ON A PERTURBATION APPROACH FOR THE ANALYSIS OF STOCHASTIC TRACKING ALGORITHMS

R. AGUECH , E. MOULINES , P. PRIOURET

Abstract

In this paper, a perturbation expansion technique is introduced to decompose the tracking error of a general adaptive tracking algorithm in a linear regression model. This method allow to obtain tracking error bound but also tight approximate expressions for the moments of the tracking error. These expressions allow to evaluate, both qualitatively and quantitatively, the impact of several factors on the tracking error performance which have been overlooked in previous contributions.

keywords 60G35, 93E10, 93E12, 93E15, 93E23

Corresponding author:
E. Moulines,
Ecole Nationale Supérieure des Télécommunications,
CNRS-URA 820,
46, rue Barrault,
75634 Paris Cedex 13,
FRANCE

R. Aguech,
Laboratoire de Probabilités, Université Pierre et Marie Curie,
4 Place Jussieu, 75252 Paris
FRANCE

P. Priouret,
Laboratoire de Probabilités, Université Pierre et Marie Curie,
4 Place Jussieu, 75252 Paris
FRANCE

1. Introduction

An important issue in system identification, signal processing, automatic control is that of tracking the parameter variations of a linear dynamical system with observable stochastic inputs and noise-corrupted outputs:

$$y_t = \phi_t^T \theta_t + v_t \,; \quad t \geq 0 \tag{1}$$

where $\{y_t\}_{t\geq 0}$ and $\{v_t\}_{t\geq 0}$ are respectively the scalar observation and noise, $\{\phi_t\}_{t\geq 0}$ and $\{\theta_t\}_{t\geq 0}$ are the d-dimensional stochastic regressors and the unknown time-varying parameter. This model encompasses many different applications, including channel equalization, time delay estimation and echo cancelation (see Solo and Kong (1995)). To track the variations of the parameter, it is customary to use a recursive algorithm for updating an estimate $\hat{\theta}_t$ of the parameter (see for example Ljung and Soderstrom (1983), Macchi (1994), Solo and Kong (1995), Kushner and Yin (1997) and the references therein). These algorithms may take many different forms depending on what was is willing to assume on the observation noise and on the parameter variations and on the amount of computations that is acceptable; e.g. standard stochastic approximation with fixed step-size, recursive least-squares with forgetting factor, or adaptations of Kalman-Bucy filters. In this work, we focus on fixed step-size stochastic approximation algorithm,

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \mu L_t (y_t - \phi_t^T \hat{\theta}_t). \tag{2}$$

where μ is referred to as the adaptation step-size and L_t is a random vector, which can be chosen in a number of different ways. By far the most popular algorithm in that class is the Least Mean Square algorithm (LMS), introduced in a landmark paper by Widrow and Hoff (1960). In that case, $L_t = \phi_t$. Such an algorithm is referred to as a gradient algorithm, because the increment of the algorithm is opposite to the (stochastic) gradient of the mean-square error, $e_t(\theta) = E(y_t - \phi_t^T \theta)^2$. A useful variant of the LMS, is the normalized LMS, for which

$$L_t = \frac{\phi_t}{\epsilon + \|\phi_t\|^2} \,; \; \epsilon > 0$$

There is a vast literature on the analysis of algorithms of type (2). In most contributions, the main goal is to obtain bounds on the tracking errors. Preliminary results in that directions have been obtained by Bitmead and Anderson (1980), Eweda and Macchi (1985); see Ljung and Gunnarson (1990) for a review of these early contributions. These results have been later significantly improved by Guo (1994) and Guo and Ljung (1995). In this contribution a different approach is pursued. Our goal is to obtain explicit expression and not only bounds for the tracking error. To that purpose, we will use a technique, referred to as *perturbation expansion*, consisting in approximating the process defined in (2) by a family of nested processes, with much simpler structure than the original error process. This decomposition enables the computation of explicit expressions for the moments of the tracking errors and other related quantities. This type of approximation technique has been introduced independently by Priouret and Ljung (1991), who considered general fixed step-size stochastic gradient algorithm, and Solo (1992), who used this method to get insights on the behavior of the tracking error variance of the LMS; however, in these papers, only first order approximations are considered. We extend this approach to obtain expansions valid up to any given order. This decomposition will enable us to obtain (under some technical conditions) an asymptotic expansion of the tracking error up to any given order in the step-size μ.

To analyze any adaptive algorithm, it is usually convenient to convert it to a so-called error form; indeed from (1) and (2), we can write

$$\tilde{\theta}_{t+1} = (I - \mu L_t \phi_t^T)\tilde{\theta}_t + \mu L_t v_t - w_{t+1}, \tag{3}$$

where $\tilde{\theta}_t \triangleq \hat{\theta}_t - \theta_t$ is the weight-error vector and $w_{t+1} \triangleq \theta_{t+1} - \theta_t$ is the lag-noise. . This is a time varying non-homogeneous difference equation. Since this equation is linear, $\tilde{\theta}_{t+1}$ can further be decomposed as

$$\begin{aligned}
\tilde{\theta}_t &= {}^u\tilde{\theta}_t + \mu\, {}^v\tilde{\theta}_t + {}^w\tilde{\theta}_t, && (4)\\
{}^u\tilde{\theta}_{t+1} &= (I - \mu L_t \phi_t^T)\, {}^u\tilde{\theta}_t \quad {}^u\tilde{\theta}_0 = \tilde{\theta}_0 = -\theta_0, && (5)\\
{}^v\tilde{\theta}_{t+1} &= (I - \mu L_t \phi_t^T)\, {}^v\tilde{\theta}_t + L_t v_t, \quad {}^v\tilde{\theta}_0 = 0 && (6)\\
{}^w\tilde{\theta}_{t+1} &= (I - \mu L_t \phi_t^T)\, {}^w\tilde{\theta}_t - w_{t+1}, \quad {}^w\tilde{\theta}_0 = 0 && (7)
\end{aligned}$$

$\{{}^u\tilde{\theta}_t\}$ is a transient term, reflecting the way the successive estimates of the regression coefficients forget the initial conditions. $\{{}^v\tilde{\theta}_t\}$ accounts for the errors introduced by the measurement noise, $\{v_t\}$; similarly, $\{{}^w\tilde{\theta}_{t+1}\}$ accounts for the errors associated lag-noise $\{w_t\}$. According to these definitions, ${}^v\tilde{\theta}_t$ and ${}^w\tilde{\theta}_t$ obey an inhomogeneous stochastic recurrence equation

$$\begin{aligned}
\delta_{t+1} &= (I - \mu F_t)\delta_t + \xi_t, \qquad \delta_0 = 0 && (8)\\
&= \sum_{s=0}^{t} \Phi(t,s)\xi_s && (9)
\end{aligned}$$

where $\{F_t\}_{t\geq 0}$ matrix valued random process, $\{\xi_t\}_{t\geq 0}$ is a $(d \times 1)$ vector-valued random process, and $\Phi(t,s)$ is defined as

$$\Phi(t,s) = \begin{cases} (I - \mu F_t)(I - \mu F_{t-1}) \cdots (I - \mu F_{s+1}) & t > s \\ I & t = s \\ 0 & \text{otherwise} \end{cases}$$

Here, the dependence of δ_t upon the step-size μ is implicit. Eqs (6) and (7) may be rewritten as (8) with $F_t = L_t \phi_t^T$ and

$$\xi_t = L_t v_t \text{ measurement noise}, \xi_t = -w_{t+1} \text{ lag noise} \tag{10}$$

2. Perturbation expansion: the method

In this section, we concentrate on the general recurrence equation (8). We apply the results to the lag-error term $\tilde{\theta}_t$ in the next section. Equations of the form (8) have received a considerable attention in the literature. The first question of interest is to derive conditions upon which $\{\delta_t\}$ is stable in L_p $(p \geq 1)$, in the sense that there exists some $\mu_0 > 0$ such that, for all $0 < \mu \leq \mu_0$

$$\sup_{t\geq 0} \|\delta_t\|_p < \infty, \ \|\delta_t\|_p = E(|\delta_t|^p)^{1/p}$$

where, for any $p \times q$ matrix $X = [X_{ij}]$, $|X|$ denotes the Euclidean norm: $|X| = (\sum_{i=1}^p \sum_{j=1}^q X_{ij}^2)^{1/2}$. L_p stability is an obvious extension of the notion of stability for deterministic linear systems. To our best knowledge, it was apparently first tackled by Eweda and Macchi (1985,1986), who provide an explicit expression for the bound under very restrictive conditions (m-independence and moment

conditions). As elaborated, e.g. in Guo (1994) and Guo and Ljung (1995) the essential key to the analysis of (8) is to prove that the product of random matrices $\Phi(t,s)$ is exponentially stable in L_q (for q sufficiently large), meaning that, there exists a constant $c_q < \infty$ and $\beta > 0$ such that $0 < \beta < 1/\mu_0$ verifying

$$\|\Phi(t,s)\|_q \leq c_q(1-\beta\mu)^{t-s}$$

for all $t > s$ and all $0 < \mu \leq \mu_0$. It is easily seen that exponential stability of this product of matrices in particular implies the L_p stability of δ_t by application of the Holder inequality. Results in that direction have been obtained recently by Guo (1994) and Guo and Ljung (1995), where practical conditions guaranteeing the exponential stability of product of random matrices are given.

As explained in the introduction, the purpose of this paper is to go one step beyond : we want to devise a general method enabling to obtain not only bounds but explicit *approximate* expressions for the moments of δ_t. As pointed out in Ljung and Gunnarson (1989,1990) these quantities are essential to understand the tracking behavior of a given algorithm or / and to design algorithm with better tracking performance. As illustrated in section 5 on a particular (simple) situation, these approximate expressions allow to evaluate the impact, both qualitatively and quantitatively, of different factors on the tracking performance (e.g. the dependence structure of the regressor sequence, of the observation and the lag noise). Results in that directions have been obtained by Ljung and Priouret (1991), who gave an exact first order approximation of the tracking error covariance of a stochastic gradient algorithm with a fixed step-size, and by Solo (1992), who worked out a first-order approximation of the tracking error covariance for the LMS algorithm (see also Solo and Kong (1995)). These results were latter extended to a more general class of algorithms (under much weaker assumptions) by Guo and Ljung (1995). However, in these works, only first order expansions of the tracking error are considered.

The approach developed in this contribution relies upon a perturbation technique. Applied to the recurrence equation (8), the whole procedure goes as follows. Denote $\bar{F}_t = E(F_t)$ (it is admitted at this point that this expectation exists) and $Z_t = \bar{F}_t - F_t$. We may decompose $(I - \mu F_t)$ according to

$$I - \mu F_t = (I - \mu \bar{F}_t) + \mu Z_t. \tag{11}$$

Now, decompose the recurrence equations (8) into two separate recursions:

$$J_{t+1}^{(0)} = (I - \mu\bar{F}_t)J_t^{(0)} + \xi_t, \quad J_0^{(0)} = 0 \tag{12}$$

$$H_{t+1}^{(0)} = (I - \mu F_t)H_t^{(0)} + \mu Z_t J_t^{(0)}, \quad H_0^{(0)} = 0 \tag{13}$$

$$\delta_t = J_t^{(0)} + H_t^{(0)}. \tag{14}$$

According to (12), $J_t^{(0)}$ satisfy a *deterministic* inhomogeneous first-order linear difference equation:

$$J_{t+1}^{(0)} = \sum_{s=0}^{t} \psi(t,s)\xi_s \tag{15}$$

where, as above,

$$\psi(t,s) = \begin{cases} (I-\mu\bar{F}_t)(I-\mu\bar{F}_{t-1})\cdots(I-\mu\bar{F}_{s+1}) & t>s \\ I & t=s \\ 0 & \text{otherwise} \end{cases}$$

Under appropriate assumptions on the matrix valued sequences $\{F_t\}_{t\geq 0}$ and on the excitation $\{\xi_t\}$, it will be shown that, for some $p>0$ there exists a constant $C<\infty$ and $\mu_0>0$ such that for all $0<\mu\leq\mu_0$

$$\sup_{t\geq 0}\|J_t^{(0)}\|_p \leq C/\sqrt{\mu} \quad \text{and} \quad \sup_{t\geq 0}\|H_t^{(0)}\|_p \leq C \tag{16}$$

where $C<\infty$ is a constant depending on $\{F_t\}$ and $\{\xi_t\}$ (see below). Thus, $J_t^{(0)}$ may be considered as the leading term in the expansion, while $H_t^{(0)}$ may be seen as a correction term. This decomposition forms the basis of the first-order approximation results obtained in Priouret and Ljung (1991), Solo (1992) and Guo and Ljung (1995) The same procedure can be iterated to obtain approximations of increased accuracy. For that purpose, it suffices to decompose (13) using (11), and iterate that decomposition up to order $n>1$. Using this technique, the weight-error vector δ_t may be decomposed as

$$\delta_t = J_t^{(0)} + J_t^{(1)} + \cdots + J_t^{(n)} + H_t^{(n)}, \tag{17}$$

where the processes $J_t^{(r)}$, $0\leq r\leq n$ and $H_t^{(n)}$ are respectively defined as

$$J_{t+1}^{(0)} = (I-\mu\bar{F}_t)J_t^{(0)} + \xi_t \;;\quad J_0^{(0)} = 0 \tag{18}$$

$$J_{t+1}^{(r)} = (I-\mu\bar{F}_t)J_t^{(r)} + \mu Z_t J_t^{(r-1)} \;;\quad J_t^{(r)} = 0,\ 0\leq t<r \tag{19}$$

$$H_{t+1}^{(n)} = (I-\mu F_t)H_t^{(n)} + \mu Z_t J_t^{(n)} \;;\quad H_t^{(n)} = 0,\ 0\leq t<n \tag{20}$$

The processes $J_t^{(r)}$ depend linearly on ξ_t and polynomially in the error $Z_t = \bar{F}_t - F_t$. It is thus feasible (examples are given below) to compute the joint moments of these processes, and to obtain expressions for the moments of $\tilde{\delta}_t^{(n)} = J_t^{(0)}+\cdots+J_t^{(n)}$. The residual term $H_t^{(n)}$ is, under appropriate conditions, uniformly bounded, *i.e.*, there exists some constant $C<\infty$ and $\mu_0>0$, such that, for all $0<\mu\leq\mu_0$, we have

$$\sup_{t\geq 0}\|H_t^{(n)}\|_p \leq C\mu^{n/2}. \tag{21}$$

Upper bounds for the constant C (depending upon the regression sequence, the observation noise and the lag noise) are given below. $\tilde{\delta}_t^{(n)}$ is thus the leading term of the expansion, whereas $H_t^{(n)}$ is a remainder, which is uniformly bounded in L_p. By computing the moments of $\tilde{\delta}_t^{(n)}$, on can obtain an approximation of the moments of the tracking error δ_t, the error between the exact and the approximate expressions being uniformly bounded by $\mu^{n/2}$.

3. Main results

3.1. Notations and Definitions. All the processes are assumed to be defined on the same basic probability space $(\Omega,\mathcal{A},P)$. For $p>0$, and $\mathcal{B}\subset\mathcal{A}$, denote $L_p(\Omega,\mathcal{B},P)$ the Banach space of $\mathcal{B}$-

measurable random variables such that $\|X\|_p < \infty$; for brevity, we set: $L_p(\Omega, \mathcal{A}, P) = L_p$. A random matrix sequence $\{X_t\}_{t\geq 0}$ is called L_p-bounded if $\sup_{t\geq 0} \|X_t\|_p < \infty$. Let $X \triangleq \{X_t\}_{t\in\mathbb{Z}}$ be a vector-valued random process. The σ-field generated by the random variables X_t, $a \leq t \leq b$ is denoted $\mathcal{M}_a^b(X)$.

3.2. Assumptions on $\{F_k\}_{k\geq 0}$. As mentionned in the previous section, the key assumption is the *exponential stability* of the products of random matrices $\prod(I - \mu F_k)$ and the corresponding products of deterministic matrices $\prod(I - \mu E(F_k))$. After Guo (1994), define, for $p \geq 1$, $\mu^* > 0$ and $0 < \beta < 1/\mu^*$, $\mathcal{S}(p, \beta, \mu^*)$ the L_p *exponentially stable familly* as

$$\mathcal{S}(p,\beta,\mu^*) = \left\{ F : \left\| \prod_{j=i+1}^{k} (I - \mu F_j) \right\|_p \leq K'_{\beta,\mu^*}(F)\ (1-\beta\mu)^{k-i} \quad \forall\, \mu \in (0, \mu^*],\ \forall\, k \geq i \geq 0 \right\}.$$

Likewise, define the *averaged exponentially stable* familly as:

$$\mathcal{S}(\beta,\mu^*) = \left\{ F : \left| \prod_{j=i+1}^{k} (I - \mu E(F_j)) \right| \leq K_{\beta,\mu^*}(F)\ (1-\beta\mu)^{k-i} \quad \forall\, \mu \in (0, \mu^*],\ \forall\, k \geq i \geq 0 \right\}.$$

Practical criteria for checking exponential stability have been proposed by Guo (1993) and Guo (1994). Some improvements of these results can be found in a series of papers co-authored by Guo and Ljung (1995). Necessary and sufficient conditions for the LMS algorithm are presented in Guo *et al* (1997). Alternate conditions implying exponential stability have been obtained in the case where the matrix valued process $\{F_t\}_{t\geq 0}$ is Markovian (Priouret and Veretenikov (1996)).

In addition to L_p- and averaged- exponential stability, additional moments and mixing conditions are required. Here, mixing refers, as usual, to some kind of 'decorrelation' between the past and the future of the process. It is convenient to use the notion of weak-mixing, introduced by Doukhan and Louhichi (1997).

Definition Let $q \geq 1$ and let $X = \{X_n\}_{n\geq 0}$ be a $(l \times 1)$ matrix-valued process. Let $\delta = (\delta(r))_{r\in\mathbb{N}}$ be a sequence of positive numbers decreasing to zero at infinity. The sequence $X = \{X_n\}_{n\geq 0}$ is said to be (δ, q)-weak dependent if there exist finite constants $C = \{C_1, \cdots, C_q\}$, such that for any $1 \leq m < s \leq q$, any m-tuple $t_1, \cdots, t_m$ and any $(s-m)$-tuple $t_{m+1}, \cdots, t_s$, with $t_1 \leq \cdots \leq t_m < t_m + r \leq t_{m+1} \leq \cdots t_s$, it holds

$$\sup_{1\leq i_1,\cdots,i_s\leq l} \left|\mathrm{cov}(X_{t_1,i_1}\cdots X_{t_m,i_m}, X_{t_{m+1},i_{m+1}}\cdots X_{t_s,i_s})\right| \leq C_s \delta(r)$$

where $X_{n,i}$ denotes the i-th component of X_n.

As shown in Doukhan and Louhichi (1997), weak-dependent processes encompass a large class of models and in particular strongly mixing processes, weak shift processes or models with a Markovian representation... Recall that the process $\{X_t\}_{t\geq 0}$ is strongly mixing if

$$\sup_{s\in\mathbb{Z}} \sup_{A\in\mathcal{M}_{-\infty}^{s}(X), B\in\mathcal{M}_{s+\tau}^{\infty}(X)} |P(A\cap B) - P(A)P(B)| = \alpha_X(\tau) \rightarrow_{\tau\rightarrow\infty} 0$$

$\alpha_X(\tau)$ is referred to as the (strong) mixing coefficient. The following lemma shows that strong-mixing processes also are weak-mixing and give the relation between the strong-mixing coefficient $\alpha_X(\tau)$ and the sequence $\delta(\tau)$ appearing in the definition of weak-mixing processes. We have:

Lemma 1. *Assume that $X = \{X_t\}_{t\geq 0}$ is a $(l \times 1)$ strongly mixing process and that:*

$$\sup_{t\geq 0} \sup_{1\leq i\leq l} \|X_{t,i}\|_{q'} < \infty$$

for some $q' > q$. Then $X = \{X_t\}_{t\geq 0}$ is (δ, q)-weak dependent, with, for $s \in \{1, \cdots, q\}$,

$$\delta(r) = \alpha(r)^{(q'-q)/q'}, \quad r \geq 1$$

$$C_s = 12 \left(\sup_{t\geq 0} \sup_{1\leq i\leq l} \|X_{t,i}\|_{s+q'-q} \right)^s$$

The proof is in section A.

3.3. Assumptions on the excitation sequence $\xi = \{\xi_k\}_{k\geq 0}$. In addition to the above stated assumptions on the matrix process $F = \{F_k\}_{k\geq 0}$, we need to impose some conditions on the excitation sequence $\xi = \{\xi_k\}_{k\geq 0}$. Typically, it is required that $\xi = \{\xi_k\}_{k\geq 0}$ is mixing (in a sense given below) and possesses a sufficient number of finite moments. To cover the application to the analysis of stochastic tracking algorithm, it is convenient to express these conditions in a slightly non conventional way, by defining a subset of process verifying a kind of generalized Rosenthal's inequality with random weights. This construction is explicited below.

Let $\mathcal{B}$ be a subfield of the basic probability space $(\Omega, \mathcal{A})$.

- Let $q \geq p > 0$ be two real numbers. For any $\eta = \{\eta_k\}_{k\geq 0}$ $(d \times 1)$ matrix-valued sequence, define

$$\mathcal{N}(p, q; \mathcal{B}) = \Bigg\{ \eta : \left\| \sum_{k=s}^{t} G_k \eta_k \right\|_p \leq \rho_{p,q}(\eta) \left(\sum_{k=s}^{t} \|G_k\|_q^2 \right)^{1/2} \quad \forall\, 0 \leq s \leq t \text{ and}$$
$$\forall\, G = \{G_k\}_{k\geq 0} \ (1 \times d) \text{ vector-valued process} \in L_q(\Omega, \mathcal{B}, P) \Bigg\} \tag{22}$$

In the previous definition, $\rho_{p,q}(\eta)$ is a finite constant, depending upon $\eta = \{\eta_k\}_{k\geq 0}$ but not on s, t or on the sequence $G = \{G_k\}_{k\geq 0}$. As above, it is of interest to state conditions upon which a process $\{\xi_k\}_{k\geq 0}$ belongs to the set $\mathcal{N}(p, q; \mathcal{B})$ and to give an upper bound for the value of the constant $\rho_{p,q}(\xi)$. To deal with the applications considered in this paper (see 10), we restrict our attention to the case where $\{\xi_t\}_{t\geq 0}$ may be decomposed as

$$\xi_t = M_t \epsilon_t \tag{23}$$

where the processes $M = \{M_t\}_{t\geq 0}$ and $\epsilon = \{\epsilon_t\}_{t\geq 0}$ verify

- **(X1)** $\{M_t\}_{t\in\mathbb{Z}}$ is a $(d \times l)$ $\mathcal{B}$-measurable matrix–valued process
- **(X2)** $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a $l \times 1$ vector–valued process,

- **(X3)** $\{M_t\}$ and $\{\epsilon_t\}$ are independent.

Recall that, in our application, M_t is either equal L_t (for the measurement noise) or to the identity I (for the lag-noise) and is measurable w.r.t to the σ-subfield generated by the regressor sequence $\sigma(\phi_s, 0 \leq s \leq t)$. The noise ϵ_t is either equal to the measurement noise v_t or to the lag-noise w_t. The assumptions (X1-X3) are thus most often met in our application.

Under these assumptions, it may be shown that $\{\xi_t\} \in \mathcal{N}(p, q; \mathcal{B})$ if $\{M_t\}$ is L_r-bounded (for large enough r) and $\{\epsilon_t\}$ is either a martingale difference or an α-mixing process, under appropriate conditions on the moment and (for α-mixing processes) on the rate of decay of the mixing coefficients. To be more specific, define the class of processes

$$\mathcal{N}'(p) = \left\{ \epsilon : \left\| \sum_{k=s}^{t} D_k \epsilon_k \right\|_p \leq \rho'_p(\epsilon) \left(\sum_{k=s}^{t} |D_k|^2 \right)^{1/2} \forall\, 0 \leq s \leq t \text{ and } \forall\, D = \{D_k\}_{k \in \mathbb{N}}\ (1 \times l) \text{ deterministic vector} \right\} \tag{24}$$

Proposition 1. *Assume that $\xi_t = M_t \epsilon_t$ and (X1-X3) . Let $q > p$. Assume in addition that:*

$$\sup_{k \geq 0} \|M_k\|_{pq/(q-p)} < \infty \text{ and } \epsilon \triangleq \{\epsilon_k\} \in \mathcal{N}'(p).$$

Then, $\{\xi_k\} \in \mathcal{N}(p, q; \mathcal{B})$ and

$$\rho_{p,q}(\xi) = \rho'_p(\epsilon) \sup_k \|M_k\|_{pq/(q-p)}. \tag{25}$$

Proof is given in section C. Some typical examples of processes belonging to $\mathcal{N}'(p)$ are given in the proposition below.

Proposition 2. *$\{\epsilon_t\}$ belongs to $\mathcal{N}'(p)$ if*

(i) *$\{\epsilon_t\}_{t \geq 0}$ is a L_p-bounded martingale increment: $\sup_{t \geq 0} \|\epsilon_t\|_p < \infty$. In such case, $\rho'_p(\epsilon)$ in (24) may be chosen to be*

$$\rho'_p(\epsilon) = B_p \sup_{s \geq 0} \|\epsilon_s\|_p \tag{26}$$

where B_p is a universal constant that depends only on p and l (the dimension of the $\{\epsilon_s\}$).

(ii) *$\{\epsilon_t\}_{t \geq 0}$ is α-mixing with mixing coefficients $\alpha_\epsilon(\tau)$. Moreover, there exists $\delta > 0$ such that:*

$$\sup_s \|\epsilon_s\|_{p+\delta} < \infty \text{ and } \sum_\tau \alpha_\epsilon(\tau)^{2\delta/(p(p+\delta))} < \infty. \tag{27}$$

In such case, the constant $\rho'_p(\epsilon)$ may be set to

$$\rho'_p(\epsilon) = B'_p \left(\sum_\tau \alpha_\epsilon(\tau)^{2\delta/p(p+\delta)} \right) \sup_s \|\epsilon_s\|_{p+\delta} \tag{28}$$

where B'_p is a universal constant which depends only on p and l.

(i) is a consequence of the Burkholder inequality for martingale increments (see Hall and Heyde (1980)). **(ii)** is a an application of Proposition 10, Corollary 11 by Kouritzin (1994). The proof is in section C.

To conclude this section, it is worthwhile to note that $\mathcal{N}(p,q;\mathcal{B})$ (resp. $\mathcal{N}'(p)$) are invariant by stable linear time-varying transformation. This means that, if $\{\eta_k\} \in \mathcal{N}(p,q;\mathcal{B})$ (resp. $\{\eta_k\} \in \mathcal{N}'(p)$), then the process $\{\epsilon_k\}$ defined as

$$\epsilon_k = \sum_{j=-\infty}^{\infty} A(k,j)\eta_{k-j} \qquad \sum_{j=-\infty}^{\infty} \sup_k |A(k,j)| < \infty$$

where $A(k,j)$ are deterministic matrices, also belongs to $\mathcal{N}(p,q;\mathcal{B})$ (resp. $\mathcal{N}'(p)$). Moreover, the constants are upper bounded by

$$\rho_{p,q}(\epsilon) \le \rho_{p,q}(\eta) \left(\sum_{j=-\infty}^{\infty} \sup |A(k,j)| \right), \tag{29}$$

$$\rho'_p(\epsilon) \le \rho'_p(\eta) \left(\sum_{j=-\infty}^{\infty} \sup |A(k,j)| \right) \tag{30}$$

Thus, random processes generated from martingale difference or α-mixing process via an infinite order time-varying linear filter can all be included in $\mathcal{N}'(p)$.

3.4. The Main Results. We may now formulate the central results of our contribution. The first result gives condition upon which $J_s^{(r)}$ is uniformly bounded in L_p and provides an expression for that bound.

Theorem 1. *Assume that, for some integer n and real numbers $q \ge p \ge 2$:*

- ***(i)*** $F = \{F_t\}_{t\ge 0}$ *is averaged exponentially stable:* $F \in \mathcal{S}(\beta, \mu_0)$,
- ***(ii)*** F *is* $(\delta, q(n+2))$ *weakly dependent, and* $\sum (r+1)^{(q(n+2)/2)-1}\delta(r) < \infty$,
- ***(iii)*** $\xi = \{\xi_t\} \in \mathcal{N}(p,q;F)$

Then, there exists a constant $K < \infty$ (depending on F and on the numerical constants p, q, n, μ_0, β but not *on $\{\xi_t\}$ or on the stepsize parameter μ), such that for all $0 < \mu \le \mu_0$, for all $0 \le r \le n$*

$$\sup_{s\ge 1} \|J_s^{(r)}\|_p \le K\, \rho_{p,q}(\xi)\mu^{(r-1)/2}. \tag{31}$$

where $\rho_{p,q}(\xi)$ is the constant defined in (22). (the precise value of constant K may be found from the proof).

Here $\mathcal{N}(p,q;F)$ is a shorthand notation for $\mathcal{N}(p,q;\mathcal{M}_0^{\infty}(F))$. The proof is given in section B. To complete our program, we need to bound the remainder term $H_s^{(n)}$. As shown below, under appropriate conditions, $H_s^{(n)}$ is uniformly bounded in L_q as soon as $J_s^{(n+1)}$ is L_r stable (for sufficiently large r) and the bound for $\sup_s \|H_s^{(n)}\|_q$ is proportional to $\sup_s \|J_s^{(n+1)}\|_r$.

Theorem 2. *Let $p \geq 2$ and let $a, b, c > 0$ such that $1/a + 1/b + 1/c = 1/p$. Let $n \in \mathbb{N}$. Assume that*

- $\{F_t\} \in \mathcal{S}(a, \beta, \mu_0)$ *for some* $\mu_0 > 0$ *and* $0 < \beta < 1/\mu_0$,
- $\sup_{s \geq 0} \|Z_s\|_b < \infty$ *and,*
- $\sup_{s \geq 0} \|J_s^{(n+1)}\|_c < \infty$.

Then, there exists a constant $K' < \infty$ (depending on the numerical constants a, b, c, β, μ_0, n and on the process $\{F_t\}$ but not on the process $\{\xi_t\}$ or on the stepsize parameter μ), such that for all $0 < \mu \leq \mu_0$,

$$\sup_{s \geq 0} \|H_s^{(n)}\|_p \leq K' \sup_{s \geq 0} \|J_s^{(n+1)}\|_c. \tag{32}$$

(the precise value of K' may be found from the proof)

4. Performance of adaptive tracking algorithms

A number of useful error bounds or expressions can be drawn from the results derived in the previous section. We use the notations introduced in section 1. Let n be a positive integer and p be a real number $p \geq 2$. Finally, let a, b, c, d be positive numbers such that $a^{-1} + b^{-1} + c^{-1} = p^{-1}$ and $d \geq c$. Denote: $F_t = L_t \phi_t^T$.

- **(H1)** $F = \{F_t\}$ is L_a and averaged exponentially stable: $F \in \mathcal{S}(\beta, \mu_0) \cap \mathcal{S}(a, \beta, \mu_0)$, for some $\mu_0 > 0$ and $0 < \beta < 1/\mu_0$.
- **(H2)** F is $(\delta, d(n+2))$-weak dependent; in addition,
$$\sup_{t \geq 0} \|F_t - E(F_t)\|_b < \infty \quad \text{and} \quad \sum (r+1)^{d(n+2)/2-1} \delta(r) < \infty.$$
- **(H3)** $\{L_t v_t\} \in \mathcal{N}(c, d; F)$ and $\{w_t\} \in \mathcal{N}(c, d; F)$.

As emphasized in proposition 2, assumption (H3) is verified when $\{w_t\}$ is a martingale increment or an α-mixing processes (under appropriate conditions on the moments and on the rate of convergence of the mixing coefficients). The parameter to track θ_t follows a random walk model, perhaps with non stationary and non independent increments. **(H3)** implies that $\|\theta_t\|_p / \sqrt{t}$ (but not $\|\theta_t\|_p$) is upper bounded.

Note that, under the additional assumption that the observation noise is independent from $\{F_t\}$, propositions 1 and 2 give practical criteria to establish that $\{L_t v_t\} \in \mathcal{N}(p, q; F)$.

According to (4), the tracking error may be expanded as $\tilde{\theta}_t = ^u \tilde{\theta}_t + \mu ^v \tilde{\theta}_t + ^w \tilde{\theta}_t$, where $^u\tilde{\theta}_t$, $^v\tilde{\theta}_t$ and $^w\tilde{\theta}_t$ are respectively defined in (5), (6) and (7). The terms $^v\tilde{\theta}_t$ and $^w\tilde{\theta}_t$ may further be decomposed

as

$$
{}^{v}\tilde{\theta}_t = \sum_{k=0}^{r_v} {}^{v}J_t^{(k)} + {}^{v}H_t^{(r_v)}
$$

$$
{}^{w}\tilde{\theta}_t = \sum_{k=0}^{r_w} {}^{w}J_t^{(k)} + {}^{w}H_t^{(r_w)}
$$

where r_v and r_w are two integers (not necessarily equal) such that $0 \leq r_v \leq n-1$ and $0 \leq r_w \leq n-1$. Under (H1)-(H3), theorems 1, 2 show that there exists a constant K (depending on $\{F_t\}$, the constants a, b, c, d, β, μ_0, but not on $\{v_t\}$, $\{w_t\}$ or on the stepsize μ), such that, for all $\mu \in (0, \mu_0]$

$$
\sup_t \|{}^{v}J_t^{(k)}\|_p \leq K\rho_{c,d}(Lv)\mu^{(k-1)/2}, \quad k \in \{0, \cdots, r_v\} \text{ and } \sup_t \|{}^{v}H_t^{(r_v)}\|_p \leq K\rho_{c,d}(Lv)\mu^{r_v/2}
$$

$$
\sup_t \|{}^{w}J_t^{(k)}\|_p \leq K\rho_{c,d}(w)\mu^{(k-1)/2}, \quad k \in \{0, \cdots, r_w\} \text{ and } \sup_t \|{}^{w}H_t^{(r_w)}\| \leq K\rho_{c,d}(w)\mu^{r_w/2}
$$

Summarizing these results, we have

Proposition 3. *Assume that (H1-H2-H3) hold, and let $r_v, r_w \in \{0, \cdots, n-1\}$. Then, there exists a constant $K < \infty$ (depending upon $a, b, c, d, n, \beta, \mu_0$ but not on $\{L_t v_t\}$ and $\{w_t\}$), such that, for all $\mu \in (0, \mu_0]$ and all $t \geq 0$*

$$
\|\tilde{\theta}_t - \mu \sum_{k=0}^{r_v} {}^{v}J_t^{(k)} - \sum_{k=0}^{r_w} {}^{w}J_t^{(k)}\|_p \leq K\left(\rho_{c,d}(Lv)\mu^{r_v/2+1} + \rho_{c,d}(w)\mu^{r_w/2}\right) + \|\Phi(t,0)\tilde{\theta}_0\|_p.
$$

Setting $r_v = r_w = 0$, one obtains the expansion of the tracking error bounds presented in Priouret and Ljung (1991), and latter extended for more general algorithms by Guo and Ljung (1995), Guo *et al* (1997). Higher-order expansions can be used to obtain refined approximation of the tracking error moments (see the discussion below). Tracking error bounds can trivially be derived from these formulas.

5. A worked-out example

In this section, approximate expressions of the tracking error covariance matrix for the LMS algorithm are derived. To illustrate our findings, it is shown in this section that first order approximation of the tracking error covariance may fail, in certain situation, to capture the behavior of the algorithm. It is argued that a second order expansion leads to significantly better approximation, in many situations of practical interest; moreover, second order approximation reveals the impact of certain factors which do not influence the first order approximation, in particular the dependence between the successive regression vectors. To illustrate these effects without obscuring the presentation with cumbersome notations and details, a very simple situation is considered. Theoretical results are validated by means of a small-scale Monte-Carlo experiment. More details will be given in a forthcoming paper.

- **(M1)** The regressor $\{\phi_t\}_{t\geq 0}$ is a strict-sense stationary vector autoregressive process

$$
\phi_{t+1} = \kappa\phi_t + u_{t+1}
$$

where κ ($-1 < \kappa < 1$) is a scalar, $\{u_t\}_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed Gaussian random vectors with zero-mean and covariance matrix $\sigma_u^2 I$.

- **(M2)** The measurement noise process $\{v_t\}_{t\geq 0}$ and the lag-noise process $\{w_t\}_{t\geq 0}$ are two sequences of zero-mean i.i.d random variables (vectors), with bounded moments of order r, where $r > 2$. Moreover: $E(w_0 w_0^T) = \gamma^2 I$.
- **(M3)** $\mathcal{M}_0^\infty(v)$, $\mathcal{M}_0^\infty(\phi)$ and $\mathcal{M}(\theta)$ are independent.

Because, our main concern in this section is the asymptotic regime, we set: $\tilde{\theta}_0 = 0$. To apply the results in section 4, we first need check that assumptions (M1-M3) imply (H1-H3). We set $b = c = 2r$, $a = 2r/r-2$ and $d = r+\delta$ where $\delta > 0$ but is otherwise arbitrary. Note that: $a^{-1}+b^{-1}+c^{-1} = p^{-1}$, with $p = 2$.

It follows from Mokkadem (1988) that, under (M1), $\{\phi_t\}$ is geometrically completely regular (see also Davydov (1973)), so that $F_t \triangleq \phi_t\phi_t^T$ is strongly mixing with exponentially decaying strong-mixing coefficient and thus, by lemma 1 (see section A), weak-dependent with exponentially decaying weak-dependent coefficient. Thus $Z_t = F_t - E(F_t)$ verifies the assumption of proposition 6, for any $n \geq 1$. It follows from Theorem 1 (Priouret and Veretenikov (1996)) that F_t is exponentially stable, *i.e.* for any $p \geq 1$ there exists $\mu_0 > 0$ and $0 < \beta < 1/\mu_0$ such that $\{F_t\} \in \mathcal{S}(\beta,\mu_0) \cap \mathcal{S}(p,\beta,\mu_0)$. Similarly, by applying propositions 1 and 2, we have: $\{\phi_t v_t\} \in \mathcal{N}(r, r+\delta; F)$ and $\{w_t\} \in \mathcal{N}(r,r,F)$, and

$$\rho_{r,r+\delta}(\phi v) = B_r\|\phi_0\|_{r(r+\delta)/\delta}\|v_0\|_r \quad\text{and}\quad \rho_{r,r}(w) = B_r\gamma\mu_r(w) \tag{33}$$

where $\mu_r(w)$ is the standardized moemt of order r of $\{w_t\}$ and B_r is a universal constant defined in (26). Hence, (H2) and (H3) are satisfied. Because $\{w_t\}_{t\geq 0}$ and $\{v_t\}_{t\geq 0}$ are independent, the processes $\{{}^v\tilde{\theta}_t\}_{t\geq 0}$ and $\{{}^w\tilde{\theta}_t\}$ are uncorrelated. Thus,

$$\Gamma = \lim_{t\to\infty} E(\tilde{\theta}_t\tilde{\theta}_t^T) = {}^v\Gamma\mu^2 + {}^w\Gamma$$

where ${}^v\Gamma = \lim_{t\to\infty} E({}^v\tilde{\theta}_t^v\tilde{\theta}_t^T)$ and ${}^w\Gamma = \lim_{t\to\infty} E({}^w\tilde{\theta}_t^w\tilde{\theta}_t^T)$. We wish to obtain approximate expressions for ${}^v\Gamma$ and ${}^w\Gamma$, denoted ${}^v\bar{\Gamma}$ and ${}^w\bar{\Gamma}$ such that, for all $\mu \in (0,\mu_0]$

$$|{}^v\Gamma -{}^v\bar{\Gamma}| \leq K\mu^{1/2} \quad\text{and}\quad |{}^w\Gamma -{}^w\bar{\Gamma}| \leq K\gamma^2\mu^{1/2}$$

where $K < \infty$ is some constant. To that purpose, we expand ${}^v\tilde{\theta}_t$ and ${}^w\tilde{\theta}_t$ to the second-order:

$$\begin{aligned}
{}^v\tilde{\theta}_t &= {}^vJ_t^{(0)} + {}^vJ_t^{(1)} + {}^vJ_t^{(2)} + {}^vH_t^{(2)}, \\
{}^w\tilde{\theta}_t &= {}^wJ_t^{(0)} + {}^wJ_t^{(1)} + {}^wJ_t^{(2)} + {}^wH_t^{(2)}
\end{aligned}$$

Under the stated assumptions, it follows from theorems 1 and 2 that, there exists some constant $C < \infty$, such that for all $\mu \in (0,\mu_0]$, we have

$$\sup_{t\geq 0}\left|E({}^vJ_t^{(1)}({}^vJ_t^{(2)} + {}^vH_t^{(2)})^T\right| \leq C\rho_{r,r+\delta}^2(\phi v)\mu^{1/2} \sup_{t\geq 0}\left|E({}^vJ_t^{(0)v}H_t^{(2)})\right| \leq C\rho_{r,r+\delta}^2(\phi v)\mu^{1/2} \tag{34}$$

$$\sup_{t\geq 0}\left|E({}^wJ_t^{(1)}({}^wJ_t^{(2)} +{}^w H_t^{(2)})^T\right| \leq C\gamma^2\mu_r^2(w)\mu^{1/2} \sup_{t\geq 0}\left|E({}^wJ_t^{(0)w}H_t^{(2)})\right| \leq C\gamma^2\mu_r^2(w)\mu^{1/2} \tag{35}$$

It remains to evaluate $\lim_{t\to\infty} E({}^vJ_t^{(0)v}J_t^{(i)})$, $\lim_{t\to\infty} E({}^wJ_t^{(0)w}J_t^{(i)})$, $i = 0,1,2$ and $E({}^vJ_t^{(1)v}J_t^{(1)})$ and $E({}^wJ_t^{(1)w}J_t^{(1)})$. Denote: $\alpha = \sigma_u^2/(1-\kappa^2)$. Tedious but straightforward calculations show that

$$\begin{aligned}
\lim_{t\to\infty} E({}^vJ_t^{(0)v}J_t^{(0)^T}) &= \frac{\sigma_v^2}{\mu(2-\mu\alpha)} I, \\
\lim_{t\to\infty} E({}^vJ_t^{(0)v}J_t^{(1)^T}) &= -\frac{\kappa^2\sigma_v^2(d+1)\alpha}{2(1-\kappa^2)} I + O(\mu), \\
\lim_{t\to\infty} E({}^vJ_t^{(0)v}J_t^{(2)^T}) &= \frac{\kappa^2\sigma_v^2\alpha(d+1)\alpha}{4(1-\kappa^2)} I + O(\mu), \\
\lim_{t\to\infty} E({}^vJ_t^{(1)v}J_t^{(1)^T}) &= \frac{(1+\kappa^2)\sigma_v^2\alpha(d+1)}{4(1-\kappa^2)} I + O(\mu)
\end{aligned}$$

yielding the following expression for ${}^v\bar{\Gamma}$

$$ {}^v\bar{\Gamma} = \frac{\sigma_v^2}{2\mu} I + \alpha\sigma_v^2 \frac{d+2}{4} I + O(\mu) \tag{36}$$

It is worthwhile to note that the first order correction does not depend upon the autoregressive coefficient κ, *i.e.* the dependence among the successive regressors does not influence the covariance ${}^v\bar{\Gamma}$ up to the second order in the stepsize μ. It may be shown that this results holds under much weaker assumptions on the regression sequence $\{\phi_t\}$ (see Perrier *et al* (1995) for a more general discussion), as long as $\{v_t\}$ is a martingale increment and is independent from $\{\phi_t\}$. Similarly, it can be shown that

$$\begin{aligned}
\lim_{t\to\infty} E({}^wJ_t^{(0)w}J_t^{(0)^T}) &= \frac{\gamma^2}{\mu\alpha(2-\mu\alpha)} I, \\
\lim_{t\to\infty} E({}^wJ_t^{(0)w}J_t^{(1)^T}) &= 0, \\
\lim_{t\to\infty} E({}^wJ_t^{(0)w}J_t^{(2)^T}) &= \frac{\gamma^2\kappa^2(d+1)}{4(1-\kappa^2)} I + O(\gamma^2\mu), \\
\lim_{t\to\infty} E({}^wJ_t^{(1)w}J_t^{(1)^T}) &= \frac{\gamma^2(1+\kappa^2)(d+1)}{4(1-\kappa^2)} I + O(\gamma^2\mu).
\end{aligned}$$

yielding the following approximate expression for ${}^w\Gamma$,

$$ {}^w\bar{\Gamma} = \frac{\gamma^2}{2\mu\alpha} I + \frac{\gamma^2}{4}\left(1 + (d+1)\frac{1+2\kappa^2}{1-\kappa^2}\right) I + O(\gamma^2\mu). \tag{37}$$

It is interesting to note that the first order correction depends upon the autoregressive coefficient κ : when κ is close to one, the correction term becomes large. This behavior is illustrated in figures 5 and 5, where the asymptotic tracking error variance $\lim_{t\to\infty} \|\tilde{\theta}_t\|^2$ is displayed as a function of the stepsize μ. In both cases, we set: $\gamma = 0.05$, $d = 10$, $\sigma_v^2 = 3$ and, for every value o κ, $\sigma_u^2 = 1-\kappa^2$ (so that $\alpha = 1$). Two values of κ are considered: $\kappa = 0$ (figure 5) and $\kappa = 0.9$ (figure 5)). On the figures, the value of the asymptotic tracking error variance obtained by Monte-Carlo simulations (solid-line) are compared with the approximate expressions obtained by (i) retaining only the first term in (36) and (37) (dashed line) or (ii) including the two terms in (36) and (37) (dashed-dotted line). As seen on these figures, the autoregressive coefficient strongly affects the asymptotic tracking error

variance, as predicted by the second-order approximation (whereas the first-order approximation does not predict any effect with respect to the variation of the autoregressive parameter). It is however interesting to note that the optimal value for the stepsize (the value which minimizes the asymptotic tracking error variance) obtained by minimizing the second-order approximation does not vary with κ and is reasonably close to the one obtained by minimizing the first order approximation.

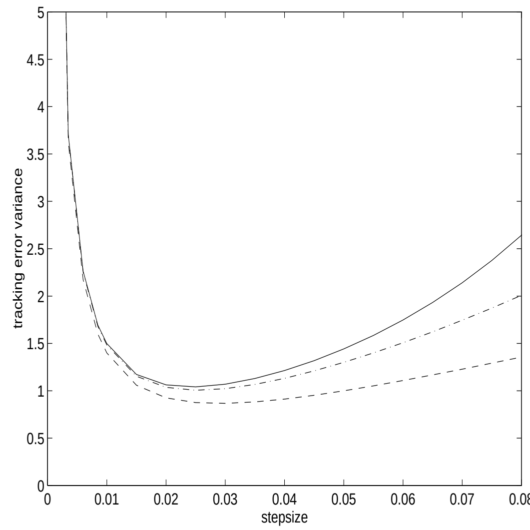

FIGURE 1. $\kappa = 0$. solid line: Monte-Carlo simulation. Dashed line: first order approximation. Dashed-dotted line: second-order approximation

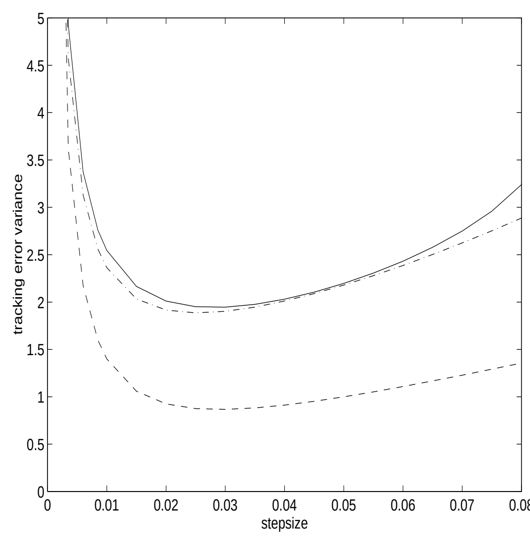

FIGURE 2. $\kappa = 0.9$. solid line: Monte-Carlo simulation. Dashed line: first order approximation. Dashed-dotted line: second-order approximation

Appendix A. Rosenthal's type inequalities for weak dependent sequences

In this section, some useful properties on weak dependent sequences are given. The most useful result in this section is the proposition 6, which can be seen as an extension of Rosenthal's inequality: this proposition is repeatedly used in the sequel. Most of the material in this section is already known. We have found however convenient to gather the most important results and proofs for completeness.

Proposition 4. *Let $G = \{G_t\}_{t\geq 0}$ be a $(d \times d)$ zero-mean matrix-valued process. Let q be an* even *integer and $j \in \mathbb{N}$. Assume that G is (δ, qj)-weak dependent. Assume in addition that*

$$\sum_{r=0}^{\infty}(r+1)^{qj/2-1}\delta(r) < \infty. \tag{38}$$

Then, there exists a finite constant $\bar{D}_{q,j}(G)$, such that:

$$\left\| \sum_{s\leq i_1<\cdots<i_j\leq t} G_{i_1}\cdots G_{i_j} \right\|_q \leq \bar{D}_{q,j}(G)\ (t-s)^{j/2} \tag{39}$$

for all $0 \leq s \leq t$.

Proof of Proposition 4: The proof is a direct application of the following result:

Proposition 5. *Let $q \geq 2$ and let $X = \{X_t\}_{t\geq 0}$ be a $(l \times 1)$ be a zero-mean ($E(X_t) = 0$) vector-valued (δ, q)-weak dependent process. Assume in addition that $\sum_0^\infty (r+1)^{q/2-1}\delta(r) < \infty$. Then, there exist finite constants $\gamma = \{\gamma_2, \cdots, \gamma_q\}$ such that, for all $1 \leq s \leq q$ and all $n \geq 1$,*

$$\sup_{(i_1,\cdots,i_s)} \left(\sum_{1\leq t_1,\cdots,t_s\leq n} |E(X_{t_1,i_1}\cdots X_{t_s,i_s})| \right) \ \leq\ \gamma_s\ s!\ n^{s/2} \tag{40}$$

Remark The constants $\gamma_2, \cdots, \gamma_q$, can be evaluated recursively as follows. Let $\sigma_s = \sum_{r=0}^\infty (r+1)^{s/2-1}\delta(r)$, for $1 < s \leq q$. Set $\gamma_1 = 0$, $\gamma_2 = C_2\sigma_2$ and, evaluate, for $2 < s \leq q$

$$\gamma_s = \sum_{m=1}^{s-1} \gamma_m \gamma_{s-m} + (s-1) C_s \sigma_s. \tag{41}$$

Proof of Proposition 5: The proof is adapted from Doukhan and Louhichi (1997). Define, for $1 < s \leq q$:

$$I(n,s) = \{(t_1, t_2, \cdots, t_s) : \ 1 \leq t_1 \leq \cdots \leq t_s \leq n\} \tag{42}$$

$$A(n,s) = \sup_{(i_1,\cdots,i_s)} \sum_{I(n,s)} |E(X_{t_1,i_1}\cdots X_{t_s,i_s})|. \tag{43}$$

Note that

$$\sup_{(i_1,\cdots,i_s)} \sum_{1\leq t_1,\cdots,t_s\leq n} |E(X_{t_1,i_1}\cdots X_{t_s,i_s})| \leq s! A(n,s)$$

Define for $1 \le m \le s-1$ and $0 \le r \le n-1$, the sets

$$I(n,s,m,r) = \{(t_1,\cdots,t_s) \in I(n,s) : \; t_{m+1} - t_m = r = \max(t_{i+1}-t_i)\},$$
$$I(n,s,m) = \bigcup_{r=0}^{n-1} I(n,s,m,r).$$

It is easily seen that

$$I(n,s) = \bigcup_{m=1}^{s-1} I(n,s,m) = \bigcup_{m=1}^{s-1}\bigcup_{r=0}^{n-1} I(n,s,m,r)$$

and the cardinal of set $I(n,s,m,r)$ is bounded by $n(r+1)^{s-2}$. Let $1 \le m \le s-1$. Note that

$$E(X_{t_1,i_1}\cdots X_{t_s,i_s}) =$$
$$E(X_{t_1,i_1}\cdots X_{t_m,i_m})E(X_{t_{m+1},i_{m+1}}\cdots X_{t_s,i_s}) + \operatorname{cov}(X_{t_1,i_1}\cdots X_{t_m,i_m}, X_{t_{m+1},i_{m+1}}\cdots X_{t_s,i_s})$$

This implies, under the weak-dependence condition, that

$$\sup_{(i_1,\cdots,i_s)} \sum_{I(n,s,m)} |E(X_{t_1,i_1}\cdots X_{t_s,i_s})| \le A(n,m)A(n,s-m) + C_s \sum_{r=0}^{n-1} \delta(r) n (r+1)^{s-2}$$

For $0 \le r \le n-1$, $n(r+1)^{s-2} \le n^{s/2}(r+1)^{s/2-1}$. Thus:

$$A(n,s) \le \sum_{m=1}^{s-1} A(n,m)A(n,s-m) + (s-1)C_s n^{s/2} \sum_{r=0}^{n-1} (r+1)^{s/2-1}\delta(r) \tag{44}$$

The proof and the expression of the constant is obtained by a straightforward induction.

□

Proposition 6. *Let $p \ge 1$ and $n \in \mathbb{N}$. Let $G = \{G_t\}_{t\ge 0}$ be a $(d\times d)$ zero-mean matrix-valued process. Assume that $(\delta, p(n+2))$−weak dependent and that*

$$\sum (r+1)^{p(n+2)/2-1}\delta(r) < \infty. \tag{45}$$

Then, there exists a finite constant $D_{p,n}(G)$, such that for all $j \in \{1,\cdots,n\}$ and all $0 \le s \le t < \infty$, we have

$$\left\| \sum_{s\le i_1<\cdots<i_j\le t} G_{i_1}\cdots G_{i_j} \right\|_{pn/j} \le D_{p,n}(G)(t-s)^{j/2}. \tag{46}$$

Proof of Proposition 6: For $j \in \{1,\cdots,n\}$, denote $q(n,j)$ the smallest even integer verifying $pn/j \le q(n,j)$. It is easily seen that $pn/j \le q(n,j) \le pn/j+2$, which implies that $\max_{j\in\{1,\cdots,n\}} jq(n,j) \le (p+2)n$. Under the stated assumption, Proposition 4 implies that, for all $0 \le s \le t$

$$\| \sum_{s\le i_1<\cdots<i_j\le t} G_{i_1}\cdots G_{i_j}\|_{pn/j} \le \| \sum_{s\le i_1<\cdots<i_j\le t} G_{i_1}\cdots G_{i_j}\|_{q(n,j)},$$
$$\le \bar{D}_{q(n,j),j}(G)(t-s)^{j/2}$$

which concludes the proof.

□

We conclude this section by the proof of lemma 1. This proof is directly adapted from the Doukhan and Louhichi (1997) (see also Yokoyama (1981) for similar results and arguments).

Proof of Lemma 1: Set $\zeta = q' - q$. Let $a, b, c \geq 1$ be constants such that $a^{-1} + b^{-1} + c^{-1} = 1$, and $\mathcal{G}$, $\mathcal{F} \subset \mathcal{A}$ be two σ-algebra. Suppose that $\xi \in \mathcal{L}_b(\Omega, \mathcal{G}, P)$ and $\eta \in \mathcal{L}_c(\Omega, \mathcal{F}, P)$. It is known that (Deo,1973),

$$|E(\xi\eta) - E(\xi)E(\eta)| \leq 12(\alpha(\mathcal{G}, \mathcal{H}))^{1/a} \|\xi\|_b \|\eta\|_c \tag{47}$$

where

$$\alpha(\mathcal{G}, \mathcal{H}) \triangleq \sup_{A \in \mathcal{G}, B \in \mathcal{H}} |P(A \cap B) - P(A)P(B)| .$$

Set $s \in \{1, \cdots, q\}$, $m \in \{1, \cdots, m-1\}$. Finally, let $1 \leq t_1 \leq \cdots \leq t_s$ be an arbitrary ordered s-uplet with $t_{m+1} - t_m = r$. Apply (47) with $\xi = X_{t_1,i_1} \cdots X_{t_m,i_m}$, $\eta = X_{t_{m+1},i_{m+1}} \cdots X_{t_s,i_s}$, $\mathcal{G} = \mathcal{M}_{-\infty}^{t_m}(X)$, $\mathcal{H} = \mathcal{M}_{t_{m+1}}^{\infty}(X)$, and $a = (s+\zeta)/\zeta$, $b = (s+\zeta)/m$ and $c = (s+\zeta)/(s-m)$:

$$\begin{aligned} &\left|\mathrm{cov}(X_{t_1,i_1} \cdots X_{t_m,i_m}, X_{t_{m+1},i_{m+1}} \cdots X_{t_s,i_s})\right| \\ &\qquad \leq 12 \|X_{t_1,i_1} \cdots X_{t_m,i_m}\|_{(s+\zeta)/m} \|X_{t_{m+1},i_{m+1}} \cdots X_{t_s,i_s}\|_{(s+\zeta)/(s-m)} \alpha(r)^{\zeta/s+\zeta} \end{aligned} \tag{48}$$

Holder inequality yields

$$\|X_{t_1,i_1} \cdots X_{t_m,i_m}\|_{(s+\zeta)/m} \leq \prod_{j=1}^{m} \|X_{t_j,i_j}\|_{s+\zeta},$$

$$\|X_{t_{m+1},i_{m+1}} \cdots X_{t_s,i_s}\|_{(s+\zeta)/(s-m)} \leq \prod_{j=m+1}^{s} \|X_{t_j,i_j}\|_{s+\zeta}.$$

which, together with (48) implies

$$\left|\mathrm{cov}(X_{t_1,i_1} \cdots X_{t_m,i_m}, X_{t_{m+1},i_{m+1}} \cdots X_{t_s,i_s})\right| \leq 12 \left(\sup_{t \geq 0} \sup_{1 \leq i \leq l} \|X_{t,i}\|_{s+\zeta} \right)^s \alpha(r)^{\zeta/s+\zeta}$$

which concludes the proof

□.

Appendix B. Proof of theorems 1 and 2

In the first part of the section, a sketch of the proof of given; technical lemmas are given at the end of the section.

Proof of Theorem 1: Before going further, we need some additional notations. Define: $S_0(t, s) = \psi(t, s)$ for $t \geq s$ and $S_0(t, s) = 0$ for $s > t$. For $k \geq 1$, define $S_k(t, s)$ recursively as

$$S_k(t, s) = \sum_{u=s}^{t} \psi(t, u) Z_u S_{k-1}(u-1, s) = \sum_{u=s}^{t} S_{k-1}(t, u) Z_u \psi(u-1, s). \tag{49}$$

Denote respectively

$$D_1(t,s) = \sum_{u=s}^{t} Z_u, \quad t \geq s, \quad D_1(t,s) = 0, \quad s > t, \tag{50}$$

$$D_k(t,s) = \sum_{u=s}^{t} D_{k-1}(t,u+1) Z_u \tag{51}$$

Note that, by construction, for $k \geq 0$

$$S_k(t,s) = 0, \quad s > t-k, \quad \text{and } D_k(t,s) = 0, \quad s > t-k+1.$$

From (18), (19) and (49), it is easily seen that $J_{t+1}^{(r)}$ may decomposed as

$$J_{t+1}^{(r)} = \mu^r \sum_{s=0}^{t} S_r(t,s) \xi_s \tag{52}$$

Let r be an integer, and c, μ_0, β some real numbers such that $c > 0$, $\mu_0 > 0$ and $0 < \beta < 1/\mu_0$. Consider $\{W(u,v;\mu)\}$ a process indexed by $(u,v) \in \mathbb{N} \times \mathbb{N}$ and $\mu \in \mathbb{R}^+$. We say that $W(u,v;\mu)$ belongs to the set $\mathcal{L}(\beta, \mu_0, c, r)$, if there exists a constant $C_{\mathcal{L}}(W) < \infty$, such that, for all $0 \leq u \leq v$, and $0 < \mu \leq \mu_0$

$$\|W(u,v;\mu)\|_c \leq C_{\mathcal{L}}(W)(1-\beta\mu)^{v-u}(v-u)^{r/2} \sum_{k=0}^{r} \mu^k (v-u)^k. \tag{53}$$

It is easily seen that the subspaces $\mathcal{L}(\beta, \mu_0, c, r)$ are nested in the sense that for $c' \geq c$ and $r' \geq r$, we have $\mathcal{L}(\beta, \mu_0, c, r) \subset \mathcal{L}(\beta, \mu_0, c', r')$.

Lemma 2. *Assume that* $S_r \triangleq \{S_r(t,s)\} \in \mathcal{L}(\beta, \mu_0, q, r)$ *and* $\xi \triangleq \{\xi_k\} \in \mathcal{N}(p,q;F)$. *Then, there exists a constant* $C(S_r) < \infty$ *(depending upon* β, μ_0, p, q, r *and* $\{F_k\}$ *but not on* $\{\xi_k\}$ *or on the stepsize* μ*) such that*

$$\sup_{t \geq 0} \|J_t^{(r)}\|_p \leq C(S_r)\ \rho_{p,q}(\xi) \mu^{(r-1)/2} \quad \forall\, \mu \ \in \ (0, \mu_0].$$

Proof of Lemma 2: Since $\{\xi_t\} \in \mathcal{N}(p,q;F)$, there exists a constant $\rho_{p,q}(\xi)$ such that

$$\|J_{t+1}^{(r)}\|_p \leq \rho_{p,q}(\xi) \mu^r \left(\sum_{s=0}^{t} \|S_r(t,s)\|_q^2 \right)^{1/2}$$

Under the stated assumptions, $S_r(t,s) \in \mathcal{L}(\beta, \mu_0, q, r)$: there exists a constant $C_{\mathcal{L}}(S_r) < \infty$ such that, for all $t \geq s \geq 0$:

$$\|S_r(t,s)\|_q \leq C_{\mathcal{L}}(S_r)\ (1-\beta\mu)^{t-s}(t-s)^{r/2} \sum_{k=0}^{r} \mu^k (t-s)^k \quad \forall\, \mu \ \in \ (0, \mu_0].$$

This latter relation implies that

$$\sum_{s=0}^{t} \|S_r(t,s)\|_q^2 \leq r C_{\mathcal{L}}^2(S_r)\ \sum_{k=0}^{r} \mu^{2k} \sum_{u=0}^{t} (1-\beta\mu)^{2u} u^{2k+r} \quad \forall\, \mu \ \in \ (0, \mu_0].$$

The proof is concluded by noting that, for all $\alpha \geq 0$, there exists a constant $D_\alpha < \infty$, such that, for all $0 < \mu \leq 1/\beta$,

$$\sum_{u=0}^{\infty} (1-\beta\mu)^{2u} u^\alpha \leq D_\alpha / (\beta\mu)^{1+\alpha}$$

The previous lemma will allow us to conclude the proof if we are able to prove that

$$\{S_r(t,s)\} \in \mathcal{L}(\beta, \mu_0, q, r), \text{ for } r \in \{1, \cdots, n\}.$$

In fact, we will prove a slightly stronger property: $S_r(t,s) \in \mathcal{L}(\beta, \mu_0, qn/r, r)$, for $r \in \{1, \cdots, n\}$. This part is more intricate, and requires some preparatory lemmas.

Lemma 3. *Let $j \in \{1, \cdots, r-1\}$. Denote: $\Delta_j(v,u) = D_j(v,u) - D_j(v, u+1)$. We have, for $t \geq s$,*

$$\sum_{u=s}^{t} \psi(t,u)\Delta_j(t,u) S_{r-j}(u-1,s) - \sum_{u=s}^{t} \psi(t,u)\Delta_{j+1}(t,u) S_{r-j-1}(u-1,s) = \mu \sum_{u=s}^{t} \psi(t,u)\bar{F}_{u-1} D_j(t,u) S_{r-j}(u-2,s) - \mu \sum_{u=s}^{t} \psi(t,u) D_j(t,u) \bar{F}_{u-1} S_{r-j}(u-2,s)$$

$$\sum_{u=s}^{t} \psi(t,u)\Delta_r(t,u)\psi(u-1,s) - \psi(t,s+1) D_r(t,s+1) = \mu \sum_{u=s}^{t} \psi(t,u)\bar{F}_{u-1} D_r(t,u)\psi(u-2,s) - \mu \sum_{u=s}^{t} \psi(t,u) D_r(t,u)\bar{F}_{u-1}\psi(u-2,s).$$

Proof of Lemma 3: The proof involves only simple algebraic manipulations. First note that

$$\sum_{u=s}^{t} \psi(t,u)\Delta_j(t,u) S_{r-j}(u-1,s) = \sum_{u=s}^{t} \left(\psi(t,u) - \psi(t,u-1)\right) D_j(t,u) S_{r-j}(u-2,s) + \sum_{u=s}^{t} \psi(t,u) D_j(t,u) \left(S_{r-j}(u-1,s) - S_{r-j}(u-2,s)\right)$$

For $t \geq s$ we have

$$\psi(t,s) - \psi(t,s-1) = \mu\psi(t,s)\bar{F}_{s-1},$$

$$\psi(t,s) - \psi(t-1,s) = \begin{cases} -\mu\bar{F}_t\psi(t-1,s) & t > s \\ I & t = s \end{cases}$$

$$S_k(t,s) - S_k(t-1,s) = Z_t S_{k-1}(t-1,s) - \mu\bar{F}_t S_k(t-1,s).$$

The proof of this lemma is concluded by noting that

$$D_j(t,u) Z_{u-1} = D_{j+1}(t,u-1) - D_{j+1}(t,u).$$

$\square$Consider $\{V(u,v)\}$ a process indexed by $(u,v) \in \mathbb{N} \times \mathbb{N}$. We say that $V(u,v)$ belongs to the set $\mathcal{M}(c,r)$, if, there exists $C_{\mathcal{M}}(V) < \infty$, such that, for all $u \leq v$, we have

$$\|V(v,u)\|_c \leq C_{\mathcal{M}}(V)(v-u)^{r/2}. \tag{54}$$

Lemma 4. *Let $r \in \mathbb{N}$ and $j \in \{1, \cdots, r-1\}$ and let $V \triangleq \{V(u,v)\}$ and $W \triangleq \{W(u,v;\mu)\}$ be two processes such that $V \in \mathcal{M}(cr/j, j)$ and $W \in \mathcal{L}(\beta, \mu_0, cr/(r-j), r-j)$, for some $c > 0$. Then, the process $\{U(t,s;\mu)\}$ defined by*

$$U(t,s;\mu) = \mu \sum_{u=s}^{t-1} \psi(t,u) V(t,u) W(u,s;\mu)$$

belongs to the set $\mathcal{L}(\beta, \mu_0, c, r)$.

Proof of Lemma 4: The proof is elementary. Let $t \geq s \geq 0$. Since $\{F_t\} \in \mathcal{S}(\beta, \mu_0)$, $|\psi(t,s)| \leq K_{\beta,\mu_0}(F)(1-\beta\mu)^{t-s}$ for all $\mu \in (0, \mu_0]$ and $0 \leq s \leq t$. This implies

$$\|U(t,s;\mu)\|_c \leq K_{\beta,\mu_0}(F)\mu \sum_{u=s}^{t} (1-\beta\mu)^{t-u} \|V(t,u)\|_{cr/j} \|W(u,s;\mu)\|_{cr/(r-j)},$$
$$\leq K_{\beta,\mu_0}(F)\mu \sup_{s \leq u \leq t} \left(\|V(t,u)\|_{cr/j} \right) \sup_{s \leq u \leq t} \left(\|W(u,s;\mu)\|_{cr/(r-j)} / (1-\beta\mu)^{u-s} \right) (t-s)(1-\beta\mu)^{t-s}.$$

The proof follows by noting that, under the stated assumptions, there exist constants $C_{\mathcal{M}}(V) < \infty$ and $C_{\mathcal{L}}(W) < \infty$ such that, for all $0 \leq s \leq t$, and all $0 < \mu \leq \mu_0$

$$\sup_{s \leq u \leq t} \|V(t,u)\|_{cr/j} \leq C_{\mathcal{M}}(V)(t-s)^{j/2},$$

$$\sup_{s \leq u \leq t} \left(\|W(u,s;\mu)\|_{cr/(r-j)} / (1-\beta\mu)^{u-s} \right) \leq C_{\mathcal{L}}(W)(t-s)^{(r-j)/2} \left(\sum_{k=0}^{r-j} \mu^k (t-s)^k \right)$$

By combining the two preceding lemmas, we obtain the following useful criterion

Lemma 5. *Assume that $\{F_t\} \in \mathcal{S}(\beta, \mu_0)$. Let $c > 0$ and $r \in \mathbb{N}$. Assume in addition that $\{D_j(t,s)\} \in \mathcal{M}(cr/j, j)$ for $1 \leq j \leq r$ and $\{S_j(t,s)\} \in \mathcal{L}(\beta, \mu_0, cr/j, j)$ for $1 \leq j < r$. Then, $\{S_r(t,s)\} \in \mathcal{L}(\beta, \mu_0, c, r)$.*

Proof of Lemma 5: By iterated application of Lemma 3, we have for $t \geq s \geq 0$

$$S_r(t,s) = \sum_{u=s}^{t} \psi(t,u) \Delta_r(t,u) \psi(u-1,s) + \sum_{j=1}^{r-1} G_j(t,s), \tag{55}$$

where the processes $G_j(t,s)$ are defined, for $1 \leq j \leq r-1$ as

$$G_j(t,s) = \mu \sum_{u=s}^{t} \psi(t,u) \bar{F}_{u-1} D_j(t,u) S_{r-j}(u-2,s) - \mu \sum_{u=s}^{t} \psi(t,u) D_j(t,u) \bar{F}_{u-1} S_{r-j}(u-2,s).$$

Under the stated assumptions, Lemma 4 shows that $G_j(t,s) \in \mathcal{L}(\beta, \mu_0, c, r)$, for $1 \leq j \leq r-1$. The first term on the RHS of (55) may be decomposed (Lemma 3) as

$$\sum_{u=s}^{t} \psi(t,u) \Delta_r(t,u) \psi(u-1,s) = \psi(t,s+1) D_r(t,s+1)$$
$$+ \mu \left(\sum_{u=s}^{t} \psi(t,u) \bar{F}_{u-1} D_r(t,u) \psi(u-2,s) - \sum_{u=s}^{t} \psi(t,u) D_r(t,u) \bar{F}_{u-1} \psi(u-2,s) \right)$$

Since $D_r(t,s) \in \mathcal{M}(c,r)$, there exists a constant $C_{\mathcal{M}}(D_r) < \infty$, such that $\|D_r(t,s)\|_c \leq C_{\mathcal{M}}(D_r)(t-s)^{r/2}$ for all $t \geq s \geq 0$. Since $\{F_t\} \in \mathcal{S}(\beta,\mu_0)$, we have

$$\left\| \sum_{u=s}^{t} \psi(t,u)\Delta_r(t,u)\psi(u-1,s) \right\|_c \leq K_{\beta,\mu_0}(F) C_{\mathcal{M}}(D_r)(1-\beta\mu)^{t-s}(t-s)^{r/2}(1+\mu \sup_{s\geq 0} |\bar{F}_s|(t-s))$$

for all $t \geq s \geq 0$ and all $0 < \mu \leq \mu_0$, which concludes the proof.

□

Under assumption **(ii)** of theorem 1, an application of proposition 6 shows that: $\{D_r(t,s)\} \in \mathcal{M}(qn/r,r)$, $r \in \{1,\cdots,n\}$; lemma 5 leads us to a condition upon which $S_r(t,s)$ belongs to $\mathcal{L}(\beta,\mu_0,qn/r,r)$, $r \in \{1,\cdots,n\}$.

Lemma 6. *Under the assumptions of theorem 1, it holds that*

$$\textit{for} \quad r \in \{1,\cdots,n\}, \quad \{S_r(t,s)\} \in \mathcal{L}(\beta,\mu_0,qn/r,r).$$

Proof of Lemma 6: The proof is by induction on r. By application of proposition 6, we have: $D_1(t,s) \in \mathcal{M}(qn,1)$. This implies by Lemma 5 that $S_1(t,s) \in \mathcal{L}(\beta,\mu_0,qn,1)$. Assume now that the property is verified up to order $r-1$, with $1 < r \leq n$. Set $c = qn/r$. By application of proposition6, we have $D_j(t,s) \in \mathcal{M}(qn/j,j) = \mathcal{M}(cr/j,j)$. The induction hypothesis implies that $S_j(t,s) \in \mathcal{L}(\beta,\mu_0,qn/j,j) = \mathcal{L}(\beta,\mu_0,cr/j,j)$, for $1 \leq j < r$. We have $S_r(t,s) \in \mathcal{L}(\beta,\mu_0,c,r) = \mathcal{L}(\beta,\mu_0,qn/r,r)$ by Lemma 5, which concludes the proof.

□

The proof of theorem 1 is concluded by applying Lemma 2 and Lemma 6.

Proof of Theorem 2: Solving recursively the difference equation (20), we may express $H_{t+1}^{(n+1)}$ as a linear combination of $J_s^{(n+1)}$, with random matrix-valued weights $\Phi(t,s)$

$$H_{t+1}^{(n+1)} = \mu \sum_{s=0}^{t} \Phi(t,s) Z_s J_s^{(n+1)}. \tag{56}$$

Since $\{F_t\} \in \mathcal{S}(a,\beta,\mu_0)$ we have, for all $0 < \mu \leq \mu_0$, and all $0 \leq s \leq t$, $\|\Phi(t,s)\|_a \leq K'_{a,\beta,\mu_0}(F)(1-\beta\mu)^{t-s}$, which implies

$$\|H_{t+1}^{(n+1)}\|_p \leq K'_{a,\beta,\mu_0}(F)\beta^{-1} \sup_{s\geq 0} \|Z_s\|_b \sup_{s\geq 0} \|J_s^{(n+1)}\|_c.$$

By construction, $H_s^{(n)}$ may be decomposed as $H_s^{(n)} = J_s^{(n+1)} + H_s^{(n+1)}$; the Minkowski inequality implies

$$\|H_s^{(n)}\|_p \leq \|J_s^{(n+1)}\|_p + \|H_s^{(n+1}\|_p \leq \left(1 + K'_{a,\beta,\mu_0}(F)\beta^{-1} \sup_{s\geq 0} \|Z_s\|_b\right) \sup_{s\geq 0} \|J_s^{(n+1)}\|_c.$$

which concludes the proof.

□

Appendix C. Proof of Propositions 1 and 2

Proof of Proposition 1: Denote $\mathcal{M}_0^\infty(\epsilon)$ and $\mathcal{M}_0^\infty(F)$ the σ-field generated by $\{\epsilon_t\}_{t\geq 0}$ and $\{F_t\}_{t\geq 0}$, respectively. Let $0 \leq s \leq t$ and $\{G_k\}$ be a process measureable w.r.t $\mathcal{M}_0^\infty(F)$ and bounded in L_q (but otherwise arbitrary). We have, under the stated assumptions

$$E\left(\left|\sum_{k=s}^{t} G_k M_k \epsilon_k\right|^p\right) = E\left(E\left(\left|\sum_{k=s}^{t} G_k M_k \epsilon_k\right|^p \mid \mathcal{M}_0^\infty(F)\right)\right), \tag{57}$$

$$E\left(\left|\sum_{k=s}^{t} G_k M_k \epsilon_k\right|^p \mid \mathcal{M}_0^\infty(F)\right) \leq [\rho'_p(\epsilon)]^p \left(\sum_{k=s}^{t} |G_k M_k|^2\right)^{p/2} \tag{58}$$

The proof is concluded by applying the Minkowski inequality and the Holder inequality

$$E\left(\left(\sum_{k=s}^{t} |G_k M_k|^2\right)^{p/2}\right) \leq \left(\sum_{k=s}^{t} \|G_k M_k\|_p^2\right)^{p/2}, \tag{59}$$

$$\leq \sup_{k\geq 0} \|M_k\|_{pq/(q-p)}^p \left(\sum_{k=s}^{t} \|G_k\|_q^2\right)^{p/2} \tag{60}$$

Proof of Proposition 2: Its is assumed (without any loss of generality) that $\{D_k\}_{k\geq 0}$ and $\{\epsilon_k\}_{k\geq 0}$ are scalar-valued. Assume first that $\{\epsilon_k\}_{k\geq 0}$ is a L_p stable martingale increment. According to the Burkholder's inequality (Hall and Heyde (1981)), there exists a universal constant $B_p < \infty$ (independent of the sequence of scalar weights $\{D_k\}_{k\geq 0}$ and on the process $\{\epsilon_k\}_{k\geq 0}$) such that

$$\begin{aligned} E\left|\sum_{k=s}^{t} D_k \epsilon_k\right|^p &\leq B_p \left\|\sum_{k=s}^{t} D_k^2 \epsilon_k^2\right\|_{p/2}^{p/2}, \\ &\leq B_p \sup_k \|\epsilon_k\|_p^p \left(\sum_{k=s}^{t} D_k^2\right)^{p/2} \end{aligned}$$

which concludes the proof for martingale increments. **(ii)** is an application of proposition 10, Corollary 11 (p. 264) by Kouritzin (1994)

$$E\left|\sum_{k=s}^{t} D_k \epsilon_k\right|^p \leq B_p \left(\sum_{k=s}^{t} \alpha_\epsilon(k)^{\frac{2\delta}{p(p+\delta)}}\right)^{p/2} \sup_{k\geq 0} \|\epsilon_k\|_{p+\delta}^p \left(\sum_{k=s}^{t} D_k^2\right)^{p/2}$$

References

[1] R.R. Bitmead and B. Anderson, and T. Ng, Convergence rate determination for gradient based adaptive estimators, *Automatica,AC*, vol. 25(4), pp.185-191, 1980

[2] A. Davydov, Mixing conditions for Markov chains, *Th. Prob. Appl.*, vol. 18, pp. 312-328, 1973

[3] C. Deo, A note on empirical processes of strong mixing sequences. *Ann. Probability*, Vol. 1, pp. 870–875, 1973

[4] P. Doukhan and S. Louhichi Weak dependence and moment inequalities Pre-print Université de Paris-Sud, 1997.

[5] E.Eweda and O.Macchi., Tracking error bouds of adaptive nonstationary filtering, *Automatica*, vol. 2(3), pp. 293-302, 1985

[6] S. Gunnarson and L. Ljung, Frequency domain tracking characteristics of adaptive algorithms, *IEEE Trans. on Acoust. Speech and Signal Proc.*, vol. 37, pp 1072–1089, 1989
[7] L. Guo and L. Ljung and G.-J. Wang, Necessary and Sufficient Conditions for Stability of LMS, *IEEE Trans. on Automatic Control*, vol. 42, N0 6, pp 761–769, 1997
[8] L. Guo and L. Ljung, Exponential Stability of General Tracking Algoritms, *IEEE Trans. on Automatic Control*, vol.40 , pp.1376-1387, 1995
[9] L.Guo, Stability of recursive stochastic tracking algorithms, *SIAM journal of control and Optimisation*, vol. 23, pp.1195-1225, 1994
[10] P.Hall and C.C.Heyde. , Martingale Limit Theory and its Applications, *Academic Press, New York, 1980*
[11] M. Kouritzin, Inductive methods and rates of r-th mean convergence in adaptive filtering, *Stochastics and Stochastices Reports*, vol. 51, pp. 241–266, 1994
[12] H.J. Kushner and G. Yin, *Stochastic approximation algorithms and applications*, Applications of Mathematics, Stochastic Modelling and Applied Probability, Springer-Verlag, 1997
[13] L. Ljung and S. Gunnarson, Adaptation and tracking in system identification: a survey *Automatica*, vol. 26, pp. 7–21, 1990.
[14] L. Ljung and P.Priouret, A result on the mean square error obtained using general tracking algorithms. *Int. Jour. on Adaptative Control and Sig. Proc.*, vol. 5, n0 4, pp. 231–250, 1991
[15] L. Ljung and T. Soderstrom, *Theory and pratice of recursive identification*, MIT press, Cambridge, MA, 1983
[16] O. Macchi, Optimisation of adaptive identification for time-varying filters *IEEE Trans. On Automatic Control*, vol. 31, pp. 283–287, 1986
[17] O. Macchi, *LMS adaptive processing with applications in transmission*, Wiley, 1994
[18] A. Mazroui *Poursuite et controle pour certains modèles de régressions.These de paris 6,1993*
[19] A. Mokkadem, Mixing properties of ARMA processes, *Th. Prob. Appl.*, vol. 29, pp. 309-315, 1988
[20] A. Perrier, B. Delyon, E. Moulines, Performance analysis of stochastic gradient identification algorithms using a perturbation expansion approach To appear in *IEEE. Trans. on Signal Proc.*, 1998
[21] P. Priouret and A. Veretennikov, A remark on the stability of the LMS. tracking algorithm, To appear in *Stochastic Analysis Appl.*, 1998
[22] V. Solo, The limit behavior of the LMS algorithm *IEEE Trans. Acoust. Speech and Signal Proc.*, vol. 37, pp.1909–1922, 1989
[23] V. Solo, The error variance of the LMS *IEEE Trans. on Signal Proc.*, vol. 40, pp. 803–813, 1990.
[24] V.Solo and X.Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*, *Prentice Hall , 1995*
[25] R.Yokoyama, Moment bouds for Stationary Mixing Sequences, *Z.Wahrscheinlichkeitstheorie.Gbiete*, vol. 52, pp. 45-57, 1980
[26] B. Widrow and M. Hoff Adaptive switching circuits, *IRE Wescon Convention Record Part IV*, pp. 96–104, 1960.

Appendix D. Detailed derivations

D.1. Observation noise. For notational convenience, we set ${}^vJ_t = J_t$. Straightforward application of the definitions yield:

$$J_{t+1}^{(1)} = \mu \sum_{s=0}^{t} (1-\alpha\mu)^{t-s-1} D_1(t, s+1)\phi_s v_s,$$

$$J_{t+1}^{(0)} = \sum_{s=0}^{t} (1-\alpha\mu)^{t-s} \phi_s v_s.$$

Thus,

$$\lim_{t\to\infty} E\left(J_t^{(1)} J_t^{(0)^T}\right) = \mu\sigma_v^2 \sum_{s=0}^{\infty} (1-\alpha\mu)^{2u-1} E\left(D_1(u,1)\phi_0\phi_0^T\right)$$

It is easily seen that, for $u \geq 1$

$$E(D_1(u,1)\phi_0\phi_0^T) = -\alpha^2(d+1)\kappa^2 \frac{1-\kappa^{2u}}{1-\kappa^2}$$

The expression follows. Similarly, we have

$$J_{t+1}^{(2)} = \mu^2 \sum_{s=0}^{t} (1-\alpha\mu)^{t-s-2} D_2(t, s+1)\phi_s v_s$$

Thus,

$$\lim_{t\to\infty} E\left(J_t^{(2)} J_t^{(0)^T}\right) = \mu^2\sigma_v^2 \sum_{s=0}^{\infty} (1-\alpha\mu)^{2s-2} E(D_2(s,1)\phi_0\phi_0^T).$$

One may show that, for $v > u > 0$

$$E(Z_v Z_u \phi_0\phi_0^T) = \alpha^3\kappa^{2v}(d^2+3d+4) + \alpha^3\kappa^{2(v-u)}(d+1).$$

The proof follows. Finally,

$$\lim_{t\to\infty} E(J_t^{(1)} J_t^{(1)^T}) = \sigma_v^2\mu^2 \sum_{u=0}^{\infty} (1-\alpha\mu)^{2u-2} E(D_1(u,1)\phi_0\phi_0^T D_1(u,1)).$$

We have

$$E(D_1(u,1)\phi_0\phi_0^T D_1(u,1)) = \alpha^3(d^2+3d+4) \sum_{v,w=1}^{u} \kappa^{|v-w|}\kappa^{v+w} + \alpha^3(d+1) \sum_{v,w=1}^{u} \kappa^{2|v-w|}$$

and the expression follows.

D.2. Lag noise. It follows from the definition that

$$J_{t+1}^{(2)} = \mu^2 \sum_{s=0}^{t} (1-\alpha\mu)^{t-s-2} D_2(t, s+1) w_s.$$

Hence,

$$\lim_{t\to\infty} E(J_t^{(2)} J_t^{(0)^T}) = \gamma^2\mu^2 \sum_{s=0}^{\infty} (1-\alpha\mu)^{2s-2} E(D_2(s,1)).$$

The proof is concluded by noting that, for $v > u \geq 0$,

$$E(Z_v Z_u) = \alpha^2 \kappa^{2(v-u)}(d+1)I$$

Similarly,

$$\lim_{t\to\infty} E(J_t^{(1)} J_t^{(1)^T}) = \gamma^2\mu^2 \sum_{s=0}^{\infty}(1-\alpha\mu)^{2s-2} E(D_1(s,1)D_1(s,1))$$

and

$$E(D_1(s,1)D_1(s,1)) = \sum_{-s+1}^{s-1}(s-|\tau|)E(Z_\tau Z_0) = \alpha^2(d+1)\sum_{-s+1}^{s-1}(s-|\tau|)\kappa^{2|\tau|}$$

which concludes the proof.

D.3. Some useful summation formulas.

$$\sum_{1\leq u<v\leq s} a^{2v} = \frac{a^2\left(-a^{2s+1} + sa^{2s}(a^2-1) + a^2\right)}{\left(a^2-1\right)^2},$$

$$\sum_{1\leq u<v\leq s} a^{2(v-u)} = -\frac{a^2\left(-\left(a^2\right)^s + a^2 s - s + 1\right)}{\left(a^2-1\right)^2},$$

$$\sum_{v,w=1}^{s} a^{2|v-w|} = \frac{s(1-a^4) - 2a^2 + 2a^{2(s+1)}}{(1-a^2)^2},$$

$$\sum_{v,w=1}^{s} a^{|v-w|}a^{v+w} = \frac{a^2}{(1-a^2)^2}\left(1 + a^2 - (2s+1)a^{2s} + (2s-1)a^{2(s+1)}\right),$$

$$\sum_{s=0}^{\infty}(1-\alpha\mu)^s = \frac{1}{\alpha\mu}$$

$$\sum_{s=0}^{\infty} s(1-\alpha\mu)^{2s} = -\frac{\left(-1+\alpha\,\mu\right)^4\left(\alpha^2\mu^2 - 2\,\alpha\,\mu - 1\right)}{\alpha^2\mu^2\left(-2+\alpha\,\mu\right)^2}$$