INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2003-10

Dependent Noise for Stochastic Algorithms

P. DOUKHAN[1]
O. BRANDIERE[2]

[1] CREST, Laboratoire de Statistique.
[2] Université Paris 11.

Dependent noise for stochastic algorithms

Paul Doukhan and Odile Brandière
LS-CREST and Université Paris 11

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 the large number for triangular arrays of weighted dependent random variables are also given.

Algorithmes stochastiques à bruit dépendant

Résumé

La dépendance du bruit d'un algorithme stochastique est modélisée de différentes manières, de sorte que la méthode de l'équation différentielle ordinaire reste applicable. Ces techniques de dépendance faible sont illustrées par des applications à un algorithme de régression linéaire et à l'étude de tableaux triangulaires de variables aléatoires pondérées dépendantes.

Key words. Stochastic approximation, ordinary differential equations, dependent noise, linear regression
AMS subject classifications : 60F99 , 60G10, 60G48, 62J05, 62L20.

1 Introduction

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

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

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

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

• (η_n) is a "small" stochastic disturbance.

The ordinary differential equation (ODE) method (see [3], [12], [17], ...) associates the possible limit sets of (1) with the properties of the associated ODE

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

If this algorithm has bounded sample paths, then these sets are compact connected invariant and "chain-recurrent" in the Benaïm sense for the ODE (cf. [2]). These sets are more or less complicated. Various situations may then happen, for example the most simple 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 ...

One often assumes the following assumptions on the "small disturbance"

$$\eta_{n+1} = c_n(\varepsilon_{n+1} + r_{n+1}), \tag{4}$$

where (c_n) denotes a nonnegative deterministic sequence such that

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

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

$$E(\varepsilon_{n+1}|\mathcal{F}_n) = 0 \qquad \forall n \geq 0, \qquad \text{and} \tag{6}$$

$$r_n \to_{n\to\infty} 0 \qquad a.s. \tag{7}$$

The aim of this paper is to consider algorithms satisfying each of those assumptions except for (6). We substitute it by some dependence conditions. In this framework, we shall be able to apply the ODE method as soon as

$$\sum c_n \varepsilon_{n+1} \text{ converges a.s.} \tag{8}$$

The paper is devoted to sufficient conditions for (8). Section 2 considers the weak dependence condition from Doukhan & Louhichi in [11]; set for this

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

$$\mathcal{L} = \{h/ \text{ such that } h : \mathbb{R}^u \to \mathbb{R}, \text{ for some } u \geq 0^*, \text{Lip }(h) + \|h\|_\infty < \infty\},$$

and consider some function from $C : \mathbb{N}^{\star 2} \to \mathbb{R}$. The sequence (ε_n) is said to be $(\theta, \mathcal{L}, C)$-weakly dependent if there exists a sequence $\theta = (\theta_r)_{r\geq 0}$ such that $\theta_r \downarrow 0$ as $n \uparrow \infty$ and statisfying, for any $(u+v)$-tuple $(t_1,\ldots,t_{u+v})$ with $t_1 \leq \cdots \leq t_u < t_u + r \leq t_{u+1} \leq \cdots \leq t_{u+v}$,

$$\left|\text{Cov}\left(h\left(\varepsilon_{t_1},\ldots\varepsilon_{t_u}\right), k\left(\varepsilon_{t_{u+1}},\ldots,\varepsilon_{t_{u+v}}\right)\right)\right| \leq C(u,v)(\text{Lip}(h)+\text{Lip}(k))\theta_r. \tag{9}$$

Various examples of this situation may be found in [10]; they include

- general Bernoulli shifts, $\varepsilon_t = \sum_{\ell=1}^{\infty} \sum_{k_1,\ldots,k_\ell} a^{(\ell)}_{k_1,\ldots,k_\ell} \eta_{t-k_1} \cdots \eta_{t-k_\ell}$,
- stable Markov chains such as, $\varepsilon_t = G(\varepsilon_{t-1},\ldots,\varepsilon_{t-p}) + \eta_t$,
- or ARCH(∞) models, $\varepsilon_t = \left(a_0 + \sum_{j\geq 1} a_j \varepsilon_{t-j}\right)\eta_t$

generated by some i.i.d. sequence (η_t). In the first example, the situation of an infinite moving average, for which $\ell = 1$, is of a special interest and $\theta_r \leq 4E|\eta_0|\sum_{2|k|\geq r}|a^{(1)}_k|$. Now $\theta_r \downarrow 0$ (geometrically) in the second case, if $|G(x_1,\ldots,x_p) - G(y_1,\ldots,y_p)| \leq \sum_j b_j|x_j - y_j|$ with $\sum_j a_j < 1$ and $E|\eta_0| < \infty|$. In the last, non-Markov and non-linear, example a chaotic expansion holds if $\sum_{j\geq 1}|a_j|E|\eta_0| < 1$ and then any class of rate may be obtained for θ_r. Note that r always denotes the gap in time between"past" and "future". A generalization to the vector $\mathbb{R}^d$-case is also provided below.

Section 3 considers a weakly dependent noise in the sense of the γ-weak coefficients in Dedecker & Doukhan, [8]. The mixingale-type coefficients, defined for the sequence $(\varepsilon_n)_{n\geq 0}$, used there is defined as $\gamma_r = \sup_{k\geq 0}||E(\varepsilon_{k+r}|\sigma(\varepsilon_i, i \leq k)) - E(\varepsilon_{k+r})||_1$. The sequence (ε_n) is said to be γ-weakly dependent if $\gamma_r \to 0$ as $r \to \infty$. In [8], this is proved that a causal version of $(\theta, \mathcal{L})$-weak dependence implies γ-weak dependence, where le right left hand side in eqn. (9) writes $\leq C(v)\text{Lip}(k)\theta_r$ only depends on k. Counter-examples of γ-weakly dependent sequences which are not $(\theta, \mathcal{L})$-weakly dependent may also be found there. We first settle an immediate **extension** of this notion to $\mathbb{R}^d$ valued sequences. The definition of γ-weak dependence extends to $\mathbb{R}^d$:

Proposition 1 *The two following assertions are equivalent :*
(i) A $\mathbb{R}^d$-valued sequence (X_n) is γ-weakly dependent,
(ii) Each component (X^ℓ_n) $(\ell = 1,\ldots,d)$ of (X_n) is γ-weakly dependent.

Proof. Clearly, $\|E(X^\ell_{n+r} - E(X^\ell_{n+r})|\mathcal{F}_n)\|_1 \le \|E(X_{n+r} - E(X_{n+r})|\mathcal{F}_n)\|_1$ and (i) implies (ii). On another hand,

$\|E(X_{n+r} - E(X_{n+r})|\mathcal{F}_n)\|_1 = E\sqrt{\sum_{\ell=1}^d (E(X^\ell_{n+r} - E(X^\ell_{n+r})|\mathcal{F}_n))^2}$ hence
$\|E(X_{n+r} - E(X_{n+r})|\mathcal{F}_n)\|_1 \le \sqrt{d}\max_{1\le\ell\le d}\theta^\ell_r$. ∎

In the frame of the $(\theta,\mathcal{L})$-weak dependence we say that the $\mathbb{R}^d$-valued sequence (X_n) is $(\theta,\mathcal{L})$-weakly dependent if each component (X^ℓ_n) is $(\theta^\ell,\mathcal{L})$-weakly dependent.

The two forthcoming sections are devoted to provide moment inequalities of the Marcinkiewicz-Zygmund type adapted to deduce the relation (8) in those two frames. The last section 4 is devoted to apply the study to the examples of Robbins-Monro algorithm and to strong laws of the large numbers for triangular arrays.

2 Weakly dependent noise

Let (ε_n) be a sequence of centered random variables. satisfying a $(\theta,\mathcal{L})$-weak dependence as described in eqn (9). We denote by S_n, the sum $\sum_{i=1}^n \varepsilon_i$ and $C_q = \max_{u+v\le q} C(u,v)$,

$$\sup|\mathrm{Cov}(\varepsilon_{t_1}\cdots\varepsilon_{t_m},\varepsilon_{t_{m+1}}\cdots\varepsilon_{t_q})| \le C_q q^\gamma M^{q-2}\theta_r. \tag{10}$$

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

$$|\mathrm{Cov}(\varepsilon_{t_1}\cdots\varepsilon_{t_m},\varepsilon_{t_{m+1}}\cdots\varepsilon_{t_q})| \le M_q\int_0^{\min(\theta_r,1)} Q_{\varepsilon_{t_1}}(x)\cdots Q_{\varepsilon_{t_q}}(x)dx. \tag{11}$$

where we denote by Q_X 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 (10) is mainly suitable for bounded sequences while (11) holds for more general r.v.s, using moment or tail assumptions. Various examples for which one of these two bounds holds in [11]. Moreover, let p be some fixed integer not less than 2.

If (10) holds for all $q\le p$, then, for any $n\ge 2$

$$|ES_n^p| \le \frac{(2p-2)!}{(p-1)!}p^\gamma\left\{\left(C_2 n\sum_{r=0}^{n-1}\theta_r\right)^{p/2}\vee\left(C_pM^{p-2}n\sum_{r=0}^{n-1}(r+1)^{p-2}\theta_r\right)\right\}. \tag{12}$$

If, now, (11) holds for all $q\le p$, then, for any $n\ge 2$

$$|ES_n^p| \le \frac{(2p-2)!}{(p-1)!}\left\{\left(C_p\sum_{i=1}^n\int_0^1[\min(\theta^{-1}(u),n)]^{p-1}Q^p_{\varepsilon_i}(u)du\right)\right. \tag{13}$$

$$\vee \left(\left(C_2 \sum_{i=1}^{n} \int_0^1 [\min(\theta^{-1}(u), n)]^{p-1} Q^p_{\varepsilon_i}(u) du \right)^{p/2} \right) \Bigg\}.$$

Hence, if $\Sigma_n = \sum_1^n c_{n-1}\varepsilon_n$, and

$$\sup |\mathrm{Cov}(\varepsilon_{t_1} \cdots \varepsilon_{t_m}, \varepsilon_{t_{m+1}} \cdots \varepsilon_{t_q})| =: C_{rq}(t_1),$$

where the supremum is considered, for $t_1 \geq 1$ fixed, over all $\{t_2, \cdots, t_q\}$ such that $t_1 \leq \cdots \leq t_m < t_{m+1} \cdots \leq t_q$ and $r = t_{m+1} - t_m$.

Using similar techniques as in [11], we derive the following result

Proposition 2 *Let $p \geq 2$ be some fixed integer and let (ε_n) be a centered $(\theta, \mathcal{L})$-weakly dependent sequence of real random variables. Assume, for all $1 \leq i \leq n$ and $2 \leq q \leq p$, that*

$$C_{rq}(i) \leq C_{rp}^{\frac{q-2}{p-2}}(i) C_{r2}^{\frac{p-q}{p-2}}(i). \tag{14}$$

Then for $n \geq 2$,

$$\begin{aligned} |E\Sigma_n^p| \quad &\leq \quad \frac{(2p-2)!}{(p-1)!} \left\{ \left(\sum_{i=1}^{n} c_{i-1}^p \sum_{r=0}^{n-1} C_{rp}(i)(r+1)^{p-2} \right) \right. && (15) \\ &\vee \quad \left. \left(\sum_{i=1}^{n} c_{i-1}^2 \sum_{r=0}^{n-1} C_{rp}(i) \right)^{p/2} \right\}. \end{aligned}$$

Note that (14) is satisfied as soon as (10) holds. And in this case, as in the most of examples in [11], we obtain :

$$\begin{aligned} |E\Sigma_n^p| \quad &\leq \quad \frac{(2p-2)!}{(p-1)!} \left\{ \left(C_p p^\gamma M^{p-2} \sum_{i=1}^{n} c_{i-1}^p \sum_{r=0}^{n-1} (r+1)^{p-2} \theta_r \right) \right. && (16) \\ &\vee \quad \left. \left(C_2 2^\gamma \sum_{i=1}^{n} c_{i-1}^2 \sum_{r=0}^{n-1} \theta_r \right)^{p/2} \right\}. \end{aligned}$$

This result is mainly adapted to the bounded sequence. The following result is appropriate to more general r.v.s but require a moment assumption.

Proposition 3 *Let p be some fixed integer not less than 2 and (ε_n) a centered $(\theta, \mathcal{L})$-weakly dependent sequence of r.v.s. Assume that for all $2 < q \leq p$, eqn. (11) holds and there exists a constant $c > 0$ with,*

$$M_q \quad \leq \quad M_p^{\frac{q-2}{p-2}} M_2^{\frac{p-q}{p-2}}, \qquad \textit{and,} \tag{17}$$

$$\exists \quad k > p, \quad \forall i \geq 0 : \qquad P(|\varepsilon_i| > t) \leq \frac{c}{t^k}. \tag{18}$$

Then for $n \geq 2$,

$$|E\Sigma_n^p| \leq \frac{(2p-2)!}{(p-1)!} \cdot c^{1/k} \left\{ \left(M_p \sum_{i=1}^{n} c_{i-1}^p \sum_{r=0}^{n-1} (r+1)^{p-2} \theta_r^{\frac{k-p}{k}} \right) \vee \left(M_2 \sum_{i=1}^{n} c_{i-1}^2 \sum_{r=0}^{n-1} \theta_r^{\frac{k-2}{k}} \right)^{p/2} \right\} \tag{19}$$

Note that (18) holds as soon as the ε_n's have a k-th order moment such that $\forall i \geq 0$ $E|\varepsilon_i|^k \leq c$.

Arguing as in Billingsley ([4]), if (15) holds for some p such that

$$\left(\sum_{i=1}^{\infty} c_{i-1}^p \sum_{r=0}^{\infty} C_{rp}(i)(r+1)^{p-2} \right) \vee \left(\sum_{i=1}^{\infty} c_{i-1}^2 \sum_{r=0}^{\infty} C_{rp}(i) \right)^{p/2} < \infty, \tag{20}$$

hence $\forall \varepsilon > 0$, $\lim_{n\to\infty} P(\sup_{k\geq 1} |\Sigma_{n+k} - \Sigma_n| > \varepsilon) = 0$, thus (Σ_n) is *a.s.* a Cauchy sequence, hence it converges. In the same way, if (19) holds for some p such that

$$\left(\sum_{i=1}^{n} c_{i-1}^p \sum_{r=0}^{\infty} (r+1)^{p-2} \theta_r^{\frac{k-p}{k}} \right) \vee \left(\sum_{i=1}^{n} c_{i-1}^2 \sum_{r=0}^{n-1} \theta_r^{\frac{k-2}{k}} \right)^{p/2} < \infty, \tag{21}$$

then, (Σ_n) converges with probability 1.

Equip $\mathbb{R}^d$ with its p-norm $\|(x_1, \ldots, x_d)\|_p^p = x_1^p + \cdots + x_d^p$. Let the sequence $(\varepsilon_n)_{n\geq 0}$ be an $\mathbb{R}^d$-valued and $(\theta, \mathcal{L})$- weakly dependent sequence. Set $\varepsilon_n = (\varepsilon_n^1, \ldots, \varepsilon_n^d)$ then $\|\sum_{i=1}^n c_i \varepsilon_i\|_p^p = \sum_{\ell=1}^d (\sum_{i=1}^n c_i \varepsilon_i^\ell)^p$. And if each component $(\varepsilon_n^\ell)_{n\geq 0}$ is $(\theta^\ell, \mathcal{L})$-weakly dependent and such that relations (20) or (21) hold, then $E\|\Sigma_n\|_p^p < \infty$. Arguing as before, we deduce that the sequence $(\Sigma_n)_{n\geq 0}$ converges with probability 1.

The Proofs of the propositions 2 and 3 is in section **7**.

3 $\gamma-$weakly dependent noise

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

Proposition 4 *Let $p > 2$ and $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence of centered random variables such that (18) holds. Then for any $n \geq 2$,*

$$|E\Sigma_n^p| \leq \left(2pK_1 \sum_{i=1}^{n} c_i^{2-\frac{2(k-p)}{p(k-1)}} \sum_{j=0}^{n-1} \gamma_j^{\frac{2(k-p)}{p(k-1)}} \right)^{p/2}, \tag{22}$$

where K_1 depends of r, p and c.

Quote that here $p \in \mathbb{R}$, and is not necessarily an integer. If, now, (22) holds for some p such that

$$\sum_{i=1}^{\infty} c_i^{2-m} \sum_{j=0}^{\infty} \gamma_j^m < \infty, \tag{23}$$

where $m = \frac{2(k-p)}{p(k-1)} < 1$, then (Σ_n) converges with probability 1. The proof of this proposition is in section **7**.

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

$$E\|\Sigma_n\|^p = E\sum_{\ell=1}^{d}\left(\sum_{i=1}^{n} c_i \varepsilon_i^{\ell}\right)^p,$$

and if each component $(\varepsilon_n^{\ell})_{n\geq 0}$ $(\ell = 1, \ldots, d)$ are γ^{ℓ}-weakly dependent and verifies (23), $E\|\Sigma_n\|^p < \infty$ and we conclude as before that $(\Sigma_n)_{n\geq 0}$ 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 is also used in mechanics, for adjustments, as well as in statistics, to fit a median ([12], page 50)... It writes

$$Z_{n+1} = Z_n - c_n(f(Z_n) - a) + c_n \varepsilon_{n+1}. \tag{24}$$

It is usually assumed that the prediction error (ε_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 convergence a.s. of this algorithm, under the usual assumptions and the conditions yielding the a.s. convergence of $\sum_0^n c_n \varepsilon_{n+1}$.
Under the assumptions of the proposition 2, if for some $p > 2$

$$\forall i \geq 0, \qquad \left(\sum_{r=0}^{\infty} C_{rp}(i)(r+1)^{p-2}\right) \vee \left(\sum_{r=0}^{\infty} C_{rp}(i)\right) < \infty, \tag{25}$$

the algorithm (24) converges a.s.

If the assumptions of proposition 3 hold, then the convergence a.s. of the algorithm (24) are ensured as soon as, for some $p > 2$,

$$\left(\sum_{r=0}^{\infty}(r+1)^{p-2}\theta_r^{\frac{k-p}{k}}\right) \vee \left(\sum_{r=0}^{\infty}\theta_r^{\frac{k-2}{k}}\right) < \infty.$$

Under the assumptions of proposition 4, as soon as (23) is satisfied, algorithm (24) converges with probability one.

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 $\mathcal{C}^2$ on an open set G of $\mathbb{R}^d$. The Kiefer-Wolfowiftz algorithm ([12], page 53) is stated as :

$$Z_{n+1} = Z_n - 2c_n \nabla V(Z_n) - \eta_{n+1} \tag{26}$$

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

Usually, the prediction error (ε_n) is assumed to 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 $\sum \frac{c_n}{b_n}\varepsilon_{n+1}$. The $(\theta, \mathcal{L})$-weak dependence assumptions are the same as for the Robbins-Monro algorithm. Concerning the γ-weak dependence, the condition (23) is replaced by

$$\sum_{i=1}^{\infty} \left(\frac{c_i}{b_i}\right)^{2-m} \sum_{i=1}^{\infty} \gamma_i^m < \infty.$$

5 Weighted weakly dependent variables triangular arrays

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

Theorem (Chow, [5]) *Let $\varepsilon_1, \ldots, \varepsilon_i, \ldots$ be independent and identically distributed random variables with $E(\varepsilon_i) = 0$ and $E|\varepsilon_i|^q < \infty$ for some $q \ge 2$. If for some real constant K, non depending on n, $\sum_{i=1}^n c_{ni}^2 \le K$, and $n^{1/q} \max_{1\le i\le n} |c_{ni}| \le K$, then,*

$$\forall \eta > 0 \qquad \sum_{n=1}^{\infty} P(n^{-1/q}|U_n| \ge \eta) < \infty.$$

The last inequality is a result of complete convergence of $n^{-1/q}|U_n|$ to 0. This notion was introduced by Hsu and Robbins [15]. Complete convergence implies the almost sure convergence from the Borel-Cantelli Lemma.

Li *et al.* [18] extend this result to arrays $(c_{ni})_{\{n\ge 1\ i\in Z\}}$ for $q = 2$. Quote also Yu, [22], who obtains a result analogue to Chow's for martingale differences. Ghosal and Chandra [13] extend the previous results and prove some similar results to these of Li *et al.* for martingales differences. As in [18], their main tool is the Hoffmann-Jorgensen inequality ([14]). Peligrad and Utev [20] propose a central limit theorem for partial sums of a sequence $U_n = \sum_{i=1}^n c_{ni}\varepsilon_i$ where $\sup_n c_{ni}^2 < \infty$, $\max_{1\le i\le n} |c_{ni}| \to 0$ as $n \to \infty$ and ε_i's are in turn, pairwise mixing martingale difference, mixing sequences or associated sequences. Mcleish, [19], De Jong, [7], and, more recently Shinxin [21], extend the previous results in the case of L_q-mixingale arrays. Those results have various applications. They are used for the proof strong convergence of kernel estimators. In this paper we extend Li *et al.* results to our weak dependent frame. A straightforward consequence of proposition 3 is the following result :

Corollary 1 *Under the assumptions of proposition 3, if q is an even integer such that $k > q > p$, and if for some real constant K, non depending on n $\sum_{i=1}^n c_{n,i-1}^2 < K$, and if $\theta_r = \mathcal{O}(r^{-\alpha})$, with $\alpha > (\frac{q-1}{k-q})k$, or $\theta_r = \mathcal{O}(e^{-r})$, then*

$$\forall \eta > 0 \quad \sum_n P(n^{-1/p}|U_n| \ge \eta) < \infty,$$

Proof. Proposition 3 implies

$$\begin{aligned} E|U_n|^q &\le \frac{(2q-2)!}{(q-1)!} c^{1/k} \left(\left(M_q \sum_{i=1}^n c_{n,n-i}^q \sum_{r=0}^{n-1} (r+1)^{q-2} \theta_r^{(k-q)/k} \right) \right. \\ &\vee \left. \left(M_2 \left(\sum_{i=1}^n c_{n,n-i}^2 \sum_{r=0}^{n-1} \theta_r^{(k-2)/k} \right)^{p/2} \right) \right). \end{aligned}$$

If $\sum_{i=1}^n c_{n,i-1}^2 < K$ and $\theta_r = \mathcal{O}(r^{-\alpha})$, with $\alpha > (\frac{q-1}{k-q})k$, then there exists a real constant K_1 $E|U_n|^q < K_1$, and the result follows from $P(n^{-1/p}|U_n| > \eta) \le \frac{E|U_n|^q}{\eta^q n^{q/p}}$. If $\sum_{i=1}^n c_{n,i-1}^2 < K$ and $\theta_r = \mathcal{O}(e^{-r})$, $E|U_n|^q < K_2$ for a real constant K_2 and $\sum_n P\left(n^{-1/p}|U_n| > \eta\right) < \infty$. ∎

As a straight consequence of proposition 4, we obtain the following result:

Corollary 2 *Under the assumptions of proposition 4, if $q > p$, $k > q > 1$, and $\sum_{i=1}^\infty c_{ni}^{2-n} \sum_{j=0}^\infty \gamma_j^m < \infty$ where $m = \frac{2}{q}(\frac{k-q}{k-1})$, then*

$$\forall \eta > 0 \quad \sum_n P(n^{-1/p}|U_n| \ge \eta) < \infty.$$

Proof. $E|U_n|^q \le \left(2qK_1 \sum_{i=1}^n c_{ni}^{2-m} \sum_{j=0}^{n-1} \gamma_j^m \right)^{q/2}$ from proposition 4, and the relation $\sum_{i=1}^\infty c_{ni}^{2-n} \sum_{j=0}^\infty \gamma_j^m < \infty$ implies $\sum_n P(n^{-1/p}|U_n| > \eta) < \infty$. This concludes. ∎

6 Linear regression

We observe a stationary (bounded) sequence, $(y_n, x_n) \in \mathbb{R} \times \mathbb{R}^d$, defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$. Let $M = \sup_n \|y_n x_n\|$.

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

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

This problem leads to study the gradient algorithm

$$Z_{n+1} = Z_n + c_n(y_{n+1} - x_{n+1}^T Z_n) x_{n+1}, \tag{27}$$

where $c_n = \mathcal{O}(n^{-1})$ (so (c_n) verifies (2) and (5)). Let $C_{n+1} = x_{n+1} x_{n+1}^T$, we obtain:

$$Z_{n+1} = Z_n + c_n(y_{n+1} x_{n+1} - C_{n+1} Z_n)). \tag{28}$$

Let $U = E(y_{n+1} x_{n+1})$, $C = E(C_{n+1})$ and $h(Z) = U - CZ$, then (28) becomes :

$$\begin{aligned} Z_{n+1} &= Z_n + c_n h(Z_n) + c_n \eta_{n+1}, \qquad \text{with} && (29) \\ \eta_{n+1} &= (y_{n+1} x_{n+1} - U) + (C - C_{n+1}) Z_n. && (30) \end{aligned}$$

Denoting $\mathcal{F}_n = (\sigma(y_i, x_i); i \le n)$ and we define the following assumption **A-lr**:

> *C is not singular and for any $i, j \in \{1, \dots, d\}$, $(x_n^i x_n^j)$ and $(y_n x_n)$ are γ-weakly dependent, such that the γ-weakly dependent coefficient γ_r is $\mathcal{O}(a^{-r})$ with $a > 1$.*

Note. If $(y_n, x_n)_{n \in \mathbb{N}}$ is θ-weakly dependent in the Dedecker and Doukhan sense ([8]), then **A-lr** is satisfied. This is proved in annex.
Denoting $\lambda_{max}(C)$ and $\lambda_{min}(C)$, respectively the largest and the smallest eigenvalue of the matrix C, we now claim:

Proposition 5 *Under assumption* **A-lr** *:*
(i) if (Z_n) is a.s. bounded then the perturbation (η_n) of algorithm (29) splits into three terms of which two are γ-weakly dependent and one is a rest leading a.s. to zero.
(ii) if a.s.

$$\limsup_n \|C - C_n\| \le \frac{\lambda_{min}(C)^{3/2}}{\lambda_{max}(C)^{1/2}}, \tag{31}$$

then (Z_n) is bounded and the result of (i) follows.

We thus can use the previous proposition 4 and the tools of the ODE method in order to study the algorithm (29).

Proof of the proposition 5. To start with, we prove the result of (i) and assume that (Z_n) is a.s. bounded. Then we prove that this assumption is justified.

The perturbation (η_n) splits into two terms : $(y_{n+1}x_{n+1} - U)$ and $(C - C_{n+1})Z_n$. The first term is obvious γ-weakly dependent thanks to assumption **A-lr**. It remains to study $(C - C_{n+1})Z_n$.

Study of $(C - C_{n+1})Z_n$: write $(C - C_{n+1})Z_n = \varepsilon_{n+1} + r_{n+1}$ with $\varepsilon_{n+1} = (C - C_{n+1})Z_n - E[(C - C_{n+1})Z_n]$ and $r_{n+1} = E[(C - C_{n+1})Z_n]$. We will prove that the sequence (ε_n) is γ-weakly dependent and that $r_n \to 0$ a.s.

Notice that $r_{n+1} = E[(C - C_{n+1})\sum_{j=\frac{n}{2}}^{n-1}(Z_{j+1} - Z_j)] + E[(C - C_{n+1})Z_{\frac{n}{2}}]$, and since $Z_{j+1} - Z_j = c_j(y_{j+1}x_{j+1} - C_{j+1}Z_j)$,

$$\begin{aligned} r_{n+1} &= \sum_{j=\frac{n}{2}}^{n-1} E[(C - C_{n+1})c_j y_{j+1}x_{j+1}] \\ &- \sum_{j=\frac{n}{2}}^{n-1} E[(C - C_{n+1})c_j C_{j+1}Z_j] \\ &+ E[(C - C_{n+1})Z_{\frac{n}{2}}]. \end{aligned}$$

If $\frac{n}{2}$ is not an integer, we replace it by $\frac{n-1}{2}$. Expectations conditionally with respect to $\mathcal{F}_{j+1}$ of each term of the first and the second sum and with respect $\mathcal{F}_{\frac{n}{2}}$ of the last term give, by assuming (Z_n) bounded :

$$r_{n+1} \leq K_1 \sum_{j=\frac{n}{2}}^{n-1} \gamma^1_{n-j}c_j + K_2 \sum_{j=\frac{n}{2}}^{n-1} \gamma^1_{n-j}c_j + \gamma^1_{\frac{n}{2}-1},$$

where K_1 and K_2 are non-negative constants and γ^1_r denotes the γ-weakly dependent coefficient of the sequence $(C - C_n)$ which is γ-weakly dependent, thanks to assumption **A-lr** ($\gamma^1_r \leq dM\gamma_r$). And since $c_j = \mathcal{O}(j^{-1})$ and $\gamma_r = \mathcal{O}(a^{-r})$ with $a > 1$, $r_n = \mathcal{O}(n^{-1})$ and converges a.s. to zero. On the other hand, for $r \geq 6$:

$$\begin{aligned} E(\varepsilon_{n+r}|\mathcal{F}_n) &= E[(C - C_{n+r})Z_{n+r-1}|\mathcal{F}_n] - E(\varepsilon_{n+r}), \\ &= \sum_{j=n+\frac{r}{2}}^{n+r-2} E[(C - C_{n+r})(Z_{j+1} - Z_j)|\mathcal{F}_n] \\ &\quad + E[(C - C_{n+r})Z_{n+\frac{r}{2}}|\mathcal{F}_n] - r_{n+r}. \end{aligned}$$

Note also that if $\frac{r}{2}$ is not an integer, we replace it by $\frac{r+1}{2}$. Conditionally with respect to $\mathcal{F}_{j+1}$, the expectation of each term in the sum (and with respect $\mathcal{F}_{n+\frac{r}{2}}$ for the the last term) yields, for some non-negative constant K,

$$E\|E(\varepsilon_{n+r}|\mathcal{F}_n)\| \leq K\left(\sum_{j=n+\frac{r}{2}}^{n+r-2} c_j\gamma^1_{n+r-j-1} + \gamma^1_{r/2}\right) + \mathcal{O}((n+r)^{-1}) \text{ hence,}$$

$E\|E(\varepsilon_{n+r}|\mathcal{F}_n)\| \leq \mathcal{O}((n+\frac{r}{2})^{-1}) + K\gamma^1_{\frac{r}{2}} + \mathcal{O}((n+r)^{-1})$, and $E\|E(\varepsilon_{n+r}|\mathcal{F}_n)\| = \gamma^2_r$, with $\lim_{r\to+\infty}\gamma^2_r = 0$. So (ε_n) is γ-weakly dependent.
We now prove *(ii)*. Let $V(Z) = Z^TCZ = \|\sqrt{C}Z\|^2$. Since C is not singular, V is a Lyapounov function and $\nabla V(Z) = CZ$ is a Lipschitz function, we have

$$V(Z_{n+1}) \leq V(Z_n) + (Z_{n+1}-Z_n)^T\nabla V(Z_n) + K\|Z_{n+1}-Z_n\|^2,$$

where $K > 0$ is a constant. Furthermore $\|Z_{n+1}-Z_n\|^2 \leq c_n^2(\|y_{n+1}x_{n+1}\|^2 + \|C_{n+1}Z_n\|^2)$. Since (y_n, x_n) is bounded, (C_n) is also bounded and $\|C_{n+1}Z_n\|^2 \leq K_1\|Z_n\|^2$, where K_1 is a non-negative constant. Moreover,

$$\|Z_n\|^2 \leq \frac{\|Z_n\|^2}{\|\sqrt{C}Z_n\|^2}V(Z_n) \leq \frac{1}{\lambda_{min}(C)}V(Z_n). \tag{32}$$

So that, $\|C_{n+1}Z_n\|^2 \leq K_2V(Z_n)$, where $K_2 \geq 0$ is a constant, and
$V(Z_{n+1}) \leq V(Z_n)(1+K_2c_n^2) + Mc_n^2 + (Z_{n+1}-Z_n)^TCZ_n$.
The last term gives

$$\begin{aligned}
(Z_{n+1}-Z_n)^TCZ_n &= c_n(y_{n+1}x_{n+1}^TCZ_n - Z_n^TC_{n+1}CZ_n)\\
&= c_n(y_{n+1}x_{n+1}^TCZ_n - Z_n^T(C_{n+1}-C)CZ_n - Z_n^TC^2Z_n),\\
&\leq c_n(M\|CZ_n\| + Z_n^T(C-C_{n+1})CZ_n - \lambda_{min}(C)V(Z_n))\\
&\leq c_n(M\sqrt{\lambda_{max}(C)V(Z_n)} + Z_n^T(C-C_{n+1})CZ_n - \lambda_{min}(C)V(Z_n)).
\end{aligned}$$

The last inequality follows from:

$$\|CZ_n\| \leq \sqrt{(\sqrt{C}Z_n)^TC(\sqrt{C}Z_n)} \leq \sqrt{\lambda_{max}(C)V(Z_n)}. \tag{33}$$

On the other hand, $Z_n^T(C-C_{n+1})CZ_n \leq \|C-C_{n+1}\|\|Z_n\|\|CZ_n\|$, and we deduce $Z_n^T(C-C_{n+1})CZ_n \leq \|C-C_{n+1}\|\sqrt{\frac{\lambda_{max}(C)}{\lambda_{min}(C)}}V(Z_n)$, from (32) and (33), and

$$\begin{aligned}
(Z_{n+1}-Z_n)^TCZ_n &\leq -\lambda_{min}(C)V(Z_n)\\
&\times \left(1 - \|C-C_{n+1}\|\frac{\sqrt{\lambda_{max}(C)}}{\lambda_{min}(C)^{3/2}} - M\frac{\sqrt{\lambda_{max}(C)}}{\lambda_{min}(C)\sqrt{V(Z_n)}}\right).
\end{aligned}$$

Thanks to the assumption (31), for n large enough ($n > N$), and for $V(Z_n) > M_1$, $\left(\text{where } M_1 = M^2\frac{\lambda_{max}(C)}{\lambda^2_{min}(C)}\left(1-\|C-C_{n+1}\|\frac{\sqrt{\lambda_{max}(C)}}{\lambda^{3/2}_{min}(C)}\right)^{-2}\right)$

$$(Z_{n+1}-Z_n)^TCZ_n < 0.$$

And if $T = \inf\{n > N / V(Z_n) \leq M_1\}$. By the Robbins-Sigmund theorem, $V(Z_n)$ converges a.s. to a finite limit on $\{T = +\infty\}$, so (Z_n) is bounded since V is a Lyapounov function.
On $\{\liminf_n V(Z_n) \leq M_1\}$, $V(Z_n)$ doesn't converge to ∞ and using Delyon ([9], Theorem 2), we deduce that $V(Z_n)$ converges to a finite limit, as soon as :

$$\forall k > 0, \quad \sum c_n^2 \, \|h(Z_n) + \eta_{n+1}\|^2 \, \mathbb{1}_{\{V(Z_n)<k\}} < \infty \tag{34}$$

$$\forall k > 0, \quad \sum c_n \langle \eta_{n+1}, \nabla V(Z_n)\rangle \, \mathbb{1}_{\{V(Z_n)<k\}} < \infty. \tag{35}$$

Using relations $\sum c_n^2 < \infty$ and the fact that $\{V(Z_n) < k\}$, $\|h(Z_n) + \eta_{n+1}\|^2$ is bounded, we thus deduce (34). To prove (35), it is enough, by proposition 4, to prove that
$\langle \eta_{n+1}, \nabla V(Z_n)\rangle \, \mathbb{1}_{\{V(Z_n)<k\}} = e_{n+1}$, is a γ-weakly dependent sequence. But to use the result of proposition 4, it is necessary to center e_n. So we are going to prove that $\sum c_n E e_{n+1} < \infty$ and that $(e_n - Ee_n)$ is a γ-weakly dependent sequence.
Study of $E(e_n)$. First of all, we must note a few elements. Denoting I the unit matrix of $\mathbb{R}^d$, $Z_n = (I - c_{n-1}C_n)Z_{n-1} + c_{n-1}x_n y_n$. If $\lambda = \sup_n \lambda_{max}(C_n)$, then λ is a.s. finite since (x_n) is a.s. bounded. So, for n large enough $c_{n-1}\lambda < 1$ and $(I - c_{n-1}C_n)$ is not singular. So we obtain

$$\begin{aligned} \|Z_{n-1}\| &\leq \frac{1}{1 - c_{n-1}\lambda}(\|Z_n\| + c_{n-1}M) \\ &\leq (1 + bc_{n-1})(\|Z_n\| \wedge M) \end{aligned}$$

where b is some non-negative constant, non depending on n. Moreover

$$\begin{aligned} V(Z_n) &< k \Longrightarrow \|Z_n\|^2 < \frac{k}{\lambda_{min}(C)}, \qquad \text{and} \\ \|Z_n\| &< k' \Longrightarrow V(Z_n) < \lambda_{max}(C) k'^2. \qquad \text{So that,} \\ \mathbb{1}_{\{V(Z_n)<k\}} &= \mathbb{1}_{\{\|Z_n\|<k_n\}} = \mathbb{1}_{\{\|Z_{n-j}\|<k_{n-j}\}}, \qquad \text{where} \\ k_{n-j} &\leq (1 + c_{n-1})^j \left(\sqrt{\frac{k}{\lambda_{min}(C)}} \wedge M \right). \end{aligned}$$

And since $c_n = \mathcal{O}(n^{-1})$, for any $0 \leq j \leq n$, $(1 + ac_{n-1})^j$ is bounded independently of n, so is k_{n-j}.

$$\begin{aligned} E(e_n) &= E(x_{n+1}y_{n+1} - U)^T Z_n \, \mathbb{1}_{\{V(Z_n)<k\}} \\ &+ EZ_n^T(C - C_{n+1})CZ_n \, \mathbb{1}_{\{V(Z_n)<k\}} = A_n + B_n, \qquad \text{and,} \end{aligned}$$

$$A_n = \sum_{j=\frac{n}{2}}^{n-1} E(y_{n+1}x_{n+1} - U)^T C(Z_{j+1} - Z_j) \, \mathbb{1}_{\{\|Z_j\|<k_j\}} + E(y_{n+1}x_{n+1} - U)^T CZ_{\frac{n}{2}}.$$

Note that if $\frac{n}{2}$ is not an integer, we replace it by $\frac{n-1}{2}$. Expectations conditionally with respect to $\mathcal{F}_{j+1}$ of each term of the sum and with respect to $\mathcal{F}_{\frac{n}{2}}$ of the last term give us : $A_n \leq K_1 \sum_{j=\frac{n}{2}}^{n-1} c_j \gamma_{n-j} + K_2 \gamma_{\frac{n}{2}+1}$, where K_1 and K_2 are non-negative constants. So, since $\gamma_r = \mathcal{O}(a^{-r})$ with $a > 1$,

$$A_n = \mathcal{O}(n^{-1}) + \mathcal{O}(a^{-\frac{n}{2}}), \tag{36}$$

and $\sum c_n A_n < \infty$.

$$\begin{aligned}
B_n &= \sum_{j=\frac{n}{2}}^{n-1} E(Z_{j+1} - Z_j)^T (C - C_{n+1}) C (Z_{j+1} - Z_j)\, \mathbb{1}_{\{\|Z_j\| < k_j\}} \\
&+ 2 \sum_{j=\frac{n}{2}}^{n-1} \sum_{i=j+1}^{n-2} E(Z_{j+1} - Z_j)^T (C - C_{n+1}) C (Z_{i+1} - Z_i)\, \mathbb{1}_{\{\|Z_j\| < k_j\} \cap \{\|Z_i\| \leq k_i\}} \\
&+ 2 \sum_{j=\frac{n}{2}}^{n-1} E Z_{n/2}^T (C - C_{n+1}) C (Z_{j+1} - Z_j)\, \mathbb{1}_{\{\|Z_j\| < k_j\} \cap \{\|Z_{n/2}\| < k_{n/2}\}} \\
&+ E Z_{n/2}^T (C - C_{n+1}) C Z_{n/2}\, \mathbb{1}_{\{\|Z_{n/2}\| < k_{n/2}\}}.
\end{aligned}$$

Expectations conditionally with respect to $\mathcal{F}_{j+1}$ of each term of the first and the third sums, with respect to $\mathcal{F}_{i+1}$ of the second term, and with respect to $\mathcal{F}_{\frac{n}{2}}$ of the last term give us :

$$B_n \leq K_3 \sum_{j=\frac{n}{2}}^{n-1} c_j^2 \gamma_{n-j}^1 + K_4 \sum_{j=\frac{n}{2}}^{n-1} c_j \sum_{i=j+1}^{n-2} c_i \gamma_{n-i}^1 + K_5 \sum_{j=\frac{n}{2}}^{n-1} c_j \gamma_{n-j}^1 + K_6 \gamma_{n/2+1}^1,$$

where $K_3,\ K_4,\ K_5$ and K_6 are non-negative constants. And

$$B_n = \mathcal{O}(n^{-2}) + \mathcal{O}(n^{-2}) + \mathcal{O}(n^{-1}) + \mathcal{O}(a^{-n/2-1}). \tag{37}$$

Hence $\sum c_n B_n < \infty$ and $\sum c_n E e_{n+1} < \infty$.
Study of $(e_n - E e_n)$. We now prove that this sequence is γ-weakly dependent. Write $E(e_{n+r} - E e_{n+r} | \mathcal{F}_n) = D_{n+r} + G_{n+r} - E e_{n+r}$, with

$$\begin{aligned}
D_{n+R} &= E[(y_{n+r} x_{n+r} - U)^T C Z_{n+r-1}\, \mathbb{1}_{\{V(Z_{n+r-1}) < k\}}\ |\mathcal{F}_n], \\
G_{n+r} &= E(Z_{n+r-1}^T (C - C_{n+r}) C Z_{n+r-1}\, \mathbb{1}_{\{V(Z_{n+r-1}) < k\}}\ |\mathcal{F}_n] \\
D_{n+r} &= \sum_{j=n+\frac{r}{2}}^{n+r-2} E[((y_{n+r} x_{n+r} - U)^T C (Z_{j+1} - Z_j)\, \mathbb{1}_{\{\|Z_j\| < k_j\}}\ |\mathcal{F}_n] \\
&+ E[((y_{n+r} x_{n+r} - U)^T C Z_{n+\frac{r}{2}}\, \mathbb{1}_{\{\|Z_{n+\frac{r}{2}}\| < k_{n+\frac{r}{2}}\}}\ |\mathcal{F}_n],
\end{aligned}$$

Here again, if $\frac{r}{2}$ is not a integer, we replace it by $\frac{r-1}{2}$. Expectations conditionally with respect to $\mathcal{F}_{j+1}$ of each term of the sum and with respect $\mathcal{F}_{n+\frac{r}{2}}$ of the last term give us :

$$\begin{aligned} E\|D_{n+r}\| &\leq \sum_{j=n+\frac{r}{2}}^{n+r-2} c_j \gamma^1_{n+r-j-1} + \gamma^1_{n+\frac{r}{2}}, \\ &= \mathcal{O}(\frac{1}{n+r}) + \mathcal{O}(a^{-n-\frac{r}{2}}) \end{aligned}$$

Hence $\lim_{r\to+\infty} E\|D_{n+r}\| = 0$. We study G_{n+r} in the same way and $E\|G_{n+r}\| = \mathcal{O}(\frac{1}{n+r})$, and by (36) and (37), since $Ee_{n+1} = A_n + B_n$, we obtain $Ee_{n+r} = \mathcal{O}(\frac{1}{n+r})$, the result is proved. ∎

7 Proofs

7.1 Proof of proposition 2

We use a sketch similar to Doukhan and Louhichi's proof in [11].

$$E(\sum_{i=1}^{n} c_i \varepsilon_i)^p \leq p! \sum_{1\leq t_1 \leq \cdots \leq t_p \leq n} c_{t_1} \ldots c_{t_p} |E(\varepsilon_{t_1} \cdots \varepsilon_{t_p})|. \tag{38}$$

Denote

$$A_p(n) = \sum_{1\leq t_1 \leq \cdots \leq t_p \leq n} c_{t_1} \cdots c_{t_p} |E(\varepsilon_{t_1} \cdots \varepsilon_{t_p})|,$$

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

$$\begin{aligned} A_p(n) &\leq \sum_{1\leq t_1 \leq \cdots \leq t_p \leq n} c_{t_1} \cdots c_{t_p} |E(\varepsilon_{t_1} \cdots \varepsilon_{t_m}) E(\varepsilon_{t_{m+1}} \cdots \varepsilon_{t_p})| \\ &+ \sum_{1\leq t_1 \leq \cdots \leq t_p \leq n} c_{t_1} \cdots c_{t_p} |\mathrm{cov}(\varepsilon_{t_1} \cdots \varepsilon_{t_m}, \varepsilon_{t_{m+1}} \cdots \varepsilon_{t_p})|. \end{aligned}$$

Denote

$$\begin{aligned} A^1_p(n) &= \sum_{1\leq t_1 \leq \ldots \leq t_p \leq n} c_{t_1} \ldots c_{t_p} |E(\varepsilon_{t_1} \cdots \varepsilon_{t_m}) E(\varepsilon_{t_{m+1}} \ldots \varepsilon_{t_p})|, \\ A^2_p(n) &= \sum_{1\leq t_1 \leq \ldots \leq t_p \leq n} c_{t_1} \cdots c_{t_p} |\mathrm{cov}(\varepsilon_{t_1} \cdots \varepsilon_{t_m}, \varepsilon_{t_{m+1}} \cdots \varepsilon_{t_p})|. \end{aligned}$$

Since the sequence (c_n) is decreasing to 0, we deduce, as in [11],

$$A^1_p(n) \leq A_m(n) A_{p-m}(n). \tag{39}$$

By the definition of $C_{rq}(t_1)$, using lemma 15 in [11], we obtain

$$A_p^2(n) \leq \sum_{t_1=1}^{n} c_{t_1}^p \sum_{r=0}^{n-1} C_{rp}(t_1)(r+1)^{p-2}.$$

And by (14), the expression $\sum_{i=1}^n c_i^p \sum_{r=0}^{n-1} C_{rp}(i)(r+1)^{p-2} = V_p(n)$, verifies, for any integers $2 \leq q \leq p-1$:

$$V_q(n) \leq V_p^{\frac{q-2}{p-2}}(n) V_2^{\frac{p-q}{p-2}}(n).$$

Now, lemma 12 of [11] leads to $A_p(n) \leq \frac{1}{p} \begin{pmatrix} 2p-2 \\ p-1 \end{pmatrix} (V_2^{\frac{p}{2}}(n) \vee V_p(n))$, hence

$$E\left(\sum_{i=1}^{n} c_i \varepsilon_i\right)^p \leq \frac{(2p-2)!}{(p-1)!}(V_2^{\frac{p}{2}}(n) \vee V_p(n)).$$

This ensures the result. ∎

7.2 Proof of proposition 3

Using the same denotations as in the previous proof, by (11)

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

where $\theta(u) = \theta_{[u]}([u]$ denotes the integer part of $u)$. Denote

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

If (17) is verified,

$$W_q(n) \leq W_p(n)^{\frac{q-2}{p-2}}(n) W_2^{\frac{p-q}{p-2}}(n).$$

So we can conclude as in the previous proof. ∎

7.3 Proof of proposition 4

Proceeding as in [8], we deduce

$$|E(\Sigma_n^p)| \leq \left(2p \sum_{i=1}^{n} b_{i,n}\right)^{\frac{p}{2}},$$

where

$$b_{i,n} = \max_{i \leq \ell \leq n} \left\| c_i\varepsilon_i \sum_{k=0}^{\ell-i} E(c_{i+k}\varepsilon_{i+k}|\mathcal{F}_i) \right\|_{\frac{p}{2}}.$$

Let $q = \frac{p}{p-2}$, then there exists Y such that $\|Y\|_q = 1$.
Applying proposition 1 of [8], we obtain

$$b_{i,n} \leq \sum_{k=0}^{n-i} \int_0^{\gamma_k} Q_{\{Yc_i\varepsilon_i\}} \circ G_{\{c_{i+k}\varepsilon_{i+k}\}}(u)du,$$

where G_X is the inverse of $x \to \int_0^x Q_X(u)du$. Since $G_{\{c_i\varepsilon_i\}}(u) = G_{\varepsilon_i}(\frac{u}{c_i}) = G(\frac{u}{c_i})$, we get

$$b_{i,n} \leq \sum_{k=0}^{n-i} \int_0^{\gamma_k} Q_{\{Yc_i\varepsilon_i\}} \circ G(\frac{u}{c_{i+k}})du \leq \sum_{k=0}^{n-i} c_{i+k} \int_0^{\frac{\gamma_k}{c_{i+k}}} Q_{\{Yc_i\varepsilon_i\}} \circ G(u)du,$$

and the Fréchet inequality (1957) yields

$$\begin{aligned} b_{i,n} &\leq \sum_{k=0}^{n-i} c_{i+k} \int_0^{G(\frac{\gamma_k}{c_{i+k}})} Q_Y(u)Q_{\{c_i\varepsilon_i\}}(u)Q(u)du \\ &\leq \sum_{k=0}^{n-i} c_i c_{i+k} \int_0^1 \mathbb{1}_{\{u \leq G(\frac{\gamma_k}{c_{i+k}})\}} Q^2(u)Q_Y(u)du \end{aligned}$$

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

$$b_{i,n} \leq c_i \sum_{k=0}^{n-i} c_{i+k} \left(\int_0^1 \mathbb{1}_{\{u \leq G(\frac{\gamma_k}{c_{i+k}})\}} Q^p(u)du \right)^{\frac{2}{p}}.$$

By (18), $Q(u) \leq c^{\frac{1}{r}} u^{-\frac{1}{r}}$ and setting $K = \frac{r-1}{rc^{\frac{1}{r}}}$ yields

$$\begin{aligned} b_{i,n} &\leq c_i \sum_{k=0}^{n-i} c_{i+k} \left(\int_0^1 \mathbb{1}_{\left\{u \leq G(\frac{\gamma_k}{c_{i+k}})\right\}} c^{\frac{p}{r}} u^{-\frac{p}{r}} du \right)^{\frac{2}{p}} \\ &\leq c_i \sum_{k=0}^{n-i} c_{i+k} \left(K \frac{\gamma_k}{c_{i+k}} \right)^{\frac{r}{r-1}(1-\frac{p}{r})\frac{2}{p}}. \end{aligned}$$

Noting that $(c_n)_{n\geq 0}$ is decreasing, the result follows with $K_1 = K^{\frac{2(r-p)}{p(r-1)}}$. ∎

8 Annex

8.1 Proof of the note in the section 6

This note claims that if $(y_n, x_n)_{n\in\mathbb{N}}$ is θ-weakly dependent in the Dedecker and Douhkan sense ([8]), then **A-lr** is satisfied. Let us remind the definition of a θ-weakly dependent $\mathbb{R}^d$-valued sequence which is used in ([8]):

If $\mathcal{L}_1$ is the space of bounded 1-Lipschitz real valued functions defined on $\mathbb{R}^d$, (X_n) is θ-weakly dependent as soon as

$$\theta_r = \sup_{n\geq 0}\{\sup_{f\in\mathcal{L}_1}(\|E[f(X_{r+n})|\sigma(X_i, i\leq n)] - E[f(X_{k+n})]\|_1)\}$$

tends to zero as r tends to infinity.

For any $f\in\mathcal{L}_1$, $|f(x)-f(y)|\leq |x^1-y^1|+\cdots+|x^d-y^d|$, where the x^j's $(j=1,\ldots,d)$ are the components of x.

First, note that if a $\mathbb{R}^d$-valued sequence (X_n) is θ-weakly dependent, any $\mathbb{R}^j$-valued sequence $(j=1,\ldots,d-1)$ $(Y_n)=(X_n^{t_1},\ldots,X_n^{t_j})$ 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 a bounded 1-Lipschitz function, defined on $\mathbb{R}$ and g the function defined on $\mathbb{R}^2$ by $g(x,y)=f(xy)$. It is enough to prove that g is a Lipschitz function defined on $\mathbb{R}^2$.

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

and g is Lipschitz as soon as x and y are bounded.

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

References

[1] M. BENAÏM and M.W. HIRSCH (1995) *Dynamics of Morse-Smale urn processes*, Ergodic Theory and Dynamical Systems, 15, pp.1005-1030.

[2] M. BENAÏM (1996) *A dynamical system approach to stochastic approximation*, SIAM J. control and optimization, vol. 34, 2, pp. 437-472.

[3] A. BENVENISTE and M. MÉTIVIER and P. PRIOURET (1987) *Algorithmes adaptatifs et approximations stochastiques*, Masson.

[4] P. BILLINGSLEY (1985) *Probability and measure*, second edition, John Wiley & Sons.

[5] Y.S. CHOW (1966) *Some convergence theorems for independent random variables*, Ann. Math. Stat. 37, 1482-1493 36, No. 4, pp. 1293-1314.

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

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

[8] P. Dedecker P. Doukhan (2002) *An new covariance inequality and applications*, to appear.

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

[10] P. Doukhan (2002) *Models inequalities and limit theorems for stationary sequences*, in Theory and applications of long range dependence (Doukhan *et alii* ed.) Birkhäuser, 43-101.

[11] P. Doukhan and S. Louhichi (1999), *A new weak dependence condition and applications to moment inequalities*, Stochastic Process and their applications 84, 313-342.

[12] M. Duflo (1996), *Algorithmes Stochastiques*, Collection Math. et Appl. 23, Springer.

[13] S. Ghosal, T.K. Chandra (1998), *Complete convergence of martingale arrays*, J. Theor. Probab. Vol 11, No. 3, 621-631.

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

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

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

[17] H.J. Kushner and D.S. Clark (1978) *Stochastic approximation for constrained and uncontrained systems*, Applied math. science series, vol. 26, Springer.

[18] D. LI, M.B. Rao, T. Jiang, X. Wang (1995) *Complete convergence and almost sure convergence of weighted sums of random variables*,J. Theor. Prob. 8, 49-76.

[19] D.L. Macleish (1975) *A maximal inequality and dependent strong laws*, Ann. Probab. 3, 829-339.

[20] M. Peligrad, S. Utev (1997) *Central limit theorem for linear processes*, The Ann. of Probab., Vol. 25, N3, 443-456.

[21] EG. Shixin (1999) *On the convergence of weighted sums of L_q-mixingale arrays*, Acta Math. Hungar. 82(1-2), 113-120.

[22] K.F. Yu (1990) *Complete convergence of weighted sums of martingale differences*, J. Theor. Prob. 3, 339-347.

[1] ENSAE and Laboratory of Statistics, CREST.
Timbre J340, 3, avenue Pierre Larousse,
92240 Malakoff, France
doukhan@ensae.fr

[2] Université Paris 11
Laboratoire de Mathématiques
Bat 425 91405 Orsay, France.
odile.brandiere@math.u-psud.fr