,

$$\lim_n \mathbb{E}\left[\mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} \exp\left(i\sqrt{n}\, t^T (\bar{\theta}_n - \theta_\star)\right)\right] = \mathbb{E}\left[\mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} \exp\left(-\frac{1}{2} t^T \nabla h(\theta_\star)^{-1} U_\star (\nabla h(\theta_\star)^{-1})^T t\right)\right].$$

We write

$$\bar{\theta}_n - \theta_\star = -\frac{\nabla h(\theta_\star)^{-1}}{n+1} \sum_{k=0}^n e_{k+1} + Z_n.$$

We show that $\sqrt{n} Z_n \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m}$ converges to zero in probability for any $m \geq 1$; for this step, the main tool is Proposition 3.1. The proof is then concluded by AVER1c.

4. APPLICATION TO SA WITH CONTROLLED MARKOV CHAIN DYNAMICS

Let $\{\mathcal{K}_n, n \geq 0\}$ be a sequence of compact subsets of $\Theta \subseteq \mathbb{R}^d$ such that

$$\mathcal{K}_n \subseteq \mathcal{K}_{n+1}, \qquad \bigcup_{n \geq 0} \mathcal{K}_n = \Theta.$$

Let $\{Q_\theta, \theta \in \Theta\}$ be a family of Markov transition kernels onto $(\mathbb{X}, \mathcal{X})$. We consider the following SA algorithm with truncation at randomly varying bounds: $\theta_0 \in \mathcal{K}_0, \sigma_0 = 0$ and for $n \geq 0$,

set $\theta_{n+1/2} = \theta_n + \gamma_{n+1} H(\theta_n, X_{n+1})$.
update

$$(\theta_{n+1}, \sigma_{n+1}) = \begin{cases} (\theta_{n+1/2}, \sigma_n), & \text{if } \theta_{n+1/2} \in \mathcal{K}_{\sigma_n}, \\ (\theta_0, \sigma_n + 1) & \text{otherwise,} \end{cases}$$

where $\{X_n, n \geq 0\}$ is a controlled Markov chain on $(\Omega, \mathcal{A}, \mathbb{P})$ with conditional distribution given by

$$\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) = Q_{\theta_n}(X_n, A), \qquad \mathcal{F}_n = \sigma(\theta_0, X_0, \cdots, X_n). \tag{4.1}$$

The random sequence $\{\sigma_n, n \geq 0\}$ is a non-negative integer-valued sequence counting the number of truncations. Such a truncated SA was introduced by Chen *et al.* [10] (see also Chen [9], Chap. 2) to address the boundedness problem of the SA sequence $\{\theta_n, n \geq 0\}$. A more general truncated SA algorithm with controlled Markov chain dynamics is introduced in Andrieu *et al.* [2]: when truncation occurs, both the parameter $\theta_{n+1/2}$ and the draw X_n used to obtain the next point X_{n+1} are modified.

The key point of the proof of convergence of this algorithm is to show that the number of truncations is finite with probability one, so that after some random time, the sequence $\{\theta_n, n \geq 0\}$ is almost-surely bounded and obeys the update rule $\theta_{n+1} = \theta_n + \gamma_{n+1} H(\theta_n, X_{n+1})$. Conditions implying almost-sure boundedness and almost-sure convergence of the sequence $\{\theta_n, n \geq 0\}$ when $\{X_n, n \geq 0\}$ is a controlled Markov chain can be found in Andrieu *et al.* ([2], Sect. 3). We assume

A1 For any $\theta \in \Theta$, there exists a probability distribution π_θ on $(\mathbb{X}, \mathcal{X})$ such that $\pi_\theta Q_\theta = \pi_\theta$.

For simplicity, we consider the case when H is bounded and the step-size sequence is polynomially decreasing. Extensions to the case H is unbounded can be done along the same lines as in Andrieu *et al.* [2].

A2 (a) for any compact set $\mathcal{K} \subseteq \Theta$, $\sup_{\theta \in \mathcal{K}} \sup_{x \in \mathbb{X}} |H(\theta, x)| < \infty$. Set

$$h(\theta) = \int H(\theta, x) \, \pi_\theta(\mathrm{d}x). \tag{4.2}$$

(b) There exists $a \in (1/2, 1]$ such that $\gamma_n = \gamma_\star / n^a$. When $a = 1$, $\gamma_\star$ satisfies the condition C4b.

Since in this paper we are interested in CLT's, the stability and the convergence of the algorithm is also assumed:

A3 the number of truncations is finite with probability one: $\mathbb{P}(\limsup_n \sigma_n < \infty) = 1$; and there exists $\theta_\star \in \Theta$ satisfying C1 such that $\mathbb{P}(\lim_n \theta_n = \theta_\star) > 0$.

For a function $W : \mathbb{X} \to [1, \infty)$, define the W-norm of a measurable function $f : \mathbb{X} \to \mathbb{R}$ by $\|f\|_W = \sup_{\mathbb{X}} |f|/W$. We assume that the transition kernels $\{Q_\theta, \theta \in \Theta\}$ satisfy

A4 (a) For any $\theta \in \Theta$, there exists a measurable function $\widehat{H}_\theta : (\mathbb{X}, \mathcal{X}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that

$$H(\theta, x) - h(\theta) = \widehat{H}_\theta(x) - Q_\theta \widehat{H}_\theta(x). \tag{4.3}$$

There exists a function $V_1 : \mathcal{X} \to [1, \infty)$ such that for any compact subset $\mathcal{K} \subseteq \Theta$,

$$\sup_{\theta \in \mathcal{K}} \left(\|\widehat{H}_\theta\|_{V_1} + \|Q_\theta \widehat{H}_\theta\|_{V_1} \right) < \infty. \tag{4.4}$$

(b) For any $\theta \in \Theta$, there exists a measurable function $U_\theta : (\mathbb{X}, \mathcal{X}) \to (\mathbb{R}^{d^2}, \mathcal{B}(\mathbb{R}^{d^2}))$ such that

$$F_\theta(x) - \int F_\theta(x)\, \pi_\theta(\mathrm{d}x) = U_\theta(x) - Q_\theta U_\theta(x), \tag{4.5}$$

where $F_\theta(x) = \int Q_\theta(x, dy)\, \widehat{H}_\theta(y) \widehat{H}_\theta(y)^T - Q_\theta \widehat{H}_\theta(x) \left(Q_\theta \widehat{H}_\theta(x) \right)^T$. There exists a function $V_2 : \mathcal{X} \to [1, \infty)$ such that for any compact subset $\mathcal{K} \subseteq \Theta$,

$$\sup_{\theta \in \mathcal{K}} \left(\|U_\theta\|_{V_2} + \|Q_\theta U_\theta\|_{V_2} \right) < \infty. \tag{4.6}$$

(c) There exist $\delta, \tau > 0$ and $\bar{\tau} > 1/a - 1$ such that for any $m \geq 1$,

$$\sup_{k \geq m} \mathbb{E}\left[\left(V_1^{2+\tau}(X_{k+1}) + V_2^{1+\bar{\tau}}(X_{k+1}) \right) \mathbb{1}_{\bigcap_{m \leq j \leq k} \{|\theta_j - \theta_\star| \leq \delta\}} \right] < \infty,$$
$$\mathbb{E}\left[V_1^{2+\tau}(X_m) + V_2^{1+\bar{\tau}}(X_m) \right] < \infty.$$

(d) For any compact subset $\mathcal{K} \subseteq \Theta$, there exist $C > 0$ and $b > (1/a - 1) \vee (1/2)$ such that

$$\left\| Q_\theta \widehat{H}_\theta - Q_{\theta'} \widehat{H}_{\theta'} \right\|_{V_1} + \|U_\theta - U_{\theta'}\|_{V_2} \leq C\, |\theta - \theta'|^b.$$

Furthermore, almost-surely

$$\lim_n \left(\int F_{\theta_n}(x)\, \pi_{\theta_n}(\mathrm{d}x) - \int F_{\theta_\star}(x)\, \pi_{\theta_\star}(\mathrm{d}x) \right) \mathbb{1}_{\lim_q \theta_q = \theta_\star} = 0.$$

Conditions implying the existence of π_θ and solutions to the Poisson equations (4.3) and (4.5) can be found *e.g.* in Hernandez−Lerma and Lasserre ([18], Chap. 8) or in Meyn and Tweedie ([24], Chap. 17). When the transition kernel Q_θ is uniformly ergodic, then $V_1 = V_2$ and is equal to the constant function 1. When the kernel is V-geometrically ergodic, we can choose $V_1 = V^{1/p}, V_2 = V^{2/p}$ for any $p \geq 2$. Sufficient conditions for (4.4) and (4.6) based on Lyapunov drift inequalities when the chain is geometrically ergodic (resp. subgeometrically ergodic) are given by Fort *et al.* ([16], Lem. 2.3) (resp. Andrieu *et al.* [1]). Andrieu *et al.* ([2], Prop. 6.1). gives sufficient conditions to check A4c (compare this assumption with the condition A3(ii) of Andrieu *et al.*) when the kernels are V-geometrically ergodic: in this case, for any $p \geq 2$ we can choose $V_1 = V^{1/p}, V_2 = V^{2/p}$ and $\bar{\tau}$ such that $2(1+\bar{\tau})/p = 1$. The first set of conditions in A4d is an assumption on the regularity-in-θ of the solution to the Poisson equation. Andrieu *et al.* ([2], Prop. 6.1). give sufficient conditions in terms of the regularity-in-θ of the transition kernels Q_θ. When $\pi_\theta = \pi$ for any θ, the second set of conditions can be established by combining smoothness-in-θ properties of the function F_θ and the dominated convergence theorem. When π_θ depends on θ, Fort *et al.* ([16], Thm. 2.11 and Prop. 4.3) give sufficient conditions for this condition to hold.

The following proposition provides a set of conditions implying A1 and A4. Its proof is in Section 5.7.

Proposition 4.1. *Assume that*

(i) *for any $\theta \in \Theta$, Q_θ is phi-irreducible and aperiodic.*

(ii) *there exists a measurable function $V :\to [1, \infty)$ and for any $\theta \in \Theta$, there exist constants $\lambda_\theta \in (0, 1)$ and $b_\theta \in [1, \infty)$ such that $Q_\theta V \leq \lambda_\theta V + b_\theta$.*

(iii) *there exists $\alpha \in (0, 1/2)$ and for any $\theta \in \Theta$, there exist $\delta_\theta \in (0, 1)$ and a probability measure ν_θ on $(\mathbb{X}, \mathcal{X})$ such that $Q_\theta(x, \cdot) \geq \delta_\theta \nu_\theta$ for any $x \in \{V^\alpha \leq 2b_\theta/(1 - \lambda_\theta^\alpha) - 1\}$.*

(iv) *for any compact set $\mathcal{K} \subseteq \Theta$, $\sup_{\theta \in \mathcal{K}} \left(b_\theta \vee (1 - \lambda_\theta^\alpha)^{-1} \vee \delta_\theta^{-1} \right) < \infty$.*

(v) *there exists* $\beta \in (1/2, 1]$ *and for any compact set* $\mathcal{K} \subseteq \Theta$ *there exists a constant* C *such that for any* $\theta, \theta' \in \mathcal{K}$,

$$\sup_{\theta \in \mathcal{K}} \sup_{x \in \mathbb{X}} |H(\theta, x)| < \infty,$$

$$|H(\theta, x) - H(\theta', x)| \leq C|\theta - \theta'|^{\beta},$$

$$\sup_{f, \|f\|_{V^\alpha} \leq 1} \|Q_\theta f - Q_{\theta'} f\|_{V^\alpha} + \sup_{f, \|f\|_{V^{2\alpha}} \leq 1} \|Q_\theta f - Q_{\theta'} f\|_{V^{2\alpha}} \leq C |\theta - \theta'|^{\beta}.$$

Fix $\theta_0 \in \Theta$. *Then the conditions* A1 *and* A4 *hold provided* $\gamma_n \sim \gamma_\star / n^a$ *for* $a > 2\alpha \vee 1/(1 + \beta)$.

Let us now prove how the assumptions A1 to A4 imply the conditions C1 to C4. Under A3, the condition C1 holds; note also that the conditional probability $\mathbb{P}(\cdot | \lim_q \theta_q = \theta_\star)$ is well defined. By using (4.2) and (4.3), we write the truncated SA algorithm on the form (2.1) by setting

$$e_{n+1} = \widehat{H}_{\theta_n}(X_{n+1}) - Q_{\theta_n} \widehat{H}_{\theta_n}(X_n),$$

$$r_{n+1} = Q_{\theta_n} \widehat{H}_{\theta_n}(X_n) - Q_{\theta_n} \widehat{H}_{\theta_n}(X_{n+1}) + (\theta_0 - \theta_{n+1/2}) \mathbb{1}_{\theta_{n+1/2} \notin \mathcal{K}_{\sigma_n}}.$$

Let us prove that the condition C2 holds. Since $\theta_n \in \mathcal{F}_n$, equation (4.1) implies C2a. Fix δ such that $B(\theta_\star, \delta) = \{\theta \in \mathbb{R}^d, |\theta - \theta_\star| \leq \delta\} \subseteq \Theta$. Set

$$\mathcal{A}_{m,k} = \begin{cases} \emptyset & \text{if } k < m, \\ \bigcap_{m \leq j \leq k} \{|\theta_j - \theta_\star| \leq \delta, \theta_j = \theta_{j-1/2}\} & \text{otherwise.} \end{cases}$$

Then for any k, m, $\mathcal{A}_{m,k} \in \mathcal{F}_k$; $\lim_k \mathcal{A}_{m,k} = \mathcal{A}_m$ where $\mathcal{A}_m = \bigcap_{j \geq m} \{|\theta_j - \theta_\star| \leq \delta, \theta_j = \theta_{j-1/2}\}$; and $\lim_m \mathbb{P}(\mathcal{A}_m | \lim_q \theta_q = \theta_\star) = 1$ by A3. Fix $m \geq 1$; by (4.4) applied with $\mathcal{K} = B(\theta_\star, \delta)$, there exists a constant C such that for any $k \geq m$

$$\mathbb{E}\left[|e_{k+1}|^{2+\tau} \mathbb{1}_{\mathcal{A}_{m,k}}\right] \leq C \mathbb{E}\left[\left(V_1^{2+\tau}(X_k) + V_1^{2+\tau}(X_{k+1})\right) \mathbb{1}_{\mathcal{A}_{m,k}}\right].$$

A4c concludes the proof of C2b. Observe that $\mathbb{E}\left[e_{k+1} e_{k+1}^T | \mathcal{F}_k\right] = F_{\theta_k}(X_k)$. By using (4.5), we write $\mathbb{E}\left[e_{k+1} e_{k+1}^T | \mathcal{F}_k\right] = U_\star + D_k^{(1)} + D_k^{(2,a)} + D_k^{(2,b)}$ with

$$U_\star = \int F_{\theta_\star}(x) \, \pi_{\theta_\star}(\mathrm{d}x),$$

$$D_k^{(1)} = \int F_{\theta_k}(x) \, \pi_{\theta_k}(\mathrm{d}x) - \int F_{\theta_\star}(x) \, \pi_{\theta_\star}(\mathrm{d}x),$$

$$D_k^{(2,a)} = U_{\theta_k}(X_{k+1}) - Q_{\theta_k} U_{\theta_k}(X_k),$$

$$D_k^{(2,b)} = U_{\theta_k}(X_k) - U_{\theta_k}(X_{k+1}).$$

By A4d, $D_k^{(1)} \xrightarrow{\text{a.s.}} 0$ on the set $\{\lim_q \theta_q = \theta_\star\}$. By (4.1), $\mathbb{E}\left[D_k^{(2,a)} | \mathcal{F}_{k-1}\right] = 0$; by application of the Burkholder inequality (see *e.g.* Hall and Heyde [17], Thm. 2.10), it holds for any $A_k \in \mathcal{F}_k$ such that $\lim_k A_k = \{\lim_q \theta_q = \theta_\star\}$

$$\mathbb{E}\left[\left|\sum_{k=1}^n D_k^{(2,a)}\right| \mathbb{1}_{A_k} \mathbb{1}_{\mathcal{A}_{m,k}}\right] \leq \left(\mathbb{E}\left[\left|\sum_{k=1}^n D_k^{(2,a)}\right|^{1+\bar{\tau}} \mathbb{1}_{A_k} \mathbb{1}_{\mathcal{A}_{m,k}}\right]\right)^{1/(1+\bar{\tau})} \leq C \, n^{1/((1+\bar{\tau}) \wedge 2)}.$$

The constant C is finite since under (4.6) and A4c, $\sup_k \mathbb{E}\left[|D_k^{(2,a)}|^{1+\bar{\tau}} \mathbb{1}_{\mathcal{A}_{m,k}}\right] < \infty$. Furthermore,

$$\sum_{k=m}^n D_k^{(2,b)} = U_{\theta_m}(X_m) - U_{\theta_n}(X_{n+1}) + \sum_{k=m+1}^n \left(U_{\theta_k}(X_k) - U_{\theta_{k-1}}(X_k)\right)$$

so that by A4c-d, there exists a constant C such that

$$\mathbb{E}\left[\left|\sum_{k=1}^{n} D_k^{(2,b)}\right| \mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star}\right] \leq C\left(1 + \sum_{k=m+1}^{n} \gamma_k^b\right).$$

The above discussion shows that C2c is verified since $\bar{\tau} > 1/a - 1$ and $b > 1/a - 1$.

Finally, let us study r_n. We write $r_{n+1} = r_{n+1}^{(1)} + r_{n+1}^{(2)}$ with

$$r_{n+1}^{(1)} = \left(\theta_0 - \theta_{n+1/2}\right) \mathbb{1}_{\theta_{n+1/2} \notin \mathcal{K}_{\sigma_n}} + Q_{\theta_{n+1}} \widehat{H}_{\theta_{n+1}}(X_{n+1}) - Q_{\theta_n} \widehat{H}_{\theta_n}(X_{n+1}).$$

By A3 and A4d, $\gamma_n^{-1/2} r_n^{(1)} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} = o_{w.p.1}(1) + o_{L^1}(1)$ for any fixed $m \geq 1$. In addition, by (4.4), there exists a constant C such that

$$C\,\mathbb{E}\left[\left|\sum_{k=1}^{n} r_k^{(2)}\right| \mathbb{1}_{\mathcal{A}_m}\right] \leq \mathbb{E}\left[V_1(X_1)\right] + \mathbb{E}\left[V_1(X_{n+1})\mathbb{1}_{\mathcal{A}_m}\right];$$

it follows by A4c that the condition C3 is verified.

The above discussion is summarized in the following proposition

Proposition 4.2. *Assume* A1, A2, A3 *and* A4. *Then, the conditions* C1 *to* C4 *are satisfied and*

$$U_\star = \int \pi_{\theta_\star}(\mathrm{d}x) \left(\widehat{H}_{\theta_\star}(x)\,\widehat{H}_{\theta_\star}(x)^T - Q_{\theta_\star}\widehat{H}_{\theta_\star}(x)\,\left(Q_{\theta_\star}\widehat{H}_{\theta_\star}(x)\right)^T\right).$$

By application of Theorem 2.1, we obtain a CLT for randomly truncated SA with controlled Markov chain dynamics.

Our result improves on Delyon ([11], Thm. 25). Under stronger conditions (for example, it is assumed that V_1 and V_2 are bounded functions; there is a single target $\theta_\star$ and $b = 1$ in the regularity-in-θ assumption Cbd), Delyon [11] establishes a CLT in the case $\gamma_n = \gamma_\star/n^a$ with the condition $a \in (2/3, 1]$. Note that if V_1, V_2 are bounded then A4c holds with any $\bar{\tau} > 0$ and if $b = 1$ then $b > (1/2) \vee (1/a - 1)$; hence, our approach only requires $a \in (1/2, 1]$ which is the usual range of values for SA algorithms.

Our result also improves on Bouton ([7], Cor. of Thm. 2): our assumptions only require Hölder-continuity of some quantities with respect to θ, and the Lipschitz-continuity of $\theta \mapsto \pi_\theta(F_\theta)$ is not required as in Bouton (see [7], assumption (H.7)) which is a strong improvement especially in situations when we do not have an explicit expression of π_θ.

Using similar tools, the conditions of Theorem 3.2 can be verified; details are left to the interested reader.

5. PROOF

5.1. Definitions and notations

Let $\{A_n, n \geq 0\}$ be a sequence of sets such that

$$A_n \in \mathcal{F}_n, \qquad \lim_n \mathbb{1}_{A_n} = \mathbb{1}_{\lim_q \theta_q = \theta_\star} \qquad \text{w.p.1.} \tag{5.1}$$

Such a sequence exists by Lemma 5.6. Define recursively two sequences

$$\mu_{n+1} = (\mathrm{Id} + \gamma_{n+1}\nabla h(\theta_\star))\mu_n + \gamma_{n+1}e_{n+1}, \qquad \mu_0 = 0; \tag{5.2}$$

$$\rho_{n+1} = \theta_{n+1} - \theta_\star - \mu_{n+1}, \qquad \rho_0 = \theta_0 - \theta_\star; \tag{5.3}$$

and the matrices $\psi_\star(n,k)$ for $1 \le k \le n$,

$$\psi_\star(n,k) = \prod_{j=k}^{n}(\mathrm{Id} + \gamma_j \nabla h(\theta_\star)). \tag{5.4}$$

By convention, $\psi_\star(n, n+1) = \mathrm{Id}$. Under C1a–b, there exist a set of random $d \times d$ symmetric matrices $\{R_i^{(n)}, \ i \le d\}$ such that the entry i of the column vector $\{h(\theta_n) - \nabla h(\theta_\star)(\theta_n - \theta_\star)\}$ is equal to $(\theta_n - \theta_\star)^T R_i^{(n)}(\theta_n - \theta_\star)$. More precisely,

$$R_i^{(n)}(k, \ l) = \int_0^1 \frac{1}{2}(1-t)^2 \frac{\partial^2 h_i}{\partial \theta_k \partial \theta_l}(\theta_n + t(\theta_n - \theta_\star)) \ \mathrm{d}t. \tag{5.5}$$

Let $R_\bullet^{(n)}$ be the tensor such that

$$h(\theta_n) = \nabla h(\theta_\star)(\theta_n - \theta_\star) + (\theta_n - \theta_\star)^T R_\bullet^{(n)}(\theta_n - \theta_\star). \tag{5.6}$$

Finally, for $1 \le k \le n$, define the $d \times d$ matrices

$$\psi(n,k) = \prod_{j=k}^{n}\left(\mathrm{Id} + \gamma_j \left\{\nabla h(\theta_\star) + 2\mu_{j-1}^T R_\bullet^{(j-1)} + \rho_{j-1}^T R_\bullet^{(j-1)}\right\}\right), \tag{5.7}$$

with the convention that $\psi(n, n+1) = \mathrm{Id}$.

5.2. Preliminary results on the sequence $\{\mu_n, n \ge 0\}$

By iterating (5.2), we have by definition of $\psi_\star$ (see (5.4))

$$\mu_{n+1} = \sum_{k=1}^{n+1} \gamma_k \psi_\star(n+1, k+1)e_k. \tag{5.8}$$

Proposition 5.1. *Assume* C1b–c, C2a–b *and* C4. *Then*

(i) $\mu_n \mathbb{1}_{\lim_q \theta_q = \theta_\star} \xrightarrow{a.s.} 0$ *when* $n \to \infty$.
(ii) *for any* $m \ge 1$, $\gamma_k^{-1}|\mu_k|^2 \ \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} = O_{L^1}(1) + o_{w.p.1}(1)$.

The proof is on the same lines as the proof of ([12], Lem. 6) and is omitted here; a detailed proof is given in the supplementary material.

5.3. Preliminary results on the sequence $\{\rho_n, n \ge 0\}$

By (5.3) and (5.6),

$$\begin{aligned}
\rho_{n+1} &= (\mathrm{Id} + \gamma_{n+1}\nabla h(\theta_\star))\rho_n + \gamma_{n+1}r_{n+1} + \gamma_{n+1}(\theta_n - \theta_\star)^T R_\bullet^{(n)}(\theta_n - \theta_\star) \\
&= (\mathrm{Id} + \gamma_{n+1}\nabla h(\theta_\star))\rho_n + \gamma_{n+1}r_{n+1} + \gamma_{n+1}(\mu_n + \rho_n)^T R_\bullet^{(n)}(\mu_n + \rho_n) \\
&= \left(\mathrm{Id} + \gamma_{n+1}\nabla h(\theta_\star) + 2\gamma_{n+1}\mu_n^T R_\bullet^{(n)} + \gamma_{n+1}\rho_n^T R_\bullet^{(n)}\right)\rho_n \\
&\quad + \gamma_{n+1}r_{n+1} + \gamma_{n+1}\mu_n^T R_\bullet^{(n)}\mu_n.
\end{aligned}$$

By induction, this yields

$$\rho_n = \psi(n,1)\rho_0 + \sum_{k=1}^{n} \gamma_k \psi(n, k+1)\left(r_k + \mu_{k-1}^T R_\bullet^{(k-1)}\mu_{k-1}\right), \tag{5.9}$$

where $\psi(n,k)$ is given by (5.7).

Proposition 5.2. *Assume* C1, C2a–b *and* C4. *Let* $\theta_0 \in \Theta$. *Then, for any* $m \geq 1$,

$$\left\{ \rho_n - \sum_{k=1}^{n} \gamma_k \psi(n, k+1) r_k \right\} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} = \gamma_n^{1 \wedge (1/2 + \kappa)} \, O_{w.p.1}(1) O_{L^1}(1),$$

with $\kappa = 1/2$ *under* C4a *and* $\kappa \in (0, L\gamma_\star - 1/2)$ *under* C4b.
  Assume in addition C3. *Then, for any* $m \geq 1$,

$$\sum_{k=1}^{n} \gamma_k \psi(n, k+1) r_k \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} = \gamma_n^{1/2} \, O_{w.p.1}(1) o_{L^1}(1).$$

The proof is on the same lines as the proof of ([12], Lem. 6) and is omitted here; a detailed proof is given in the supplementary material.

5.4. Proof of Theorem 2.1

By (5.3), $\gamma_n^{-1/2} (\theta_n - \theta_\star) = \gamma_n^{-1/2} \mu_n + \gamma_n^{-1/2} \rho_n$. We first prove that on $\{\lim_q \theta_q = \theta_\star\}$, the second term tends to zero in probability. By C2b, for any $\epsilon > 0$ there exists $m \geq 1$ such that $\mathbb{P}(\mathcal{A}_m | \lim_q \theta_q = \theta_\star) \geq 1 - \epsilon$. Therefore, it is sufficient to prove that for any $m \geq 1$, $\gamma_n^{-1/2} \rho_n \mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \xrightarrow{\mathbb{P}} 0$ when $n \to \infty$. This property holds by Proposition 5.2.

We now prove a CLT for the sequence $\{\gamma_n^{-1/2} \mu_n, n \geq 0\}$. It is readily seen that

$$\lim_n \mathbb{E}\left[\exp(i\gamma_n^{-1/2} t^T \mu_n) \mathbb{1}_{\lim_q \theta_q = \theta_\star} \right] = \mathbb{E}\left[\exp(-0.5 t^T V t) \mathbb{1}_{\lim_q \theta_q = \theta_\star} \right]$$

if and only if

$$\lim_n \mathbb{E}\left[\exp(i\gamma_n^{-1/2} t^T \mu_n \mathbb{1}_{\lim_q \theta_q = \theta_\star}) \right] = \mathbb{E}\left[\exp(-0.5 t^T V t \mathbb{1}_{\lim_q \theta_q = \theta_\star}) \right]$$

Furthermore, by C4 and Lemma 5.7, for any fixed $\ell \geq 1$, $\lim_n \gamma_n^{-1/2} |\psi_\star(n, \ell)| = 0$ (where $\psi_\star$ is given by (5.4)); this property, together with (5.8) and (5.1) imply that

$$\lim_n \mathbb{E}\left[\exp(i\gamma_n^{-1/2} t^T \mu_n \mathbb{1}_{\lim_q \theta_q = \theta_\star}) \right] = \lim_n \mathbb{E}\left[\exp\left(it^T \sum_{k=1}^{n} X_{n+1,k} \mathbb{1}_{A_{k-1}} \right) \right]$$

where $X_{n+1,k} = \gamma_{n+1}^{-1/2} \gamma_k \psi_\star(n+1, \ k+1) e_k$. By C2a and (5.1), $\mathbb{E}\left[X_{n+1,k} \mathbb{1}_{A_{k-1}} | \mathcal{F}_{k-1} \right] = 0$ and the limit in distribution is obtained by standard results on CLT for martingale-arrays (see *e.g.* Hall and Heyde [17], Cor. 3.1).

Lindeberg condition. We have to prove that for any $\epsilon > 0$,

$$\sum_{k=1}^{n} \mathbb{E}\left[|X_{n+1,k}|^2 \mathbb{1}_{|X_{n+1,k}| \geq \epsilon} \, | \mathcal{F}_{k-1} \right] \mathbb{1}_{A_{k-1}} \xrightarrow{\mathbb{P}} 0.$$

Following the same lines as above, it can be proved that equivalently, we have to prove for any $m \geq 1$,

$$\mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \sum_{k=1}^{n} \mathbb{E}\left[|X_{n+1,k}|^2 \mathbb{1}_{|X_{n+1,k}| \geq \epsilon} \, | \mathcal{F}_{k-1} \right] \xrightarrow{\mathbb{P}} 0.$$

Let $m \geq 1$ be fixed and set $X_{n+1,k} = X_{n+1,k}^{(1)} + X_{n+1,k}^{(2)}$ with

$$X_{n+1,k}^{(1)} = X_{n+1,k} \mathbb{1}_{\mathcal{A}_{m,k-1}}, \qquad X_{n+1,k}^{(2)} = X_{n+1,k} \left(1 - \mathbb{1}_{\mathcal{A}_{m,k-1}} \right).$$

We can assume without loss of generality that τ given by C2b is small enough so that $(2+\tau)L\gamma_\star > 1+\tau$. Then,

$$
\sum_{k=1}^{n+1} \mathbb{E}\left[\left|X_{n+1,k}^{(1)}\right|^{2+\tau}\right] = \sum_{k=1}^{n+1} \mathbb{E}\left[\left|\gamma_{n+1}^{-1/2}\gamma_k\psi_\star(n+1,k+1)e_k\mathbb{1}_{\mathcal{A}_{m,k-1}}\right|^{2+\tau}\right]
$$

$$
\leq \sup_k \mathbb{E}\left[|e_k\mathbb{1}_{\mathcal{A}_{m,k-1}}|^{2+\tau}\right] \ \gamma_{n+1}^{-1-\tau/2}\sum_{k=1}^{n+1}\gamma_k^{2+\tau}|\psi_\star(n+1,k+1)|^{2+\tau}.
$$

Under C1b–c, C2b and C4, Lemmas 5.7 and 5.8 imply

$$
\limsup_n \gamma_{n+1}^{-(1+\tau)}\sum_{k=1}^{n+1}\gamma_k^{2+\tau}|\psi_\star(n+1,k+1)|^{2+\tau} < +\infty
$$

since $(2+\tau)L\gamma_\star > 1+\tau$, Lemma 5.8 applies even in the case C4b). Hence,

$$
\sum_{k=1}^{n+1} \mathbb{E}\left[\left|X_{n+1,k}^{(1)}\right|^{2+\tau}\right] = o(\gamma_n^{\tau/2}).
$$

Consider now $X_{n+1,k}^{(2)}$. Since there exists a random variable K such that $\mathbb{1}_{\mathcal{A}_m}(1 - \mathbb{1}_{\mathcal{A}_{m,k-1}})\mathbb{1}_{\lim_q \theta_q = \theta_\star} = 0$ for any $k \geq K$, it holds for any $n \geq K$,

$$
\mathbb{1}_{\lim_q \theta_q = \theta_\star}\mathbb{1}_{\mathcal{A}_m}\sum_{k=1}^n \mathbb{E}\left[\left|X_{n+1,k}^{(2)}\right|^2 \mathbb{1}_{|X_{n+1,k}|\geq\epsilon}\ |\mathcal{F}_{k-1}\right]
$$

$$
= \mathbb{1}_{\lim_q \theta_q = \theta_\star}\mathbb{1}_{\mathcal{A}_m}\sum_{k=1}^K \mathbb{E}\left[|X_{n+1,k}|^2\mathbb{1}_{|X_{n+1,k}|\geq\epsilon}\ |\mathcal{F}_{k-1}\right]\left(1 - \mathbb{1}_{\mathcal{A}_{m,k-1}}\right)
$$

$$
\leq \mathbb{1}_{\lim_q \theta_q = \theta_\star}\mathbb{1}_{\mathcal{A}_m}\gamma_n^{-1}\sum_{k=1}^K \gamma_k^2|\psi_\star(n+1,k+1)|^2\mathbb{E}\left[|e_k|^2\ |\mathcal{F}_{k-1}\right]\left(1 - \mathbb{1}_{\mathcal{A}_{m,k-1}}\right).
$$

Under C4, this term is $o_{w.p.1}(1)$. Therefore, the first condition of ([17], Cor. 3.1). is satisfied.

Limiting variance. We prove the second condition of ([17], Cor. 3.1). Set

$$
V_n^{(1)} \stackrel{\text{def}}{=} \gamma_n^{-1}\sum_{k=1}^n \gamma_k^2\psi_\star(n,\ k+1)U_\star\psi_\star(n,\ k+1)^T\mathbb{1}_{\lim_q \theta_q = \theta_\star},
$$

$$
\overline{V}_n^{(2)} \stackrel{\text{def}}{=} \gamma_n^{-1}\sum_{k=1}^n \gamma_k^2\psi_\star(n,\ k+1)\left(\mathbb{E}[e_k e_k^T|\mathcal{F}_{k-1}]\mathbb{1}_{A_{k-1}} - U_\star\mathbb{1}_{\lim_q \theta_q = \theta_\star}\right)\ \psi_\star(n,\ k+1)^T;
$$

We prove that $V_n^{(1)} \stackrel{\mathbb{P}}{\longrightarrow} V\mathbb{1}_{\lim_q \theta_q = \theta_\star}$ and $\overline{V}_n^{(2)} \stackrel{\mathbb{P}}{\longrightarrow} 0$. It holds on $\{\lim_q \theta_q = \theta_\star\}$,

$$
V_{n+1}^{(1)} = \gamma_{n+1}U_\star + \frac{\gamma_n}{\gamma_{n+1}}\left(\text{Id} + \gamma_{n+1}\nabla h(\theta_\star)\right)\ V_n^{(1)}\ \left(\text{Id} + \gamma_{n+1}\nabla h(\theta_\star)\right)^T
$$

$$
= V_n^{(1)} + \gamma_n(U_\star + \nabla h(\theta_\star)V_n^{(1)} + V_n^{(1)}\nabla h(\theta_\star)^T) + \frac{\gamma_n - \gamma_{n+1}}{\gamma_{n+1}}V_n^{(1)}
$$

$$
+ (\gamma_{n+1} - \gamma_n)U_\star + \gamma_n\gamma_{n+1}\nabla h(\theta_\star)V_n^{(1)}\nabla h(\theta_\star)^T
$$

and by Lemma 5.9, $\lim_n V_n^{(1)} = V\mathbb{1}_{\lim_q \theta_q = \theta_\star}$ almost-surely. Following the same lines as above, it can be proved that $\overline{V}_n^{(2)}$ and $V_n^{(2)}$ given by

$$
V_n^{(2)} = \mathbb{1}_{\lim_q \theta_q = \theta_\star}\ \gamma_n^{-1}\sum_{k=1}^n \gamma_k^2\psi_\star(n,\ k+1)\left(\mathbb{E}\left[e_k e_k^T|\mathcal{F}_{k-1}\right] - U_\star\right)\ \psi_\star(n,\ k+1)^T
$$

have the same limit in probability. By C2c, we write $V_n^{(2)} = \left(V_n^{(2,a)} + V_n^{(2,b)}\right) \mathbb{1}_{\lim_q \theta_q = \theta_\star}$ with

$$V_n^{(2,a)} = \gamma_n^{-1} \sum_{k=1}^{n} \gamma_k^2 \psi_\star(n,\ k+1) D_{k-1}^{(1)} \psi_\star(n,\ k+1)^T$$

$$V_n^{(2,b)} = \gamma_n^{-1} \sum_{k=1}^{n} \gamma_k^2 \psi_\star(n,\ k+1) D_{k-1}^{(2)} \psi_\star(n,\ k+1)^T.$$

We have $\left|V_n^{(2,a)}\right| \leq \gamma_n^{-1} \sum_{k=1}^{n} \gamma_k^2 |\psi_\star(n,\ k+1)|^2\ |D_{k-1}^{(1)}|$. By Lemma 5.8, there exists a constant C such that on $\{\lim_q \theta_q = \theta_\star\}$

$$\limsup_n \left|V_n^{(2,a)}\right| \leq C\ \limsup_k \left|D_k^{(1)}\right|,$$

where we used (5.1). The rhs tends to zero w.p.1. by C2c. We now consider $V_n^{(2,b)}$. Since $\lim_m \mathbb{P}(\mathcal{A}_m | \lim_q \theta_q = \theta_\star) = 1$, it is sufficient to prove that for any $m \geq 1$, $V_n^{(2,b)} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} \xrightarrow{\mathbb{P}} 0$ when $n \to \infty$. Let $m \geq 1$. Set

$$\Xi_n \overset{\text{def}}{=} \sum_{j=0}^{n} D_j^{(2)} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m}.$$

By the Abel transform, we have

$$V_{n+1}^{(2,b)} \mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star} = \gamma_{n+1} \Xi_n + \gamma_{n+1}^{-1} \sum_{k=0}^{n-1} \{\gamma_{k+1}^2 \psi_\star(n+1, k+2) \Xi_k \psi_\star(n+1, k+2)^T$$
$$- \gamma_{k+2}^2 \psi_\star(n+1, k+3) \Xi_k \psi_\star(n+1, k+3)^T\}$$

Under C2c, $\gamma_n \Xi_n \xrightarrow{\mathbb{P}} 0$. For the second term, following the same lines as in Delyon ([11], Proof of Thm. 24, Chap. 4), it can be proved that the expectation of the second term is upper bounded by

$$C\ \gamma_{n+1}^{-1} \sum_{k=0}^{n-1} \gamma_{k+1}^2\ |\psi_\star(n+1, k+2)|^2\ (\gamma_k \mathbb{E}\left[|\Xi_k|\right]).$$

Since $\lim_k \gamma_k \mathbb{E}\left[|\Xi_k|\right] = 0$, Lemma 5.8 implies that $V_n^{(2,b)} \mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \xrightarrow{\mathbb{P}} 0$. This concludes the proof.

Remark 5.3. From the proof above, it can be seen that the assumption on the r.v. $D_n^{(2)}$ can be relaxed in

$$\lim_n \gamma_n \mathbb{E}\left[\left|\sum_{k=1}^{n} D_k^{(2)} \mathbb{1}_{A_k} \mathbb{1}_{\mathcal{A}_{m,k}}\right|\right] = 0.$$

Observe indeed that in probability,

$$\lim_n V_n^{(2,b)} \mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star} = \lim_n \gamma_n^{-1} \sum_{k=1}^{n} \gamma_k^2 \psi_\star(n,\ k+1) D_{k-1}^{(2)} \psi_\star(n,\ k+1)^T \mathbb{1}_{\mathcal{A}_{m,k-1}} \mathbb{1}_{A_{k-1}}.$$

5.5. Proof of Proposition 3.1

The proof is prefaced with a preliminary lemma.

Lemma 5.4. *Let $\{\gamma_n, n \geq 1\}$ is a (deterministic) positive sequence satisfying* C4a *and A be a (deterministic) $d \times d$ Hurwitz matrix. Let $\{x_n, n \geq 0\}$ be a sequence of $\mathbb{R}^d$-valued r.v. satisfying*

$$x_{n+1} = x_n + \gamma_{n+1} A x_n + \gamma_{n+1} \zeta_{n+1}^{(1)} + \gamma_{n+1} \zeta_{n+1}^{(2)}, \qquad n \geq 0,$$

where

$$\sum_{k=1}^{n} \gamma_k \left(\prod_{j=k+1}^{n+1} (\mathrm{Id} + \gamma_j A) \right) \zeta_k^{(1)} \mathbb{1}_{\lim_q x_q = 0} = \sqrt{\gamma_n} O_{w.p.1}(1) O_{L^2}(1),$$

$$|\zeta_n^{(2)}| \mathbb{1}_{\lim_q x_q = 0} = |x_n|^2 \ O_{w.p.1.}(1).$$

Then

$$\gamma_n^{-1} |x_n|^2 \mathbb{1}_{\lim_q x_q = 0} = O_{w.p.1.}(1) O_{L^1}(1).$$

The proof can be easily adapted from Delyon ([11], Thms. 20 and 23) and is omitted here.

Proof of Proposition 3.1. By (5.6)

$$\theta_{n+1} - \theta_\star = \theta_n - \theta_\star + \gamma_{n+1} \nabla h(\theta_\star) \ (\theta_n - \theta_\star) + \gamma_{n+1} (e_{n+1} + r_{n+1}) + \gamma_{n+1} (\theta_n - \theta_\star)^T R_\bullet^{(n)} (\theta_n - \theta_\star)$$

Let $m \geq 1$. We apply Lemma 5.4 with $x_n \leftarrow (\theta_n - \theta_\star) \mathbb{1}_{\mathcal{A}_m}$, $A \leftarrow \nabla h(\theta_\star)$, $\zeta_{n+1}^{(1)} = (e_{n+1} + r_{n+1}) \mathbb{1}_{\mathcal{A}_m}$ and $\zeta_{n+1}^{(2)} = (\theta_n - \theta_\star)^T R_\bullet^{(n)} (\theta_n - \theta_\star) \mathbb{1}_{\mathcal{A}_m}$. Under C1c, A is a Hurwitz matrix and $|\zeta_{n+1}^{(2)}| \mathbb{1}_{\lim_q \theta_q = \theta_\star} = O_{w.p.1}(1) |x_n|^2$.

We write $\zeta_{n+1}^{(1)} = \left(e_{n+1} \mathbb{1}_{\mathcal{A}_{m,n}} + e_{n+1} \left(1 - \mathbb{1}_{\mathcal{A}_{m,n}} \right) + r_{n+1} \right) \mathbb{1}_{\mathcal{A}_m}$. Under C4, AVER1a–b, Lemmas 5.7 and 5.8 imply

$$\sum_{k=1}^{n} \gamma_k \psi_\star(n+1, k+1) \ e_k \mathbb{1}_{\mathcal{A}_{m,k-1}} = \sqrt{\gamma_n} O_{L^2}(1).$$

Upon noting that $\mathbb{1}_{\mathcal{A}_m} \left(1 - \mathbb{1}_{\mathcal{A}_{m,k}} \right) = 0$ for all $k \geq K$ where K is a r.v. finite w.p.1.

$$\left(\sum_{k=1}^{n} \gamma_k \psi_\star(n+1, k+1) \ e_k \left(1 - \mathbb{1}_{\mathcal{A}_{m,k-1}} \right) \right) \mathbb{1}_{\mathcal{A}_m} = \left(\sum_{k=1}^{K} \gamma_k \psi_\star(n+1, k+1) \ e_k \left(1 - \mathbb{1}_{\mathcal{A}_{m,k-1}} \right) \right) \mathbb{1}_{\mathcal{A}_m}.$$

Therefore, by Lemma 5.8, this second term is $\sqrt{\gamma_n} O_{w.p.1}(1)$. Finally, Lemma 5.8 and AVER2a–b imply that the last term is $\sqrt{\gamma_n} O_{w.p.1}(1) O_{L^2}(1)$ (the proof is on the same lines as the proof of Prop. 5.2 and details are omitted).

5.6. Proof of Theorem 3.2

The proof is adapted from the proof of Delyon ([11], Thm. 26). Under C1c, $\nabla h(\theta_\star)$ is invertible. By (2.1) and Lemma 5.10 applied with $x_k \leftarrow \theta_k - \theta_\star$ and $A \leftarrow \nabla h(\theta_\star)$, we have

$$\sqrt{n} \left(\bar{\theta}_n - \theta_\star \right) = -\nabla h(\theta_\star)^{-1} \frac{\sqrt{n}}{n+1} \sum_{k=0}^{n} e_{k+1} + \sqrt{n} Z_n$$

where

$$\nabla h(\theta_\star) Z_n \stackrel{\text{def}}{=} -\frac{1}{n+1} \sum_{k=0}^{n} r_{k+1} - \frac{1}{n+1} \sum_{k=0}^{n} (h(\theta_k) - \nabla h(\theta_\star)(\theta_k - \theta_\star))$$

$$+ \frac{1}{n+1} \left(\frac{\theta_{n+1} - \theta_\star}{\gamma_{n+1}} - \frac{\theta_0 - \theta_\star}{\gamma_1} \right) + \frac{1}{n+1} \sum_{k=1}^{n} \left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k+1}} \right) (\theta_k - \theta_\star).$$

We prove that $\sqrt{n}Z_n \mathbb{1}_{\lim_q \theta_q = \theta_\star} \xrightarrow{\mathbb{P}} 0$; combined with AVER1c, this will conclude the proof. Since $\lim_m \mathbb{P}(\mathcal{A}_m | \lim_q \theta_q = \theta_\star) = 1$, it is sufficient to prove that for any $m \geq 1$, $\sqrt{n}Z_n \mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \xrightarrow{\mathbb{P}} 0$. Let $m \geq 1$. By AVER2c, it holds $n^{-1/2} \sum_{k=0}^n r_{k+1} \mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \xrightarrow{\mathbb{P}} 0$. By (5.6),

$$\frac{1}{n+1} \sum_{k=0}^n (h(\theta_k) - \nabla h(\theta_\star)(\theta_k - \theta_\star)) = \frac{1}{n+1} \sum_{k=0}^n (\theta_k - \theta_\star)^T R_\bullet^{(k)} (\theta_k - \theta_\star),$$

and by C1b, $R_\bullet^{(k)} \mathbb{1}_{\lim_q \theta_q = \theta_\star} = O_{w.p.1}(1)$. Therefore, by Proposition 3.1,

$$\frac{\sqrt{n}}{n+1} \sum_{k=0}^n (h(\theta_k) - \nabla h(\theta_\star)(\theta_k - \theta_\star)) \mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star} = \left(\frac{\sqrt{n}}{n+1} \sum_{k=0}^n \gamma_k W_k \overline{W}_k \right),$$

where $W_k = O_{w.p.1}(1)$ and $\overline{W}_k = O_{L^1}(1)$. AVER3 implies that this term tends to zero in probability. Proposition 3.1 and AVER3 imply that

$$\mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \frac{\sqrt{n}}{n+1} \left(\frac{\theta_{n+1} - \theta_\star}{\gamma_{n+1}} - \frac{\theta_0 - \theta_\star}{\gamma_1} \right) = \frac{O_{L^1}(1) O_{w.p.1}(1)}{\sqrt{(n+1)\gamma_{n+1}}} + o_{w.p.1}(1) \xrightarrow{\mathbb{P}} 0.$$

Finally, Proposition 3.1 and AVER3 also imply that

$$\mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \frac{\sqrt{n}}{n+1} \sum_{k=1}^n \left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k+1}} \right) (\theta_k - \theta_\star) = \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n \left| \frac{1}{\gamma_k} - \frac{1}{\gamma_{k+1}} \right| \gamma_k^{1/2} W_k \sqrt{\overline{W}_k} \right)$$

where $W_k = O_{w.p.1}(1)$ and $\overline{W}_k = O_{L^1}(1)$. This term tends to zero in probability.

Lemma 5.5. C2 *and* $\lim_n n\gamma_n > 0$ *imply* AVER1.

Proof. C2 implies trivially AVER1a–b. We only have to check AVER1c, or equivalently, prove that for any $m \geq 1$,
$$\lim_n \mathbb{E}\left[\exp\left(it^T \mathcal{E}_{n+1} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} \right) \right] = \mathbb{E}\left[\exp\left(it^T U_\star t \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} \right) \right].$$

Write $\mathcal{E}_{n+1} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} = T_{1,n} + T_{2,n}$ with $T_{1,n} = (n+1)^{-1/2} \sum_{k=0}^n e_{k+1} \mathbb{1}_{\mathcal{A}_{m,k}} \mathbb{1}_{A_k}$. By (5.1) and C2b, $T_{2,n} = o_{w.p.1}(1)$. Observe that $\mathbb{E}[e_{k+1} \mathbb{1}_{\mathcal{A}_{m,k}} \mathbb{1}_{A_k} | \mathcal{F}_k] = 0$ so that the convergence in distribution of $T_{1,n}$ will be established by applying results on martingale-arrays: we check the assumptions of Hall and Heyde ([17], Cor. 3.1). By C2b, it is easily checked that for any $\epsilon > 0$, there exists a constant C such that for any n,

$$\mathbb{E}\left[\frac{1}{n} \sum_{k=0}^n \mathbb{E}\left[|e_{k+1}|^2 \mathbb{1}_{|e_{k+1}| \geq \epsilon \sqrt{n}} | \mathcal{F}_k \right] \mathbb{1}_{\mathcal{A}_{m,k}} \right] \leq \frac{C}{n^{\tau/2}}.$$

Hence, $n^{-1} \sum_{k=0}^n \mathbb{E}\left[|e_{k+1}|^2 \mathbb{1}_{|e_{k+1}| \geq \epsilon \sqrt{n}} | \mathcal{F}_k \right] \mathbb{1}_{\mathcal{A}_{m,k}} \mathbb{1}_{A_k} \xrightarrow{\mathbb{P}} 0$. We now prove that

$$\frac{1}{n+1} \sum_{k=0}^n \mathbb{E}\left[e_{k+1} e_{k+1}^T | \mathcal{F}_k \right] \mathbb{1}_{\mathcal{A}_{m,k}} \mathbb{1}_{A_k} \xrightarrow{\mathbb{P}} U_\star \mathbb{1}_{\mathcal{A}_m} \mathbb{1}_{\lim_q \theta_q = \theta_\star}. \tag{5.10}$$

As above, we claim that this is equivalent to the proof that for any $m \geq 1$,

$$\mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m} \frac{1}{n+1} \sum_{k=0}^n \left(\mathbb{E}\left[e_{k+1} e_{k+1}^T | \mathcal{F}_k \right] - U_\star \right) \xrightarrow{\mathbb{P}} 0.$$

C2c and the Cesaro lemma imply that w.p.1, on the set $\mathcal{A}_m \cap \{\lim_q \theta_q = \theta_\star\}$, $(n+1)^{-1} \sum_{k=0}^n D_k^{(1)} \xrightarrow{\text{a.s.}} 0$. Finally, under C2c,

$$\frac{1}{n+1} \mathbb{E}\left[\left|\sum_{k=0}^n D_k^{(2)} \mathbb{1}_{\lim_q \theta_q = \theta_\star} \mathbb{1}_{\mathcal{A}_m}\right|\right] = \frac{o(1)}{n\gamma_n}$$

and the rhs tends to zero since $\lim_n n\gamma_n > 0$. This concludes the proof of (5.10) and the proof of the Lemma. $\quad\square$

5.7. Proof of Proposition 4.1

Let $\alpha \in (0, 1/2)$ be given by the assumption (4.1). By the Jensen inequality, for any $\theta \in \Theta$ and $\gamma \in [\alpha, 1]$, it holds

$$Q_\theta V^\gamma \leq \lambda_\theta^\alpha V^\gamma + b_\theta. \tag{5.11}$$

In addition, since $V \geq 1$, $\{V^\gamma \leq 2b_\theta/(1-\lambda_\theta^\alpha) - 1\} \subset \{V^\alpha \leq 2b_\theta/(1-\lambda_\theta^\alpha) - 1\}$. Therefore, by ([16], Lem. 2.3), there exists an unique probability measure π_θ invariant for Q_θ, thus implying A1.

By ([16], Lem. 2.3). again, there exist constants $C_\theta, \rho_\theta \in (0, 1)$ such that for any $\gamma \in [\alpha, 1]$, $x \in \mathbb{X}$, $\theta \in \Theta$ and any compact set $\mathcal{K} \subset \Theta$,

$$\sup_{\{f : \|f\|_{V^\gamma} \leq 1\}} |Q_\theta^n f(x) - \pi_\theta(f)| \leq C_\theta \, \rho_\theta^n V^\gamma(x); \tag{5.12}$$

$$\sup_{\theta \in \mathcal{K}} \left(C_\theta \vee (1 - \rho_\theta)^{-1} \vee \pi_\theta(V)\right) < \infty. \tag{5.13}$$

Set $\widehat{H}_\theta(x) = \sum_{n \geq 0} (Q_\theta^n H(\theta, \cdot)(x) - h(\theta))$. By (5.12), upon noting that $\|H(\theta, \cdot)\|_{V^\alpha} \leq \sup_{x \in \mathbb{X}} |H(\theta, x)|$ which is finite by assumption (v)

$$\left|\widehat{H}_\theta(x)\right| \vee \left|Q_\theta \widehat{H}_\theta(x)\right| \leq \|H(\theta, \cdot)\|_{V^\alpha} \, C_\theta (1 - \rho_\theta)^{-1} V^\alpha(x).$$

Then, (5.13) and the assumptions (iv)$-$(v) imply that A4a holds with $V_1(x) \leftarrow V^\alpha(x)$.
Set $U_\theta(x) = \sum_{n \geq 0} (Q_\theta^n F_\theta(x) - \pi_\theta(F_\theta))$. Observe that there exists a constant C_θ' such that

$$|F_\theta(x)| \leq C_\theta' \left(Q_\theta V^{2\alpha}(x) + (Q_\theta V^\alpha(x))^2\right),$$

so that by (5.11), $\|F_\theta\|_{V^{2\alpha}} < \infty$. As above, it can be proved that A4b holds with $V_2(x) \leftarrow V^{2\alpha}(x)$.
Choose a compact $\mathcal{K}$ of Θ such that $\mathcal{K} \supseteq \{\theta \in \Theta : |\theta - \theta_\star| \leq \delta\}$; $\tau \in (0, 1/\alpha - 2)$ and $\bar{\tau} = \tau/2$. Since $V_1 = V^\alpha$ and $V_2 = V^{2\alpha}$, it holds by iterating the drift condition (5.11) that

$$\sup_{k \geq m} \mathbb{E}\left[\left(V_1^{2+\tau}(X_{k+1}) + V_2^{1+\bar{\tau}}(X_{k+1})\right) \mathbb{1}_{\bigcap_{m \leq j \leq k} \{\theta_j \in \mathcal{K}\}}\right] \leq 2 \mathbb{E}\left[V^{2\alpha+\tau\alpha}(X_m)\right] + 2 \sup_{\theta \in \mathcal{K}} b_\theta (1 - \lambda_\theta)^{-1}.$$

Under the assumption (iv), the second term in the RHS is finite. Since θ_0 is fixed and $H(\theta, x)$ is bounded uniformly in θ for θ in a compact set, it can be proved by a simple induction that the random variables $\theta_1, \cdots, \theta_{m-1}$ are in a (non random) compact subset $\mathcal{K}'$ of Θ. Therefore, by iterating the drift again, we prove that $\mathbb{E}\left[V^{2\alpha+\tau\alpha}(X_m)\right] < \infty$. This concludes the proof of A4c.

Following the same lines as in the proof of ([2], Prop. 6.1), it can be proved by using (5.11) and the assumption (v) that for any compact $\mathcal{K} \subseteq \Theta$ and any $\beta' < \beta$, there exists C such that for any $\theta, \theta' \in \mathcal{K}$,

$$|h(\theta) - h(\theta)| + \|\widehat{H}_\theta - \widehat{H}_{\theta'}\|_{V^\alpha} + \|Q_\theta \widehat{H}_\theta - Q_{\theta'} \widehat{H}_{\theta'}\|_{V^\alpha} \leq C \, |\theta - \theta'|^{\beta'}. \tag{5.14}$$

Similarly, it can be proved along the same lines as the proof of ([2], Prop. 6.1). that $\|U_\theta - U_{\theta'}\|_{V^{2\alpha}} \leq C|\theta - \theta'|^{\beta'}$ provided one has

$$\sup_{\theta, \theta' \in \mathcal{K}} |\theta - \theta'|^{-\beta} \, \|F_\theta - F_{\theta'}\|_{V^{2\alpha}} < \infty. \tag{5.15}$$

Let us prove this property. Note that $F_\theta(x) - F_{\theta'}(x)$ is of the form $\mu_1 H_1 H_1^T - \mu_2 H_2 H_2^T - (\mu_1 H_1)(\mu_1 H_1)^T + (\mu_2 H_2)(\mu_2 H_2)^T$ for some probability measures μ_1, μ_2. Writing

$$\mu_1 H_1 H_1^T - \mu_2 H_2 H_2^T = (\mu_1 - \mu_2)(H_1 H_1^T) + \mu_2 H_1 (H_1 - H_2)^T + \mu_2 (H_1 - H_2) H_2^T,$$

$$(\mu_1 H_1)(\mu_1 H_1^T) - (\mu_2 H_2)(\mu_2 H_2^T) = \mu_1 H_1 (\mu_1 H_1 - \mu_2 H_2)^T + (\mu_1 H_1 - \mu_2 H_2) \mu_2 H_2^T,$$

and using the assumption (v) and (5.14), the property (5.15) holds. This concludes the proof of the first statement of A4d with $b = \beta'$.

Finally, let us consider the second statement. Since $\sup_{\theta \in \mathcal{K}} \pi_\theta(V) < \infty$ for any compact $\mathcal{K}$ of Θ (see (5.13)), the dominated convergence theorem and (5.15) imply that with probability one,

$$\mathbb{1}_{\lim_q \theta_q = \theta_\star} \lim_n \int \pi_{\theta_\star}(dy) \, (F_{\theta_n}(y) - F_{\theta_\star}(y)) = 0. \tag{5.16}$$

In addition, for any $k \geq 1$, $x \in \mathbb{X}$ and $\delta > 0$,

$$|\pi_{\theta_n} F_{\theta_n} - \pi_{\theta_\star} F_{\theta_n}| \, \mathbb{1}_{|\theta_n - \theta_\star| \leq \delta} \leq \left| Q_{\theta_n}^k F_{\theta_n}(x) - Q_{\theta_\star}^k F_{\theta_n}(x) \right| \mathbb{1}_{|\theta_n - \theta_\star| \leq \delta} + 2V^{2\alpha}(x) \sup_{|\theta - \theta_\star| \leq \delta} C_\theta \|F_\theta\|_{V^{2\alpha}} \rho_\theta^k. \tag{5.17}$$

By assumption (v), for any $\delta > 0$, there exists a constant C_1 such that for any x,

$$|Q_{\theta_n} F_{\theta_n}(x) - Q_{\theta_\star} F_{\theta_n}(x)| \, \mathbb{1}_{|\theta_n - \theta_\star| \leq \delta} \leq C_1 \sup_{|\theta - \theta_\star| \leq \delta} \|F_\theta\|_{V^{2\alpha}} V^{2\alpha}(x) |\theta_n - \theta_\star|^\beta;$$

and by a trivial induction on k, there exists a constant C_k such that for any x,

$$\left| Q_{\theta_n}^k F_{\theta_n}(x) - Q_{\theta_\star}^k F_{\theta_n}(x) \right| \mathbb{1}_{|\theta_n - \theta_\star| \leq \delta} \leq C_k \sup_{|\theta - \theta_\star| \leq \delta} \|F_\theta\|_{V^{2\alpha}} V^{2\alpha}(x) |\theta_n - \theta_\star|^\beta. \tag{5.18}$$

The controls (5.16) to (5.18) imply $\lim_n \pi_{\theta_n} F_{\theta_n} = \pi_{\theta_\star} F_{\theta_\star}$ almost-surely, on the set $\{\lim_q \theta_q = \theta_\star\}$.

5.8. Technical lemmas

Results below are classical and the proofs are omitted here; they are provided in the supplementary material.

Lemma 5.6. *Let $(\Omega, \mathcal{A}, \mathbb{P}, \{\mathcal{F}_n, n \geq 0\})$ be a filtered probability space and set $\mathcal{F}_\infty = \sigma(\mathcal{F}_n, n \geq 1)$. Let $B \in \mathcal{F}_\infty$. There exists a $\mathcal{F}_n$-adapted sequence $\{A_n, n \geq 0\}$ such that $\lim_n \mathbb{1}_{A_n} = \mathbb{1}_B$ $\mathbb{P}$-a.s.*

Lemma 5.7. *Let $|\cdot|$ be any matrix norm. Let $\{A_k, \ k \geq 0\}$ be a sequence of square matrix such that $\lim_k |A_k - A| = 0$ where A is a Hurwitz matrix. Denote by $-L$, $L > 0$, the largest real part of its eigenvalues. Let $\{\gamma_k, \ k \geq 0\}$ be a positive sequence such that $\lim_k \gamma_k = 0$. For any $0 < L' < L$, there exists a positive constant C such that for any $k \leq n$*

$$|(\mathrm{Id} + \gamma_n A_n) \cdots (\mathrm{Id} + \gamma_{k+1} A_{k+1})(\mathrm{Id} + \gamma_k A_k)| \leq C \exp\left(-L' \sum_{j=k}^n \gamma_j\right).$$

Lemma 5.8. *Let γ_k be a positive sequence such that $\lim_k \gamma_k = 0$ and $\sum_k \gamma_k = \infty$. Let $\{e_k, \ k \geq 0\}$ be a non-negative sequence. Then*

$$\limsup_n \ \gamma_n^{-p} \sum_{k=1}^n \gamma_k^{p+1} \ e_k \ \exp\left(-b \sum_{j=k+1}^n \gamma_j\right) \leq \frac{1}{C(b,p)} \limsup_n e_n,$$

(i) *with $C(b,p) = b$, for any $b > 0, p \geq 0$ if $\log(\gamma_{k-1}/\gamma_k) = o(\gamma_k)$.*
(ii) *with $C(b,p) = b - p/\gamma_\star$, for any $b\gamma_\star > p \geq 0$ if there exists $\gamma_\star > 0$ such that $\log(\gamma_{k-1}/\gamma_k) \sim \gamma_k/\gamma_\star$.*

By convention, $\sum_{j=n+1}^n \gamma_j = 0$.

Lemma 5.9. *Let $U_\star$ be a positive definite matrix.*

(a) *Assume* C1b–c *and* C4a. *Consider the equation*

$$v_{n+1} = v_n + \gamma_n f(v_n) + \frac{\gamma_n - \gamma_{n+1}}{\gamma_{n+1}} v_n + (\gamma_{n+1} - \gamma_n) U_\star + \gamma_n \gamma_{n+1} \nabla h(\theta_\star) v_n \nabla h(\theta_\star)^T,$$

where $f(v) \stackrel{\text{def}}{=} U_\star + \nabla h(\theta_\star) v + v \nabla h(\theta_\star)^T$. Then there exists an unique positive definite matrix V such that $f(V) = 0$ and $\lim_n v_n = V$.

(b) *Assume* C1b–c *and* C4b. *Consider the equation*

$$v_{n+1} = v_n + \gamma_n f(v_n) + (\gamma_{n+1} - \gamma_n) U_\star + \gamma_n \gamma_{n+1} \nabla h(\theta_\star) v_n \nabla h(\theta_\star)^T,$$

where $f(v) \stackrel{\text{def}}{=} U_\star + \nabla h(\theta_\star) v + v \nabla h(\theta_\star)^T + \gamma_\star^{-1} v$. Then there exists an unique positive definite matrix V such that $f(V) = 0$ and $\lim_n v_n = V$.

Lemma 5.10. *Define the sequence $\{x_n, n \geq 0\}$ by*

$$x_{n+1} = x_n + \gamma_{n+1} A x_n + \gamma_{n+1} \zeta_{n+1}, \qquad x_0 \in \mathbb{R}^d,$$

where $\{\gamma_n, n \geq 1\}$ is a positive sequence, $\{\zeta_n, n \geq 1\}$ is a $\mathbb{R}^d$-valued sequence and A is a $d \times d$ matrix. Then

$$A \sum_{k=0}^n x_k = -\sum_{k=0}^n \zeta_{k+1} + \left(\frac{x_{n+1}}{\gamma_{n+1}} - \frac{x_0}{\gamma_1} \right) + \sum_{k=1}^n \left(\frac{1}{\gamma_k} - \frac{1}{\gamma_{k+1}} \right) x_k.$$

References

[1] C. Andrieu, G. Fort and M. Vihola, *Quantitative Convergence Rates for sub-geometric Markov chains*. Technical report `arXiv:1309.0622v2` (2014).

[2] C. Andrieu, E. Moulines and P. Priouret, Stability of Stochastic Approximation under Verifiable Conditions. *SIAM J. Control Optim.* **44** (2005) 283–312.

[3] M. Benaim, Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, vol. 1709 of *Lect. Notes Math.* Springer, Berlin (1999) 1–68.

[4] A. Benveniste, M. Metivier and P. Priouret, Adaptive Algorithms and Stochastic Approximations. Springer Verlag (1987).

[5] P. Bianchi, G. Fort and W. Hachem, Performance of a Distributed Stochastic Approximation Algorithm. *IEEE Trans. Inform. Theory* **59** (2013) 391–405.

[6] V.S. Borkar, Stochastic Approximation: A Dynamical Systems Viewpoint. Cambridge University Press (2008).

[7] C. Bouton, Approximation gaussienne d'algorithmes stochastiques à dynamique markovienne. *Ann. Inst. Henri Poincaré* **24** (1988) 131–155.

[8] R. Buche and H.J. Kushner, Rate of Convergence for Constrained Stochastic Approximation Algorithms. *SIAM J. Control Optim.* **40** (2001) 1011–1041.

[9] H. Chen, Stochastic Approximation and its Applications. Kluwer Academic Publishers (2002).

[10] H.F. Chen, L. Guo and A.J. Gao, Convergence and robustness of the Robbins−Monro algorithms truncated at randomly varying bounds. *Stoch. Proc. Appl.* (1988).

[11] B. Delyon, *Stochastic Approximation with Decreasing Gain: Convergence and Asymptotic Theory*. Technical report. Publication interne 952, IRISA (2000).

[12] B. Delyon, M. Lavielle and E. Moulines, Convergence of a Stochastic Approximation Version of the EM Algorithm. *Ann. Statist.* **27** (1999) 94–128.

[13] M. Duflo, Algorithmes stochastiques. Springer (1996).

[14] V. Fabian, On asymptotically efficient recursive estimation. *Ann. Statist.* **6** (1978) 854–866.

[15] G. Fort, B. Jourdain, E. Kuhn, T. Lelièvre and G. Stoltz, *Convergence of the Wang-Landau algorithm*. Technical report. LTCI and CERMICS (2013). Submitted in Math. Comput. (2014).

[16] G. Fort, E. Moulines and P. Priouret, Convergence of Adaptive and Interacting Markov Chain Monte Carlo Algorithms. *Ann. Statist.* **39** (2012) 3262–3289.

[17] P. Hall and C.C. Heyde, Martingale Limit Theory and its Application. Academic Press, New York, London (1980).

[18] O. Hernandez−Lerma and J.B. Lasserre, Markov Chains and Invariant Probabilities. Birkhäuser (2003).

[19] R.A. Horn and C.R. Johnson, Topics in matrix analysis. Cambridge University Press (1994).

[20] H. Kushner, Stochastic approximation: a survey. *Wiley Interdisciplinary Reviews: Comput. Statist.* **2** (2010) 87–96.

[21] H. Kushner and H. Huang, Rates of convergence for Stochastic Approximation type algorithms. *SIAM J. Control Optim.* **17** (1979) 607–617.

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

[23] J. Lelong, Asymptotic normality of randomly truncated stochastic algorithms. *ESAIM: PS* **17** (2013) 105–119.

[24] S.P. Meyn and R.L. Tweedie, Markov Chains and Stochastic Stability. Cambridge University Press (2009).

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

[26] B.T. Polyak and A.B. Juditsky, Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30** (1992) 838–855.

[27] D. Ruppert, Handbook of Sequential Analysis, chapter Stochastic Approximation. Marcel Decker (1991).

[28] J.C. Spall, Introduction to Stochastic Search and Optimization. Wiley-Interscience (2003).

[29] Y. Zhu, Asymptotic Normality for a Vector Stochastic Difference Equation with Applications in Stochastic Approximation. *J. Multivariate Anal.* **57** (1996) 101–118.