Serdica

Mathematical Journal

Сердика

Математическо списание

Serdica Math. J. **34** (2008), 651–688

Serdica
Mathematical Journal

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

ON THE RECURSIVE ESTIMATION OF THE LOCATION AND OF THE SIZE OF THE MODE OF A PROBABILITY DENSITY

Khédidja Djeddour, Abdelkader Mokkadem, Mariane Pelletier

Communicated by N. M. Yanev

Abstract. Tsybakov [31] introduced the method of stochastic approximation to construct a recursive estimator of the location θ of the mode of a probability density. The aim of this paper is to provide a companion algorithm to Tsybakov's algorithm, which allows to simultaneously recursively approximate the size μ of the mode. We provide a precise study of the joint weak convergence rate of both estimators. Moreover, we introduce the averaging principle of stochastic approximation algorithms to construct asymptotically efficient algorithms approximating the couple (θ, μ).

1. Introduction. The most famous use of stochastic approximation algorithms in the framework of nonparametric statistics is the work of Kiefer and Wolfowitz [14], who built up an algorithm which allows the approximation of the

2000 *Mathematics Subject Classification:* 62G07, 62L20.
Key words: Location and size of the mode, nonparametric estimation, stochastic approximation algorithms, averaging principle.

maximizer of a regression function, which is observable at any level. Their well-known algorithm was widely discussed and extended in many directions (see, among many others, [1], [10], [15], [12], [27], [3], [29], [24], [6], [30], [4], [7]). In particular, Mokkadem and Pelletier [20] provided a companion algorithm to Kiefer-Wolfowitz's algorithm in order to simultaneously approximate the location and the size of the mode of the regression function. Stochastic approximation algorithms were also introduced by Révész [25, 26] to estimate a regression function from a sample of random variables, and by Tsybakov [31] to approximate the mode of a probability density. The aim of this paper is to provide a companion algorithm to Tsybakov's algorithm in order to simultaneously approximate the location and the size of the mode of a probability density.

Let us recall Robbins-Monro's scheme to construct approximation algorithms searching the zero z^* of an unknown function $h \ : \mathbb{R}^d \to \mathbb{R}^d$, which is observable at any level. First, $Z_0 \in \mathbb{R}^d$ is arbitrarily chosen, and then the sequence (Z_n) is recursively defined by setting

$$Z_n = Z_{n-1} + \gamma_n W_n,$$

where W_n is an observation of the function h at the point Z_{n-1}, and where the stepsize (γ_n) is a sequence of positive real numbers going to zero.

Let $X_1, \ldots, X_n$ be independent, identically distributed $\mathbb{R}^d$-valued random vectors, let f denote the probability density of X_1, and assume that f has a unique maximizer θ. To construct a stochastic algorithm approximating the maximizer θ of f, Tsybakov [31] defines an algorithm searching the zero of ∇f, the gradient of f, in the following way. First $\theta_0 \in \mathbb{R}^d$ is arbitrarily chosen, and then, for $n \geq 1$, θ_n is recursively defined by setting

$$\theta_n = \theta_{n-1} + \gamma_n W_n^{(\theta)},$$

where, following Robbins-Monro's procedure, $W_n^{(\theta)}$ must be an "observation" of the function ∇f at the point θ_{n-1}. Now, contrary to Robbins-Monro's framework, the function ∇f is not observable at any level, the only available observations being the random vectors X_i. In order to build up the "observation" $W_n^{(\theta)}$, Tsybakov [31] follows the method used by Révész [25], and introduces a kernel K (that is, a function satisfying $\int_{\mathbb{R}^d} K(x)dx = 1$), and a bandwidth (h_n) (that is, a sequence of positive real numbers going to zero); noting that $h_n^{-(d+1)} \nabla K(h_n^{-1}[x - X_n])$ can be regarded as an "observation" of the function ∇f at the point x, Tsybakov [31] sets $W_n^{(\theta)} = h_n^{-(d+1)} \nabla K \left(h_n^{-1}[\theta_{n-1} - X_n] \right)$, so that his algorithm approximating θ is defined by the recursive relation

$$\theta_n = \theta_{n-1} + \gamma_n \frac{1}{h_n^{d+1}} \nabla K \left(\frac{\theta_{n-1} - X_n}{h_n} \right). \tag{1}$$

Tsybakov [31] proves the srong consistency of θ_n, and establishes an upper bound of its mean squared error, as well as a minimax result.

In order to construct a companion algorithm to Tsybakov's algorithm (1), which approximates the size μ of the mode θ of the probability density f (in other words, which approximates $\mu = f(\theta)$), we define an algorithm searching the zero of the function $g \ : y \mapsto f(\theta) - y$. Following Robbins-Monro's scheme, we set $\mu_0 \in \mathbb{R}$, and, for $n \geq 1$,

$$\mu_n = \mu_{n-1} + \gamma_n W_n^{(\mu)},$$

where $W_n^{(\mu)}$ is an "observation" of the function g at the point μ_{n-1}. Let $(\tilde{h}_n)$ be a bandwidth (which may be different from (h_n)); noting that $\widetilde{h}_n^{-d} K(\widetilde{h}_n^{-1}[x - X_n])$ can be regarded as an "observation" of the function f at the point x, we set $W_n^{(\mu)} = \widetilde{h}_n^{-d} K \left(\widetilde{h}_n^{-1}[\theta_{n-1} - X_n] \right) - \mu_{n-1}$. The stochastic approximation algorithm we introduce to estimate μ is thus defined by the recursive relation

$$\mu_n = \mu_{n-1} - \gamma_n \mu_{n-1} + \gamma_n \frac{1}{\widetilde{h}_n^d} K \left(\frac{\theta_{n-1} - X_n}{\widetilde{h}_n} \right). \tag{2}$$

We prove that μ_n is strongly consistent, and we establish the weak convergence rate of (θ_n, μ_n) defined by the algorithms (1) and (2). We prove in particular that, for (θ_n) and (μ_n) to converge simultaneously at the optimal rate, the stepsize (γ_n) must be chosen such that $\lim_{n\to\infty} n\gamma_n = \gamma_0 \in (0, \infty)$, and the bandwidths (h_n) and $(\tilde{h}_n)$ must converge to zero at different rates. Now, as it is often the case in the framework of stochastic approximation algorithms, the choice of a stepsize satisfying $\lim_{n\to\infty} n\gamma_n = \gamma_0$ induces conditions on γ_0, which are difficult to handle because depending on an unknown parameter (in the present framework, γ_0 must be larger than a quantity involving the Hessian $D^2 f(\theta)$ of f at θ). The famous approach to obtain optimal convergence rates for stochastic approximation algorithms without tedious condition on the stepsize is to use the averaging principle independently introduced by Ruppert [28] and Polyak [22]. Their averaging procedure, which was widely discussed and extended (see, among many others, [32], [5], [23], [16], and [6]) allows to obtain asymptotically efficient algorithms, that is, algorithms which not only converge at the optimal rate, but which also have a minimal asymptotic covariance matrix. This procedure consists in: (i) running the approximation algorithm by using slower stepsizes; (ii) computing a suitable average of the approximations obtained in (i).

To apply the averaging principle to the approximating algorithms (1) and (2), we proceed as follows. First, we run the algorithms (1) and (2) with a slower stepsize satisfying $\lim_{n\to\infty} n\gamma_n = \infty$. Then, we define the average $\overline{\theta}_n$ of the θ_k and the average $\overline{\mu}_n$ of the μ_k by setting

$$\overline{\theta}_n = \frac{1}{\sum_{k=1}^n h_k^{d+2}} \sum_{k=1}^n h_k^{d+2}\theta_k, \tag{3}$$

$$\overline{\mu}_n = \frac{1}{\sum_{k=1}^n \widetilde{h}_k^{d}} \sum_{k=1}^n \widetilde{h}_k^{d}\mu_k. \tag{4}$$

We establish the weak convergence rate of $(\overline{\theta}_n, \overline{\mu}_n)$. We prove in particular that adequate choices of the bandwidths (h_n) and $(\widetilde{h}_n)$ allow to obtain simultaneously the asymptotic efficiency of both sequences $(\overline{\theta}_n)$ and $(\overline{\mu}_n)$.

To conclude this introduction, let us underline that the proof of the asymptotic behaviour of the sequences (θ_n), (μ_n), $(\overline{\theta}_n)$, and $(\overline{\mu}_n)$ deeply relies on the application of asymptotic properties of a general stochastic approximation algorithm. Our paper is thus organized as follows. Our main results on (θ_n), (μ_n), $(\overline{\theta}_n)$, and $(\overline{\mu}_n)$ are stated in Section 2. In Section 3, we state some asymptotic properties of a general stochastic approximation algorithm, and prove them in Section 4. Finally, Section 5 is reserved to the proof of our main results.

2. Assumptions and main results. Throughout this paper, $\|.\|$ denotes the Euclidean norm. For any function ϕ, we set $\|\phi\|_\infty = \sup_x \|\phi(x)\|$. For any matrix A, A^T denotes the transpose of A, and I_d denotes the $d \times d$ identity matrix. Moreover, we consider the following class of regularly varying sequences.

Definition 1. *Let $\gamma \in \mathbb{R}$ and $(v_n)_{n\geq 1}$ be a nonrandom positive sequence. We say that $(v_n) \in \mathcal{GS}(\gamma)$ if*

$$\lim_{n\to+\infty} n\left[1 - \frac{v_{n-1}}{v_n}\right] = \gamma. \tag{5}$$

Condition (5) was introduced by Galambos and Seneta [11] (see also [2]); it was used in [20] in the context of stochastic approximation algorithms. Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^{\gamma}(\log n)^{b}$, $n^{\gamma}(\log\log n)^{b}$, and so on.

We can now state our assumptions.

(H1) (i) K is continuously differentiable, $\int_{\mathbb{R}^d} K(x)dx = 1$, $\int_{\mathbb{R}^d} \|x\|^2 |K(x)|dx < \infty$, and $\lim_{\|x\|\to\infty} K(x) = 0$.

(ii) K is even in each of its coordinates.

(iii) $\int_{\mathbb{R}^d} \|x\| \|\nabla K(x)\| dx < \infty$ and $\|\nabla K\|_{\infty} < \infty$.

(H2) (i) f is three times continuously differentiable, $\|D^2 f\|_{\infty} < \infty$ and $\|D^3 f\|_{\infty} < \infty$.

(ii) $[\nabla f(x)]^T (x - \theta) < 0$ for all $x \neq \theta$.

(iii) The largest eigenvalue $-L^{(\theta)}$ of $D^2 f(\theta)$ is negative.

(H3) (i) $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha \in (1/2, 1]$.

(ii) $(h_n) \in \mathcal{GS}(-a)$ with $a \in \left(\frac{1-\alpha}{4}, \frac{2\alpha-1}{d+2}\right)$.

(iii) $\lim_{n\to\infty} n\gamma_n \in \left(\min\left\{\frac{1-a(d+2)}{2L^{(\theta)}}\ ;\ \frac{2a}{L^{(\theta)}}\right\}\ ;\infty\right]$.

(iv) $\left(\tilde{h}_n\right) \in \mathcal{GS}(-\tilde{a})$ with $\tilde{a} \in \left(\frac{1-\alpha}{4}, \frac{2\alpha-1}{d}\right)$.

Remark 1. Assumption (H1)(ii) implies in particular that $\int_{\mathbb{R}^d} x_i K(x)dx = 0$ for all $i \in \{1, \ldots, d\}$ and $\int_{\mathbb{R}^d} x_i x_j K(x)dx = 0$ for all $i \neq j$. Moreover, assumptions (H1)(i) and (H1)(iii) imply that $\int_{\mathbb{R}^d} \|\nabla K(x)\| dx < \infty$ and $\int_{\mathbb{R}^d} \|\nabla K(x)\|^2 dx < \infty$.

Remark 2. (H3)(ii) and (H3)(iv) imply that $a < \alpha/(d+2)$ and $\tilde{a} < \alpha/d$, respectively.

Our first result is the following proposition.

Proposition 1. *Let* (μ_n) *be the sequence defined by the stochastic approximation algorithms* (1) *and* (2). *Under* (H1)–(H3), $\lim_{n\to\infty} \mu_n = \mu$ *a.s.*

Remark 3. The assumptions, which ensure the strong consistency of the sequence (θ_n) defined by the stochastic approximation algorithms (1) are: (H1), (H2), (H3)(i)–(ii), together with the condition $\sum \gamma_n = \infty$ (see Section 5.1.1). Note that this lattest condition is weaker than (H3)(iii).

To establish the weak convergence rate of (θ_n, μ_n), we need the following additional assumption.

(H4) (i) $\lim_{n\to\infty} n\gamma_n \in \left(\min\left\{ \frac{1-\tilde{a}d}{2}; 2\tilde{a} \right\} ; \infty \right]$.

(ii) $\tilde{a} < 2a$ and $a(d+2) + 2\tilde{a} < \alpha$.

We also need to introduce the following notations.

$$\xi = \lim_{n\to\infty} (n\gamma_n)^{-1},$$

$$R^{(\theta)} = \frac{1}{2}\nabla \left(\sum_{i=1}^{d} \left[\int_{\mathbb{R}^d} x_i^2 K(x)dx \right] \frac{\partial^2 f}{\partial^2 z_i} \right)(\theta), \tag{6}$$

$$R^{(\mu)} = \frac{1}{2}\sum_{i=1}^{d} \left(\left[\int_{\mathbb{R}^d} x_i^2 K(x)dx \right] \frac{\partial^2 f}{\partial^2 z_i}(\theta) \right), \tag{7}$$

$$\Sigma^{(\mu)} = [2 - \xi(1-\tilde{a}d)]^{-1} f(\theta) \int_{\mathbb{R}^d} K^2(z)dz, \tag{8}$$

$$G = \int_{\mathbb{R}^d} \nabla K(z)[\nabla K(z)]^T dz, \tag{9}$$

and $\Sigma^{(\theta)}$ is the solution of Lyapounov's equation

$$\left(D^2 f(\theta) + \frac{[1-a(d+2)]\xi}{2} I_d \right) \Sigma^{(\theta)} + \Sigma^{(\theta)} \left(D^2 f(\theta) + \frac{[1-a(d+2)]\xi}{2} I_d \right) = -f(\theta)G. \tag{10}$$

The following theorem gives the weak convergence rate of (θ_n, μ_n).

Theorem 1. *Let* (θ_n, μ_n) *be defined by the stochastic approximation algorithms* (1) *and* (2), *and assume that* $(H1)$–$(H4)$ *hold.*

- *If* $\lim_{n\to\infty}\gamma_n^{-1}h_n^{d+6}=0$ *and* $\lim_{n\to\infty}\gamma_n^{-1}\widetilde{h}_n^{d+4}=0$*, then*
$$\begin{pmatrix}\sqrt{\gamma_n^{-1}h_n^{d+2}}\;(\theta_n-\theta)\\ \sqrt{\gamma_n^{-1}\widetilde{h}_n^{d}}\;(\mu_n-\mu)\end{pmatrix}\overset{\mathcal{D}}{\to}\mathcal{N}\left(0,\begin{pmatrix}\Sigma^{(\theta)} & 0\\ 0 & \Sigma^{(\mu)}\end{pmatrix}\right),$$
where $\overset{\mathcal{D}}{\to}$ *denotes the convergence in distribution.*
- *If there exist* $c>0$ *and* $\widetilde{c}>0$ *such that* $\lim_{n\to\infty}\gamma_n^{-1}h_n^{d+6}=c$ *and* $\lim_{n\to\infty}\gamma_n^{-1}\widetilde{h}_n^{d+4}=\widetilde{c}$*, then*
$$\begin{pmatrix}\sqrt{\gamma_n^{-1}h_n^{d+2}}\;(\theta_n-\theta)\\ \sqrt{\gamma_n^{-1}\widetilde{h}_n^{d}}\;(\mu_n-\mu)\end{pmatrix}$$
$$\overset{\mathcal{D}}{\to}\mathcal{N}\left(\begin{matrix}-\sqrt{c}\,\left[D^2f(\theta)+2a\xi I_d\right]^{-1}R^{(\theta)}\\ \sqrt{\widetilde{c}}\,(1-2\tilde{a}\xi)^{-1}R^{(\mu)}\end{matrix},\begin{pmatrix}\Sigma^{(\theta)} & 0\\ 0 & \Sigma^{(\mu)}\end{pmatrix}\right).$$
- *If* $\lim_{n\to\infty}\gamma_n^{-1}h_n^{d+6}=\infty$ *and* $\lim_{n\to\infty}\gamma_n^{-1}\widetilde{h}_n^{d+4}=\infty$*, then*
$$\begin{pmatrix}h_n^{-2}\;(\theta_n-\theta)\\ \widetilde{h}_n^{-2}\;(\mu_n-\mu)\end{pmatrix}\overset{\mathbb{P}}{\to}\begin{pmatrix}-\left[D^2f(\theta)+2a\xi I_d\right]^{-1}R^{(\theta)}\\ (1-2\tilde{a}\xi)^{-1}R^{(\mu)}\end{pmatrix}.$$

Remark 4. In the framework of Parts 1 and 2 of Theorem 1, that is, when $\lim_{n\to\infty}\gamma_n^{-1}h_n^{d+6}\in[0,\infty)$ and $\lim_{n\to\infty}\gamma_n^{-1}\widetilde{h}_n^{d+4}\in[0,\infty)$, we have $\alpha\le a(d+6)$ and $\alpha\le\widetilde{a}(d+4)$. In view of (H4)(i), it follows that $\xi<2[1-\widetilde{a}d]^{-1}$, so that $\Sigma^{(\mu)}>0$. In view of (H3)(iii), it follows that $\xi<2[1-a(d+2)]^{-1}L^{(\theta)}$, so that the matrix $D^2f(\theta)+\dfrac{[1-a(d+2)]\xi}{2}I_d$ is negative definite. Proposition 1 in [18] ensuring that G is positive definite, $\Sigma^{(\theta)}$ is thus positive definite. Now, in the framework of Parts 2 and 3 of Theorem 1, that is, when $\lim_{n\to\infty}\gamma_n^{-1}h_n^{d+6}\in(0,\infty]$ and $\lim_{n\to\infty}\gamma_n^{-1}\widetilde{h}_n^{d+4}\in(0,\infty]$, we have $\alpha\ge a(d+6)$ and $\alpha\ge\widetilde{a}(d+4)$. In view of (H3)(iii) and (H4)(i), it follows that $2a\xi<L^{(\theta)}$ and $2\widetilde{a}\xi<1$, which ensures that the limits in Parts 2 and 3 of Theorem 1 are well defined.

A stochastic approximation algorithm is said to be asymptotically efficient if it converges at the optimal rate and if its asymptotic covariance matrix is minimum (with respect to the order of symmetric matrices). In view of Theorem 1, the couple (θ_n,μ_n) converges at the optimal rate if the stepsize (γ_n) is chosen in $\mathcal{GS}(-1)$ and such that $\lim_{n\to\infty}n\gamma_n=\gamma_0$ with, in view of assumptions (H3)(iii) and (H4)(i),

$$\gamma_0>\max\left\{\min\left\{\frac{1-a(d+2)}{2L^{(\theta)}}\;;\;\frac{2a}{L^{(\theta)}}\right\}\;;\;\min\left\{\frac{1-\tilde{a}d}{2};2\tilde{a}\right\}\right\},\tag{11}$$

and if the bandwidths (h_n) and $(\tilde{h}_n)$ are chosen such that $\lim_{n\to\infty} \gamma_n^{-1} h_n^{d+6} = c > 0$ and $\lim_{n\to\infty} \gamma_n^{-1} \widetilde{h}_n^{d+4} = \widetilde{c} > 0$, respectively. We then have:

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}\ (\theta_n - \theta) \\ \sqrt{n\widetilde{h}_n^{d}}\ (\mu_n - \mu) \end{pmatrix} \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(\begin{matrix} -\sqrt{c\gamma_0}\ \left[D^2 f(\theta) + 2a\xi I_d\right]^{-1} R^{(\theta)} \\ \sqrt{\widetilde{c}\gamma_0}\ (1-2\tilde{a}\xi)^{-1} R^{(\mu)} \end{matrix} , \begin{pmatrix} \gamma_0 \Sigma^{(\theta)} & 0 \\ 0 & \gamma_0 \Sigma^{(\mu)} \end{pmatrix} \right).$$

Now, for (θ_n) (respectively, (μ_n)) to be asymptotically efficient, the asymptotic covariance matrix $\gamma_0 \Sigma^{(\theta)}$ (respectively, $\gamma_0 \Sigma^{(\mu)}$) must also be minimum. The following proposition is proved in Section 5.3.

Proposition 2.

1. *For the algorithm* (1) *to be asymptotically efficient, the stepsize* (γ_n) *must equal the matricial sequence* $(-[1-a(d+2)][D^2 f(\theta)]^{-1} n^{-1})$, *the bandwidth* (h_n) *must satisfy* $\lim_{n\to\infty} nh_n^{d+6} = c > 0$ *(in which case* $a = [d+6]^{-1}$*), and we then have*

$$\sqrt{nh_n^{d+2}}\,(\theta_n - \theta) \ \stackrel{\mathcal{D}}{\to}\ \mathcal{N}\left(-2\sqrt{c}[D^2 f(\theta)]^{-1} R^{(\theta)}, \Sigma_{\mathrm{opt}}^{(\theta)}\right),$$

   with $\Sigma_{\mathrm{opt}}^{(\theta)} = f(\theta)[1-a(d+2)][D^2 f(\theta)]^{-1} G [D^2 f(\theta)]^{-1}$.

2. *For the algorithm* (2) *to be asymptotically efficient, the stepsize* (γ_n) *must equal* $([1-\widetilde{a}d]n^{-1})$, *the bandwidth* $(\tilde{h}_n)$ *must satisfy* $\lim_{n\to\infty} n\widetilde{h}_n^{d+4} = \widetilde{c} > 0$ *(in which case* $\widetilde{a} = [d+4]^{-1}$*), and we then have*

$$\sqrt{n\widetilde{h}_n^{d}}\,(\mu_n - \mu) \ \stackrel{\mathcal{D}}{\to}\ \mathcal{N}\left(2\sqrt{\widetilde{c}} R^{(\mu)}, \Sigma_{\mathrm{opt}}^{(\mu)}\right),$$

   with $\Sigma_{\mathrm{opt}}^{(\mu)} = f(\theta)[1-\widetilde{a}d] \int_{\mathbb{R}^d} K^2(z)dz$.

In view of Proposition 2 and condition (11), it is possible to choose the stepsize (γ_n) leading to the asymptotic efficiency of the algorithm (2) only in the case when $4[d+4]^{-1} > \min\{[1-a(d+2)]/\ [2L^{(\theta)}]; 2a/L^{(\theta)}\}$. On the other hand, since the matrix $D^2 f(\theta)$ is unknown, it is not possible to choose the stepsize (γ_n) leading to the asymptotic efficiency of the algorithm (1). The following theorem, giving the weak convergence rate of the averaged algorithms (3) and (4), shows that $(\overline{\theta}_n)$ and $(\overline{\mu}_n)$ can be simultaneously asymptotically efficient, and this without any tedious condition on the stepsize (γ_n); to state it, we need the following additional assumption.

(H5) $\lim_{n\to\infty} n\gamma_n[\log(\sum_{k=1}^n \gamma_k)]^{-1} = \infty$.

Theorem 2. *Let (θ_n, μ_n) be defined by the stochastic approximation algorithms* (1) *and* (2)*, let $(\overline{\theta}_n, \overline{\mu}_n)$ be the averaged algorithms defined by* (3) *and* (4)*, and assume that $(H1)$–$(H5)$ hold.*

- *If $\lim_{n\to\infty} nh_n^{d+6} = 0$ and $\lim_{n\to\infty} n\widetilde{h}_n^{d+4} = 0$, then*

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}\,(\overline{\theta}_n - \theta) \\ \sqrt{n\widetilde{h}_n^{d}}\,(\overline{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0\,,\, \begin{pmatrix} \Sigma_{\text{opt}}^{(\theta)} & 0 \\ 0 & \Sigma_{\text{opt}}^{(\mu)} \end{pmatrix}\right). \tag{12}$$

- *If there exist $c > 0$ and $\widetilde{c} > 0$ such that $\lim_{n\to\infty} nh_n^{d+6} = c$ and $\lim_{n\to\infty} n\widetilde{h}_n^{d+4} = \widetilde{c}$, then*

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}\,(\overline{\theta}_n - \theta) \\ \sqrt{n\widetilde{h}_n^{d}}\,(\overline{\mu}_n - \mu) \end{pmatrix} \tag{13}$$
$$\xrightarrow{\mathcal{D}} \mathcal{N}\left(\begin{pmatrix} -2\sqrt{c}[D^2f(\theta)]^{-1}R^{(\theta)} \\ 2\sqrt{\widetilde{c}}R^{(\mu)} \end{pmatrix}, \begin{pmatrix} \Sigma_{\text{opt}}^{(\theta)} & 0 \\ 0 & \Sigma_{\text{opt}}^{(\mu)} \end{pmatrix}\right),$$

  and $(\overline{\theta}_n)$ and $(\overline{\mu}_n)$ are simultaneously asymptotically efficient.

- *If $\lim_{n\to\infty} nh_n^{d+6} = \infty$ and $\lim_{n\to\infty} n\widetilde{h}_n^{d+4} = \infty$, then*

$$\begin{pmatrix} h_n^{-2}\,(\overline{\theta}_n - \theta) \\ \widetilde{h}_n^{-2}\,(\overline{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathbb{P}} \begin{pmatrix} -\dfrac{[1-a(d+2)]}{1-a(d+4)}[D^2f(\theta)]^{-1}R^{(\theta)} \\ \dfrac{1-\widetilde{a}d}{1-\widetilde{a}(d+2)}R^{(\mu)} \end{pmatrix}. \tag{14}$$

Remark 5. In the case when $\lim_{n\to\infty} nh_n^{d+6} = \infty$ and $\lim_{n\to\infty} n\widetilde{h}_n^{d+4} = \infty$, we have $a \leq (d+6)^{-1}$ and $\widetilde{a} \leq (d+4)^{-1}$, so that the limit term in (14) is well defined.

3. Some preliminary results on stochastic approximation algorithms. As mentioned in the introduction, the proof of our main results deeply relies on the application of asymptotic properties of a general stochastic approximation algorithm searching the zero z^* of an unknown function $h : \mathbb{R}^d \to \mathbb{R}^d$. This algorithm is defined by setting $Z_0 \in \mathbb{R}^d$, and, for $n \geq 1$,

$$Z_n = Z_{n-1} + \gamma_n\left[h\left(Z_{n-1}\right) + R_n + c_n^{-1}\varepsilon_n\right], \tag{15}$$

where (γ_n) and (c_n) are nonrandom positive sequences going to zero, and where the random sequences (R_n) and (ε_n) are defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ equipped with a filtration $\mathcal{F} = (\mathcal{F}_n)$.

The algorithm (15) was widely studied under various assumptions; see, among many others, [21], [17], [8], [20], and the references therein. However, the results obtained in these references do not apply in the present framework. The aim of this section is to state the different properties, which will enable us to establish the asymptotic behaviour of the algorithms (1), (2), (3) and (4). To this end, we consider the algorithm (15) under the assumptions (A1)–(A7) stated below. (A1) says that the algorithm converges strongly to the zero z^* of the function h; this consistency property will be proved for (θ_n) and (μ_n) by applying Robbins-Monro's theorem (see Section 5.1). Assumptions (A1)–(A6) are classical conditions in the framework of stochastic approximation algorithms, and are adequate for the study of (1), (2), (3) and (4); (A2) requires that the unknown function $h : \mathbb{R}^d \to \mathbb{R}^d$ is smooth enough at the neighbourhood of its zero z^*, and that its Jacobian H at the point z^* is negative definite; (A4) requires that ε_n is a noise with finite conditional covariance matrix Γ_n satisfying $\lim_{n\to\infty} \Gamma_n = \Gamma$ a.s.; (A5) gives the convergence rate of the term (R_n). A contrario, (A7) is unusual in the framework of stochastic approximation algorithms; it replaces a condition on the moments of (ε_n), which is not fulfilled in our framework.

(A1) $\lim_{n\to\infty} Z_n = z^*$ a.s.

(A2) (i) There exist $\eta > 1$ and a neighbourhood of z^* on which $h(z) = H(z - z^*) + O(\|z - z^*\|^\eta)$.

(ii) H is diagonalizable and its largest eigenvalue $-L$ is negative.

(A3) Either $(c_n) \in \mathcal{GS}(-\tau)$ with $\tau \in (0, 1/2)$ or $(c_n) = 1$, in which case we set $\tau = 0$.

(A4) (i) $\mathbb{E}\left(\varepsilon_{n+1} | \mathcal{F}_n\right) = 0$.

(ii) There exists a nonrandom, positive definite matrix Γ such that $\lim_{n\to\infty} \mathbb{E}(\varepsilon_{n+1}\varepsilon_{n+1}^T | \mathcal{F}_n) = \Gamma$ a.s.

(A5) There exist $\rho \in \mathbb{R}^d$ and $(v_n) \in \mathcal{GS}(v^*)$, $v^* > 0$, such that $\lim_{n\to\infty} v_n R_n = \rho$ a.s.

(A6) (i) $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha \in \left(\max\{\frac{1}{2}, 2\tau\}, 1\right]$.

(ii) $\lim_{n\to\infty} n\gamma_n \in \left(\min\left\{\frac{1-2\tau}{2L}, \frac{v^*}{L}\right\}, \infty\right]$ where L and v^* are defined in (A2) and (A5) respectively.

(A7) There exists a sequence (w_n) adapted to $\mathcal{F}_n$ such that $\|\varepsilon_{n+1}\| \leq w_n$ for all n and such that $\lim_{n\to\infty} \gamma_n w_{n+1}^2 \log(\sum_{k=1}^n \gamma_k) = 0$.

Section 3.1 is reserved to the results on (15), which will enable us to establish the asymptotic behaviour of (1) and (2); Section 3.2 is devoted to the results on the averaged version of (15), which will enable us to establish the asymptotic behaviour of (3) and (4).

3.1. On the asymptotic behaviour of the stochastic approximation algorithm (15). The asymptotic behaviour of the algorithm (15) is given by those of the sequences (L_n) and (Δ_n) defined by:

$$\begin{aligned} L_n &= e^{(\sum_{k=1}^n \gamma_k)H} \sum_{k=1}^n e^{-(\sum_{j=1}^k \gamma_j)H} \gamma_k c_k^{-1} \varepsilon_k, \\ \Delta_n &= (Z_n - z^*) - L_n. \end{aligned}$$

Let Σ be the solution of Lyapounov's equation

$$\left(H + \frac{\xi(1-2\tau)}{2} I_d\right)\Sigma + \Sigma\left(H^T + \frac{\xi(1-2\tau)}{2} I_d\right) = -\Gamma,$$

where $\xi = \lim_{n\to\infty}(n\gamma_n)^{-1}$. The following three lemmas are proved in Section 4.

Lemma 1. *Let* (A2)–(A6) *hold. Moreover, assume that* $v^* \geq (\alpha - 2\tau)/2$ *and that there exists* $m^* > 2$ *such that* $\lim_{n\to\infty} \gamma_n^{-1+m^*/2}\mathbb{E}[\|\varepsilon_n\|^{m^*}|\mathcal{F}_{n-1}] = 0$ *a.s. Then, we have*

$$\sqrt{\gamma_n^{-1} c_n^2}\, L_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma).$$

Lemma 2. *Let* (A2)–(A7) *hold.*

- *If* $v^* \geq (\alpha - 2\tau)/2$, *then* $\|L_n\| = O\left(\sqrt{\gamma_n c_n^{-2} \log\left(\sum_{k=1}^n \gamma_k\right)}\right)$ *a.s.*
- *If* $v^* < (\alpha - 2\tau)/2$, *then* $\|L_n\| = o\left(v_n^{-1}\right)$ *a.s.*

Lemma 3. *Let* (A1)–(A7) *hold.*

- *If* $v^* \leq (\alpha - 2\tau)/2$, *then* $\lim_{n\to\infty} v_n \Delta_n = -[H + v^*\xi I_d]^{-1}\rho$ *a.s.*

- *If* $v^* > (\alpha - 2\tau)/2$, *then* $\lim_{n\to\infty} \sqrt{\gamma_n^{-1} c_n^2}\ \Delta_n = 0$ *a.s.*

The combination of Lemmas 2 and 3 gives the following upper bound of the strong convergence rate of (Z_n), which is usefull in the study of the averaged version of the algorithm (15) (see Section 4.6): under assumptions (A1)–(A7),

$$\|Z_n - z^*\| = O\left(\sqrt{\gamma_n c_n^{-2} \log\left(\sum_{k=1}^{n} \gamma_k\right)} + v_n^{-1}\right) \quad a.s. \tag{16}$$

The combination of Lemmas 1, 2 and 3 also gives, under assumptions (A1)–(A7), the weak convergence rate of (Z_n), which is the first step in the proof of the following lemma (see Section 4.4).

Lemma 4. *For the algorithm* (15) *to be asymptotically efficient, the stepsize* (γ_n) *must be chosen equal to* $(-[1-2\tau]H^{-1}n^{-1})$, *the sequences* (c_n) *and* (v_n) *must be such that* $\lim_{n\to\infty} nc_n^2 v_n^{-2} = c > 0$ *(in which case* $v^* = [1-2\tau]/2$*), and we then have*

$$\sqrt{nc_n^2}\,(Z_n - z^*) \ \stackrel{\mathcal{D}}{\to}\ \mathcal{N}\left(-2\sqrt{c}H^{-1}\rho, (1-2\tau)H^{-1}\Gamma[H^{-1}]^T\right).$$

3.2. On the averaged version of the stochastic approximation algorithm (15). As mentioned in the introduction, the averaging procedure introduced by Ruppert [28] and Polyak [22] consists in: (i) running the approximation algorithm (15) by using a slower stepsize; (ii) computing an average of the approximations obtained in (i). If (15) is Robbins-Monro's algorithm, then the average leading to the asymptotic efficiency is known to be the arithmetic one (see, for instance, [28] and [22]); if (15) is Kiefer-Wolfowitz's algorithm, then a weighted average must be used to get the efficiency (see, for instance, [6]). In this section, we establish in particular that, in order to get the asymptotic efficiency of the averaged version of the stochastic approximation algorithm (15), the average of the Z_k must be weighted by the c_k^2. We set

$$\overline{Z}_n \ = \ \frac{1}{\sum_{k=1}^n c_k^2} \sum_{k=1}^{n} c_k^2 Z_k,$$

and assume that the following additional conditions hold.

(A8) (i) $\lim_{n\to\infty} n\gamma_n \left[\log\left(\sum_{k=1}^n \gamma_k\right)\right]^{-1} = \infty$.

(ii) $\lim_{n\to\infty} n\gamma_n^{\eta} c_n^{2(1-\eta)} [\log(\sum_{k=1}^n \gamma_k)]^{\eta} = 0$, where η is defined in (A2)(i).

The asymptotic behaviour of $(\overline{Z}_n)$ is given by those of the sequences (Λ_n) and (Ξ_n) defined by

$$\begin{aligned}\Lambda_{n+1} &= \frac{-1}{\sum_{k=1}^n c_k^2} H^{-1} \sum_{k=1}^n c_k^2 c_{k+1}^{-1} \varepsilon_{k+1}, \\ \Xi_{n+1} &= \left(\overline{Z}_n - z^*\right) - \Lambda_{n+1}.\end{aligned}$$

The following lemmas are proved in Section 4.

Lemma 5. *Let* (A2)–(A6) *hold. Moreover, assume that there exists* $m^* > 2$ *such that* $\lim_{n\to\infty} n^{1-m^*/2}\mathbb{E}[\|\varepsilon_{n+1}\|^{m^*}|\mathcal{F}_n] = 0$ *a.s. Then, we have*

$$\sqrt{nc_n^2}\Lambda_{n+1}\xrightarrow{\mathcal{D}}\mathcal{N}\left(0\ ,\ (1-2\tau)H^{-1}\Gamma[H^{-1}]^T\right).$$

Lemma 6. *Assume that* (A1)–(A8) *hold.*

- *If* $\lim_{n\to\infty} v_n^{-1}\sqrt{nc_n^2} \in (0,\infty]$*, then* $\lim_{n\to\infty} v_n\ \Xi_{n+1} = \frac{-(1-2\tau)}{1-2\tau-v^*}H^{-1}\rho$ *a.s.*
- *If* $\lim_{n\to\infty} v_n^{-1}\sqrt{nc_n^2} = 0$*, then* $\lim_{n\to\infty} \sqrt{nc_n^2}\ \Xi_{n+1} = 0$ *a.s.*

The combination of Lemmas 5 and 6 gives the weak convergence rate of $(\overline{Z}_n)$ under assumptions (A1)–(A8):

- If $\lim_{n\to\infty} nc_n^2v_n^{-2} = 0$, then
$$\sqrt{nc_n^2}\left(\overline{Z}_n - z^*\right) \ \xrightarrow{\mathcal{D}}\ \mathcal{N}\left(0\ ,\ (1-2\tau)H^{-1}\Gamma[H^{-1}]^T\right).$$
- If there exists $c > 0$ such that $\lim_{n\to\infty} nc_n^2v_n^{-2} = c$ (and thus $v^* = (1-2\tau)/2$), then
$$\sqrt{nc_n^2}\left(\overline{Z}_n - z^*\right) \ \xrightarrow{\mathcal{D}}\ \mathcal{N}\left(-2\sqrt{c}H^{-1}\rho\ ,\ (1-2\tau)H^{-1}\Gamma[H^{-1}]^T\right),$$
and $(\overline{Z}_n)$ is asymptotically efficient.
- If $\lim_{n\to\infty} nc_n^2v_n^{-2} = \infty$, then
$$v_n\left(\overline{Z}_n - z^*\right) \ \xrightarrow{\mathbb{P}}\ \frac{-(1-2\tau)}{1-2\tau-v^*}\ H^{-1}\rho.$$

4. Proof of the preliminary results on stochastic approximation algorithms. Throughout the proofs, we set $s_n = \sum_{k=1}^n \gamma_k$.

4.1. Proof of Lemma 1. Let us recall that, if $U = (U_n)$ is a sequence of random vectors adapted to the filtration $\mathcal{F}$, then a predictable quadratic variation of U is a random sequence $\langle U\rangle = \langle U\rangle_n$ defined by setting $\langle U\rangle_0 = 0$ and $\langle U\rangle_n - \langle U\rangle_{n-1} = \mathbb{E}[(U_n - U_{n-1})(U_n - U_{n-1})^T|\mathcal{F}_{n-1})$ (see for instance [9, Definition 2.1.8]). Now, set

$$M_j^{(n)} = \sqrt{\gamma_n^{-1}c_n^2}e^{s_nH}\sum_{k=1}^{j} e^{-s_kH}\gamma_k c_k^{-1}\varepsilon_k.$$

$M^{(n)} = (M_j^{(n)})_{1\leq j\leq n}$ is a martingale triangular array whose predictable quadratic variation satisfies

$$\langle M\rangle_n^{(n)} = \gamma_n^{-1}c_n^2 e^{s_nH}\left[\sum_{k=1}^{n} e^{-s_kH}\gamma_k^2 c_k^{-2}\mathbb{E}(\varepsilon_k\varepsilon_k^T|\mathcal{F}_{k-1})e^{-s_kH^T}\right]e^{s_nH^T},$$

and the application of Lemma 4 in [19] ensures that

$$\lim_{n\to\infty}\langle M\rangle_n^{(n)} = \Sigma \ a.s.$$

Moreover, we have

$$\begin{aligned}
&\sum_{k=1}^{n}\mathbb{E}\left[\left\|M_k^{(n)} - M_{k-1}^{(n)}\right\|^{m^*}\middle|\mathcal{F}_{k-1}\right]\\
&= \sum_{k=1}^{n}\mathbb{E}\left[\left\|(\gamma_n^{-1}c_n^2)^{1/2}e^{(s_n-s_k)H}\gamma_k c_k^{-1}\varepsilon_k\right\|^{m^*}\middle|\mathcal{F}_{k-1}\right]\\
&= O\left((\gamma_n^{-1}c_n^2)^{m^*/2}e^{-Lm^*s_n}\sum_{k=1}^{n} e^{Lm^*s_k}\gamma_k^{m^*}c_k^{-m^*}\mathbb{E}\left[\|\varepsilon_k\|^{m^*}\middle|\mathcal{F}_{k-1}\right]\right)\\
&= O\left((\gamma_n^{-1}c_n^2)^{m^*/2}e^{-Lm^*s_n}\sum_{k=1}^{n} e^{Lm^*s_k}\gamma_k^{m^*}c_k^{-m^*}o(\gamma_k^{1-m^*/2})\right) \ a.s.\\
&= O\left((\gamma_n^{-1}c_n^2)^{m^*/2}e^{-Lm^*s_n}\sum_{k=1}^{n} e^{Lm^*s_k}\gamma_k o([\gamma_k c_k^{-2}]^{m^*/2})\right) \ a.s.\\
&= O\left((\gamma_n^{-1}c_n^2)^{m^*/2}o([\gamma_n c_n^{-2}]^{m^*/2})\right) \ a.s.\\
&= o(1) \ a.s.,
\end{aligned}$$

which ensures that Lyapounov's condition is fulfilled. Lemma 1 follows.

4.2. Proof of Lemma 2. Let $-\lambda$ be an eigenvalue of H^T, let w be an eigenvector associated with $-\lambda$, and let M_n be the martingale defined by

$$M_n = \sum_{k=1}^{n} e^{\lambda s_k} \gamma_k c_k^{-1} w^T \varepsilon_k. \tag{17}$$

Let us at first assume that either $\lim_{n\to\infty} n\gamma_n = \infty$ or $\alpha - 2\tau \leq v^*$. The predictable quadratic variation of (M_n) equals

$$\langle M \rangle_n = \sum_{k=1}^{n} e^{2\lambda s_k} \gamma_k^2 c_k^{-2} w^T \mathbb{E}(\varepsilon_k \varepsilon_k^T | \mathcal{F}_{k-1}) w,$$

and the application of Lemma 4 in [19] ensures that

$$\lim_{n\to\infty} \gamma_n^{-1} c_n^2 e^{-2\lambda s_n} \langle M \rangle_n = w^T \Sigma w \quad a.s. \tag{18}$$

Since $(\gamma_n^{-1} c_n^2) \in \mathcal{GS}(\alpha - 2\tau)$, we have

$$\begin{aligned}
\ln(\gamma_n^{-1} c_n^2) &= \ln(\gamma_0^{-1} c_0^2) + \sum_{k=1}^{n} \ln\left(\frac{\gamma_k^{-1} c_k^2}{\gamma_{k-1}^{-1} c_{k-1}^2}\right) \\
&= \ln(\gamma_0^{-1} c_0^2) + \sum_{k=1}^{n} \ln\left(1 + \frac{\alpha - 2\tau}{k} + o\left(\frac{1}{k}\right)\right) \\
&= \ln(\gamma_0^{-1} c_0^2) + \sum_{k=1}^{n} \ln\left(1 + [\alpha - 2\tau]\xi\gamma_k + o(\gamma_k)\right) \\
&= [\alpha - 2\tau]\xi s_n + o(s_n). \qquad (19)
\end{aligned}$$

It follows that

$$\begin{aligned}
\ln\left[\gamma_n^{-1} c_n^2 \exp(-2\lambda s_n)\right] &= [\alpha - 2\tau]\xi s_n + o(s_n) - 2\lambda s_n \\
&= ([1 - 2\tau]\xi - 2\lambda + o(1)) s_n.
\end{aligned}$$

Since $\lim_{n\to\infty} s_n = \infty$ and since $2\lambda > 2L > [1 - 2\tau]\xi$, we deduce that

$$\lim_{n\to\infty} \ln\left[\gamma_n^{-1} c_n^2 \exp(-2\lambda s_n)\right] = -\infty$$

i.e.

$$\lim_{n\to\infty} \gamma_n^{-1} c_n^2 \exp(-2\lambda s_n) = 0,$$

which proves that $\lim_{n\to\infty} \langle M \rangle_n = \infty$ *a.s.*

Let η be the funtion defined by $\eta(x) = \sqrt{2x \ln\ln x}$; we have

$$\begin{aligned}
\frac{|M_n - M_{n-1}|}{\langle M\rangle_n[\eta(\langle M\rangle_n)]^{-1}} &= \frac{e^{\lambda s_n}\gamma_n c_n^{-1}\left|w^T\varepsilon_n\right|\sqrt{2\ln\ln\langle M\rangle_n}}{\sqrt{\langle M\rangle_n}} \\
&\leq C_n \frac{e^{\lambda s_n}\gamma_n c_n^{-1} w_n}{\sqrt{2e^{2\lambda s_n}\gamma_n c_n^{-2} w^T\Sigma w}}\sqrt{\ln\ln(e^{2\lambda s_n}\gamma_n c_n^{-2} w^T\Sigma w)} \\
&\leq C_n'\sqrt{\gamma_n}w_n\sqrt{\ln s_n}
\end{aligned}$$

where (C_n) and (C_n') are a.s. bounded adapted sequences. Thus, there exists an adapted sequence $(\widetilde{C}_n)$ going to zero and such that

$$|M_{n+1} - M_n| \leq \widetilde{C}_n\langle M\rangle_n[\eta(\langle M\rangle_n)]^{-1}$$

The application of Theorem 6.4.24 in [9] then gives

$$\limsup_{n\to\infty}\frac{|M_n|}{\eta(\langle M\rangle_n)} \leq 1 \quad a.s.$$

In view of (18), we thus have

$$\begin{aligned}
|M_n| &= O\left(e^{\lambda s_n}\sqrt{\gamma_n c_n^{-2}\ln\ln(e^{2\lambda s_n}\gamma_n c_n^{-2})}\right) \quad a.s. \\
&= O\left(e^{\lambda s_n}\sqrt{\gamma_n c_n^{-2}\ln s_n}\right) \quad a.s.
\end{aligned}$$

Since $w^T L_{n+1} = e^{-\lambda s_n}M_{n+1}$, it follows that, for any eigenvector w of H^T,

$$\left|w^T L_n\right| = O\left(\sqrt{\gamma_n c_n^{-2}\ln s_n}\right) \quad a.s.$$

Part 1 of Lemma 2 and Part 2 of Lemma 2 in the case when $\lim_{n\to\infty} n\gamma_n = \infty$ follow. It remains to prove Part 2 of Lemma 2 in the case when $\lim_{n\to\infty} n\gamma_n < \infty$. The application of Theorem 1.3.24 in [9] ensures that the martingale (M_n) defined in (17) satisfies, for all $\gamma > 0$,

$$\begin{aligned}
|M_n|^2 &= O\left(\sum_{k=1}^{n} e^{2\lambda s_k}\gamma_k^2 c_k^{-2}\left[\ln\left(\sum_{k=1}^{n} e^{2\lambda s_k}\gamma_k^2 c_k^{-2}\right)\right]^{1+\gamma}\right) \quad a.s. \\
&= O\left(\sum_{k=1}^{n} e^{2\lambda s_k}\gamma_k o\left([v_k^2(\ln k)^{1+\gamma}]^{-1}\right)\left[\ln\left(\sum_{k=1}^{n} e^{2\lambda s_k}\gamma_k v_k^{-2}\right)\right]^{1+\gamma}\right) \quad a.s. \\
&= o\left(e^{2\lambda s_n}[v_n^2(\ln n)^{1+\gamma}]^{-1}\left[\ln\left(e^{2\lambda s_n}v_n^{-2}\right)\right]^{1+\gamma}\right) \quad a.s. \\
&= o\left(e^{2\lambda s_n}[v_n^2(\ln n)^{1+\gamma}]^{-1}[\ln n]^{1+\gamma}\right) \quad a.s. \\
&= o\left(e^{2\lambda s_n}v_n^{-2}\right) \quad a.s.
\end{aligned}$$

It follows that, for any eigenvector w of H^T, $\left|w^T L_n\right| = o\left(v_n^{-1}\right)$ a.s., which concludes the proof of Lemma 2.

4.3. Proof of Lemma 3. Set $r_n = R_n + \tilde{R}_n$ with $\|\tilde{R}_n\| = O(\|Z_n - z^*\|^\eta)$, and note that (15) can be rewritten as

$$Z_n - z^* = Z_{n-1} - z^* + \gamma_n H\left(Z_{n-1} - z^*\right) + \gamma_n r_n + \gamma_n c_n^{-1}\varepsilon_n. \tag{20}$$

Noting that

$$L_n = \gamma_n c_n^{-1}\varepsilon_n + e^{\gamma_n H}L_{n-1} = \gamma_n c_n^{-1}\varepsilon_n + \left[I_d + \gamma_n H + O\left(\gamma_n^2\right)\right]L_{n-1},$$

we get

$$\Delta_n = \left(I_d + \gamma_n H\right)\Delta_{n-1} + \gamma_n\left[O\left(\gamma_n\right)L_{n-1} + r_n\right]. \tag{21}$$

Set $A \in (0, L)$; in view of Proposition 3.I.2 in [8] there exist a matrix norm $\|\!|.|\!\|_A$ and $a \in (0, 1/A)$ such that, for all $\gamma \leq a$, $\|\!|I_d + \gamma H|\!\|_A \leq 1 - \gamma A$. Now, for $x \in \mathbb{R}^d$, define $M(x) = [xx \ldots x]$ the matrix in $M_d(\mathbb{R})$ all of whose columns are x; the function $\|.\|_A$ defined on $\mathbb{R}^d$ by $\|x\|_A = \|\!|M(x)|\!\|_A$ is then a vector norm compatible with the matrix norm $\|\!|.|\!\|_A$ (*see* [13, pp. 297]). For n large enough, we thus obtain

$$\|\Delta_n\|_A \leq \left(1 - A\gamma_n\right)\|\Delta_{n-1}\|_A + \gamma_n\left[O\left(\gamma_n\right)\|L_{n-1}\|_A + \|r_n\|_A\right].$$

Since $\lim_{n\to\infty}\left[O\left(\gamma_n\right)\|L_{n-1}\|_A + \|r_n\|_A\right] = 0$ a.s., the application of Lemma 4.I.1 in [8] ensures that

$$\lim_{n\to\infty}\|\Delta_n\|_A = 0 \quad a.s.$$

Noting that

$$r_n \;=\; R_n + O\left(\|L_{n-1}\|^\eta\right) + O\left(\|\Delta_{n-1}\|^\eta\right) \quad a.s.,$$

we rewrite (21) as

$$\begin{aligned}\Delta_n &= [I_d + \gamma_n H]\,\Delta_{n-1} + \gamma_n\left[O(\gamma_n)\,L_{n-1} + O(\|L_{n-1}\|^\eta) + O(\|\Delta_{n-1}\|^\eta) + R_n\right]\\ &= [I_d + \gamma_n H + o(\gamma_n)]\,\Delta_{n-1} + \gamma_n\left[O(\gamma_n)\,L_{n-1} + O(\|L_{n-1}\|^\eta) + R_n\right].\end{aligned}$$

Now, let (u_n) be the sequence defined as

$$(u_n) = \begin{cases} (v_n) & \text{if } \alpha - 2\tau \geq 2v^*,\\ \left(\sqrt{\gamma_n^{-1}c_n^2}\right) & \text{if } \alpha - 2\tau < 2v^*,\end{cases}$$

and note that $(u_n) \in \mathcal{GS}(u^*)$ with $u^* = \min\{v^*, (\alpha - 2\tau)/2\}$. In particular, we have

$$\frac{u_n}{u_{n-1}} = 1 + u^*\xi\gamma_n + o(\gamma_n).$$

It follows that

$$\begin{aligned}u_n\Delta_n &= \frac{u_n}{u_{n-1}}\left(I_d + \gamma_n H + o(\gamma_n)\right)u_{n-1}\Delta_{n-1}\\ &\quad + \gamma_n u_n\left[O(\gamma_n)\,L_{n-1} + O(\|L_{n-1}\|^\eta) + R_n\right]\\ &= \left(I_d + \gamma_n\left[H + u^*\xi I_d\right] + o(\gamma_n)\right)u_{n-1}\Delta_{n-1}\\ &\quad + \gamma_n u_n\left[O(\gamma_n)\,L_{n-1} + O(\|L_{n-1}\|^\eta) + R_n\right].\end{aligned}$$

Set $\tilde{m} = -[H + u^*\xi I_d]^{-1}\rho\mathbb{1}_{u^*=v^*}$ and $\delta_n = u_n\Delta_n - \tilde{m}$. We have:

$$\begin{aligned}\delta_n &= \left(I_d + \gamma_n\left[H + u^*\xi I_d\right] + o(\gamma_n)\right)\delta_{n-1} + \left(\gamma_n\left[H + u^*\xi I_d\right] + o(\gamma_n)\right)\tilde{m}\\ &\quad + \gamma_n u_n\left[O(\gamma_n)\,L_{n-1} + O(\|L_{n-1}\|^\eta) + R_n\right]\\ &= \left(I_d + \gamma_n\left[H + v^*\xi I_d\right] + o(\gamma_n)\right)\delta_{n-1} + \gamma_n\left[u_n B_n + \widetilde{B}_n\right],\end{aligned}$$

with

$$\begin{aligned}B_n &= O(\gamma_n)\,L_{n-1} + O(\|L_{n-1}\|^\eta),\\ \widetilde{B}_n &= u_n R_n + \left[H + v^*\xi I_d\right]\tilde{m} + o(1).\end{aligned}$$

Set $\widetilde{A} \in (v^*\xi, L)$; there exist a matrix norm $|||.|||_{\widetilde{A}}$ and $\widetilde{a} \in (0, 1/\widetilde{A})$ such that, for all $\gamma \leq \widetilde{a}$, $|||I_d + \gamma\left[H + v^*\xi I_d\right]|||_{\widetilde{A}} \leq 1 - \gamma\widetilde{A}$. Let $\|.\|_{\widetilde{A}}$ be the vector norm compatible with the matrix norm $|||.|||_{\widetilde{A}}$ (for all $x \in \mathbb{R}^d$, $\|x\|_{\widetilde{A}} = |||M(x)|||_{\widetilde{A}}$). For n large enough, we have

$$\|\delta_n\|_{\widetilde{A}} \leq \left(1 - \widetilde{A}\gamma_n + o(\gamma_n)\right)\|\delta_{n-1}\|_{\widetilde{A}} + \gamma_n\left[u_n B_n + \widetilde{B}_n\right]$$

Set $B \in (v^*\xi, \widetilde{A})$; for n large enough, we get

$$\|\delta_n\|_{\widetilde{A}} \leq (1 - B\gamma_n)\,\|\delta_{n-1}\|_{\widetilde{A}} + \gamma_n\left[u_n B_n + \widetilde{B}_n\right] \quad a.s.$$

Since $\lim_{n\to\infty} u_n B_n + \widetilde{B}_n = 0$ a.s., the application of Lemma 4.I.1 in [8] then ensures that $\lim_{n\to\infty} \delta_n = 0$ a.s., which concludes the proof of Lemma 3.

4.4. Proof of Lemma 4. The combination of Lemmas 1, 2 and 3 ensures that, under assumptions (A1)–(A7):

- If $\lim_{n\to\infty} \gamma_n^{-1} c_n^2 v_n^{-2} = c \geq 0$, then $\sqrt{\gamma_n^{-1} c_n^2}(Z_n - z^*) \xrightarrow{\mathcal{D}} \mathcal{N}(-\sqrt{c}[H + v^* \xi I_d]^{-1}\rho, \Sigma)$.

- If $\lim_{n\to\infty} \gamma_n^{-1} c_n^2 v_n^{-2} = \infty$, then $v_n(Z_n - z^*) \xrightarrow{\mathbb{P}} -[H + v^* \xi I_d]^{-1}\rho$.

It follows that, for (Z_n) to converge at the optimal rate, (γ_n) must be chosen such that $\lim\limits_{n\to\infty} n\gamma_n = \gamma_0 \in \left(\min\left\{\frac{1-2\tau}{2L}, \frac{v^*}{L}\right\}, \infty\right)$ and the sequences (c_n) and (v_n) must be such that $\lim_{n\to\infty} n c_n^2 v_n^{-2} = c > 0$ (in which case $v^* = (1-2\tau)/2$). We then have

$$\sqrt{n c_n^2}(Z_n - z^*) \xrightarrow{\mathcal{D}} \mathcal{N}\left(-\sqrt{c\gamma_0}\left[H + \frac{1-2\tau}{2}\xi I_d\right]^{-1}\rho, \gamma_0 \Sigma\right).$$

For (Z_n) to be asymptotically efficient, the asymptotic covariance matrix $\gamma_0\Sigma$ must also be minimum. To find this minimum covariance matrix, we allow the stepsize (γ_n) to be matricial. In other words, we consider the stochastic approximation algorithm defined as

$$Y_n = Y_{n-1} + \frac{A}{n}\left[h(Y_{n-1}) + R_n + c_n^{-1}\varepsilon_n\right],$$

where A is a $d \times d$ nonsingular matrix. Following the proof of Lemmas 1–3 (by replacing γ_n, H, R_n, and ε_n by n^{-1}, AH, AR_n, and $A\varepsilon_n$, respectively) we show that, under assumptions (A1)–(A7), if the matrix $AH + [(1-2\tau)/2]I_d$ is negative definite, and if $\lim_{n\to\infty} n c_n^2 v_n^{-2} = c > 0$, then (Y_n) satisfies the central limit theorem

$$\sqrt{n c_n^2}(Y_n - z^*) \xrightarrow{\mathcal{D}} \mathcal{N}\left(-\sqrt{c}\left[AH + \frac{1-2\tau}{2} I_d\right]^{-1} A\rho, \Sigma(A)\right), \tag{22}$$

where $\Sigma(A)$ is the solution of Lyapounov's equation

$$\left(AH + \frac{1-2\tau}{2} I_d\right)\Sigma(A) + \Sigma(A)\left(H^T A^T + \frac{1-2\tau}{2} I_d\right) = -A\Gamma A^T. \tag{23}$$

Now, set $\Delta(A) = \Sigma(A) - (1-2\tau)H^{-1}\Gamma[H^{-1}]^T$; in view of (23), we have

$$\begin{aligned}
&\left(AH + \frac{1-2\tau}{2}I_d\right)\Delta(A) + \Delta(A)\left(H^TA^T + \frac{1-2\tau}{2}I_d\right)\\
&= -A\Gamma A^T - \left(AH + \frac{1-2\tau}{2}I_d\right)([1-2\tau]H^{-1}\Gamma[H^{-1}]^T)\\
&\quad -([1-2\tau]H^{-1}\Gamma[H^{-1}]^T)\left(H^TA^T + \frac{1-2\tau}{2}I_d\right)\\
&= -A\Gamma A^T - (1-2\tau)A\Gamma[H^{-1}]^T - (1-2\tau)^2H^{-1}\Gamma[H^{-1}]^T - (1-2\tau)H^{-1}\Gamma A^T\\
&= -[A + (1-2\tau)H^{-1}]\Gamma[A + (1-2\tau)H^{-1}]^T.
\end{aligned}$$

It follows that the matrix $\Delta(A)$ is nonnegative. Moreover, $\Delta(A) = 0$ if and only if $A = -(1-2\tau)H^{-1}$. We thus deduce that the matrix $\Sigma(-(1-2\tau)H^{-1}) = (1-2\tau)H^{-1}\Gamma[H^{-1}]^T$ is minimum. Now, if $A = -(1-2\tau)H^{-1}$, then (22) becomes

$$\sqrt{nc_n^2}(Y_n - z^*) \xrightarrow{\mathcal{D}} \mathcal{N}\left(-2\sqrt{c}H^{-1}\rho, (1-2\tau)H^{-1}\Gamma[H^{-1}]^T\right).$$

Lemma 4 thus follows.

4.5. Proof of Lemma 5. Set $M_{n+1} = \sum_{k=1}^n c_k^2 c_{k+1}^{-1}\varepsilon_{k+1}$; (M_n) is a martingale whose predictable quadratic variation satisfies

$$\langle M\rangle_{n+1} = \sum_{k=1}^n c_k^4 c_{k+1}^{-2}\mathbb{E}(\varepsilon_{k+1}\varepsilon_{k+1}^T|\mathcal{F}_k) = \sum_{k=1}^n c_k^2\Gamma(1+o(1)) \quad a.s.$$

Since $(c_n^2) \in \mathcal{GS}(-2\tau)$ with $1-2\tau > 0$, we have

$$\lim_{n\to\infty}\frac{nc_n^2}{\sum_{k=1}^n c_k^2} = 1-2\tau, \tag{24}$$

and thus

$$\lim_{n\to\infty}[nc_n^2]^{-1}\langle M\rangle_{n+1} = (1-2\tau)^{-1}\Gamma \ a.s.$$

Moreover, we have

$$\begin{aligned}
&[nc_n^2]^{-m^*/2}\sum_{k=1}^n \mathbb{E}\left[\|M_{k+1} - M_k\|^{m^*}\middle|\mathcal{F}_k\right]\\
&\qquad = O\left([nc_n^2]^{-m^*/2}\sum_{k=1}^n c_k^{m^*}\mathbb{E}\left[\|\varepsilon_{k+1}\|^{m^*}\middle|\mathcal{F}_k\right]\right)
\end{aligned}$$

$$= O\left([nc_n^2]^{-m^*/2}\sum_{k=1}^n c_k^{m^*} o\left(k^{m^*/2-1}\right)\right) \ a.s.$$
$$= O\left([nc_n^2]^{-m^*/2}\sum_{k=1}^n k^{-1} o\left([kc_k^2]^{m^*/2}\right)\right) \ a.s.$$
$$= o(1) \ a.s.,$$

which ensures that Lyapounov's condition is fulfilled. It follows that

$$[nc_n^2]^{-1/2} M_{n+1} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0\ ,\ (1-2\tau)^{-1}\Gamma\right).$$

Noting that $\Lambda_{n+1} = -[\sum_{k=1}^n c_k^2]^{-1} M_{n+1}$, and applying (24) again, we obtain

$$\sqrt{nc_n^2}\Lambda_{n+1} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0\ ,\ (1-2\tau)H^{-1}\Gamma[H^{-1}]^T\right).$$

4.6. Proof of Lemma 6. In view of (20), we have

$$Z_{n-1} - z^* = \gamma_n^{-1} H^{-1}\left[(Z_n - z^*) - (Z_{n+1} - z^*)\right] - H^{-1} r_n - c_n^{-1} H^{-1}\varepsilon_n,$$

and thus

$$\begin{aligned} \overline{Z}_n - z^* &= \frac{1}{\sum_{k=1}^n c_k^2} H^{-1}\sum_{k=1}^n c_k^2 \gamma_{k+1}^{-1}\left[(Z_{k+1} - z^*) - (Z_k - z^*)\right] \\ &\quad - \frac{1}{\sum_{k=1}^n c_k^2} H^{-1}\sum_{k=1}^n c_k^2 r_{k+1} - \frac{1}{\sum_{k=1}^n c_k^2} H^{-1}\sum_{k=1}^n c_k^2 c_{k+1}^{-1}\varepsilon_{k+1}. \end{aligned}$$

It follows that

$$\Xi_{n+1} = -H^{-1}\left[\mathcal{R}_{n+1}^{(1)} + \mathcal{R}_{n+1}^{(2)} + \mathcal{R}_{n+1}^{(3)}\right] \tag{25}$$

with

$$\begin{aligned} \mathcal{R}_{n+1}^{(1)} &= \frac{1}{\sum_{k=1}^n c_k^2}\sum_{k=1}^n c_k^2 R_{k+1}, \\ \mathcal{R}_{n+1}^{(2)} &= \frac{1}{\sum_{k=1}^n c_k^2}\sum_{k=1}^n c_k^2 \gamma_{k+1}^{-1}\left[(Z_k - z^*) - (Z_{k+1} - z^*)\right], \\ \mathcal{R}_{n+1}^{(3)} &= \frac{1}{\sum_{k=1}^n c_k^2}\sum_{k=1}^n c_k^2 O(\|Z_k - z^*\|^\eta). \end{aligned}$$

We now successively establish the almost sure asymptotic behaviour of $\mathcal{R}_n^{(i)}$, $i \in \{1, 2, 3\}$.

• In view of assumption (A5), we have:

$$\mathcal{R}_{n+1}^{(1)} = \frac{1}{\sum_{k=1}^{n} c_k^2} \sum_{k=1}^{n} c_k^2 v_k^{-1} \rho[1+o(1)] \quad a.s.$$

In the case $\lim_{n\to\infty} v_n^{-1}\sqrt{nc_n^2} \in (0,\infty]$, we have $1/2 - \tau - v^* \geq 0$; hence $1-2\tau - v^* > 0$, and thus

$$\lim_{n\to\infty} \frac{nc_n^2 v_n^{-1}}{\sum_{k=1}^{n} c_k^2 v_k^{-1}} = 1 - 2\tau - v^*.$$

Applying (24), we deduce that

$$\lim_{n\to\infty} v_n \mathcal{R}_{n+1}^{(1)} = \frac{1-2\tau}{1-2\tau-v^*}\rho \quad a.s. \tag{26}$$

In the case $\lim_{n\to\infty} v_n^{-1}\sqrt{nc_n^2} = 0$, we have $v_n^{-1} = o([nc_n^2]^{-1/2})$, and thus

$$\begin{aligned} \lim_{n\to\infty} \sqrt{nc_n^2}\mathcal{R}_{n+1}^{(1)} &= \lim_{n\to\infty} \frac{\sqrt{nc_n^2}}{\sum_{k=1}^{n} c_k^2} \sum_{k=1}^{n} o\left(k^{-1/2}c_k\right) \quad a.s. \\ &= 0 \quad a.s. \end{aligned} \tag{27}$$

• Since $(c_n^2\gamma_{n+1}^{-1}) \in \mathcal{GS}(\alpha - 2\tau)$, we have

$$\begin{aligned} \mathcal{R}_{n+1}^{(2)} &= \frac{1}{\sum_{k=1}^{n} c_k^2}\Bigg[c_1^2\gamma_2^{-1}(Z_1 - z^*) - c_n^2\gamma_{n+1}^{-1}(Z_{n+1} - z^*) \\ &\qquad + \sum_{k=2}^{n}[c_k^2\gamma_{k+1}^{-1} - c_{k-1}^2\gamma_k^{-1}](Z_k - z^*)\Bigg] \\ &= \frac{1}{\sum_{k=1}^{n} c_k^2}\Bigg[c_1^2\gamma_2^{-1}(Z_1 - z^*) - c_n^2\gamma_{n+1}^{-1}(Z_{n+1} - z^*) \\ &\qquad + \sum_{k=2}^{n} c_k^2\gamma_{k+1}^{-1}\left[1 - \frac{c_{k-1}^2\gamma_k^{-1}}{c_k^2\gamma_{k+1}^{-1}}\right](Z_k - z^*)\Bigg] \\ &= \frac{1}{\sum_{k=1}^{n} c_k^2}\Bigg[c_1^2\gamma_2^{-1}(Z_1 - z^*) - c_n^2\gamma_{n+1}^{-1}(Z_{n+1} - z^*) \\ &\qquad + \sum_{k=2}^{n} c_k^2\gamma_{k+1}^{-1}O(k^{-1})(Z_k - z^*)\Bigg] \end{aligned}$$

Now, let (m_n) be the sequence defined as

$$(m_n) = \begin{cases} \left(\sqrt{n^{-1}c_n^{-2}}\right) & \text{if } \lim_{n\to\infty} v_n\sqrt{n^{-1}c_n^{-2}} = \infty, \\ (v_n^{-1}) & \text{otherwise,} \end{cases}$$

and note that $(m_n) \in \mathcal{GS}(-m^*)$ with $m^* = \min\{v^*, (1-2\tau)/2\}$. Applying (16) and (24), we obtain

$$
\begin{aligned}
\left\|\mathcal{R}_{n+1}^{(2)}\right\| &= O\left(\frac{1}{nc_n^2} + \frac{c_n^2\gamma_{n+1}^{-1}\left[\sqrt{\gamma_n c_n^{-2}\log(s_n)} + v_n^{-1}\right]}{nc_n^2}\right. \\
&\qquad \left. + \frac{\sum_{k=2}^n c_k^2\gamma_{k+1}^{-1}k^{-1}\left[\sqrt{\gamma_k c_k^{-2}\log(s_k)} + v_k^{-1}\right]}{nc_n^2}\right) \; a.s. \\
&= O\left(\frac{1}{nc_n^2} + \frac{c_n^2\gamma_{n+1}^{-1}\left[\sqrt{\gamma_n c_n^{-2}\log(s_n)} + v_n^{-1}\right]}{nc_n^2}\right. \\
&\qquad \left. + \frac{\sum_{k=2}^n c_k^2\gamma_{k+1}^{-1}k^{-1}\left[\sqrt{\gamma_k c_k^{-2}\log(s_k)} + v_k^{-1}\right]}{nc_n^2}\right) \; a.s. \\
&= O\left(\frac{1}{nc_n^2} + \frac{\sqrt{n^{-1}\gamma_n^{-1}\log(s_n)}}{\sqrt{nc_n^2}} + \frac{v_n^{-1}}{n\gamma_n}\right. \\
&\qquad \left. + \frac{1}{nc_n^2}\sum_{k=2}^n [c_k k^{-1/2}\sqrt{k^{-1}\gamma_k^{-1}\log(s_k)} + c_k^2\gamma_k^{-1}k^{-1}v_k^{-1}]\right) \; a.s. \\
&= O\left(\frac{1}{nc_n^2} + \frac{\sqrt{n^{-1}\gamma_n^{-1}\log(s_n)}}{\sqrt{nc_n^2}} + \frac{v_n^{-1}}{n\gamma_n}\right. \\
&\qquad \left. + \frac{1}{nc_n^2}\sum_{k=2}^n c_k^2[o(c_k^{-1}k^{-1/2}) + o(v_k^{-1})]\right) \; a.s. \\
&= o\left(\frac{1}{\sqrt{nc_n^2}}\right) + o\left(v_n^{-1}\right) + O\left(\frac{1}{nc_n^2}\sum_{k=2}^n c_k^2 o(m_k)\right) \; a.s. \\
&= o(m_n) \; a.s. \qquad (28)
\end{aligned}
$$

• In view of (16), (24), and (A8), we have

$$\begin{aligned}\left\|\mathcal{R}_{n+1}^{(3)}\right\| &= O\left(\frac{1}{nc_n^2}\sum_{k=1}^{n} c_k^2\left[(\gamma_k c_k^{-2}\log s_k)^{\eta/2}+v_k^{-\eta}\right]\right) \quad a.s.\\ &= O\left(\frac{1}{nc_n^2}\sum_{k=1}^{n} c_k^2\left[o([k^{-1}c_k^{-2}]^{1/2})+o(v_k^{-1})\right]\right) \quad a.s.\\ &= O\left(\frac{1}{nc_n^2}\sum_{k=1}^{n} c_k^2 o(m_k)\right) \quad a.s.\\ &= o(m_n) \quad a.s. \qquad (29)\end{aligned}$$

Part 1 (respectively, Part 2) of Lemma 6 is then a straightforward consequence of the combination of (25), (26), (28) and (29) (respectively, of (25), (27), (28) and (29)).

5. Proof of the main results.

From now on, $\mathcal{F} = (\mathcal{F}_n)$ denotes the σ-field spanned by $(X_1, \ldots, X_n)$.

5.1. Proof of Proposition 1.

We first establish an upper bound of the strong convergence rate of θ_n, and then prove the consistency of μ_n.

5.1.1. Upper bound of the strong convergence rate of $\boldsymbol{\theta_n}$.

To prove an upper bound of the strong convergence rate of θ_n, we apply Lemmas 2 and 3. To this end, we first rewrite (1) as

$$\theta_n = \theta_{n-1} + \gamma_n\left[\nabla f(\theta_{n-1}) + R_n^{(\theta)} + h_n^{-(d+2)/2}\varepsilon_n^{(\theta)}\right] \qquad (30)$$

with

$$\begin{aligned} R_n^{(\theta)} &= \frac{1}{h_n^{d+1}}\mathbb{E}\left[\nabla K\left(\frac{\theta_{n-1}-X_n}{h_n}\right)\middle|\mathcal{F}_{n-1}\right] - \nabla f(\theta_{n-1}),\\ \varepsilon_n^{(\theta)} &= \frac{1}{\sqrt{h_n^d}}\left\{\nabla K\left(\frac{\theta_{n-1}-X_n}{h_n}\right) - \mathbb{E}\left[\nabla K\left(\frac{\theta_{n-1}-X_n}{h_n}\right)\middle|\mathcal{F}_{n-1}\right]\right\}.\end{aligned}$$

Moreover, we note that, under (H1) and (H2), we have

$$\begin{aligned} R_n^{(\theta)} &= \frac{1}{h_n^{d+1}}\int_{\mathbb{R}^d}\nabla K\left(\frac{\theta_{n-1}-x}{h_n}\right)f(x)dx - \nabla f(\theta_{n-1})\\ &= \frac{1}{h_n}\int_{\mathbb{R}^d}\nabla K(z)f(\theta_{n-1}-h_n z)dz - \nabla f(\theta_{n-1})\end{aligned}$$

$$= \int_{\mathbb{R}^d} K(z)\nabla f(\theta_{n-1} - h_n z)dz - \nabla f(\theta_{n-1})$$

$$= \frac{h_n^2}{2}\left(\nabla \sum_{i=1}^{d}\left[\int_{\mathbb{R}^d} x_i^2 K(x)dx\right]\frac{\partial^2}{\partial^2 z_i}\right) f(\theta_{n-1}) + o(h_n^2), \tag{31}$$

and

$$\mathbb{E}\left[\nabla K\left(\frac{\theta_{n-1} - X_n}{h_n}\right)\left[\nabla K\left(\frac{\theta_{n-1} - X_n}{h_n}\right)\right]^T \middle| \mathcal{F}_{n-1}\right]$$

$$= \int_{\mathbb{R}^d} \nabla K\left(\frac{\theta_{n-1} - x}{h_n}\right)\left[\nabla K\left(\frac{\theta_{n-1} - x}{h_n}\right)\right]^T f(x)dx$$

$$= h_n^d \int_{\mathbb{R}^d} \nabla K(z)[\nabla K(z)]^T f(\theta_{n-1} - h_n z)dz$$

$$= h_n^d\left(f(\theta_{n-1})\left[\int_{\mathbb{R}^d} \nabla K(z)[\nabla K(z)]^T dz\right] + o(1)\right). \tag{32}$$

We now check that assumptions (A1)–(A7) required in Lemmas 2 and 3 are fufilled by the stochastic approximation algorithm (30). To this end, we need the following Robbins-Monro's Theorem (see for instance [8, page 61]).

Theorem 3 (Robbins-Monro)**.** *Let* (Z_n) *be defined by the stochastic approximation algorithm* (15)*, and assume that*

- $c_n = O(1)$*,* $\gamma_n = o(c_n^2)$*,* $\sum \gamma_n = \infty$ *and* $\sum \gamma_n^2 c_n^{-2} < \infty$*;*
- *There exists a continuously differentiable function* $V : \mathbb{R}^d \to \mathbb{R}^+$*, such that* ∇V *is Lipschitz-continuous, and such that, for all* $x \in \mathbb{R}^d$*,* $\|h(x)\|^2 \leq cte(1 + V(x))$ *and* $[\nabla V(x)]^T h(x) \leq 0$*;*
- $\mathbb{E}(\varepsilon_n|\mathcal{F}_{n-1}) = 0$*,* $\mathbb{E}(\|\varepsilon_n\|^2|\mathcal{F}_{n-1}) = O(1 + V(Z_{n-1}))$ *and* $\sum \gamma_n \|R_n\|^2 < \infty$ *a.s.*

Then, the sequence $V(Z_n)$ *converges a.s. and* $\sum \gamma_n |[\nabla V(Z_n)]^T h(Z_n)| < \infty$ *a.s.*

To apply Theorem 3, Lemma 2, and Lemma 3, we set $V : z \mapsto \|z - \theta\|^2$, $h = \nabla f$, $(c_n) = (h_n^{(d+2)/2})$, $\varepsilon_n = \varepsilon_n^{(\theta)}$, and $R_n = R_n^{(\theta)}$.

- Let us first note that it follows from (32) that $\sup_n \mathbb{E}(\|\varepsilon_n^{(\theta)}\|^2|\mathcal{F}_{n-1}) < \infty$. Moreover, (31) implies that $\|R_n^{(\theta)}\| = O(h_n^2)$, so that, in view of (H3), we have $\sum\gamma_n\|R_n^{(\theta)}\|^2 < \infty$. In view of (H2) and (H3), Theorem 3 applies. Thus, the sequence $(\|\theta_n-\theta\|)$ converges a.s., and $\sum\gamma_n|(\theta_n-\theta)^T\nabla f(\theta_n)| < \infty$ a.s. Assumption (H2) together with the condition $\sum\gamma_n=\infty$ (which is implied by (H3)(iii)) imply that $\lim_{n\to\infty}\theta_n=\theta$ a.s. Assumption (A1) is thus fulfilled and the claim in Remark 3 is proved.

- (H2) ensures that (A2) holds with $H=D^2f(\theta)$ and $L=L^{(\theta)}$.
- In view of (H3), (A3) holds with $\tau=a(d+2)/2$.
- It follows from (32) and (A1) that (A4) is satisfied with $\Gamma=f(\theta)\int_{\mathbb{R}^d}\nabla K(z)[\nabla K(z)]^T dz.$
- Note that $R_n^{(\theta)}=\int_{\mathbb{R}^d}K(z)\left[\nabla f(\theta_{n-1}-h_nz)-\nabla f(\theta_{n-1})\right]dz$. A two-order Taylor's development and the application of Lebesgues' Theorem give
$$\lim_{n\to\infty}h_n^{-2}R_n^{(\theta)}=\frac{1}{2}\nabla\left(\sum_{i=1}^d\sum_{k=1}^d\frac{\partial^2 f}{\partial z_i\partial z_k}(z)\int_{\mathbb{R}^d}z_iz_kK(z)dz\right)(\theta).$$
In view of (H1)(ii), (A5) thus holds with $(v_n)=(h_n^{-2})$ (and thus $v^*=2a$) and $\rho=R^{(\theta)}$ ($R^{(\theta)}$ being defined in (6)).
- (A6) follows from (H3) (see Remark 2).
- (H1)(iii) and (H3) ensure that (A7) holds with $w_n=2\|\nabla K\|_\infty h_n^{-d/2}$.

Set

$$L_n^{(\theta)} = e^{s_n[D^2f(\theta)]}\sum_{k=1}^n e^{-s_k[D^2f(\theta)]}\gamma_k h_k^{-(d+2)/2}\varepsilon_k^{(\theta)}, \tag{33}$$

$$\Delta_n^{(\theta)} = (\theta_n-\theta)-L_n^{(\theta)}. \tag{34}$$

The combination of Lemmas 2 and 3 ensures that

$$\|\theta_n-\theta\| = O\left(\max\left\{h_n^2\ ;\ \sqrt{\gamma_n h_n^{-(d+2)}\ln s_n}\right\}\right)\quad a.s. \tag{35}$$

5.1.2. Proof of Proposition 1. To prove Proposition 1, we apply Theorem 3. To this end, we rewrite (2) as

$$\mu_n = \mu_{n-1}+\gamma_n\left[h(\mu_{n-1})+R_n^{(\mu)}+\widetilde{h}_n^{-d/2}\varepsilon_n^{(\mu)}\right] \tag{36}$$

with

$$\begin{aligned}
h(z) &= f(\theta) - z, \\
R_n^{(\mu)} &= f(\theta_{n-1}) - f(\theta) + B_n^{(\mu)}, \\
B_n^{(\mu)} &= \frac{1}{\widetilde{h}_n^d}\mathbb{E}\left[K\left(\frac{\theta_{n-1}-X_n}{\widetilde{h}_n}\right)\middle|\mathcal{F}_{n-1}\right] - f(\theta_{n-1}), \\
\varepsilon_n^{(\mu)} &= \frac{1}{\sqrt{\widetilde{h}_n^d}}\left\{K\left(\frac{\theta_{n-1}-X_n}{\widetilde{h}_n}\right) - \mathbb{E}\left[K\left(\frac{\theta_{n-1}-X_n}{\widetilde{h}_n}\right)\middle|\mathcal{F}_{n-1}\right]\right\}.
\end{aligned}$$

To apply Theorem 3, we set $V : z \mapsto (z-\mu)^2$, $(c_n) = (\widetilde{h}_n^{d/2})$, $\varepsilon_n = \varepsilon_n^{(\mu)}$, and $R_n = R_n^{(\mu)}$. We first note that

$$\begin{aligned}
\mathbb{E}\left[K^2\left(\frac{\theta_{n-1}-X_n}{\widetilde{h}_n}\right)\middle|\mathcal{F}_{n-1}\right] &= \int_{\mathbb{R}^d} K^2\left(\frac{\theta_{n-1}-x}{\widetilde{h}_n}\right) f(x)dx \\
&= \widetilde{h}_n^d \int_{\mathbb{R}^d} K^2(z) f(\theta_{n-1} - \widetilde{h}_n z)dz \\
&= \widetilde{h}_n^d \left(f(\theta_{n-1}) \int_{\mathbb{R}^d} K^2(z)dz + o(1)\right), \qquad (37)
\end{aligned}$$

which implies that $\sup_n \mathbb{E}(|\varepsilon_n^{(\mu)}|^2|\mathcal{F}_{n-1}) < \infty$ a.s. Now, we have

$$\begin{aligned}
B_n^{(\mu)} &= \frac{1}{\widetilde{h}_n^d}\int_{\mathbb{R}^d} K\left(\frac{\theta_{n-1}-x}{\widetilde{h}_n}\right) f(x)dx - f(\theta_{n-1}) \\
&= \int_{\mathbb{R}^d} K(z) f(\theta_{n-1} - \widetilde{h}_n z)dz - f(\theta_{n-1}) \\
&= \frac{\widetilde{h}_n^2}{2}\sum_{i=1}^d \left(\left[\int_{\mathbb{R}^d} x_i^2 K(x)dx\right]\frac{\partial^2 f}{\partial^2 z_i}(\theta)\right) + o(\widetilde{h}_n^2). \qquad (38)
\end{aligned}$$

Moreover, the application of (35) ensures that

$$\begin{aligned}
|f(\theta_n) - f(\theta)| &= O(\|\theta_n - \theta\|^2) \quad a.s. \\
&= O\left(\max\left\{\ h_n^4\ ;\ \gamma_n h_n^{-(d+2)} \ln s_n\right\}\right) \quad a.s. \qquad (39)
\end{aligned}$$

In view of (H3), we deduce from the combination of (35) and (39) that $\sum \gamma_n [R_n^{(\mu)}]^2 < \infty$ a.s. The application of Theorem 3 then ensures that

$\sum \gamma_n |(\mu_n - \mu)[f(\theta_n) - f(\theta)]| < \infty$ a.s., that is, $\sum \gamma_n(\mu_n - \mu)^2 < \infty$ a.s. Proposition 1 follows from the fact that $\sum \gamma_n = \infty$.

5.2. Proof of Theorem 1. We have seen in Section 5.1.1 that Lemmas 2 and 3 can be applied to the stochastic approximation algorithm (30) with $h = \nabla f$, $H = D^2 f(\theta)$, $(c_n) = (h_n^{(d+2)/2})$, $\tau = a(d+2)/2$, $\varepsilon_n = \varepsilon_n^{(\theta)}$, $\Gamma = f(\theta) \int_{\mathbb{R}^d} \nabla K(z)[\nabla K(z)]^T dz$, $R_n = R_n^{(\theta)}$, $(v_n) = (h_n^{-2})$, $v^* = 2a$, and $\rho = R^{(\theta)}$ ($R^{(\theta)}$ being defined in (6)). Let $(L_n^{(\theta)})$ and $(\Delta_n^{(\theta)})$ by defined by (33) and (34), respectively. The following properties thus hold.

(40) • If $\alpha > a(d+6)$, then $\left\| L_n^{(\theta)} \right\| = o\left(h_n^2\right)$ *a.s.*

(41) • If $\alpha \geq a(d+6)$, then $\lim_{n\to\infty} h_n^{-2} \Delta_n^{(\theta)} = -\left[D^2 f(\theta) + 2a\xi I_d\right]^{-1} R^{(\theta)}$ *a.s.*

(42) • If $\alpha < a(d+6)$, then $\lim_{n\to\infty} \sqrt{\gamma_n^{-1} h_n^{d+2}} \Delta_n^{(\theta)} = 0$ *a.s.*

Now, set

$$\begin{aligned} L_n^{(\mu)} &= e^{-s_n} \sum_{k=1}^{n} e^{s_k} \gamma_k \tilde{h}_k^{-d/2} \varepsilon_k^{(\mu)}, \\ \Delta_n^{(\mu)} &= (\mu_n - \mu) - L_n^{(\mu)}. \end{aligned}$$

We apply Lemmas 2 and 3 to the stochastic approximation algorithm (36). To this end, we set $(c_n) = (\widetilde{h}_n^{d/2})$, $\varepsilon_n = \varepsilon_n^{(\mu)}$, and $R_n = R_n^{(\mu)}$, and check that assumptions (A1)–(A7) required in Lemmas 2 and 3 are fulfilled.

- (A1) follows from the application of Proposition 1.
- (A2) clearly holds with $H = -1$ and thus $L = 1$.
- In view of (H3), (A3) holds with $\tau = \tilde{a}d/2$.
- Since $\lim_{n\to\infty} \theta_n = \theta$ a.s. (see Remark 3), (37) ensures that (A4) holds with $\Gamma = f(\theta) \int_{\mathbb{R}^d} K^2(z) dz$.
- The combination of (39) and (H4)(ii) ensures that $|f(\theta_n) - f(\theta)| = o(\tilde{h}_n^2)$ a.s. It then follows from (38) that (A5) holds with $(v_n) = (\tilde{h}_n^{-2})$ (and thus $v^* = 2\tilde{a}$) and $\rho = R^{(\mu)}$ ($R^{(\mu)}$ being defined in (7)).
- (A6) follows from (H3) and (H4) (see Remark 2).

- (H1)(iii) and (H3) ensure that (A7) holds with $w_n = 2\|K\|_\infty \widetilde{h}_n^{-d/2}$.

The application of Lemmas 2 and 3 then ensures that the following properties hold.

(43) • If $\alpha > \widetilde{a}(d+4)$, then $\left\| L_n^{(\mu)} \right\| = o\left(\widetilde{h}_n^2\right)$ *a.s.*

(44) • If $\alpha \geq \widetilde{a}(d+4)$, then $\lim_{n\to\infty} \widetilde{h}_n^{-2} \Delta_n^{(\mu)} = [1 - 2\tilde{a}\xi]^{-1} R^{(\mu)}$ *a.s.*

(45) • If $\alpha < \widetilde{a}(d+4)$, then $\lim_{n\to\infty} \sqrt{\gamma_n^{-1} \widetilde{h}_n^d} \Delta_n^{(\mu)} = 0$ *a.s.*

Theorem 1 follows the combination of Properties (40)–(45) together with the following lemma.

Lemma 7. *Let the assumptions of Theorem* 1 *hold. If* $\alpha \leq \min\{a(d+6), \widetilde{a}(d+4)\}$, *then we have*

$$\begin{pmatrix} \sqrt{\gamma_n^{-1} h_n^{d+2}}\ L_n^{(\theta)} \\ \sqrt{\gamma_n^{-1} \widetilde{h}_n^d}\ L_n^{(\mu)} \end{pmatrix} \stackrel{\mathcal{D}}{\rightarrow} \mathcal{N}\left(0\ ,\ \begin{pmatrix} \Sigma^{(\theta)} & 0 \\ 0 & \Sigma^{(\mu)} \end{pmatrix}\right).$$

We now prove Lemma 7. Set $H = D^2 f(\theta)$, $\sigma_k^{(\theta)} = \gamma_k h_k^{-(d+2)/2}$, $\sigma_k^{(\mu)} = \gamma_k \tilde{h}_k^{-d/2}$, and

$$M_j^{(n)} = \begin{pmatrix} \sqrt{\gamma_n^{-1} h_n^{d+2}} e^{s_n H} & 0 \\ 0 & \sqrt{\gamma_n^{-1} \tilde{h}_n^d} e^{-s_n} \end{pmatrix} \sum_{k=1}^{j} \begin{pmatrix} e^{-s_k H} \sigma_k^{(\theta)} \varepsilon_k^{(\theta)} \\ e^{s_k} \sigma_k^{(\mu)} \varepsilon_k^{(\mu)} \end{pmatrix}.$$

For a given n, $M^{(n)} = (M_j^{(n)})_{1\leq j\leq n}$ is a martingale whose predictable quadratic variation satisfies

$$\langle M \rangle_n^{(n)} = \begin{pmatrix} \sqrt{\gamma_n^{-1} h_n^{d+2}} e^{s_n H} & 0 \\ 0 & \sqrt{\gamma_n^{-1} \tilde{h}_n^d} e^{-s_n} \end{pmatrix} \times C_n \begin{pmatrix} \sqrt{\gamma_n^{-1} h_n^{d+2}} e^{s_n H^T} & 0 \\ 0 & \sqrt{\gamma_n^{-1} \tilde{h}_n^d} e^{-s_n} \end{pmatrix}$$

with

$$C_n = \sum_{k=1}^{n} \mathbb{E}\left[\left(\begin{array}{c} e^{-s_k H} \sigma_k^{(\theta)} \varepsilon_k^{(\theta)} \\ e^{s_k} \sigma_k^{(\mu)} \varepsilon_k^{(\mu)} \end{array} \right) \left(\begin{array}{cc} \sigma_k^{(\theta)} [\varepsilon_k^{(\theta)}]^T e^{-s_k H^T} & e^{s_k} \sigma_k^{(\mu)} \varepsilon_k^{(\mu)} \end{array} \right) \middle| \mathcal{F}_{k-1} \right]$$

$$=\sum_{k=1}^{n}\begin{pmatrix}[\sigma_k^{(\theta)}]^2e^{-s_kH}\mathbb{E}\left[\varepsilon_k^{(\theta)}[\varepsilon_k^{(\theta)}]^T\middle|\mathcal{F}_{k-1}\right]e^{-s_kH^T} & \sigma_k^{(\theta)}\sigma_k^{(\mu)}e^{s_k}e^{-s_kH}\mathbb{E}\left[\varepsilon_k^{(\theta)}\varepsilon_k^{(\mu)}\middle|\mathcal{F}_{k-1}\right]\\ \sigma_k^{(\theta)}\sigma_k^{(\mu)}\mathbb{E}\left[[\varepsilon_k^{(\theta)}]^T\varepsilon_k^{(\mu)}\middle|\mathcal{F}_{k-1}\right]e^{s_k}e^{-s_kH^T} & [\sigma_k^{(\mu)}]^2e^{2s_k}\mathbb{E}\left[[\varepsilon_k^{(\mu)}]^2\middle|\mathcal{F}_{k-1}\right]\end{pmatrix}.$$

It follows that

$$\langle M\rangle_n^{(n)} = \begin{pmatrix} A_{1,n} & A_{2,n}\\ A_{3,n} & A_{4,n}\end{pmatrix}$$

with

$$\begin{aligned}
A_{1,n} &= \gamma_n^{-1}h_n^{d+2}e^{s_nH}\left\{\sum_{k=1}^{n}[\sigma_k^{(\theta)}]^2e^{-s_kH}\mathbb{E}\left[\varepsilon_k^{(\theta)}[\varepsilon_k^{(\theta)}]^T\middle|\mathcal{F}_{k-1}\right]e^{-s_kH^T}\right\}e^{s_nH^T},\\
A_{2,n} &= \gamma_n^{-1}\sqrt{h_n^{d+2}\tilde{h}_n^d}e^{s_nH}e^{-s_n}\left\{\sum_{k=1}^{n}\sigma_k^{(\theta)}\sigma_k^{(\mu)}e^{s_kH}e^{-s_k}\mathbb{E}\left[\varepsilon_k^{(\theta)}\varepsilon_k^{(\mu)}\middle|\mathcal{F}_{k-1}\right]\right\},\\
A_{3,n} &= A_{2,n}^T,\\
A_{4,n} &= \gamma_n^{-1}\tilde{h}_n^de^{-2s_n}\left\{\sum_{k=1}^{n}[\sigma_k^{(\mu)}]^2e^{2s_k}\mathbb{E}\left[[\varepsilon_k^{(\mu)}]^2\middle|\mathcal{F}_{k-1}\right]\right\}.
\end{aligned}$$

Since (32) implies that

$$\lim_{k\to\infty}\mathbb{E}\left[\varepsilon_k^{(\theta)}[\varepsilon_k^{(\theta)}]^T\middle|\mathcal{F}_{k-1}\right]=f(\theta)G,$$

and since (37) implies that

$$\lim_{k\to\infty}\mathbb{E}\left[[\varepsilon_k^{(\mu)}]^2\middle|\mathcal{F}_{k-1}\right]=f(\theta)\int_{\mathbb{R}^d}K^2(z)dz,$$

the application of Lemma 4 in [19] ensures that

$$\lim_{n\to\infty}A_{1,n}=\Sigma^{(\theta)}\quad\text{and}\quad\lim_{n\to\infty}A_{4,n}=\Sigma^{(\mu)},$$

where $\Sigma^{(\mu)}$ and $\Sigma^{(\theta)}$ are defined in (8) and (10), respectively. Now, setting $h_k^*=\min(h_k,\widetilde{h}_k)$, we note that

$$\mathbb{E}\left[\varepsilon_k^{(\theta)}\varepsilon_k^{(\mu)}\middle|\mathcal{F}_{k-1}\right]$$

$$\begin{aligned}
&= \frac{1}{\sqrt{h_n^d \widetilde{h}_n^d}} \left[\int_{\mathbb{R}^d} \nabla K\left(\frac{\theta_{k-1}-x}{h_k}\right) K\left(\frac{\theta_{k-1}-x}{h_k}\right) f(x)dx \right] + o(1) \\
&= \frac{h_n^{*d}}{\sqrt{h_n^d \widetilde{h}_n^d}} \left[\int_{\mathbb{R}^d} \nabla K\left(\frac{h_k^*}{h_k}z\right) K\left(\frac{h_k^*}{\widetilde{h}_k}z\right) f\left(\theta_{k-1}-h_k^* z\right) dz \right] + o(1) \\
&= f(\theta_{k-1}) \frac{h_n^{*d}}{\sqrt{h_n^d \widetilde{h}_n^d}} \left[\int_{\mathbb{R}^d} \nabla K\left(\frac{h_k^*}{h_k}z\right) K\left(\frac{h_k^*}{\widetilde{h}_k}z\right) dz \right] + o(1)
\end{aligned}$$

(46) $$= o(1)$$

since the function $z \mapsto [\nabla K(z)]K(z)$ is odd in each of its coordinates. We thus get

$$\begin{aligned}
\|A_{2,n}\|_2 &\leq \gamma_n^{-1} \sqrt{h_n^{d+2} \tilde{h}_n^d} \sum_{k=1}^n \left\| e^{(s_n-s_k)(H-I_d)} \right\|_2 o\left(\gamma_k^2 \sqrt{h_k^{-(d+2)} \tilde{h}_k^{-d}}\right) \\
&\to 0 \text{ as } n \to \infty
\end{aligned}$$

by application of Lemma 3 in [19]. It follows that

(47) $$\lim_{n\to\infty} \langle M \rangle_n^{(n)} = \begin{pmatrix} \Sigma^{(\theta)} & 0 \\ 0 & \Sigma^{(\mu)} \end{pmatrix} \quad a.s.$$

Let us now check that $(M_j^{(n)})_{1\leq j\leq n}$ satisfies Lindeberg's condition. Set $b > 2$; we first note that

$$\begin{aligned}
&\sum_{k=1}^n \mathbb{E}\left[\left\| \sqrt{\gamma_n^{-1} h_n^{d+2}} e^{s_n H} e^{-s_k H} \sigma_k^{(\theta)} \varepsilon_k^{(\theta)} \right\|^b \middle| \mathcal{F}_{k-1} \right] \\
&= O\left(\left[\gamma_n^{-1} h_n^{d+2}\right]^{b/2} e^{-bLs_n} \sum_{k=1}^n e^{bLs_k} \gamma_k^b h_k^{-b(d+2)/2} \mathbb{E}\left[\|\varepsilon_k^{(\theta)}\|^b \middle| \mathcal{F}_{k-1} \right] \right) \\
&= O\left(\left[\gamma_n^{-1} h_n^{d+2}\right]^{b/2} e^{-bLs_n} \sum_{k=1}^n e^{bLs_k} \gamma_k a_k \right)
\end{aligned}$$

with

$$\begin{aligned}
a_k = \gamma_k^{b-1} h_k^{-b(d+2)/2-db/2+d} &= \gamma_k^{b-1} h_k^{-db-b+d} o(h_k^{-b+2}) \\
= o([\gamma_k h_k^{-(d+2)}]^{b-1}) &= o([\gamma_k h_k^{-(d+2)}]^{b/2}).
\end{aligned}$$

It follows that

(48) $$\sum_{k=1}^n \mathbb{E}\left[\left\| \sqrt{\gamma_n^{-1} h_n^{d+2}} e^{s_n H} e^{-s_k H} \sigma_k^{(\theta)} \varepsilon_k^{(\theta)} \right\|^b \middle| \mathcal{F}_{k-1} \right] = o(1).$$

Similarly, we have

$$\begin{aligned}
&\sum_{k=1}^{n} \mathbb{E}\left[\left|\sqrt{\gamma_n^{-1}\widetilde{h}_n^d} e^{-s_n} e^{s_k} \sigma_k^{(\mu)} \varepsilon_k^{(\mu)}\right|^b \middle| \mathcal{F}_{k-1}\right] \\
&\quad = O\left(\left[\gamma_n^{-1}\widetilde{h}_n^d\right]^{b/2} e^{-bs_n} \sum_{k=1}^{n} e^{bs_k} \gamma_k^b \widetilde{h}_k^{-bd/2} \mathbb{E}\left[\left|\varepsilon_k^{(\mu)}\right|^b \middle| \mathcal{F}_{k-1}\right]\right) \\
&\quad = O\left(\left[\gamma_n^{-1}\widetilde{h}_n^d\right]^{b/2} e^{-bs_n} \sum_{k=1}^{n} e^{bs_k} \gamma_k^b \widetilde{h}_k^{-bd+d}\right) \\
&\quad = O\left(\left[\gamma_n^{-1}\widetilde{h}_n^d\right]^{b/2} e^{-bs_n} \sum_{k=1}^{n} e^{bs_k} \gamma_k o([\gamma_k \widetilde{h}_k^{-d}]^{b/2})\right) \\
&\quad = o(1). \qquad (49)
\end{aligned}$$

Lemma 7 then follows from (47), (48) and (49).

5.3. Proof of Proposition 2 and Theorem 2. To prove Proposition 2 (respectively, Theorem 2), we need to apply Lemma 4 (respectively, Lemma 6) to the stochastic approximation algorithms (30) and (36). To this end, we set

$$\begin{aligned}
\Lambda_{n+1}^{(\theta)} &= \frac{-1}{\sum_{k=1}^{n} h_k^{d+2}} [D^2 f(\theta)]^{-1} \sum_{k=1}^{n} h_k^{d+2} h_{k+1}^{-(d+2)/2} \varepsilon_{k+1}^{(\theta)}, \\
\Xi_{n+1}^{(\theta)} &= (\overline{\theta}_n - \theta) - \Lambda_{n+1}^{(\theta)}, \\
\Lambda_{n+1}^{(\mu)} &= \frac{1}{\sum_{k=1}^{n} \widetilde{h}_k^d} \sum_{k=1}^{n} \widetilde{h}_k^d \widetilde{h}_{k+1}^{-d/2} \varepsilon_{k+1}^{(\mu)}, \\
\Xi_{n+1}^{(\mu)} &= (\overline{\mu}_n - \mu) - \Lambda_{n+1}^{(\mu)}.
\end{aligned}$$

- We have seen in Section 5.1.1 that assumptions (A1)–(A7) stated in Section 3 are fulfilled by the stochastic approximation algorithm (30) with $h = \nabla f$,

$$H = D^2 f(\theta), \quad (c_n) = (h_n^{(d+2)/2}), \quad \tau = a(d+2)/2, \quad \varepsilon_n = \varepsilon_n^{(\theta)},$$

$\Gamma = f(\theta) \int_{\mathbb{R}^d} \nabla K(z) [\nabla K(z)]^T dz$, $R_n = R_n^{(\theta)}$, $(v_n) = (h_n^{-2})$, $v^* = 2a$, and $\rho = R^{(\theta)}$ ($R^{(\theta)}$ being defined in (6)). The first part of Proposition 2 is thus a straightforward consequence of the application of Lemma 4. At this first step, we note that (H2), (H3)(i) and (H5) ensure that (A8) holds with $\eta = 2$, and it thus follows from Lemma 6 that:

(50) if $\lim_{n\to\infty} nh_n^{d+6} \in (0,\infty]$,

$$\text{then } \lim_{n\to\infty} h_n^{-2}\ \Xi_{n+1}^{(\theta)} = \frac{-[1-a(d+2)]}{1-a(d+4)}[D^2 f(\theta)]^{-1} R^{(\theta)} \text{ a.s.},$$

(51) if $\lim_{n\to\infty} nh_n^{d+6} = 0$, then $\lim_{n\to\infty} \sqrt{nh_n^{d+2}}\ \Xi_{n+1}^{(\theta)} = 0$ a.s.

• We have seen in Section 5.2 that assumptions (A1)–(A7) stated in Section 3 are fufilled by the stochastic approximation algorithm (36) with $H = -1$, $(c_n) = (\widetilde{h}_n^{d/2})$, $\tau = \tilde{a}d/2$, $\varepsilon_n = \varepsilon_n^{(\mu)}$, $\Gamma = f(\theta)\int_{\mathbb{R}^d} K^2(z)dz$, $R_n = R_n^{(\mu)}$, $(v_n) = (\tilde{h}_n^{-2})$, $v^* = 2\tilde{a}$, and $\rho = R^{(\mu)}$ ($R^{(\mu)}$ being defined in (7)). The second part of Proposition 2 is thus a straightforward consequence of the application of Lemma 4. To start the proof of Theorem 2, let us note that (H5) ensures that (A8) holds with $\eta = 2$. It thus follows from Lemma 6 that:

(52) if $\lim_{n\to\infty} n\widetilde{h}_n^{d+4} \in (0,\infty]$, then $\lim_{n\to\infty} \widetilde{h}_n^{-2}\ \Xi_{n+1}^{(\mu)} = \dfrac{1-\widetilde{a}d}{1-\widetilde{a}(d+2)} R^{(\mu)}$ a.s.,

(53) if $\lim_{n\to\infty} n\widetilde{h}_n^{d+4} = 0$, then $\lim_{n\to\infty} \sqrt{n\widetilde{h}_n^{d}}\ \Xi_{n+1}^{(\mu)} = 0$ a.s.

Properties (12), (13), and (14) in Theorem 2 straightforwardly follow from the application of (50), (51), (52), (53), together with the following lemma.

Lemma 8. *Under* (H1)–(H5),

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}\Lambda_{n+1}^{(\theta)} \\ \sqrt{n\widetilde{h}_n^{d}}\Lambda_{n+1}^{(\mu)} \end{pmatrix} \overset{\mathcal{D}}{\to} \mathcal{N}\left(0\ ,\ \begin{pmatrix} \Sigma_{\text{opt}}^{(\theta)} & 0 \\ 0 & \Sigma_{\text{opt}}^{(\mu)} \end{pmatrix}\right).$$

We now prove Lemma 8. Set

$$W_j^{(n)} = \begin{pmatrix} \left[nh_n^{d+2}\right]^{-1/2} \sum_{k=1}^{j} h_k^{d+2} h_{k+1}^{-(d+2)/2} \varepsilon_{k+1}^{(\theta)} \\ \left[n\widetilde{h}_n^{d}\right]^{-1/2} \sum_{k=1}^{j} \widetilde{h}_k^{d} \widetilde{h}_{k+1}^{-d} \varepsilon_{k+1}^{(\mu)} \end{pmatrix}.$$

$W^{(n)} = (W_j^{(n)})_{1\leq j\leq n}$ is a martingale triangular array whose predictable quadratic variation satisfies

$$\langle W\rangle_n^{(n)} = \begin{pmatrix} w_{1,n} & w_{2,n} \\ w_{3,n} & w_{4,n} \end{pmatrix}$$

with

$$\begin{aligned}
w_{1,n} &= \frac{1}{nh_n^{d+2}} \sum_{k=1}^{n} h_k^{2(d+2)} h_{k+1}^{-(d+2)} \mathbb{E}\left[\varepsilon_{k+1}^{(\theta)} [\varepsilon_{k+1}^{(\theta)}]^T \middle| \mathcal{F}_k \right], \\
w_{4,n} &= \frac{1}{n\widetilde{h}_n^d} \sum_{k=1}^{n} \widetilde{h}_k^{2d} \widetilde{h}_{k+1}^{-d} \mathbb{E}\left[[\varepsilon_{k+1}^{(\mu)}]^2 \middle| \mathcal{F}_k \right], \\
w_{2,n} &= \frac{1}{\sqrt{(nh_n^{d+2})(n\widetilde{h}_n^d)}} \sum_{k=1}^{n} h_k^{d+2} h_{k+1}^{-(d+2)/2} \widetilde{h}_k^d \widetilde{h}_{k+1}^{-d/2} \mathbb{E}\left[\varepsilon_{k+1}^{(\theta)} \varepsilon_{k+1}^{(\mu)} \middle| \mathcal{F}_k \right], \\
w_{3,n} &= w_{2,n}^T.
\end{aligned}$$

We have

$$\lim_{n\to\infty} w_{1,n} = \lim_{n\to\infty} \frac{1}{nh_n^{d+2}} \sum_{k=1}^{n} h_k^{(d+2)} f(\theta) G[1+o(1)] = [1-a(d+2)]^{-1} f(\theta) G,$$

and

$$\begin{aligned}
\lim_{n\to\infty} w_{4,n} &= \lim_{n\to\infty} \frac{1}{n\widetilde{h}_n^d} \sum_{k=1}^{n} \widetilde{h}_k^d f(\theta) \int_{\mathbb{R}^d} K^2(z) dz [1+o(1)] \\
&= [1-\widetilde{a}d]^{-1} f(\theta) \int_{\mathbb{R}^d} K^2(z) dz.
\end{aligned}$$

Moreover, (46) ensures that

$$\begin{aligned}
w_{2,n} &= O\left(\frac{1}{\sqrt{(nh_n^{d+2})(n\widetilde{h}_n^d)}} \right) \sum_{k=1}^{n} o\left(h_k^{(d+2)/2} \widetilde{h}_k^{d/2} \right) \quad a.s. \\
&= o(1) \quad a.s.
\end{aligned}$$

It follows that

(54) $$\lim_{n\to\infty} \langle W \rangle_n^{(n)} = \begin{pmatrix} [1-a(d+2)]^{-1} f(\theta) G & 0 \\ 0 & [1-\widetilde{a}d]^{-1} f(\theta) \int_{\mathbb{R}^d} K^2(z) dz \end{pmatrix} \quad a.s.$$

Let us now check that $(W_j^{(n)})_{1\leq j\leq n}$ satisfies Lindeberg's condition. For $b>2$, we have

$$\sum_{k=1}^{n}\mathbb{E}\left[\left|(nh_n^{d+2})^{-1/2}h_k^{d+2}h_{k+1}^{-(d+2)/2}\varepsilon_{k+1}^{(\theta)}\right|^b\middle|\mathcal{F}_k\right] = (nh_n^{d+2})^{-b/2}\sum_{k=1}^{n}O\left(h_k^{b+d}\right)$$

$$= (nh_n^{d+2})^{-b/2}\sum_{k=1}^{n}o\left(h_k^{d+2}\right)$$

$$= o(1), \tag{55}$$

and

$$\sum_{k=1}^{n}\mathbb{E}\left[\left|(n\widetilde{h}_n^d)^{-1/2}\widetilde{h}_k^d\widetilde{h}_{k+1}^{-d/2}\varepsilon_{k+1}^{(\mu)}\right|^b\middle|\mathcal{F}_k\right] = (n\widetilde{h}_n^d)^{-b/2}\sum_{k=1}^{n}O\left(\widetilde{h}_k^d\right)$$

$$= O\left([n\widetilde{h}_n^d]^{1-b/2}\right)$$

$$= o(1). \tag{56}$$

We deduce from (54), (55) and (56) that

$$W_n^{(n)} \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(0\ ,\ \begin{pmatrix} [1-a(d+2)]^{-1}f(\theta)G & 0 \\ 0 & [1-\widetilde{a}d]^{-1}f(\theta)\int_{\mathbb{R}^d}K^2(z)dz\end{pmatrix}\right).$$

Lemma 8 then follows from the fact that

$$\begin{pmatrix}\sqrt{nh_n^{d+2}}\Lambda_{n+1}^{(\theta)} \\ \sqrt{n\widetilde{h}_n^d}\Lambda_{n+1}^{(\mu)}\end{pmatrix} = A_nW_n^{(n)}$$

with

$$A_n = \begin{pmatrix}\frac{-nh_n^{d+2}}{\sum_{k=1}^n h_k^{d+2}}[D^2f(\theta)]^{-1} & 0 \\ 0 & \frac{n\widetilde{h}_n^d}{\sum_{k=1}^n\widetilde{h}_k^d}\end{pmatrix}$$

$$\to \begin{pmatrix}-[1-a(d+2)][D^2f(\theta)]^{-1} & 0 \\ 0 & 1-\widetilde{a}d\end{pmatrix}.$$

Acknowledgments. We deeply thank an anonymous Referee for his helpful comments, which led to a substantial improvement of the original version of this paper.

REFERENCES

[1] J. R. Blum. Multidimensional stochastic approximation methods. *Ann. Math. Statist.* **25** (1954), 737–744

[2] R. Bojanic, E. Seneta. A unified theory of regularly varying sequences. *Math. Z.* **134** (1973), 91–106.

[3] H. Chen. Lower rate of convergence for locating a maximum of a function. *Ann. Statist.* **16** (1988), 1330–1334.

[4] H. F. Chen, T. E. Duncan, B. Pasik-Duncan. A Kiefer-Wolfowitz algorithm with randomized differences. *IEEE Trans. Automat. Control* **44** (1999), 442-453

[5] B. Delyon, A. B. Juditsky. Stochastic optimization with averaging of trajectories. *Stochastics Stochastic Rep.* **39** (1992), 107–118.

[6] J. Dippon, J. Renz. Weighted means in stochastic approximation of minima. *SIAM J. Control Optim.,* **35** (1997), 1811–1827.

[7] J. Dippon. Accelerated randomized stochastic optimization. *Ann. Statist.,* **31** (2003), 1260–1281.

[8] M. Duflo. Algorithmes stochastiques. Collection Applications of mathematics, Springer, 1996.

[9] M. Duflo. Random Iterative Models. Collection Applications of mathematics, Springer, 1997.

[10] V. Fabian. Stochastic approximation of minima with improved asymptotic speed. *Ann. Math. Statist.* **38** (1967), 191–200.

[11] J. Galambos, E. Seneta. Regularly varying sequences. *Proc. Amer. Math. Soc.* **41** (1973), 110–116.

[12] P. Hall, C. C. Heyde. Martingale limit theory and its application. Academic Press, Inc., New York-London, 1980.

[13] R. A. Horn, C. R. Johnson. Matrix analysis. Cambridge University Press, 1985.

[14] J. Kiefer, J. Wolfowitz. Stochastic approximation of the maximum of a regression functions. *Ann. Math. Statist.* **23** (1952), 462–466.

[15] H. J. Kushner, D. S. Clark. Stochastic approximation methods for constrained and unconstrained systems. Springer, New York, 1978.

[16] H. J. Kushner, J. Yang. Stochastic approximation with averaging of the iterates: Optimal asymptotic rate of convergence for general processes. *SIAM J. Control Optim.* **31** (1993), 1045–1062.

[17] L. Ljung, G. Pflug, H. Walk. Stochastic approximation and optimization of random systems. Birkhäuser, 1992.

[18] A. Mokkadem, M. Pelletier. The law of the iterated logarithm for the multivariable kernel mode estimator. *ESAIM Probab. Stat.* **7** (2003), 1–21.

[19] A. Mokkadem, M. Pelletier. The compact law of the iterated logarithm for multivariate stochastic approximation algorithms. *Stoch. Anal. Appl.* **23** (2005), 181–203.

[20] A. Mokkadem, M. Pelletier. A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm. *Annals of Statistics* **35** (2007), 1749–1772.

[21] M. B. Nevels'on, R. Z. Has'minskii. Stochastic approximation and recursive estimation. Amer. Math. Soc, Providence, RI, 1976.

[22] B. T. Polyak. New method of stochastic approximation type. *Automat. Remote Control* **51**, (1990), 937–946.

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

[24] B. T. Polyak, A. B. Tsybakov. Optimal orders of accuracy for search algorithms of stochastic optimization. *Problems Inform. Transmission* **26** (1990), 126–133.

[25] P. Révész. Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. *Studia Sci. Math. Hung.* **8** (1973), 391–398.

[26] P. Révész. How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statistics* **8** (1977), 119–126.

[27] D. Ruppert. Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise. *Ann. of Probab.* **10** (1982), 178–187.

[28] D. Ruppert. Stochastic approximation. Handbook of Sequential Analysis. Statist. Textbooks Monogr., vol. **118**, Dekker, New York, 1991, 503–529.

[29] J. C. Spall. A stochastic approximation algorithm for large-dimensional systems in the Kiefer-Wolfowitz setting. In: Proc. Conference on Decision and Control, IEEE, New York, 1988, 1544–1548.

[30] J. C. Spall. A one-measurement form of simultaneous perturbation stochastic approximation. *Automatica J. IFAC* **33** (1997), 109–112.

[31] A. B. Tsybakov. Recurrent estimation of the mode of a multidimensional distribution. *Problems Inform. Transmission* **26** (1990), 31–37.

[32] G. Yin. On extensions of Polyak's averaging approach to stochastic approximation. *Stochastics Stochastic Rep.* **33** (1991), 245–264.

Université de Versailles-Saint-Quentin
Département de Mathématiques
45, Avenue des Etats-Unis
78035 Versailles Cedex
France
e-mail: djeddour@math.uvsq.fr
mokkadem@math.uvsq.fr
pelletier@math.uvsq.fr

Received March 11, 2008
Revised April 29, 2008