Systems & Control Letters 19 (1992) 157–163
North-Holland

Almost sure convergence of least squares estimates for regular multivariate ARX systems

Mohamed Boutahar
Université de Provence, Laboratoire de Mathématiques Appliquées et Informatique, Marseille, France

Claude Deniau
'Statistique appliquée' (Orsay) U.R.A. D.0743 et Faculté des Sciences de Luminy G.R.I.M., Département de Mathématiques, Marseille, France

Received 23 November 1991
Revised 21 February 1992

Abstract: We establish almost sure convergence of least squares estimates for general multivariate ARX(p, s) systems, with stochastic input signal. Results of strong consistency and speed of convergence are obtained with a regularity assumption on the AR part of the system.

Keywords: ARX systems; least squares; martingale transform; regularity; speed of convergence.

1. Introduction

Consider the multivariate linear autoregressive with exogenous variables ($\mathrm{ARX}_d(p, s)$) system:

$$A(z)Y_n = B(z)U_n + \varepsilon_n, \quad n \in \mathcal{N} - \{0\}, \tag{1}$$

where Y_n is an observed output, U_n is an observed input and ε_n is an unobserved random perturbation at stage n. Here z denotes the shift operator; A and B are matrix polynomials with known degrees p and s respectively:

$$A(z) = \boldsymbol{I}_d + \sum_{j=1}^{p} \boldsymbol{A}_j z^j, \qquad B(z) = \sum_{j=1}^{s} \boldsymbol{B}_j z^j.$$

Correspondence to: C. Deniau, Fac. des Sciences de Luminy, Dépt. de Mathématiques, Case 901, 163 Av. de Luminy, 13288 Marseille Cedex 9, France.

The parameters to be estimated are the matrix coefficients of these polynomials:

$$\theta = (-\boldsymbol{A}_1, \ldots, -\boldsymbol{A}_p, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_s),$$

where $(\boldsymbol{A}_i, 1 \leqslant i \leqslant p)$ are (d, d) matrices with $(\det(\boldsymbol{A}_p) \neq 0)$, $(\boldsymbol{B}_j, 1 \leqslant j \leqslant s)$ are (d, r) matrices. We use the Least Squares (LS) estimator denoted by θ_n.

The strong consistency of θ_n has been studied by many authors. In particular results about speed of convergence were recently obtained in different cases by Lai and Wei [7], Chen and Guo [2], and Viano [10] for example. More recently Duflo et al. [4] generalized Lai and Wei's results for multivariate $\mathrm{AR}_d(p)$ systems and show that θ_n is strongly consistent provided that the system is regular.

In this paper, we will study the LS estimator θ_n in the multivariate general autoregressive with exogenous variables with the regularity hypothesis first introduced in [4]. We extended Lai and Wei, and Duflo's results obtained in AR case, by adding an input signal (U_n) to the system. Assuming that the input signal (U_n) is stochastic and independent of the noise (ε_n), we prove the strong consistency of θ_n without any stability assumption; we also obtain results about the speed of convergence of θ_n toward θ. Without the regularity assumption, it's not sufficient to choose, like here, a signal with an intensity of excitation equivalent to n, to prove strong consistency of θ_n; this problem is studied by Boutahar [1].

2. Notations and definitions

2.1.

Denote by z_i, $i = 1, \ldots, dp$, the roots of the polynomial $z^{dp} \det(A(1/z))$ associated to the system (1). The system (1) is:

- *stable*, if $\forall i$, $|z_i| < 1$ ($|\cdot|$ denotes the modulus),
- *unstable*, if $\forall i$, $|z_i| \leqslant 1$,
- *explosive*, if $\forall i$, $|z_i| > 1$,
- *general*, if no assumption on the roots location of z_i is made,
- *regular*, if no proper subspace associated to an eigenvalue λ of A_c, such that $|\lambda| > 1$, is of dimension higher than 1 (A_c is the companion matrix of $A(z)$).

The scalar product $\mathscr{C}^d$ is denoted by $\langle \cdot, \cdot \rangle$; $\|X\|$, X', and X^* denote respectively the Euclidian norm, the transpose and the transpose conjugate of X.

For a given (d, d) matrix A, denote $\lambda_{\min}(A)$ (resp. $\lambda_{\max}(A)$) the minimum (resp. the maximum) eigenvalue of A; we write $A > 0$ to say that A is positive definite. If A is symmetric and non-negative definite, we denote by $A^{1/2}$ an arbitrary square root of A, i.e: $A = A^{1/2}(A^{1/2})^*$. I_d denotes the (d, d) identity matrix, diag$(G_1, \dots, G_p)$ is the matrix where the diagonal blocks are $G_1, \dots, G_p$.

2.2.

We recall that we can write the system (1) into its regression form:

$$Y_n = \theta \Phi_{n-1} + \varepsilon_n \tag{2}$$

where

$$\Phi_n = \left(Y_n', \dots, Y_{n-p+1}', U_n', \dots, U_{n-s+1}'\right)'. \tag{3}$$

From (1) and (2), we deduce that:

$$\Phi_n = \mathscr{A}\Phi_{n-1} + e_n, \tag{4}$$

where

$$\mathscr{A} = \begin{pmatrix} A_c & B \\ 0 & K_1 \end{pmatrix},$$

$$A_c = \begin{pmatrix} -A_1 & -A_2 & \cdots & -A_{p-1} & -A_p \\ I_d & 0 & \cdots & 0 & 0 \\ 0 & I_d & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & I_d & 0 \end{pmatrix}$$

being the companion matrix of $A(z)$,

$$B = \begin{pmatrix} B_1 & B_2 & \cdots & B_{s-1} & B_s \\ 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix},$$

$$K_1 = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ I_r & 0 & \cdots & 0 & 0 \\ 0 & I_r & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & I_r & 0 \end{pmatrix}$$

and $e_n = (\varepsilon_n', 0, \dots, 0, U_n', 0, \dots, 0)'$. Consequently

$$\det(\mathscr{A} - zI_{dp+rs}) = z^{rs} \det(A_c - zI_{dp}) = z^{rs+dp} \det\left(A\left(\frac{1}{z}\right)\right).$$

2.3.

Finally we recall that the LS estimator θ_n of θ is given by

$$\theta_n^* = P_n^{-1} \sum_{k=1}^{n} \Phi_{k-1} Y_k^* \tag{5}$$

where $P_n = \sum_{k=1}^{n} \Phi_{k-1}\Phi_{k-1}^*$.

3. General hypothesis ($\mathscr{H}$)

0. All the random variables considered here are of complex values and defined in the same probability space $(\Omega, \mathscr{F}, \mathscr{P})$.

1. (ε_n) is a *martingale difference* sequence with respect to an increasing sequence of σ-fields $\mathbb{F} = (\mathscr{F})_n$ (i.e. ε_n is $\mathscr{F}_n$-measurable and $E(\varepsilon_n / \mathscr{F}_{n-1}) = 0$, $\mathscr{P}$-a.s. for every n) such that

(a) $\sup_n E(\|\varepsilon_n\|^\alpha / \mathscr{F}_{n-1}) < \infty$, $\mathscr{P}$-a.s. for some real constant $\alpha > 2$;

(b) for every n, $E(\varepsilon_n \varepsilon_n^* / \mathscr{F}_{n-1}) = \Gamma_\varepsilon > 0$, $\mathscr{P}$-a.s.

2. (U_n) satisfies the same hypothesis as (ε_n) with covariance Γ_u, a constant β instead of α, and is independent of (ε_n).

3. The system (1) is *regular*.

Remark. In the multivariate case, when the system has an explosive part, we cannot remove the regularity assumption, as it is shown in [3], p. 60.

4. The main result

The main result of this paper is:

Theorem. *Suppose that in the multivariate system* (1), *the hypothesis* ($\mathscr{H}$) *is satisfied, then the* LS *estimator defined by* (5) *is strongly consistent*; *the speed of convergence is*

$$\| \theta_n - \theta \|^2 = O\left(\frac{\log(n)}{n} \right) \quad \mathscr{P}\text{-a.s.} \tag{6}$$

The proof of this theorem is given in Section 5; before this, we present two lemmas, the first one describes the asymptotic behaviour of θ_n in the stable or unstable case; the second one gives some tools to prove the same result in the explosive case.

Lemma 1 (stable or unstable case). *Under the hypothesis* ($\mathscr{H}$), *if we assume also that the system* (1) *is stable or unstable then* θ_n *is strongly consistent and satisfies* (6).

Proof. (a) Let us first prove that

$$\log(\lambda_{\max}(P_n)) = O(\log(n)) \quad \mathscr{P}\text{-a.s.}, \tag{7}$$

where P_n is defined in (5). It is easy to see that the random sequence (e_n) is a martingale difference sequence with respect to $(\mathscr{F}_n)$ and by the conditional Minkowski inequality, if we denote $\gamma = \min(\alpha, \beta)$, we have:

$$\begin{aligned} E(\| e_n \|^\gamma / \mathscr{F}_{n-1}) &= E\left(\left(\| \varepsilon_n \|^2 + \| U_n \|^2 \right)^{\gamma/2} / \mathscr{F}_{n-1} \right) \\ &\leqslant \left\{ \left(E(\| \varepsilon_n \|^\gamma / \mathscr{F}_{n-1}) \right)^{2/\gamma} + \left(E(\| U_n \|^\gamma / \mathscr{F}_{n-1}) \right)^{2/\gamma} \right\}^{\gamma/2} \end{aligned}$$

and

$$E(\| \varepsilon_n \|^\gamma / \mathscr{F}_{n-1}) \leqslant E(\| \varepsilon_n \|^\alpha / \mathscr{F}_{n-1}),$$
$$E(\| U_n \|^\gamma / \mathscr{F}_{n-1}) \leqslant E(\| U_n \|^\beta / \mathscr{F}_{n-1}).$$

Hence $\sup_n E(\| e_n \|^\gamma / \mathscr{F}_{n-1}) < \infty$, and consequently by [8], Theorem 1(i), we conclude that

$$\| \Phi_n \| = O\left(n^{\rho - 1/2} (\log(\log(n)))^{1/2} \right) \quad \mathscr{P}\text{-a.s.}, \tag{8}$$

where ρ is the largest order of multiplicity of distinct unitary eigenvalues of $\mathscr{A}$. Since

$$\lambda_{\max}(P_n) \leqslant \operatorname{trace}(P_n) = \sum_1^n \| \Phi_{k-1} \|^2,$$

the equality (7) is proved.

(b) Now we prove the following inequality:

$$\liminf \lambda_{\min}\left(\frac{1}{n} P_n \right) > 0 \quad \mathscr{P}\text{-a.s.} \tag{9}$$

By [9], Theorem 2 there exists $\rho > 0$ such that

$$\lambda_{\min}(P_n) \geqslant \rho \lambda_{\min}(V_n) \tag{10}$$

where $V_n = \sum_{k=1}^n v_k v_k^*$ and $v_n = (U'_{n-1}, \ldots, U'_{n-dp-s}, \varepsilon'_{n-1}, \ldots, \varepsilon'_{n-dp})'$. Applying the martingale strong law to each block of the matrix V_n, we conclude that

$$\lim_{n \to \infty} \frac{1}{n} V_n = \operatorname{diag}(\Gamma_u, \ldots, \Gamma_u; \Gamma_\varepsilon, \ldots, \Gamma_\varepsilon) > 0 \quad \mathscr{P}\text{-a.s.} \tag{11}$$

By (10) and (11) the inequality (9) holds.

(c) For a fixed unitary, arbitrary vector $v \in \mathscr{C}^d$, let

$$Q_n(v) = e_n^*(v) X_n (X_n^* X_n)^{-1} X_n^* e_n(v)$$

where

$$e_n(v) = (\langle \varepsilon_1, v \rangle, \ldots, \langle \varepsilon_n, v \rangle)',$$
$$X_n^* = (\Phi_0, \ldots, \Phi_{n-1}).$$

Since $X_n^* X_n = P_n$, we can write

$$Q_n(v) = \sum_1^n \overline{\langle \varepsilon_k, v \rangle} \Phi_{k-1}^* P_n^{-1} \sum_1^n \Phi_{k-1} \langle \varepsilon_k, v \rangle$$

and using (2) and (5), we obtain

$$\begin{aligned} &\| (\theta_n - \theta)^* v \|^2 \\ &= \left\| P_n^{-1} \sum_1^n \Phi_{k-1} \langle \varepsilon_k, v \rangle \right\|^2 \\ &\leqslant \| P_n^{-1/2} \|^2 \left\| P_n^{-1/2} \sum_1^n \Phi_{k-1} \langle \varepsilon_k, v \rangle \right\|^2 \\ &= \| P_n^{-1/2} \|^2 Q_n(v) \leqslant \lambda_{\max}(P_n^{-1}) Q_n(v) \\ &= [\lambda_{\min}(P_n)]^{-1} Q_n(v). \end{aligned}$$

In view of (9), to prove the result of Lemma 1, it suffices to show that

$$Q_n(v) = O(\log(n)) \quad \mathscr{P}\text{-a.s.} \tag{12}$$

Since $(\langle \varepsilon_n, v\rangle)$ is a martingale difference sequence with a finite order $\alpha > 2$ moment, it follows by [6], Lemma 1, (2.4) that

$$Q_n(v) = O(\log(\lambda_{\max}(P_n))) \quad \mathscr{P}\text{-a.s.} \tag{13}$$

on the event $\Gamma = \{\lim_{n\to\infty}\lambda_{\max}(P_n) = \infty\}$. By (9) we get that $\mathscr{P}(\Gamma) = 1$ and by (7) the equality (12) holds. □

Before proving Lemma 2, we need some linear transformations. If system (1) is general, so is the $AR_{dp+rs}(1)$ system (Φ_n), given by (4). We will decompose it into two autoregressive systems $\Phi_n^{(i)}$, $i = 1, 2$, such that $\Phi_n^{(1)}$ is an explosive system and $\Phi_n^{(2)}$ is a stable or unstable system.

We can write $\det(A(z))$ as a product of two polynomials:

$$\det(A(z)) = a_1(z)a_2(z),$$
$$\deg(a_i(z)) = d_i, \quad i = 1, 2,$$

where the roots of $a_1(z)$ (resp. $a_2(z)$) are strictly inside (resp. outside) the unit disk of $\mathscr{C}$. Denote by $A_c(1)$ [resp. $A_c(2)$] the companion matrix of $a_1(z)$ [resp. $a_2(z)$], hence $A_c(1)$ is a (d_1, d_1) explosive matrix and $A_c(2)$ is a (d_2, d_2) stable or unstable matrix. There exists a complex non-singular matrix T such that:

$$T\mathscr{A} = \operatorname{diag}(A_c(1), A_c(2), K_1)T. \tag{14}$$

Define $\Phi_n^{(i)} = T_i\Phi_n$, where $T^{\mathrm{T}} = (T_1 \quad T_2)$, T_1 is a $(d_1, dp+rs)$ full rank matrix and T_2 is a $(d_2 + rs, dp+rs)$ full rank matrix. From (4) and (14) it follows that

$$\Phi_n^{(1)} = A_c(1)\Phi_{n-1}^{(1)} + T_1e_n, \tag{15}$$

$$\Phi_n^{(2)} = R\Phi_{n-1}^{(2)} + T_2e_n, \tag{16}$$

where $R = \operatorname{diag}(A_c(2), K_1)$; thus $(\Phi_n^{(1)})$ is an explosive $AR_{d_1}(1)$ and $(\Phi_n^{(2)})$ is a stable or unstable $AR_{d_2+rs}(1)$.

The asymptotic behaviour of the explosive system $(\Phi_n^{(1)})$ remains to be described.

Lemma 2 (explosive case). *Consider the system (15). Under the hypothesis ($\mathscr{H}$) we have*:

(i) $A_c(1)^{-n}\Phi_n^{(1)} \to \eta = \Phi_0^{(1)} + \sum_1^\infty A_c(1)^{-k}T_1e_k$ $\mathscr{P}$-a.s.,

(ii) *for every $x \in \mathscr{C}^{d_1}$ the probability distribution of $\langle x, \eta\rangle$ is continuous,*

(iii) *with $P_n(1) = \sum_1^n \Phi_{k-1}^{(1)}\Phi_{k-1}^{(1)*}$ we have*

$$A_c(1)^{-n}P_n(1)A_c(1)^{-n*} \to F_1 = \sum_1^\infty A_c(1)^{-k}\eta\eta^*A_c(1)^{-k*} \quad \mathscr{P}\text{-a.s.},$$

and moreover the matrix F_1 is positive definite.

Proof. (i) Let $M_n = A_c(1)^{-n}\Phi_n^{(1)} - \Phi_0^{(1)}$; using (15) we have $M_n = \sum_1^n A_c(1)^{-k}T_1e_k$, hence $(M_n, \mathbb{F})$ is a martingale such that:

$$E(\| M_n \|^2) = \operatorname{trace}\left(\sum_{k=1}^n A_c(1)^{-k}T_1K_{\varepsilon u}T_1^*A_c(1)^{-k*}\right),$$

where

$$K_{\varepsilon u} = \operatorname{diag}(\Gamma_\varepsilon, \mathbf{0}, \ldots, \mathbf{0}, \Gamma_u, \mathbf{0}, \ldots, \mathbf{0});$$

since $A_c(1)$ is an explosive matrix it follows that: $E(\| M_n \|^2) < \infty$, and hence (i) holds.

(ii) By application of Lemma 2 of [8], it is sufficient to show that

$$\liminf \lambda_{\min}(C_n(r)) > 0 \quad \mathscr{P}\text{-a.s, for some } r \geqslant 1, \tag{17}$$

where

$$C_n(r) = E\left(\sum_{k=1}^r A_c(1)^{r-k}T_1e_{nr+k}e_{nr+k}^* \times T_1^*\left(A_c(1)^{r-k}\right)^* / \mathscr{F}_{nr}\right).$$

But

$$C_n(r) = \sum_{k=1}^r A_c(1)^{r-k}T_1K_{\varepsilon u}T_1^*\left(A_c(1)^{r-k}\right)^* = \sum_{k=0}^{r-1} A_c(1)^kT_1K_{\varepsilon u}T_1^*\left(A_c(1)^k\right)^* = C(r). \tag{18}$$

To obtain the definite positiveness of $C(r)$, it suffices to show (cf. [3], Proposition 2.III.6) that

$$\liminf \lambda_{\min}\left(\frac{1}{n}(P_n(1))\right) > 0 \quad \mathscr{P}\text{-a.s.} \tag{19}$$

Note that the inequality (9) remains true for the general system (1). Moreover $P_n(1)=T_1P_nT_1^*$ implies that

$$\lambda_{\min}(P_n(1)) \geqslant \lambda_{\min}(T_1T_1^*)\lambda_{\min}(P_n). \tag{20}$$

Thus the desired conclusion (19) follows from (9), (20) and the fact that $\lambda_{\min}(T_1T_1^*)>0$.

(iii) A similar argument to the proof of Theorem 2(ii) of [7] can be applied to show the almost sure convergence of the matrix $A_c(1)^{-n}P_n(1)A_c(1)^{-n*}$ to F_1. The regularity assumption implies that the minimal polynomial of matrix A_c is equal to its characteristic polynomial, and this also holds for matrix $A_c(1)$; making use of (ii) it can be shown that $\eta, A_c(1)^{-1}\eta, \ldots, A_c(1)^{-(d_1-1)}\eta$ are linearly independent $\mathscr{P}$-a.s. (cf [3], Theorem 5.I.3), and hence

$$\sum_0^{d_1-1} A_c(1)^{-k}\eta\eta^*A_c(1)^{-k*}>0 \quad \mathscr{P}\text{-a.s.},$$

consequently F_1 is positive definite $\mathscr{P}$-a.s. □

5. Proof of the Theorem

Define $\Delta_n=\text{diag}(A_c(1)^nF_1^{1/2}, P_n(2)^{1/2})$, $P_n(2)=\sum_1^n\Phi_{k-1}^{(2)}\Phi_{k-1}^{(2)*}$, $\tilde{K}_n=\Delta_n^{-1}TP_nT^*\Delta_n^{-1*}$, where it is easy to prove that Δ_n^{-1} exists almost surely for large values of n. Let us prove that:

$$\tilde{K}_n \to I_{dp+rs} \quad \mathscr{P}\text{-a.s.} \tag{21}$$

We have

$$\tilde{K}_n=\begin{pmatrix} F_1^{-1/2}A_c(1)^{-n}P_n(1)A_c(1)^{-n*}(F_1^{-1/2})^* & D_n^* \\ D_n & I_{rs}\end{pmatrix},$$

where

$$D_n=P_n(2)^{-1/2}\sum_{k=1}^n \Phi_{k-1}^{(2)}\Phi_{k-1}^{(1)*}\left(A_c(1)^{-n}\right)^* \times\left(F_1^{-1/2}\right)^*.$$

But

$$\|D_n\|^2 \leqslant \text{ct.}\max_{k\leqslant n}\left(\Phi_{k-1}^{(2)*}(P_n(2))^{-1}\Phi_{k-1}^{(2)}\right) \times\left(\sum_{k=1}^n\left\|A_c(1)^{-n}\Phi_{k-1}^{(1)}\right\|\right)^2, \tag{22}$$

ct. a positive constant. Since $(\Phi_n^{(2)})$ is stable or unstable and its noise (T_2e_n) has a finite moment of order $\gamma>2$, it can be shown that

$$\max_{k\leqslant n}\left(\Phi_{k-1}^{(2)*}(P_n(2))^{-1}\Phi_{k-1}^{(2)}\right)=\text{o}(1) \quad \mathscr{P}\text{-a.s.} \tag{23}$$

(cf. [7], Theorem 4). Making use of Lemma 2(i) and the fact that the matrix $A_c(1)$ is explosive, we can obtain, in the same way as in [7], Corollary 1(i), the following equality

$$\sum_1^n\left\|A_c(1)^{-n}\Phi_{k-1}^{(1)}\right\|=\text{O}(1) \quad \mathscr{P}\text{-a.s.} \tag{24}$$

From (22)–(24) it follows that

$$D_n \to \mathbf{0} \quad \mathscr{P}\text{-a.s.} \tag{25}$$

Using Lemma 2(iii) and (25) we get (21).

Now we denote by $\tilde{\theta}_n=\theta_n-\theta$ the estimation error at stage n. Then using the nonsingular linear transformation T defined in (14), we can write

$$T^{*-1}\tilde{\theta}_n^*=\left(\theta_n^{(1)\prime}, \theta_n^{(2)\prime}\right)', \tag{26}$$

where $\theta_n^{(1)}$ is a (d_1, d) matrix and $\theta_n^{(2)}$ is a (d_2+rs, d) matrix. In view of (2) and (5), it is easy to show that

$$\Delta_n^*T^{*-1}\tilde{\theta}_n^*=\tilde{K}_n^{-1}\Delta_n^{-1}T\sum_1^n\Phi_{k-1}\varepsilon_k^*,$$

and making use of the $\tilde{K}_n$ convergence, we conclude that

$$\left\|F_1^{1/2*}A_c(1)^{*n}\theta_n^{(1)}\right\| =\text{O}\left(\left\|F_1^{-1/2}A_c(1)^{-n}\sum_1^n\Phi_{k-1}^{(1)}\varepsilon_k^*\right\|\right) \quad \mathscr{P}\text{-a.s.},$$

and

$$\left\|\left(P_n(2)^{1/2}\right)^*\theta_n^{(2)}\right\| =\text{O}\left(\left\|(P_n(2))^{-1/2}\sum_1^n\Phi_{k-1}^{(2)}\varepsilon_k^*\right\|\right) \quad \mathscr{P}\text{-a.s.} \tag{27}$$

Thus

$$\begin{aligned}\left\|F_1^{1/2*}A_c(1)^{*n}\theta_n^{(1)}\right\| &=\text{O}\left(\max_{1\leqslant k\leqslant n}\|\varepsilon_k\|\sum_1^n\left\|A_c(1)^{-n}\Phi_{k-1}^{(1)}\right\|\right)\\ &=\text{O}\left(\max_{1\leqslant k\leqslant n}\|\varepsilon_k\|\right) \quad (\text{by equality (24)})\\ &=\text{o}(\sqrt{n}) \quad \mathscr{P}\text{-a.s.},\end{aligned}$$

since $\varepsilon_n = o(\sqrt{n})$ $\mathscr{P}$-a.s., because $(1/n)\Sigma_1^n \| \varepsilon_k \|^2 \to \text{trace}(\Gamma_\varepsilon)$ $\mathscr{P}$-a.s.

Hence,

$$\| A_c(1)^{*n} \theta_n^{(1)} \| = o(\sqrt{n}) \quad \mathscr{P}\text{-a.s.} \tag{28}$$

Since $A_c(1)$ is an explosive matrix, the above equality implies that for every real δ, $\lambda^{-1} < \delta < 1$, where λ is the modulus of the smallest eigenvalue of $A_c(1)$:

$$\| \theta_n^{(1)} \| = o(\sqrt{n}\, \delta^n) \quad \mathscr{P}\text{-a.s.} \tag{29}$$

Moreover the random matrix $\Sigma_1^n \Phi_{k-1}(2) \varepsilon_k^*$ is a martingale transform. Then a similar argument as (13) can be used to obtain that

$$\left\| (P_n(2))^{-1/2} \sum_1^n \Phi_{k-1}^{(2)} \varepsilon_k^* \right\|^2 = O(\log(\lambda_{\max}(P_n(2)))) \tag{30}$$

almost surely on the event

$\Gamma_2 = \{ \lim_{n\to\infty} \lambda_{\max}(P_n(2)) = \infty \}$.

We have, as for (19),

$$\liminf \lambda_{\min}(P_n(2))/n > 0 \quad \mathscr{P}\text{-a.s.}, \tag{31}$$

and hence $\mathscr{P}(\Gamma_2) = 1$. The system $(\Phi_n(2))$ is an $\text{AR}_{d_2+rs}(1)$, stable or unstable. Then we can obtain, in the same way as (7),

$$\log(\lambda_{\max}(P_n(2))) = O(\log(n)) \quad \mathscr{P}\text{-a.s.} \tag{32}$$

From (27), (30), (31) and (32), it follows that

$$\| \theta_n^{(2)} \|^2 = O\left(\frac{\log(n)}{n} \right) \quad \mathscr{P}\text{-a.s.} \tag{33}$$

From (26), (29) and (33) we obtain (6). □

6. Particular case: Explosive system

We assume that system (1) is explosive, and we denote

$$\theta = (\theta^A, \theta^B), \qquad \theta_n = (\theta_n^A, \theta_n^B),$$

where $\theta^A = (-A_1, \ldots, -A_p)$, $\theta^B = (B_1, \ldots, B_s)$ and θ_n^A (resp. θ_n^B) is the part of θ_n corresponding to θ^A (resp. to θ^A).

Like in (14), we choose here a particular and non-singular matrix $\tilde{T}$ given by

$$\tilde{T} = \begin{pmatrix} I_{dp} & T_{12} \\ 0 & I_{rs} \end{pmatrix},$$

such that

$$\tilde{T}\mathscr{A} = \text{diag}(A_c, K_1)\tilde{T}.$$

Indeed the matrix T_{12} can be constructed as a solution of the following matrix equation:

$$XK_1 - A_c X = B \tag{34}$$

and noting that A_c and K_1 do not have common eigenvalues, implies that the last equation has a single solution (see Gantmacher [5]). Let

$$\tilde{\theta}_n^A = \theta_n^A - \theta^A, \qquad \tilde{\theta}_n^B = \theta_n^B - \theta^B.$$

Then

$$T^{*-1}\tilde{\theta}_n^* = \begin{pmatrix} \tilde{\theta}_n^A \\ -T_{12}^* \tilde{\theta}_n^A + \tilde{\theta}_n^B \end{pmatrix}.$$

Thus (29) and (33) become

$$\| \tilde{\theta}_n^A \| = o(\sqrt{n}\, \delta^n) \quad \mathscr{P}\text{-a.s.} \tag{35}$$

and

$$\| -T_{12}^* \tilde{\theta}_n^A + \tilde{\theta}_n^B \|^2 = O\left(\frac{\log(n)}{n} \right) \quad \mathscr{P}\text{-a.s.} \tag{36}$$

Consequently by (35) and (36) we conclude that

$$\| \tilde{\theta}_n^B \|^2 = O\left(\frac{\log(n)}{n} \right) \quad \mathscr{P}\text{-a.s.}$$

Hence, we deduce:

Corollary. *Consider the* LS *estimator in* (1); *under hypothesis* ($\mathscr{H}$), *and if we assume also that the system is explosive then* θ_n *is strongly consistent. The autoregressive part of* θ_n *has an exponential speed of convergence, i.e.*

$$\| \tilde{\theta}_n^A \| = o(\sqrt{n}\, \delta^n) \quad \mathscr{P}\text{-a.s.},$$

and the signal part of θ_n *has a speed of convergence given by*

$$\| \tilde{\theta}_n^B \|^2 = O\left(\frac{\log(n)}{n} \right) \quad \mathscr{P}\text{-a.s.}$$

References

[1] M. Boutahar, Strong consistency of least squares estimates in a general $\text{ARX}_d(p, s)$ system, *Stochastics and Stochastics Rep.* (to appear).

[2] H.F. Chen and L. Guo, Convergence rate for least-squares identification and adaptive control for stochastic systems, *Internat. J. Control* **44** (5) (1986) 1459–1476.

[3] M. Duflo, *Méthodes Récursives Aléatoires* (Masson, Paris, 1990).

[4] M. Duflo, R. Sennoussi and A. Touati, Propriétées asymptotiques presques sûres de l'estimateur des moindres carrés d'un modèle autorégressif vectoriel, *Ann. Inst. H. Poincaré Sér. B* **27** (1) (1991).

[5] F.R. Gantmacher, *Théorie des Matrices*, Tome 1 (Dunod, Paris 1966).

[6] T.L. Lai and C.Z. Wei, Least squares estimates in stochastic regression models with applications to identification and control of dynamics systems, *Ann. Statist.* **10** (1) (1982) 154–166.

[7] T.L. Lai and C.Z. Wei, Asymptotic properties of general autoregressive models and strong consistency of least squares estimates of their parameters, *J. Multivariate Anal.* **13** (1983) 1–13.

[8] T.L. Lai and C.Z. Wei, Asymptotic properties of weighted sums with applications to stochastic regression in linear dynamic systems, In: P.R. Krishnaiah, Ed., *Multivariate Analysis VI* (Elsevier Science Publishers, Amsterdam, 1985) 375–393.

[9] T.L. Lai and C.Z. Wei, On the concept of excitation in least square identification and adaptive control, *Stochastics* **16** (1986) 227–254.

[10] M.C. Viano, Partial convergence of extended least squares for insufficiently excited linear systems, *Internat. J. Adaptive Control Signal Process.* **4** (1990) 207–217.