PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 24, Fasc. 2 (2004), pp. 381–399

DEPENDENT NOISE FOR STOCHASTIC ALGORITHMS

BY

PAUL DOUKHAN (MALAKOFF) AND ODILE BRANDIÈRE (ORSAY)

Abstract. We introduce different ways of being dependent for the input noise of stochastic algorithms. We are aimed to prove that such innovations allow to use the ODE (ordinary differential equation) method. Illustrations to the linear regression frame and to the law of large numbers for triangular arrays of weighted dependent random variables are also given.

2000 AMS Subject Classification: Primary 62L20; Secondary 62J05.

Key words and phrases: Stochastic approximation, ordinary differential equations, dependent noise, linear regression.

1. INTRODUCTION

We consider the $\boldsymbol{R}^d$-valued stochastic algorithm, defined on a probability space $(\Omega, \mathscr{A}, \mathscr{P})$ and driven by the recurrence equation

$$Z_{n+1} = Z_n + \gamma_n h(Z_n) + \xi_{n+1}, \tag{1}$$

where h is a continuous function from an open set $G \subseteq \boldsymbol{R}^d$ to $\boldsymbol{R}^d$, (γ_n) is a deterministic real sequence decreasing to zero and satisfying

$$\sum_{n \geqslant 0} \gamma_n = \infty, \tag{2}$$

and (ξ_n) is a "small" stochastic disturbance.

The ordinary differential equation (ODE) method (see e.g. [3], [13], [19]) associates the possible limit sets of (1) with the properties of the associated ODE

$$dz/dt = h(z). \tag{3}$$

These sets are compact connected invariant and "chain-recurrent" in the Benaïm sense for the ODE (cf. [1]). These sets are more or less complicated. Various situations may then happen, for example the simplest case is an equilibrium: z is a solution of $h(z) = 0$, an equilibria cycle, or a finite set of equilibria is linked to the ODE's trajectories, connected sets of equilibria, or periodic cycles for the ODE, etc.

To use the ODE method, we need some conditions: (Z_n) is bounded and

$$\xi_{n+1} = c_n(\zeta_{n+1} + r_{n+1}), \tag{4}$$

where (c_n) denotes a nonnegative deterministic sequence such that

$$\gamma_n = O(c_n), \quad \sum c_n^2 < \infty, \tag{5}$$

(ζ_n) and (r_n) are $\boldsymbol{R}^d$-valued random vector sequences, defined on $(\Omega, \mathscr{A}, \mathscr{P})$, and adapted with respect to an increasing sequence of σ-fields $(\mathscr{F}_n)_{n \geqslant 0}$ and satisfying almost surely (a.s.) on $A \subset \Omega$:

$$\sum c_n \zeta_{n+1} < \infty \text{ a.s.} \tag{6}$$

and

$$r_n \to 0 \text{ a.s.} \quad \text{as } n \to \infty. \tag{7}$$

The classic algorithms theory relates to a noise (ζ_n) which is a martingale difference sequence. The aim of this paper is to replace this condition on the noise by a weakly dependence condition.

The paper is devoted to sufficient conditions for (6). Section 2 considers the weak dependence condition from Doukhan and Louhichi in [12]; set for this

$$\mathrm{Lip}(h) = \sup_{(x_1, \ldots, x_u) \neq (y_1, \ldots, y_u)} \frac{|h(x_1, \ldots, x_u) - h(y_1, \ldots, y_u)|}{|x_1 - y_1| + \ldots + |x_u - y_u|} \quad \text{if } h: \boldsymbol{R}^u \to \boldsymbol{R},$$

$$\Lambda = \{h \text{ such that } h: \boldsymbol{R}^u \to \boldsymbol{R} \text{ for some } u > 0 \text{ and } \mathrm{Lip}(h) + \|h\|_\infty < \infty\},$$

and consider some function $C: \boldsymbol{N}^{*2} \to \boldsymbol{R}$. The sequence (ζ_n) is said to be $(\varepsilon, \Lambda, C)$-*weakly dependent* if there exists a sequence $\varepsilon = (\varepsilon_r)_{r \geqslant 0}$ such that $\varepsilon_r \downarrow 0$ as $r \uparrow \infty$ and satisfying, for any $(u+v)$-tuple $(t_1, \ldots, t_{u+v})$ with $t_1 \leqslant \ldots \leqslant t_u < t_u + r \leqslant t_{u+1} \leqslant \ldots \leqslant t_{u+v}$,

$$\left|\mathrm{Cov}\left(h(\zeta_{t_1}, \ldots, \zeta_{t_u}), k(\zeta_{t_{u+1}}, \ldots, \zeta_{t_{u+v}})\right)\right| \leqslant C(u, v)\left(\mathrm{Lip}(h) + \mathrm{Lip}(k)\right)\varepsilon_r. \tag{8}$$

Various examples of this situation may be found in [11]; they include

- general Bernoulli shifts, $\zeta_t = \sum_{l=1}^{\infty} \sum_{k_1, \ldots, k_l} a_{k_1, \ldots, k_l}^{(l)} \xi_{t-k_1} \cdots \xi_{t-k_l}$,
- stable Markov chains such as $\zeta_t = G(\zeta_{t-1}, \ldots, \zeta_{t-p}) + \xi_t$, or
- ARCH(∞) models, $\zeta_t = \left(a_0 + \sum_{j \geqslant 1} a_j \zeta_{t-j}\right) \xi_t$

generated by some i.i.d. sequence (ξ_t). In the first example, the situation of an infinite moving average, for which $l = 1$, is of a special interest and $\varepsilon_r \leqslant 4E|\xi_0| \sum_{2|k| \geqslant r} |a_k^{(1)}|$. Now $\varepsilon_r \downarrow 0$ (geometrically) in the second case if

$$|G(x_1, \ldots, x_p) - G(y_1, \ldots, y_p)| \leqslant \sum_j b_j |x_j - y_j|$$

with $\sum_j b_j < 1$ and $E|\xi_0| < \infty$. In the last, non-Markov and non-linear example, a chaotic expansion holds if $\sum_{j\geqslant 1} |a_j| E|\xi_0| < 1$, and then any class of rate may be obtained for ε_r. Note that r always denotes the gap in time between "past" and "future". A generalization to the vector $\mathbf{R}^d$-case is also provided below.

Section 3 considers a weakly dependent noise in the sense of the γ-weak coefficients in Dedecker and Doukhan [9]. The mixingale-type coefficient used there, determined for the sequence $(\zeta_n)_{n\geqslant 0}$, is defined as

$$\gamma_r = \sup_{k\geqslant 0} \left\| E\left(\zeta_{k+r} | \sigma(\zeta_i, i \leqslant k)\right) - E(\zeta_{k+r}) \right\|_1 .$$

The sequence (ζ_n) is said to be *γ-weakly dependent* if $\gamma_r \downarrow 0$ as $r \uparrow \infty$. In [9], this is proved that a causal version of $(\varepsilon, \Lambda, C)$-weak dependence implies γ-weak dependence, where the right-hand side in (8) takes the form $C(v)\,\mathrm{Lip}(k)\,\theta_r$. Counterexamples of γ-weakly dependent sequences which are not θ-weakly dependent may also be found there.

We first settle an immediate extension of this notion to $\mathbf{R}^d$-valued sequences. The definition of γ-weak dependence extends to $\mathbf{R}^d$ and we have:

PROPOSITION 1.1. *The following two assertions are equivalent:*
(i) *An $\mathbf{R}^d$-valued sequence (X_n) is γ-weakly dependent.*
(ii) *Each component (X_n^l) $(l = 1, \ldots, d)$ of (X_n) is γ-weakly dependent.*

Proof. Clearly,

$$\left\| E\left(X_{n+r}^l - E(X_{n+r}^l) \,|\, \mathscr{F}_n\right) \right\|_1 \leqslant \left\| E\left(X_{n+r} - E(X_{n+r}) \,|\, \mathscr{F}_n\right) \right\|_1$$

and (i) implies (ii). On the other hand,

$$\left\| E\left(X_{n+r} - E(X_{n+r}) \,|\, \mathscr{F}_n\right) \right\|_1 = E\sqrt{\textstyle\sum_{l=1}^d \left(E\left(X_{n+r}^l - E(X_{n+r}^l) \,|\, \mathscr{F}_n\right)\right)^2},$$

whence

$$\left\| E\left(X_{n+r} - E(X_{n+r}) \,|\, \mathscr{F}_n\right) \right\|_1 \leqslant \sqrt{d} \max_{1\leqslant l\leqslant d} \gamma_r^l. \ \blacksquare$$

In the frame of the (θ, Λ, C)-weak dependence we say that the $\mathbf{R}^d$-valued sequence (X_n) is *(θ, Λ, C)-weakly dependent* if each component (X_n^l) is (θ^l, Λ, C)-weakly dependent.

The forthcoming two sections are devoted to provide moment inequalities of the Marcinkiewicz–Zygmund type adapted to deduce the relation (6) in those two frames. Section 4 is devoted to apply the study to the examples of Robbins–Monro and Kiefer–Wolfowitz algorithms. In Section 5 we obtain sufficient conditions for the complete convergence of triangular arrays, extending to Chow [6].

Finally, Section 6 is devoted to the specific algorithm of linear regression and with dependent entries. In [5] Chen has also studied this topic. He works in a more general matrix-valued framework. Assuming only the stationarity and the ergodicity of entries, he derives the a.s. convergence of the algorithm. We get the same result with a γ-weakly dependence assumption, but this assumption, more restrictive, allows us to reach, thanks to a moment technique, a precise $n^{-1/2}$-convergence rate.

The proofs are relegated to a final section.

2. WEAKLY DEPENDENT NOISE

Let (ζ_n) be a sequence of centered random variables satisfying an $(\varepsilon, \Lambda, C)$-weak dependence as described in inequality (8). We denote by S_n the sum $\sum_{i=1}^{n} \zeta_i$, put $C_q = \max_{u+v \leqslant q} C(u, v)$, and we assume that

$$\sup |\mathrm{Cov}(\zeta_{t_1} \ldots \zeta_{t_m}, \zeta_{t_{m+1}} \ldots \zeta_{t_q})| \leqslant C_q q^\gamma M^{q-2} \varepsilon_r, \tag{9}$$

where the supremum is taken over all $\{t_1, \ldots, t_q\}$ such that $1 \leqslant t_1 \leqslant \ldots \leqslant t_q$, and $1 \leqslant m < q$ such that $t_{m+1} - t_m = r$, or

$$|\mathrm{Cov}(\zeta_{t_1} \ldots \zeta_{t_m}, \zeta_{t_{m+1}} \ldots \zeta_{t_q})| \leqslant M_q \int_0^{\min(\varepsilon_r, 1)} Q_{\zeta_{t_1}}(x) \ldots Q_{\zeta_{t_q}}(x)\, dx, \tag{10}$$

where Q_X denotes the quantile function of $|X|$, which is the generalized inverse of the tail function $t \mapsto P(|X| > t)$ and $M_q = \max(C_q, 2)$.

The bound (9) is mainly suitable for bounded sequences while (10) holds for more general r.v.'s, using moment (or tail) assumptions. Various examples for which one of these two bounds holds are given in [12]. Moreover, let p be some fixed integer not less than 2.

If (9) holds for all $q \leqslant p$, then, for any $n \geqslant 2$

$$|ES_n^p| \leqslant \frac{(2p-2)!}{(p-1)!} p^\gamma \left\{\left(nC_2 \sum_{r=0}^{n-1} \varepsilon_r\right)^{p/2} \vee \left(nC_p M^{p-2} \sum_{r=0}^{n-1} (r+1)^{p-2} \varepsilon_r\right)\right\}. \tag{11}$$

If, now, (10) holds for all $q \leqslant p$, then, for any $n \geqslant 2$

$$|ES_n^p| \leqslant \frac{(2p-2)!}{(p-1)!} \left\{\left(C_p \sum_{i=1}^{n} \int_0^1 [\min(\varepsilon^{-1}(u), n)]^{p-1} Q_{\zeta_i}^p(u)\, du\right) \vee \left(\left(C_2 \sum_{i=1}^{n} \int_0^1 [\min(\varepsilon^{-1}(u), n)]^{p-1} Q_{\zeta_i}^p(u)\, du\right)^{p/2}\right)\right\}. \tag{12}$$

Putting $\Sigma_n = \sum_{i=1}^{n} c_{i-1} \zeta_i$, and using similar techniques to those in [12], we derive the following result:

PROPOSITION 2.1. *Let $p \geqslant 2$ be some fixed integer and let (ζ_n) be a centered $(\varepsilon, \Lambda, C)$-weakly dependent sequence of real random variables such that* (10) *holds* *for all* $q \leqslant p$. *Then for* $n \geqslant 2$

$$|E\Sigma_n^p| \leqslant \frac{(2p-2)!}{(p-1)!}\{(C_p p^\gamma M^{p-2}\sum_{i=1}^{n} c_{i-1}^p \sum_{r=0}^{n-1}(r+1)^{p-2}\varepsilon_r) \vee (C_2 2^\gamma \sum_{i=1}^{n} c_{i-1}^2 \sum_{r=0}^{n-1}\varepsilon_r)^{p/2}\}. \tag{13}$$

This result is mainly adapted to the bounded sequence. The following result is appropriate to more general real variables (r.v.) sequence but requires a moment assumption (or a tail) condition.

PROPOSITION 2.2. *Let p be some fixed integer not less than 2 and (ζ_n) a centered $(\varepsilon, \Lambda, C)$-weakly dependent r.v. sequence. Assume that for all $2 < q \leqslant p$ inequality* (10) *holds with*

$$M_q \leqslant M_p^{(q-2)/(p-2)} M_2^{(p-q)/(p-2)}, \tag{14}$$

and there exists a constant $c > 0$ such that

$$\exists k > p,\ \forall i \geqslant 0,\ P(|\zeta_i| > t) \leqslant c/t^k. \tag{15}$$

Then for $n \geqslant 2$

$$|E\Sigma_n^p| \leqslant \frac{(2p-2)!}{(p-1)!}c^{1/k}\{(M_p\sum_{i=1}^{n} c_{i-1}^p \sum_{r=0}^{n-1}(r+1)^{p-2}\varepsilon_r^{(k-p)/k}) \vee (M_2 \sum_{i=1}^{n} c_{i-1}^2 \sum_{r=0}^{n-1}\varepsilon_r^{(k-2)/k})^{p/2}\}. \tag{16}$$

Note that (15) holds if the ζ_n's have a k-th order moment such that, for all $i \geqslant 0$, $E|\zeta_i|^k \leqslant c$.

Arguing as in Billingsley [4], if (13) holds for some p such that

$$\{(C_p p^\gamma M^{p-2}\sum_{i=1}^{n} c_{i-1}^p \sum_{r=0}^{n-1}(r+1)^{p-2}\varepsilon_r) \vee (C_2 2^\gamma \sum_{i=1}^{n} c_{i-1}^2 \sum_{r=0}^{n-1}\varepsilon_r)^{p/2}\} < \infty, \tag{17}$$

then, for all $\zeta > 0$,

$$\lim_{n\to\infty} P(\sup_{k\geqslant 1}|\Sigma_{n+k}-\Sigma_n| > \zeta) = 0.$$

Thus (Σ_n) is a.s. a Cauchy sequence, and hence it converges. In the same way, if (16) holds for some p such that

$$(\sum_{i=1}^{n} c_{i-1}^p \sum_{r=0}^{\infty}(r+1)^{p-2}\varepsilon_r^{(k-p)/k}) \vee (\sum_{i=1}^{n} c_{i-1}^2 \sum_{r=0}^{n-1}\varepsilon_r^{(k-2)/k})^{p/2} < \infty, \tag{18}$$

then (Σ_n) converges with probability 1.

Equip $\boldsymbol{R}^d$ with its p-norm $\|(x_1, \ldots, x_d)\|_p^p = x_1^p + \ldots + x_d^p$. Let the sequence $(\zeta_n)_{n\geqslant 0}$ be an $\boldsymbol{R}^d$-valued and $(\varepsilon, \Lambda, C)$-weakly dependent sequence. Set $\zeta_n = (\zeta_n^1, \ldots, \zeta_n^d)$; then

$$\|\sum_{i=1}^{n} c_i \zeta_i\|_p^p = \sum_{l=1}^{d} (\sum_{i=1}^{n} c_i \zeta_i^l)^p.$$

And if each component $(\zeta_n^l)_{n\geqslant 0}$ is $(\varepsilon^l, \Lambda, C)$-weakly dependent and such that the relation (17) or (18) holds, then $\boldsymbol{E}\|\Sigma_n\|_p^p < \infty$. Arguing as before, we deduce that the sequence $(\Sigma_n)_{n\geqslant 0}$ converges with probability 1.

The proofs of Propositions 2.1 and 2.2 are given in Section 7.

3. γ-WEAKLY DEPENDENT NOISE

Let $(\zeta_n)_{n\geqslant 0}$ be a sequence of integrable real-valued random variables, and $(\gamma_r)_{r\geqslant 0}$ the associated mixingale-coefficients. Then we obtain the following result:

PROPOSITION 3.1. *Let $p > 2$ and $(\zeta_n)_{n\in\boldsymbol{N}}$ be a sequence of centered random variables such that* (15) *holds. Then for any $n \geqslant 2$*

$$(19) \qquad |\boldsymbol{E}\Sigma_n^p| \leqslant \big(2pK_1 \sum_{i=1}^{n} c_i \sum_{j=0}^{n-i} c_{j+i}\gamma_j^{2(k-p)/p(k-1)}\big)^{p/2},$$

where K_1 depends on k, p and c.

Observe that here $p \in \boldsymbol{R}$ and is not necessarily an integer. If, now, (19) holds for some $p > 2$ such that

$$(20) \qquad \lim_{n\to\infty} \sum_{i=1}^{n} c_i \sum_{j=0}^{n-i} c_{i+j}\gamma_j^m < \infty,$$

where $m = 2(k-p)/p(k-1) < 1$, then (Σ_n) converges with probability 1. Observe that if $\sum_{j=0}^{\infty}\gamma_j^m < \infty$, then (20) is satisfied. The proof of this proposition is in Section 7.

As in Section 2, the result extends to $\boldsymbol{R}^d$. Indeed, if we consider a centered $\boldsymbol{R}^d$-valued and γ-weakly dependent sequence $(\zeta_n)_{n\geqslant 0}$, we have, as in Section 2,

$$\boldsymbol{E}\|\Sigma_n\|^p = \boldsymbol{E}\sum_{l=1}^{d} (\sum_{i=1}^{n} c_i \zeta_i^l)^p,$$

and if each component $(\zeta_n^l)_{n\geqslant 0}$ $(l = 1, \ldots, d)$ is γ^l-weakly dependent and satisfies (15) and (20), then $\boldsymbol{E}\|\Sigma_n\|^p < \infty$, and we conclude as before that $(\Sigma_n)_{n\geqslant 0}$ converges a.s.

4. EXAMPLES OF APPLICATION

4.1. Robbins–Monro algorithm. The Robbins–Monro algorithm is used for dosage, to obtain level a of a function f which is usually unknown. It is also used in mechanics, for adjustments, as well as in statistics to fit a median ([13], p. 50). It writes

$$Z_{n+1} = Z_n - c_n\big(f(Z_n) - a\big) + c_n \zeta_{n+1} \tag{21}$$

with $\sum c_n = \infty$ and $\sum c_n^2 < \infty$. It is usually assumed that the prediction error (ζ_n) is an identically distributed and independent r.v. sequence, but this does not look natural. Weak dependence seems more reasonable. Hence the previous results ensure the a.s. convergence of this algorithm under the usual assumptions and the conditions yielding the a.s. convergence of $\sum_0^n c_n \zeta_{n+1}$.

Under the assumptions of Proposition 2.1, if for some $p > 2$

$$\sum_{r=0}^{\infty} (r+1)^{p-2} \varepsilon_r < \infty, \tag{22}$$

the algorithm (21) converges a.s.

If the assumptions of Proposition 2.2 hold, then the a.s. convergence of the algorithm (21) is ensured if for some $p > 2$

$$\Big(\sum_{r=0}^{\infty} (r+1)^{p-2} \varepsilon_r^{(k-p)/k}\Big) \vee \Big(\sum_{r=0}^{\infty} \varepsilon_r^{(k-2)/k}\Big) < \infty.$$

Under the assumptions of Proposition 3.1, if (20) is satisfied, the algorithm (21) converges with probability 1.

4.2. Kiefer–Wolfowitz algorithm. It is also a dosage algorithm. Here we want to reach the minimum z^* of a real function V which is $\mathscr{C}^2$ on an open set G of $\boldsymbol{R}^d$. The Kiefer–Wolfowitz algorithm ([13], p. 53) is stated as:

$$Z_{n+1} = Z_n - 2c_n \nabla V(Z_n) - \xi_{n+1}, \tag{23}$$

where $\xi_{n+1} = c_n b_n^{-1} \zeta_{n+1} + c_n b_n^2 q(n, Z_n)$, $\|q(n, Z_n)\| \leqslant K$ (for some $K > 0$), $\sum c_n = \infty$, $\sum c_n b_n^2 < \infty$ and $\sum (c_n/b_n)^2 < \infty$ (for instance, $c_n = 1/n$, $b_n = n^{-b}$ with $0 < b < \frac{1}{2}$).

The prediction error (ζ_n) is usually assumed to be i.i.d. centered and square integrable and independent of Z_0. The results of Sections 2 and 3 improve on this assumption until weakly dependent innovations. It is now enough to ensure the convergence a.s. of $\sum c_n b_n^{-1} \zeta_{n+1}$. The (ε, Λ)-weak dependence assumptions are the same as for the Robbins–Monro algorithm. Concerning the γ-weak dependence, the condition (20) is replaced by

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{c_i}{b_i} \sum_{j=1}^{n-i} \frac{c_{i+j}}{b_{i+j}} \gamma_j^m < \infty.$$

5. WEIGHTED WEAKLY DEPENDENT VARIABLES TRIANGULAR ARRAYS

In this section, we consider a sequence $(\zeta_i)_{i\geqslant 1}$ and a triangular array $(c_{ni})_{\{1\leqslant i\leqslant n, n\geqslant 1\}}$ of non-negative real constants. We put $U_n = \sum_{i=1}^n c_{ni}\zeta_i$. If the ζ_i's are i.i.d., Chow has established the following complete convergence result:

THEOREM (Chow [6]). *Let $\zeta_1, \ldots, \zeta_i, \ldots$ be independent and identically distributed random variables with $E(\zeta_i) = 0$ and $E|\zeta_i|^q < \infty$ for some $q \geqslant 2$. If for some real constant K, not depending on n, $\sum_{i=1}^n c_{ni}^2 \leqslant K$ and $n^{1/q}\max_{1\leqslant i\leqslant n}|c_{ni}| \leqslant K$, then*

$$\forall t > 0, \ \sum_{n=1}^{\infty} P(n^{-1/q}|U_n| \geqslant t) < \infty.$$

The last inequality is a result of the complete convergence of $n^{-1/q}|U_n|$ to 0. This notion was introduced by Hsu and Robbins [17]. Complete convergence implies the almost sure convergence from the Borel–Cantelli lemma. Li et al. [20] extend this result to arrays $(c_{ni})_{\{n\geqslant 1, i\in \mathbf{Z}\}}$ for $q = 2$. Recall also Yu [24] who obtains a result analogous to Chow's for martingale differences. Ghosal and Chandra [15] extend previous results and prove some similar results to those of Li et al. for martingale differences. As in [20], their main tool is the Hoffmann-Jørgensen inequality [16]. Peligrad and Utev [22] propose a central limit theorem for partial sums of a sequence $U_n = \sum_{i=1}^n c_{ni}\zeta_i$, where $\sup_n c_{ni}^2 < \infty$, $\max_{1\leqslant i\leqslant n}|c_{ni}| \to 0$ as $n \to \infty$ and ζ_i's are, in turn, pairwise mixing martingale difference, mixing sequences or associated sequences. Mcleish [21], De Jong [8], and, more recently, Shixin [23] extend previous results in the case of L_q-mixingale arrays. Those results have various applications. They are used for the proof of strong convergence of kernel estimators. In this paper we extend Li et al. results to our weak dependent frame. A straightforward consequence of Proposition 2.2 is the following result:

COROLLARY 5.1. *Under the assumptions of Proposition 2.2, if q is an even integer such that $k > q > p$, and if for some real constant K, not depending on n, $\sum_{i=1}^n c_{n,i-1}^2 < K$, and if $\varepsilon_r = \mathcal{O}(r^{-\alpha})$, with $\alpha > \big((q-1)/(k-q)\big)k$, or $\varepsilon_r = \mathcal{O}(e^{-r})$, then*

$$\forall t > 0, \ \sum_n P(n^{-1/p}|U_n| \geqslant t) < \infty.$$

Proof. Proposition 2.2 implies

$$|EU_n|^q \leqslant \frac{(2q-2)!}{(q-1)!}c^{1/k}\Big(\Big(M_q\sum_{i=1}^n c_{n,n-i}^q \sum_{r=0}^{n-1}(r+1)^{q-2}\varepsilon_r^{(k-q)/k}\Big)$$

$$\vee M_2\Big(\sum_{i=1}^n c_{n,n-i}^2\sum_{r=0}^{n-1}\varepsilon_r^{(k-2)/k}\Big)^{p/2}\Big).$$

If $\sum_{i=1}^{n} c_{n,i-1}^2 < K$ and $\varepsilon_r = \mathcal{O}(r^{-\alpha})$ with $\alpha > \big((q-1)/(k-q)\big)k$, then there exists a real constant K_1 such that $E|U_n|^q < K_1$, and the result follows from

$$P(n^{-1/p}|U_n| > t) \leqslant \frac{E|U_n|^q}{t^q n^{q/p}}.$$

If $\sum_{i=1}^{n} c_{n,i-1}^2 < K$ and $\varepsilon_r = \mathcal{O}(e^{-r})$, then $E|U_n|^q < K_2$ for a real constant K_2 and

$$\sum_n P(n^{-1/p}|U_n| > t) < \infty. \ ∎$$

As a straightforward consequence of Proposition 3.1, we obtain the following result:

COROLLARY 5.2. *Under the assumptions of Proposition* 3.1, *if* $q > p$, $k > q > 1$, *and*

$$\lim_{n\to\infty} \sum_{i=1}^{n} c_{n,i} \sum_{j=0}^{n-i} c_{n,i+j}\gamma_j^m < \infty, \quad \textit{where } m = \frac{2}{q}\left(\frac{k-q}{k-1}\right),$$

then

$$\forall t > 0, \sum_n P(n^{-1/p}|U_n| \geqslant t) < \infty.$$

Proof. We obtain

$$E|U_n|^q \leqslant \Big(2qK_1 \sum_{i=1}^{n} c_{n,i} \sum_{j=0}^{n-i} c_{n,i+j}\gamma_j^m\Big)^{q/2}$$

from Proposition 3.1, and the relation $\lim_{n\to\infty} \sum_{i=1}^{n} c_{n,i} \sum_{j=0}^{n-i} c_{n,i+j}\gamma_j^m < \infty$ implies $\sum_n P(n^{-1/p}|U_n| > t) < \infty$. This completes the proof. ∎

6. LINEAR REGRESSION

We consider a stationary (bounded) sequence $(y_n, x_n) \in \mathbf{R} \times \mathbf{R}^d$, defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$.

We look for the vector Z^* which minimizes the linear prediction error of y_n with x_n. We identify the $\mathbf{R}^d$-vector x_n and its column matrix in the canonical basis. Consequently,

$$Z^* = \arg\min_{Z \in \mathbf{R}^d} E\big[(y_n - x_n^T Z)^2\big].$$

This problem leads to study the gradient algorithm

$$Z_{n+1} = Z_n + c_n (y_{n+1} - x_{n+1}^T Z_n) x_{n+1},$$

where $c_n = g/n$ with $g > 0$ (so (c_n) satisfies (2) and (5)). Let $C_{n+1} = x_{n+1} x_{n+1}^T$; then we obtain

$$Z_{n+1} = Z_n + c_n (y_{n+1} x_{n+1} - C_{n+1} Z_n). \tag{24}$$

Let us put $U = E(y_{n+1} x_{n+1})$, $C = E(C_{n+1})$, $Y_n = Z_n - C^{-1} U$, and $h(Y) = -CY$; then (24) becomes

$$Y_{n+1} = Y_n + c_n h(Y_n) + c_n \xi_{n+1} \tag{25}$$

with

$$\xi_{n+1} = (y_{n+1} x_{n+1} - C_{n+1} C^{-1} U) + (C - C_{n+1}) Y_n. \tag{26}$$

Remark that here the solutions of (3) are the trajectories

$$z(t) = z_0 e^{-Ct},$$

so every trajectory converges to 0, the unique equilibrium point of the differentiable function h ($Dh(0) = -C$ and 0 is an attractive zero of h).

Putting $\mathscr{F}_n = \big(\sigma(Y_i);\ i \leqslant n\big)$ we define the following assumption:

A-lr. *C is not singular and (C_n) and $(y_n x_n)$ are γ-weakly dependent sequences such that the γ-weakly dependent coefficient γ_r is $\mathcal{O}(a^r)$ with $a < 1$.*

Note. If $(y_n, x_n)_{n\in N}$ is θ-weakly dependent in the Dedecker and Doukhan sense (see [9]), then **A-lr** is satisfied. This is proved in the Appendix.

Putting $M = \sup_n \|x_n\|^2$ now we claim:

PROPOSITION 6.1. *Under the assumption* **A-lr**, *(Y_n) is a.s. bounded and the perturbation (ξ_n) of the algorithm (25) splits into three terms, two of which are γ-weakly dependent and one is a rest leading to zero. Consequently, the ODE method assures the a.s. convergence of Y_n to zero (hence $Z^* = C^{-1} U$). Moreover, if $g < 1/2M$, then*

$$\sqrt{n}\, Y_n = \mathcal{O}(1) \ \ a.s. \tag{27}$$

Proof of Proposition 6.1. To start with, we prove that $Y_n \to 0$ a.s. by assuming that (Y_n) is a.s. bounded. Then we justify this assumption and finally we prove (27).

The perturbation ξ_{n+1} splits into two terms: $(y_{n+1} x_{n+1} - C_{n+1} C^{-1} U)$ and $(C - C_{n+1}) Y_n$. The first term is centered and obviously γ-weakly dependent with dependence coefficient γ_r. And γ_r is $\mathcal{O}(a^r)$ by the assumption **A-lr**. It remains to study $(C - C_{n+1}) Y_n$.

Study of $(C - C_{n+1}) Y_n$. Write

$$(C - C_{n+1}) Y_n = \zeta_{n+1} + r_{n+1}$$

with

$$\zeta_{n+1} = (C - C_{n+1}) Y_n - E[(C - C_{n+1}) Y_n] \quad \text{and} \quad r_{n+1} = E[(C - C_{n+1}) Y_n].$$

We will prove that the sequence (ζ_n) is γ-weakly dependent with an appropriate dependent coefficient and that $r_n \to 0$.

Notice that $r_{n+1} = E\left[(C-C_{n+1})\sum_{j=n/2}^{n-1}(Y_{j+1}-Y_j)\right]+E\left[(C-C_{n+1})\,Y_{n/2}\right]$, and since $Y_{j+1}-Y_j = -c_j C_{j+1} Y_j + c_j(y_{j+1}x_{j+1}-C_{j+1}C^{-1}U)$, we have

$$r_{n+1} = -\sum_{j=n/2}^{n-1} E(C-C_{n+1})c_j C_{j+1} Y_j + \sum_{j=n/2}^{n-1} E(C-C_{n+1})c_j(y_{j+1}x_{j+1}-C_{j+1}C^{-1}U)$$
$$+E(C-C_{n+1})\,Y_{n/2}.$$

If $n/2$ is not an integer, we replace it by $(n-1)/2$. In the same way, in the first sum we replace Y_j by $\sum_{i=j/2}^{j-1}(Y_{i+1}-Y_i)+Y_{j/2}$ with the same remark as above if $j/2$ is not an integer. Consequently,

$$(28)\qquad r_{n+1} = -\sum_{j=n/2}^{n-1} E(C-C_{n+1})c_j C_{j+1}$$
$$\times\Big[\sum_{i=j/2}^{j-1} -c_i C_{i+1} Y_i + c_i(y_{i+1}x_{i+1}-C_{i+1}C^{-1}U)+Y_{j/2}\Big]$$
$$+\sum_{j=n/2}^{n-1} E(C-C_{n+1})c_j(y_{j+1}x_{j+1}-C_{n+1}C^{-1}U)+E(C-C_{n+1})\,Y_{n/2}.$$

Expectations conditionally with respect to $\mathscr{F}_{j+1}$ of each term of the second sum and with respect $\mathscr{F}_{n/2}$ of the last term give, by assuming (Y_n) bounded,

$$(29)\qquad \|r_{n+1}\| \leqslant \|A\| + K_1 \sum_{j=n/2}^{n-1} c_j \gamma_{n+1-j}\gamma_{j+1} + K_2\gamma_{n/2+1},$$

where A denotes the first sum of (28), and K_i (for $i = 1, 2, \ldots$) is a non-negative constant. Moreover,

$$A = -\sum_{j=n/2}^{n-1} E(C-C_{n+1})c_j(C_{j+1}-C)$$
$$\times\Big[\sum_{i=j/2}^{j-1} -c_i C_{i+1} Y_i + c_i(y_{i+1}x_{i+1}-C_{i+1}C^{-1}U)\Big]$$
$$-\sum_{j=n/2}^{n-1} E(C-C_{n+1})c_j(C_{j+1}-C)\,Y_{j/2}$$
$$+\sum_{j=n/2}^{n-1} E(C-C_{n+1})c_j C\Big[\sum_{i=j/2}^{j-1} -c_i C_{i+1} Y_i + c_i(y_{i+1}x_{i+1}-C_{i+1}C^{-1}U)\Big]$$
$$+\sum_{j=n/2}^{n-1} E(C-C_{n+1})c_j C Y_{j/2}.$$

Expectations conditionally successively with respect to $\mathscr{F}_{j+1}$ and $\mathscr{F}_{i+1}$ of each term of the first and the third sum and with respect to $\mathscr{F}_{j+1}$, then $\mathscr{F}_{j/2}$ of the second and the forth sum give, by assuming (Y_n) bounded,

$$\|A\| \leqslant K_3 \sum_{j=n/2}^{n-1} c_j \gamma_{n-j} \sum_{i=j/2}^{j-1} c_i \gamma_{j-i} + K_4 \sum_{j=n/2}^{n-1} c_j \gamma_{n-j} \gamma_{j/2}. \tag{30}$$

Since $c_j = g/j$, (29) and (30) involve that r_n is $\mathcal{O}(n^{-2})$, so r_n converges to zero. On the other hand, for $r \geqslant 6$:

$$\begin{aligned} E(\zeta_{n+r} \mid \mathscr{F}_n) &= E[(C - C_{n+r}) Y_{n+r-1} \mid \mathscr{F}_n] - E(\zeta_{n+r}) \\ &= \sum_{j=n+r/2}^{n+r-2} E[(C - C_{n+r})(Y_{j+1} - Y_j) \mid \mathscr{F}_n] + E[(C - C_{n+r}) Y_{n+r/2} \mid \mathscr{F}_n] - r_{n+r}. \end{aligned}$$

Note also that if $r/2$ is not an integer, we replace it by $(r+1)/2$. Using the same techniques as above, we obtain

$$E\|E(\zeta_{n+r} \mid \mathscr{F}_n)\| \leqslant K_5 \Big(\sum_{j=n+r/2}^{n+r-2} c_j \gamma_{n+r-j-1} \sum_{i=j/2}^{j-1} c_i \gamma_{j-i} + \gamma_{r/2} \Big) + \mathcal{O}\big((n+r)^{-2}\big),$$

and hence

$$E\|E(\zeta_{n+r} \mid \mathscr{F}_n)\| \leqslant \mathcal{O}\big((n+r/2)^{-2}\big) + K_5 \gamma_{r/2} + \mathcal{O}\big((n+r)^{-2}\big),$$

and

$$\|E(\zeta_{n+r} \mid \mathscr{F}_n)\|_1 = \gamma_r^1 \quad \text{with } \gamma_r^1 = \mathcal{O}(r^{-2}).$$

Consequently, (ζ_n) is γ-weakly dependent, and since (20) is satisfied, the ODE method may be used and Y_n converges to 0 a.s.

Now we prove that Y_n is a.s. bounded. Let $V(Y) = Y^T C Y = \|\sqrt{C} Y\|^2$. Since C is not singular, V is a Lyapunov function and $\nabla V(Y) = 2CY$ is a Lipschitz function, so we have

$$V(Y_{n+1}) \leqslant V(Y_n) + (Y_{n+1} - Y_n)^T \nabla V(Y_n) + K_6 \|Y_{n+1} - Y_n\|^2.$$

Furthermore,

$$\|Y_{n+1} - Y_n\|^2 \leqslant 2c_n^2 (\|y_{n+1} x_{n+1} - C_{n+1} C^{-1} U\|^2 + 2c_n^2 \|C_{n+1} Y_n\|^2).$$

Since (y_n, x_n) is bounded, (C_n) and $\|y_{n+1} x_{n+1} - C_{n+1} C^{-1} U\|^2$ are also bounded. Moreover,

$$\|C_{n+1} Y_n\|^2 \leqslant K_7 \|Y_n\|^2 \leqslant \frac{K_7}{\lambda_{\min}(C)} V(Y),$$

where $\lambda_{\min}(C)$ is the smallest eigenvalue of C. Consequently,

$$V(Y_{n+1}) \leqslant V(Y_n)(1 + K_8 c_n^2) + K_9 c_n^2 + 2(Y_{n+1} - Y_n)^T C Y_n.$$

The last term becomes

$$\begin{aligned}2(Y_{n+1}-Y_n)^T CY_n &= -2c_n\|CY_n\|^2+2c_n(y_{n+1}x_{n+1}-C_{n+1}C^{-1}U)^T CY_n\\&\quad +2c_n Y_n^T(C-C_{n+1})CY_n\\&\leqslant -2c_n\|CY_n\|^2+c_n K_{10}\|CY_n\|+2c_n Y_n^T(C-C_{n+1})CY_n\\&\leqslant -2c_n\|CY_n\|^2+c_n K_{10}\|CY_n\|+2c_n u_{n+1}V(Y_n),\end{aligned}$$

where $u_n = \max\{X_i^T(C-C_n)X_i,\ 1\leqslant i\leqslant d\}$ and $\{X_1, \ldots, X_d\}$ is an orthogonal basis of unit eigenvectors of C.

We now obtain

$$V(Y_{n+1})\leqslant V(Y_n)(1+K_8 c_n^2+2c_n u_{n+1})+K_9 c_n^2-c_n(2\|CY_n\|^2-K_{10}\|CY_n\|). \tag{31}$$

Note that under the assumption **A-lr** (u_n) is a γ-weakly dependent sequence with a weakly dependent coefficient $\gamma_r = \mathcal{O}(a^r)$ and $\sum_{n=0}^{\infty} c_n u_{n+1} < \infty$. Moreover, if $V(Y_n)\geqslant K_{10}^2/4\lambda_{\min}(C)$, we have

$$\|CY_n\|\geqslant K_{10}/2 \quad\text{and}\quad -(2\|CY_n\|^2-K_{10}\|CY_n\|)\leqslant 0.$$

Let us put

$$T=\inf\left\{n \mid V(Y_n)\leqslant\frac{K_{10}^2}{4\lambda_{\min}(C)}\right\}.$$

By the Robbins–Sigmund theorem, $V(Y_n)$ converges a.s. to a finite limit on $\{T=+\infty\}$, so (Y_n) is bounded since V is a Lyapunov function.

On $\{\liminf_n V(Y_n)\leqslant K_{10}^2/4\lambda_{\min}(C)\}$, $V(Y_n)$ does not converge to ∞ and, using Delyon [10], Theorem 2, we deduce that $V(Y_n)$ converges to a finite limit if

$$\forall k>0,\ \sum c_n^2\|h(Y_n)+\xi_{n+1}\|^2\,\mathbf{1}_{\{V(Y_n)<k\}}<\infty, \tag{32}$$

$$\forall k>0,\ \sum c_n\langle\xi_{n+1},\ \nabla V(Z_n)\rangle\,\mathbf{1}_{\{V(Y_n)<k\}}<\infty. \tag{33}$$

Using the relation $\sum c_n^2<\infty$ and the fact that on $\{V(Y_n)<k\}$ the function $\|h(Y_n)+\xi_{n+1}\|^2$ is bounded, we deduce (32). To prove (33) it is enough, by Proposition 3.1, to show that $\langle\xi_{n+1},\ \nabla V(Y_n)\rangle\,\mathbf{1}_{\{V(Y_n)<k\}}=e_{n+1}$ is a γ-weakly dependent sequence with dependent coefficient which satisfies (20). But to use the result of Proposition 3.1 it is necessary to center e_{n+1}. So we are going to prove that $\sum c_n Ee_{n+1}<\infty$ and that $(e_{n+1}-Ee_{n+1})$ is a γ-weakly dependent sequence with a dependent coefficient γ_r^2 which is $\mathcal{O}(r^{-2})$.

Study of $E(e_{n+1})$. First of all, we must note a few elements. Denoting by I the unit matrix of $\boldsymbol{R}^d$, we obtain

$$Y_n=(I-c_{n-1}C_n)Y_{n-1}+c_{n-1}(x_n y_n-C_n C^{-1}U).$$

Let $\lambda_{\max}(C_n)$ be the largest eigenvalue of C_n. Then note that $\lambda_{\max}(C_n)=\|x_n\|^2\leqslant M$.

And for n large enough $c_{n-1} M < 1$ and $(I - c_{n-1} C_n)$ is not singular. Consequently, if $M_1 = \sup_n \{x_n y_n - C_n C^{-1} U\}$, we obtain

$$\|Y_{n-1}\| \leqslant \frac{1}{1-c_{n-1}\lambda}(\|Y_n\| + c_{n-1} M_1) \leqslant (1+bc_{n-1})(\|Y_n\| \wedge M_1),$$

where b is some non-negative constant, not depending on n. Moreover,

$$V(Y_n) < k \Rightarrow \|Y_n\|^2 < \frac{k}{\lambda_{\min}(C)}$$

and

$$\|Y_n\| < k' \Rightarrow V(Y_n) < \lambda_{\max}(C) k'^2,$$

which implies

$$\mathbf{1}_{\{V(Y_n)<k\}} = \mathbf{1}_{\{\|Y_n\|<k_n\}} = \mathbf{1}_{\{\|Y_{n-j}\|<k_{n-j}\}},$$

where

$$k_{n-j} \leqslant (1+c_{n-1})^j \left(\sqrt{\frac{k}{\lambda_{\min}(C)}} \wedge M_1 \right).$$

Moreover, since $c_n = g/n$, for any $0 \leqslant j \leqslant n$ it follows that $(1+ac_{n-1})^j$ is bounded independently of n, so is k_{n-j}. Consequently,

$$E(e_{n+1}) = E(x_{n+1} y_{n+1} - C_{n+1} C^{-1} U)^T C Y_n \mathbf{1}_{\{V(Y_n)<k\}} + E Y_n^T (C - C_{n+1}) C Y_n \mathbf{1}_{\{V(Y_n)<k\}}.$$

We have

$$\begin{aligned}
E(e_{n+1}) = & \sum_{j=n/2}^{n-1} E(y_{n+1} x_{n+1} - C_{n+1} C^{-1} U)^T C (Y_{j+1} - Y_j) \mathbf{1}_{\{\|Y_j\|<k_j\}} \\
& + E(y_{n+1} x_{n+1} - C_{n+1} C^{-1} U)^T C Z_{n/2} \\
& + \sum_{j=n/2}^{n-1} E(Y_{j+1} - Y_j)^T (C - C_{n+1}) C (Y_{j+1} - Y_j) \mathbf{1}_{\{\|Y_j\|<k_j\}} \\
& + 2 \sum_{j=n/2}^{n-1} \sum_{i=j+1}^{n-2} E(Y_{j+1} - Y_j)^T (C - C_{n+1}) C (Y_{i+1} - Y_i) \mathbf{1}_{\{\|Y_j\|<k_j\} \cap \{\|Y_i\| \leqslant k_i\}} \\
& + 2 \sum_{j=n/2}^{n-1} E Y_{n/2}^T (C - C_{n+1}) C (Y_{j+1} - Y_j) \mathbf{1}_{\{\|Y_j\|<k_j\} \cap \{\|Y_{n/2}\|<k_{n/2}\}} \\
& + E Y_{n/2}^T (C - C_{n+1}) C Y_{n/2} \mathbf{1}_{\{\|Y_{n/2}\|<k_{n/2}\}}.
\end{aligned}$$

Note that if $n/2$ is not an integer, we replace it by $(n-1)/2$. Using always the same technique, we obtain

$$Ee_{n+1} = \mathcal{O}(n^{-2}) + \mathcal{O}(a^{n/2}) + \mathcal{O}(n^{-2}) + \mathcal{O}(n^{-2}) + \mathcal{O}(a^{n/2}),$$

and hence $\sum c_n E e_n < \infty$.

Study of $(e_n - Ee_n)$. We now prove that this sequence is γ-weakly dependent with a relevant dependent coefficient. Write

$$E(e_{n+r} - Ee_{n+r} \mid \mathscr{F}_n) = D_{n+r} + G_{n+r} - Ee_{n+r}$$

with

$$D_{n+r} = E\left[(y_{n+r} x_{n+r} - C_{n+1} C^{-1} U)^T C Y_{n+r-1} \mathbf{1}_{\{V(Y_{n+r-1}) < k\}} \mid \mathscr{F}_n\right],$$

$$G_{n+r} = E\left[Y_{n+r-1}^T (C - C_{n+r}) C Y_{n+r-1} \mathbf{1}_{\{V(Y_{n+r-1}) < k\}} \mid \mathscr{F}_n\right],$$

$$D_{n+r} = \sum_{j=n+r/2}^{n+r-2} E\left[(y_{n+r} x_{n+r} - C_{n+1} C^{-1} U)^T C (Y_{j+1} - Y_j) \mathbf{1}_{\{\|Y_j\| < k_j\}} \mid \mathscr{F}_n\right]$$
$$+ E\left[(y_{n+r} x_{n+r} - C_{n+1} C^{-1} U)^T C Y_{n+r/2} \mathbf{1}_{\{\|Y_{n+r/2}\| < k_{n+r/2}\}} \mid \mathscr{F}_n\right].$$

Here again, if $r/2$ is not an integer, we replace it by $(r-1)/2$. Again, the same techniques as for r_n give

$$E\|D_{n+r}\| = \mathcal{O}\left((n+r)^{-2}\right) + \mathcal{O}\left(a^{n+r/2}\right).$$

We study G_{n+r} in the same way and $E\|G_{n+r}\| = \mathcal{O}\left((n+r)^{-2}\right)$, and since $Ee_{n+r} = \mathcal{O}\left((n+r)^{-2}\right)$, (20) is satisfied and the result is proved.

Proof of (27). For $n > N$, let us put

$$\Pi_n^N = (I - c_n C_{n+1}) \ldots (I - c_N C_{N+1}).$$

Since $g < 1/2M$, for $N \geqslant 1$ it follows that Π_n^N is not singular and we have

$$Y_{n+1} = \Pi_n^N Y_N + \sum_{j=N}^{n} c_j \Pi_n^N (\Pi_j^N)^{-1} \zeta_{j+1}^1,$$

where $\zeta_{j+1}^1 = y_{j+1} x_{j+1} - C_{j+1} C^{-1} U$. And since $Y_n \to 0$, we get

$$-Y_N = \sum_{j=N}^{\infty} c_j (\Pi_j^N)^{-1} \zeta_{j+1}^1,$$

$$(\Pi_j^N)^{-1} = (I - c_N C_{N+1})^{-1} \ldots (I - c_j C_{j+1})^{-1},$$

and

$$\|(\Pi_j^N)^{-1}\| \leqslant \frac{1}{\prod_{i=N}^{j} (1 - c_i M)}.$$

Hence

$$\|(\Pi_j^N)^{-1}\| = \mathcal{O}\left(\exp\left(M \sum_{i=N}^{j} c_i\right)\right) = \mathcal{O}\left((j/N)^{gM}\right), \tag{34}$$

$$\|\sqrt{N} Y_N\| = \left\| \sum_{j=N}^{\infty} \frac{g}{\sqrt{j}} \sqrt{\frac{N}{j}} (\Pi_j^N)^{-1} \zeta_{j+1}^1 \right\|.$$

Since $g < 1/2M$, (34) involves that the sum converges. Indeed, (20) is satisfied with $k = 5$ and $p = 3$ (so $m = \frac{1}{3}$), and since ζ_{j+1}^1 is γ-weakly dependent with a mixingale coefficient $\gamma_r = \mathcal{O}(a^r)$. Hence the result is proved. ■

7. PROOFS

7.1. Proof of Proposition 2.1. We use a sketch similar to Doukhan and Louhichi's proof in [12]. Therefore we get

$$E\Big(\sum_{i=1}^{n} c_i \zeta_i\Big)^p \leqslant p! \sum_{1 \leqslant t_1 \leqslant \ldots \leqslant t_p \leqslant n} c_{t_1} \ldots c_{t_p} |E(\zeta_{t_1} \ldots \zeta_{t_p})|. \tag{35}$$

Let us put

$$A_p(n) = \sum_{1 \leqslant t_1 \leqslant \ldots \leqslant t_p \leqslant n} c_{t_1} \ldots c_{t_p} |E(\zeta_{t_1} \ldots \zeta_{t_p})|,$$

so for any $t_2 \leqslant t_m \leqslant t_{p-1}$

$$A_p(n) \leqslant \sum_{1 \leqslant t_1 \leqslant \ldots \leqslant t_p \leqslant n} c_{t_1} \ldots c_{t_p} |E(\zeta_{t_1} \ldots \zeta_{t_m}) E(\zeta_{t_{m+1}} \ldots \zeta_{t_p})|$$

$$+ \sum_{1 \leqslant t_1 \leqslant \ldots \leqslant t_p \leqslant n} c_{t_1} \ldots c_{t_p} |\operatorname{cov}(\zeta_{t_1} \ldots \zeta_{t_m}, \zeta_{t_{m+1}} \ldots \zeta_{t_p})|.$$

Let us write

$$A_p^1(n) = \sum_{1 \leqslant t_1 \leqslant \ldots \leqslant t_p \leqslant n} c_{t_1} \ldots c_{t_p} |E(\zeta_{t_1} \ldots \zeta_{t_m}) E(\zeta_{t_{m+1}} \ldots \zeta_{t_p})|,$$

$$A_p^2(n) = \sum_{1 \leqslant t_1 \leqslant \ldots \leqslant t_p \leqslant n} c_{t_1} \ldots c_{t_p} |\operatorname{cov}(\zeta_{t_1} \ldots \zeta_{t_m}, \zeta_{t_{m+1}} \ldots \zeta_{t_p})|.$$

Since the sequence (c_n) is decreasing to 0, we deduce, as in [12],

$$A_p^1(n) \leqslant A_m(n) A_{p-m}(n). \tag{36}$$

By (9) we obtain

$$A_p^2(n) \leqslant \sum_{t_1=1}^{n} c_{t_1}^p \sum_{r=0}^{n-1} C_p p^\gamma M^{p-2} (r+1)^{p-2} \varepsilon_r,$$

and the expression $\sum_{i=1}^n c_i^p \sum_{r=0}^{n-1} C_p p^\gamma M^{p-2}(r+1)^{p-2} \varepsilon_r = V_p(n)$ satisfies, for any integer $2 \leqslant q \leqslant p-1$, the following:

$$V_q(n) \leqslant V_p^{(q-2)/(p-2)}(n)\, V_2^{(p-q)/(p-2)}(n).$$

Now, Lemma 12 of [12] leads to

$$A_p(n) \leqslant \frac{1}{p} \binom{2p-2}{p-1} \big(V_2^{p/2}(n) \vee V_p(n)\big),$$

and hence

$$E\Big(\sum_{i=1}^{n} c_i \zeta_i\Big)^p \leqslant \frac{(2p-2)!}{(p-1)!}\big(V_2^{p/2}(n) \vee V_p(n)\big).$$

This ensures the result. ∎

7.2. Proof of Proposition 2.2. Using the same notation as in the previous proof, by (10) we get

$$V_p(n) \leqslant M_p \sum_{i=1}^{n} c_i^p \int_0^1 \min\big(\varepsilon^{-1}(u), n\big)^{p-1} Q_i^p(u)\,du,$$

where $\varepsilon(u) = \varepsilon_{[u]}$ ($[u]$ denotes the integer part of u). Let us write

$$W_p(n) = M_p \sum_{i=1}^{n} c_i^p \int_0^1 \min\big(\varepsilon^{-1}(u), n\big)^{p-1} Q_i^p(u)\,du.$$

If (14) is satisfied, we obtain

$$W_q(n) \leqslant W_p(n)^{(q-2)/(p-2)}(n)\, W_2^{(p-q)/(p-2)}(n).$$

Thus we can conclude as in the previous proof. ∎

7.3. Proof of Proposition 3.1. Proceeding as in [9], we deduce that

$$|E(\Sigma_n^p)| \leqslant \Big(2p \sum_{i=1}^{n} b_{i,n}\Big)^{p/2},$$

where

$$b_{i,n} = \max_{i \leqslant l \leqslant n} \Big\|c_i \zeta_i \sum_{j=0}^{l-i} E(c_{i+j}\zeta_{i+j} \,|\, \mathscr{F}_i)\Big\|_{p/2}.$$

Let $q = p/(p-2)$; then there exists Y such that $\|Y\|_q = 1$. Applying Proposition 1 of [9], we obtain

$$b_{i,n} \leqslant \sum_{j=0}^{n-i} c_i c_{i+j} \int_0^{\gamma_j} Q_{\{Y\zeta_i\}} \circ G_{\{\zeta_{i+j}\}}(u)\,du,$$

where G_X is the inverse of $x \to \int_0^x Q_X(u)\,du$.

The Fréchet inequality [14] yields

$$b_{i,n} \leqslant \sum_{j=0}^{n-i} c_i c_{i+j} \int_0^1 \mathbf{1}_{\{u \leqslant G(\gamma_j)\}} Q^2(u)\, Q_Y(u)\,du,$$

where $Q = Q_{\zeta_i}$. Using Hölder's inequality, we also obtain

$$b_{i,n} \leqslant c_i \sum_{j=0}^{n-i} c_{i+j} \Big(\int_1^1 \mathbf{1}_{\{u \leqslant G(\gamma_j)\}} Q^p(u)\,du\Big)^{2/p}.$$

By (15), $Q(u) \leqslant c^{1/k} u^{-1/k}$, and setting $K = (k-1)/kc^{1/k}$ yields

$$b_{i,n} \leqslant c_i \sum_{j=0}^{n-i} c_{i+j} \Big(\int_0^1 \mathbf{1}_{\{u \leqslant G(\gamma_j)\}} c^{p/k} u^{-p/k} du\Big)^{2/p} \leqslant c_i \sum_{j=0}^{n-i} c_{i+j} \left(K \frac{\gamma_j}{c_{i+j}}\right)^{k(k-1)^{-1}(1-p/k)\cdot 2p^{-1}}$$

The result follows with $K_1 = K^{2(k-p)/p(k-1)}$. ∎

8. APPENDIX

Proof of the Note in Section 6. This note claims that if $(y_n, x_n)_{n \in \boldsymbol{N}}$ is θ-weakly dependent in the Dedecker and Doukhan sense [9], then **A-lr** is satisfied. Let us remind the definition of a θ-weakly dependent $\boldsymbol{R}^d$-valued sequence which is used in [9]:

If $\Lambda^{(1)}$ is a space of bounded 1-Lipschitz real-valued functions defined on $\boldsymbol{R}^d$, (X_n) is *θ-weakly dependent* if

$$\theta_r = \sup_{n \geqslant 0} \{ \sup_{f \in \Lambda^{(1)}} (\|\boldsymbol{E}[f(X_{r+n}) \mid \sigma(X_i, i \leqslant n)] - \boldsymbol{E}[f(X_{k+n})]\|_1)\}$$

tends to zero as r tends to infinity.

For any $f \in \Lambda^{(1)}$, $|f(x) - f(y)| \leqslant |x^1 - y^1| + \ldots + |x^d - y^d|$, where the x^j's $(j = 1, \ldots, d)$ are the components of x.

First, note that if an $\boldsymbol{R}^d$-valued sequence (X_n) is θ-weakly dependent, any $\boldsymbol{R}^j$-valued $(j = 1, \ldots, d-1)$ sequence $(Y_n) = (X_n^{t_1}, \ldots, X_n^{t_j})$ is θ-weakly dependent. So, if (y_n, x_n) is θ-weakly dependent, then so are (y_n) and (x_n^j) $(j = 1, \ldots, d)$.

Let f be a bounded 1-Lipschitz function, defined on $\boldsymbol{R}$, and g the function defined on $\boldsymbol{R}^2$ by $g(x, y) = f(xy)$. It is enough to prove that g is a Lipschitz function defined on $\boldsymbol{R}^2$. Indeed,

$$\frac{|g(x, y) - g(x', y')|}{|x - x'| + |y - y'|} \leqslant \frac{|xy - x'y'|}{|x - x'| + |y - y'|} \leqslant \frac{|x|\,|y - y'| + |y'|\,|x - x'|}{|x - x'| + |y - y'|} \leqslant \max(|x|, |y'|),$$

and g is Lipschitz if x and y are bounded.

Thus, since (x_n) and (y_n) are bounded, the result follows. ∎

REFERENCES

[1] M. Benaïm, *A dynamical system approach to stochastic approximation*, SIAM J. Control. Optim. 34 (2) (1996), pp. 437–472.

[2] M. Benaïm and M. W. Hirsch, *Dynamics of Morse–Smale urn processes*, Ergodic Theory Dynamical Systems 15 (1995), pp. 1005–1030.

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

[4] P. Billingsley, *Probability and Measure*, 2nd edition, Wiley, 1985.

[5] H.-F. Chen, *Stochastic Approximation and Its Applications*, Nonconvex Optimization and Its Application, Vol. 64, Kluwer Academic Publishers, 1985.

[6] Y. S. Chow, *Some convergence theorems for independent random variables*, Ann. Math. Statist. 36 (4) (1966), pp. 1293–1314; ibidem 37 (1966), pp. 1482–1493.

[7] C. Coulon-Prieur and P. Doukhan, *A triangular CLT for weakly dependent sequences*, Statist. Probab. Lett. 47 (2000), pp. 61–68.

[8] R. M. De Jong, *A strong law of large numbers for triangular mixingale arrays*, Statist. Probab. Lett. 27 (1996), pp. 1–9.

[9] P. Dedecker and P. Doukhan, *A new covariance inequality and applications*, Stochastic Process. Appl. 106 (1) (2003), pp. 63–80.

[10] B. Delyon, *General convergence result on stochastic approximation*, IEEE Trans. Automat. Control. 41 (1996), p. 9.

[11] P. Doukhan, *Models inequalities and limit theorems for stationary sequences*, in: *Theory and Applications of Long Range Dependence*, P. Doukhan et al. (Eds.), Birkhäuser, 2002, pp. 43–101.

[12] P. Doukhan and S. Louhichi, *A new weak dependence condition and applications to moment inequalities*, Stochastic Process. Appl. 84 (1999), pp. 313–342.

[13] M. Duflo, *Algorithmes stochastiques*, Collect. Math. Appl. 23, Springer, 1996.

[14] M. Fréchet, *Généralisation d'une inégalité de Minkowski*, Bull. Inst. Internat. Statist. 35 (2) (1957), pp. 99–103.

[15] S. Ghosal and T. K. Chandra, *Complete convergence of martingale arrays*, J. Theoret. Probab. 11 (3) (1998), pp. 621–631.

[16] J. Hoffmann-Jørgensen, *Sums of independent Banach space valued random variables*, Studia Math. 52 (1974), pp. 159–186.

[17] P. L. Hsu and H. Robbins, *Complete convergence and the strong law of large numbers*, Proc. Nat. Acad. Sci. 33 (1947), pp. 25–31.

[18] G. D. Kersting, *Some results on the asymptotic behavior of the Robbins–Monro procedure*, Bull. Int. Statist. Inst. 47 (1977), pp. 327–335.

[19] H. J. Kushner and D. S. Clark, *Stochastic approximation for constrained and unconstrained systems*, Appl. Math. Sci. 26, Springer, 1978.

[20] D. Li, M. B. Rao, T. Jiang and X. Wang, *Complete convergence and almost sure convergence of weighted sums of random variables*, J. Theoret. Probab. 8 (1995), pp. 49–76.

[21] D. L. Mcleish, *A maximal inequality and dependent strong laws*, Ann. Probab. 3 (1975), pp. 829–839.

[22] M. Peligrad and S. Utev, *Central limit theorem for linear processes*, Ann. Probab. 25 (3) (1997), pp. 443–456.

[23] E. G. Shixin, *On the convergence of weighted sums of L_q-mixingale arrays*, Acta Math. Hungar. 82 (1–2) (1999), pp. 113–120.

[24] K. F. Yu, *Complete convergence of weighted sums of martingale differences*, J. Theoret. Probab. 3 (1990), pp. 339–347.

Paul Doukhan
ENSAE and Laboratory of Statistics
CREST Timbre J340
3, avenue Pierre Larousse
92240 Malakoff, France
E-mail: doukhan@ensae.fr

Odile Brandière
Université de Paris Sud 11
Laboratoire de Mathématiques
Bât. 425
91405 Orsay, France
E-mail: odile.brandiere@math.u-psud.fr

Received on 14.10.2003;
revised version on 14.11.2004