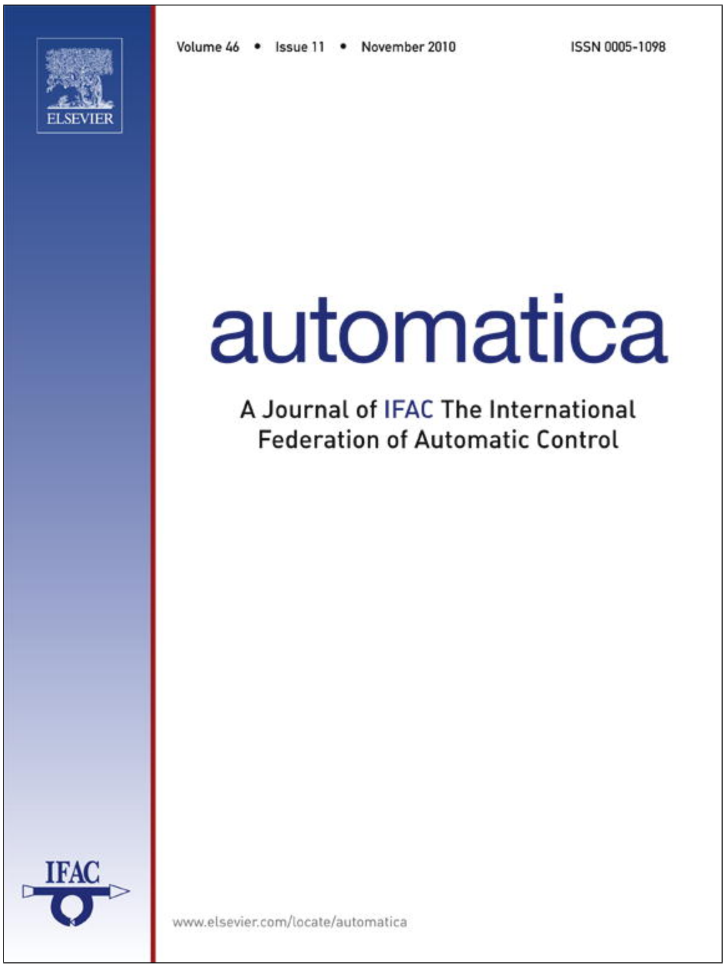
Volume 46 • Issue 11 • November 2010
ISSN 0005-1098
ELSEVIER
automatica
A Journal of IFAC The International
Federation of Automatic Control
IFAC
www.elsevier.com/locate/automatica

Automatica 46 (2010) 1799–1805

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

A new concept of strong controllability via the Schur complement for ARX models in adaptive tracking[☆]

Bernard Bercu[a,*], Victor Vazquez[b]

[a] Université Bordeaux 1, Institut de Mathématiques de Bordeaux, UMR CNRS 5251, and INRIA Bordeaux, team CQFD, 351 cours de la libération, 33405 Talence cedex, France

[b] Universidad Autónoma de Puebla, Facultad de Ciencias Fisico Matemáticas, Avenida San Claudio y Rio Verde, 72570 Puebla, Mexico

ARTICLE INFO

Article history:
Received 14 August 2008
Received in revised form
7 April 2010
Accepted 16 June 2010
Available online 3 August 2010

Keywords:
Estimation
Adaptive tracking control
Schur complement
Central limit theorem
Law of iterated logarithm

ABSTRACT

We propose a new concept of strong controllability related to the Schur complement of a suitable limiting matrix. This new notion allows us to extend the previous convergence results associated with multidimensional ARX models in stochastic adaptive tracking. On the one hand, we carry out a sharp analysis of the almost sure convergence for both least squares and weighted least squares algorithms. On the other hand, we also provide a central limit theorem and a law of iterated logarithm for these two algorithms.

1. Introduction

Multidimensional autoregressive with exogenous input models (ARX) are versatile and useful tools in many areas of applied mathematics, such as video segmentation (Vidal, 2008), financial mathematics (Fukata, Washio, & Motoda, 2006; Huang & Jane, 2009), population dynamics (Li & Zhang, 2008), robotics (Chang & Tzenog, 2008), and neurosciences (Baraldi, Manginelli, Maieron, Liberati, & Porro, 2007; Liu, Birch, & Allen, 2003). On the other hand, stochastic adaptive tracking plays a crucial role in a wide range of application areas. The goal of this paper is to investigate the asymptotic properties of estimators of the unknown parameters associated with multidimensional ARX models in stochastic adaptive tracking. We shall establish the almost sure convergence as well as the central limit theorem for both least squares and weighted least squares algorithms. An important literature is devoted to parameter estimation for ARMAX models as well as to adaptive tracking, see e.g. Aström and Wittenmark (1995), Bercu (1995), Bercu (1998), Bercu and Portier (2008), Caines (1988), Chen and Guo (1991), Guo and Chen (1991), Guo (1996), Lai and Wei (1986), and Nazin (1993). Our aim is to carry out a sharp analysis of the asymptotic behavior of the least squares estimators via the introduction of a new concept of strong controllability associated with the Schur complement of a suitable limiting matrix. This new notion is really easy to understand and it cannot be avoided when the adaptive tracking control proposed by Aström and Wittenmark (1995) is used. It allows us to extend the previous convergence results (Bercu, 1998; Bercu & Portier, 2002; Guo & Chen, 1991; Guo, 1994, 1995; Jankumas, 2000).

Consider the d-dimensional autoregressive process with adaptive control of order (p, q), $\text{ARX}_d(p, q)$ for short, given for all $n \geq 0$ by

$$A(R)X_{n+1} = B(R)U_n + \varepsilon_{n+1} \tag{1}$$

where R stands for the shift-back operator and X_n, U_n and ε_n are the system output, input and driven noise, respectively. The polynomials A and B are given for all $z \in \mathbb{C}$ by

$$A(z) = I_d - A_1 z - \cdots - A_p z^p,$$
$$B(z) = I_d + B_1 z + \cdots + B_q z^q,$$

where A_i and B_j are unknown square matrices of order d and I_d is the identity matrix. Denote by θ the unknown parameter of the model

$$\theta^t = (A_1, \ldots, A_p, B_1, \ldots, B_q).$$

☆ This work has been supported in part by INRIA, the French National Institute for Research in Computer Science and Control, by CONACYT, the Mexican National Council for Science and Technology, and by the ECOS Scientific Cooperation Programme. The material in this paper was partially presented at the 47th IEEE Conference on Decision and Control, December 9–11, 2008, Cancun, Mexico. This paper was recommended for publication in revised form by Associate Editor Giuseppe De Nicolao under the direction of Editor Ian R. Petersen.

* Corresponding author.
E-mail addresses: Bernard.Bercu@math.u-bordeaux1.fr (B. Bercu), Victor.Vazquez@siu.buap.mx (V. Vazquez).

0005-1098/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2010.06.043

Relation (1) can be rewritten as

$$X_{n+1} = \theta^t \Phi_n + U_n + \varepsilon_{n+1}, \tag{2}$$

where the regression vector is given by $\Phi_n = \left(X_n^p, U_{n-1}^q\right)^t$ with $X_n^p = (X_n^t, \dots, X_{n-p+1}^t)$, $U_n^q = (U_n^t, \dots, U_{n-q+1}^t)$. In all the sequel, we shall assume that (ε_n) is a martingale difference sequence adapted to the filtration $\mathbb{F} = (\mathcal{F}_n)$ where $\mathcal{F}_n$ stands for the σ-algebra of the events occurring up to time n. We also assume that, for all $n \geq 0$, $\mathbb{E}[\varepsilon_{n+1}\varepsilon_{n+1}^t|\mathcal{F}_n] = \Gamma$ a.s. where Γ is a positive definite deterministic covariance matrix. Our new concept of strong controllability is closely related to the almost sure convergence of the matrix $S_n = \sum_{k=0}^n \Phi_k \Phi_k^t$. In the particular case $q = 0$, it was shown in Bercu (1998) that

$$\lim_{n\to\infty} \frac{S_n}{n} = L \quad \text{a.s.}$$

where L is the block diagonal matrix of order dp given by $L = \operatorname{diag}(\Gamma, \dots, \Gamma)$. Under the classical causality assumption, we shall now prove that

$$\lim_{n\to\infty} \frac{S_n}{n} = \Lambda \quad \text{a.s.} \tag{3}$$

where Λ is the square matrix of order $\delta = d(p+q)$

$$\Lambda = \begin{pmatrix} L & K^t \\ K & H \end{pmatrix},$$

and the matrices H and K will be explicitly calculated. It is well known (Horn & Johnson, 1990) that $\det(\Lambda) = \det(L)\det(S)$ where $S = H - KL^{-1}K^t$ is the Schur complement of L in Λ. Moreover, as L is positive definite, Λ is positive definite if and only if S is positive definite. Via our new concept of strong controllability, we shall propose a suitable assumption under which S is positive definite. One can easily understand this assumption which cannot be avoided. This new notion will allow us to extend the previous convergence results (Bercu, 1998; Bercu & Portier, 2002; Guo & Chen, 1991; Guo, 1994, 1995; Jankumas, 2000) by showing a central limit theorem (CLT) and a law of iterated logarithm (LIL) for both the least squares (LS) and the weighted least squares (WLS) algorithms associated with the estimation of θ.

The paper is organized as follows. Section 2 is devoted to the introduction of our new concept of strong controllability together with some linear algebra calculations. Section 3 deals with the parameter estimation and the stochastic adaptive control. In Section 4, we establish convergence (3) and we deduce a CLT as well as a LIL for both LS and WLS algorithms. A short conclusion is given in Section 5. All technical proofs are postponed in the Appendices.

2. Strong controllability

In all the sequel, we shall make use of the well-known causality assumption on B. More precisely, we assume that for all $z \in \mathbb{C}$ with $|z| \leq 1$,

$(A_1)\quad \det(B(z)) \neq 0.$

In other words, the polynomial $\det(B(z))$ only has zeros with modulus > 1. Consequently, if $r > 1$ is strictly less than the smallest modulus of the zeros of $\det(B(z))$, then $B(z)$ is invertible in the ball with center zero and radius r and $B^{-1}(z)$ is a holomorphic function (see e.g. Duflo, 1997, page 155). For all $z \in \mathbb{C}$ such that $|z| \leq r$, we shall denote

$$P(z) = B^{-1}(z)(A(z) - I_d) = \sum_{k=1}^{\infty} P_k z^k. \tag{4}$$

All the matrices P_k may be explicitly calculated as functions of the matrices A_i and B_j. For example, we always have $P_1 = -A_1$. In addition, one can see that if $p = q = 1$ then for all $k \geq 2$, $P_k = -(-B_1)^{k-1}A_1$ while if $p = 2, q = 1$, $P_k = (-B_1)^{k-2}(B_1A_1 - A_2)$. Moreover, if $p = 1, q = 2$ then $P_2 = B_1A_1$ and $P_3 = (B_2 - B_1^2)A_1$ while if $p = 2, q = 2$, $P_2 = B_1A_1 - A_2$ and $P_3 = (B_2 - B_1^2)A_1 + B_1A_2$. We shall often make use of the square matrix of order dq given, if $p \geq q$, by

$$\Pi = \begin{pmatrix} P_p & P_{p+1} & \cdots & P_{p+q-2} & P_{p+q-1} \\ P_{p-1} & P_p & P_{p+1} & \cdots & P_{p+q-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ P_{p-q+2} & \cdots & P_{p-1} & P_p & P_{p+1} \\ P_{p-q+1} & P_{p-q+2} & \cdots & P_{p-1} & P_p \end{pmatrix},$$

while, if $p \leq q$, by

$$\Pi = \begin{pmatrix} P_p & P_{p+1} & \cdots & \cdots & P_{p+q-2} & P_{p+q-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ P_1 & P_2 & \cdots & \cdots & P_{q-1} & P_q \\ 0 & P_1 & P_2 & \cdots & P_{q-2} & P_{q-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & P_1 & \cdots & P_p \end{pmatrix}.$$

Definition 1. An $\mathrm{ARX}_d(p, q)$ model is said to be strongly controllable if B is causal and Π is invertible,

$(A_2)\quad \det(\Pi) \neq 0.$

Remark 2. The concept of strong controllability is not really restrictive. For example, if $p = q = 1$, assumption (A_2) reduces to $\det(A_1) \neq 0$, if $p = 2, q = 1$ to $\det(A_2 - B_1A_1) \neq 0$, and if $p = 1$, $q = 2$ to $\det(A_1) \neq 0$.

One can observe that our strong controllability notion is closely related to the usual concept of controllability or to the coprimness of the matrix polynomials $A - I_d$ and B. As a matter of fact, the resultant of the polynomials $A - I_d$ and B is given by

$$\operatorname{Res}(A - I_d, B) = \begin{vmatrix} -A_p & -A_{p-1} & \cdots & -A_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & -A_p & \cdots & -A_2 & -A_1 & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & -A_p & -A_{p-1} & \cdots & -A_1 & 0 \\ B_q & B_{q-1} & \cdots & B_1 & I_d & 0 & \cdots & \cdots & 0 \\ 0 & B_q & \cdots & B_2 & B_1 & I_d & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & B_q & B_{q-1} & \cdots & B_1 & I_d \end{vmatrix}.$$

This determinant involves q rows with the matrices A_i and p rows with the matrices B_j. It is not hard to see Fresnel (2001) and Gelfand, Kapranov, and Zelevinsky (1994) that $\operatorname{Res}(A - I_d, B) = \det(R)$ where R is the Sylvester matrix

$$R = \begin{pmatrix} I_d & B_1 & B_2 & \cdots & B_q & 0 & \cdots & \cdots & 0 \\ 0 & I_d & B_1 & B_2 & \cdots & B_q & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & I_d & B_1 & B_2 & \cdots & B_q \\ 0 & -A_1 & -A_2 & \cdots & -A_p & 0 & \cdots & \cdots & 0 \\ 0 & 0 & -A_1 & -A_2 & \cdots & -A_p & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & 0 & -A_1 & -A_2 & \cdots & -A_p \end{pmatrix}.$$

For all $z \in \mathbb{C}$ such that $|z| \leq r$, we can also define

$$D(z) = B^{-1}(z) = \sum_{k=0}^{\infty} D_k z^k,$$

$$Q(z) = (A(z) - I_d)B^{-1}(z) = \sum_{k=1}^{\infty} Q_k z^k,$$

where the matrices D_k and Q_k may be explicitly calculated as functions of the matrices A_i and B_j. Let Δ be the symmetric square matrix of order δ

$$\Delta = \begin{pmatrix} I_d & D_1 & \cdots & D_{p+q-2} & D_{p+q-1} \\ 0 & I_d & D_1 & \cdots & D_{p+q-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & I_d & D_1 \\ 0 & \cdots & \cdots & 0 & I_d \end{pmatrix}.$$

Moreover, let Ш be the square matrix of order dq given, if $p \geq q$, by

$$Ш = \begin{pmatrix} Q_p & Q_{p+1} & \cdots & Q_{p+q-2} & Q_{p+q-1} \\ Q_{p-1} & Q_p & Q_{p+1} & \cdots & Q_{p+q-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ Q_{p-q+2} & \cdots & Q_{p-1} & Q_p & Q_{p+1} \\ Q_{p-q+1} & Q_{p-q+2} & \cdots & Q_{p-1} & Q_p \end{pmatrix},$$

while, if $p \leq q$, by

$$Ш = \begin{pmatrix} Q_p & Q_{p+1} & \cdots & \cdots & Q_{p+q-2} & Q_{p+q-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Q_1 & Q_2 & \cdots & \cdots & Q_{q-1} & Q_q \\ 0 & Q_1 & Q_2 & \cdots & Q_{q-2} & Q_{q-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & Q_1 & \cdots & Q_p \end{pmatrix}.$$

In addition, denote by T the rectangular matrix of dimension $dq \times dp$ given, if $p \geq q$, by

$$T = \begin{pmatrix} 0 & Q_1 & Q_2 & \cdots & \cdots & Q_{p-2} & Q_{p-1} \\ 0 & 0 & Q_1 & \cdots & \cdots & Q_{p-3} & Q_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & Q_1 & Q_2 & \cdots & Q_{p-q+1} \\ 0 & \cdots & \cdots & 0 & Q_1 & \cdots & Q_{p-q} \end{pmatrix},$$

while, if $p \leq q$, by

$$T = \begin{pmatrix} 0 & Q_1 & \cdots & Q_{p-2} & Q_{p-1} \\ 0 & 0 & Q_1 & \cdots & Q_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & Q_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

We can easily see that

$$\mathrm{Res}(A - I_d, B) = \det(R) = \det(R\Delta) = \det(Ш).$$

Moreover, one can observe that the matrix Π is different from Ш, except in the particular case of dimension $d = 1$. Consequently, if $d = 1$,

$$\det(\Pi) \neq 0 \Longleftrightarrow A - I_d \text{ and } B \text{ are coprime} \tag{5}$$

which corresponds to the usual notion of controllability. Otherwise, if $d \geq 2$, it is easy to provide many counterexamples for which the equivalence (5) fails.

For $1 \leq i \leq q$, denote by H_i the square matrix of order d, $H_i = \sum_{k=i}^{\infty} P_k \Gamma P_{k-i+1}^t$. In addition, let H be the square matrix of order dq

$$H = \begin{pmatrix} H_1 & H_2 & \cdots & H_{q-1} & H_q \\ H_2^t & H_1 & H_2 & \cdots & H_{q-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ H_{q-1}^t & \cdots & H_2^t & H_1 & H_2 \\ H_q^t & H_{q-1}^t & \cdots & H_2^t & H_1 \end{pmatrix}. \tag{6}$$

For $1 \leq i \leq p$, let $K_i = P_i \Gamma$ and denote by K the rectangular matrix of dimension $dq \times dp$ given, if $p \geq q$, by

$$K = \begin{pmatrix} 0 & K_1 & K_2 & \cdots & \cdots & K_{p-2} & K_{p-1} \\ 0 & 0 & K_1 & \cdots & \cdots & K_{p-3} & K_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & K_1 & K_2 & \cdots & K_{p-q+1} \\ 0 & \cdots & \cdots & 0 & K_1 & \cdots & K_{p-q} \end{pmatrix},$$

while, if $p \leq q$, by

$$K = \begin{pmatrix} 0 & K_1 & \cdots & K_{p-2} & K_{p-1} \\ 0 & 0 & K_1 & \cdots & K_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & K_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Finally, let L be the block diagonal matrix of order dp

$$L = \mathrm{diag}(\Gamma, \ldots, \Gamma), \tag{7}$$

and denote by Λ the square matrix of order $\delta = d(p+q)$

$$\Lambda = \begin{pmatrix} L & K^t \\ K & H \end{pmatrix}. \tag{8}$$

The following lemma is the keystone of all our analysis.

Lemma 3. *Let S be the Schur complement of L in Λ*

$$S = H - KL^{-1}K^t. \tag{9}$$

If (A_1) *and* (A_2) *hold, S and Λ are invertible and*

$$\Lambda^{-1} = \begin{pmatrix} L^{-1} + L^{-1}K^tS^{-1}KL^{-1} & -L^{-1}K^tS^{-1} \\ -S^{-1}KL^{-1} & S^{-1} \end{pmatrix}. \tag{10}$$

Proof. The proof is given in Appendix A. □

3. Estimation and adaptive control

First of all, we focus our attention on the estimation of the parameter θ. We shall make use of the weighted least squares (WLS) algorithm which satisfies, for all $n \geq 0$,

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n + a_n S_n^{-1}(a)\Phi_n \left(X_{n+1} - U_n - \widehat{\theta}_n^t \Phi_n\right)^t, \tag{11}$$

where the initial value $\hat{\theta}_0$ may be arbitrarily chosen and

$$S_n(a) = \sum_{k=0}^{n} a_k \Phi_k \Phi_k^t + I_\delta,$$

where the identity matrix I_δ with $\delta = d(p+q)$ is added in order to avoid useless invertibility assumption. The choice of the weighted sequence (a_n) is crucial. If $a_n = 1$, we find the standard LS algorithm with

$$S_n = \sum_{k=0}^{n} \Phi_k \Phi_k^t + I_\delta,$$

while, if $a_n^{-1} = (\log s_n)^{1+\gamma}$ with $s_n = \sum_{k=0}^{n} \|\Phi_k\|^2$ and $\gamma > 0$, we obtain the WLS algorithm introduced by Bercu and Duflo (1992) and Bercu (1995).

Next, we are concern with the choice of the stochastic adaptive control U_n. The crucial role played by U_n is to regulate the dynamic of the process (X_n) by forcing X_n to track step by step a predictable reference trajectory (x_n). We shall make use of the adaptive tracking control proposed by Aström and Wittenmark (1995) given, for all $n \geq 0$, by

$$U_n = x_{n+1} - \widehat{\theta}_n^t \Phi_n. \tag{12}$$

By substituting (12) into (2), we obtain the closed-loop system

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1}, \tag{13}$$

where the prediction error $\pi_n = (\theta - \widehat{\theta}_n)^t \Phi_n$. In all the sequel, we assume that the reference trajectory (x_n) satisfies

$$\sum_{k=1}^{n} \|x_k\|^2 = o(n) \quad \text{a.s.} \tag{14}$$

In addition, we also assume that the driven noise (ε_n) satisfies the strong law of large numbers, i.e. if

$$\Gamma_n = \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k \varepsilon_k^t, \tag{15}$$

then Γ_n converges a.s. to Γ. That is the case if, for example, (ε_n) is a white noise or if (ε_n) has a finite conditional moment of order > 2. Finally, let (C_n) be the average cost matrix sequence defined by

$$C_n = \frac{1}{n}\sum_{k=1}^{n} (X_k - x_k)(X_k - x_k)^t.$$

Definition 4. The tracking is said to be optimal if C_n converges a.s. to Γ.

4. Main results

Our first result is related to the almost sure properties of the LS algorithm.

Theorem 5. *Assume that the* $\mathrm{ARX}_d(p, q)$ *model is strongly controllable and that* (ε_n) *has finite conditional moment of order* > 2. *Then, for the LS algorithm, we have*

$$\lim_{n\to\infty} \frac{S_n}{n} = \Lambda \quad \text{a.s.} \tag{16}$$

where the limiting matrix Λ *is given by* (8). *In addition, the tracking is optimal*

$$\|C_n - \Gamma_n\| = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{17}$$

We can sharpen (17) *by*

$$\lim_{n\to\infty} \frac{1}{\log n}\sum_{k=1}^{n} (X_k - x_k - \varepsilon_k)(X_k - x_k - \varepsilon_k)^t = \delta\Gamma \quad \text{a.s.}$$

with $\delta = d(p + q)$. *Finally,* $\widehat{\theta}_n$ *converges almost surely to* θ

$$\|\widehat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{18}$$

Our second result deals with the almost sure properties of the WLS algorithm.

Theorem 6. *Assume that the* $\mathrm{ARX}_d(p, q)$ *model is strongly controllable. In addition, suppose that either* (ε_n) *is a white noise or* (ε_n) *has finite conditional moment of order* > 2. *Then, for the WLS algorithm with* $a_n^{-1} = (\log s_n)^{1+\gamma}$ *where* $\gamma > 0$, *we have*

$$\lim_{n\to\infty} (\log n)^{1+\gamma}\frac{S_n(a)}{n} = \Lambda \quad \text{a.s.} \tag{19}$$

where the limiting matrix Λ *is given by* (8). *In addition, the tracking is optimal*

$$\|C_n - \Gamma_n\| = o\left(\frac{(\log n)^{1+\gamma}}{n}\right) \quad \text{a.s.} \tag{20}$$

Finally, $\widehat{\theta}_n$ *converges almost surely to* θ

$$\|\widehat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{(\log n)^{1+\gamma}}{n}\right) \quad \text{a.s.} \tag{21}$$

Remark 7. One can observe that Theorems 5 and 6 extend the results of Bercu (1998) and Guo (1994, 1995) previously established for controlled $\mathrm{AR}_d(p)$ models.

Theorem 8. *Assume that the* $\mathrm{ARX}_d(p, q)$ *model is strongly controllable and that* (ε_n) *has finite conditional moment of order* $\alpha > 2$. *In addition, suppose that* (x_n) *has the same regularity in norm as* (ε_n) *which means that for all* $2 < \beta < \alpha$

$$\sum_{k=1}^{n} \|x_k\|^\beta = \mathcal{O}(n) \quad \text{a.s.} \tag{22}$$

Then, the LS and WLS algorithms share the same central limit theorem

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Lambda^{-1} \otimes \Gamma) \tag{23}$$

where the inverse matrix Λ^{-1} *is given by* (10) *and the symbol* $\otimes$ *stands for the matrix Kronecker product. Moreover, for any vectors* $u \in \mathbb{R}^d$ *and* $v \in \mathbb{R}^\delta$ *with* $\delta = d(p + q)$, *they also share the same law of iterated logarithm*

$$\begin{aligned}\limsup_{n\to\infty} &\left(\frac{n}{2\log\log n}\right)^{1/2} v^t(\widehat{\theta}_n - \theta)u \\ &= -\liminf_{n\to\infty}\left(\frac{n}{2\log\log n}\right)^{1/2} v^t(\widehat{\theta}_n - \theta)u \\ &= \left(v^t\Lambda^{-1}v\right)^{1/2}\left(u^t\Gamma u\right)^{1/2} \quad \text{a.s.}\end{aligned} \tag{24}$$

In particular,

$$\left(\frac{\lambda_{\min}(\Gamma)}{\lambda_{\max}(\Lambda)}\right) \le \limsup_{n\to\infty}\left(\frac{n}{2\log\log n}\right)\|\hat{\theta}_n - \theta\|^2 \quad \text{a.s.}$$

$$\limsup_{n\to\infty}\left(\frac{n}{2\log\log n}\right)\|\hat{\theta}_n - \theta\|^2 \le \left(\frac{\lambda_{\max}(\Gamma)}{\lambda_{\min}(\Lambda)}\right) \quad \text{a.s.}$$

where $\lambda_{\min}(\Gamma)$ *and* $\lambda_{\max}(\Gamma)$ *are the minimum and the maximum eigenvalues of* Γ.

Proof. The proofs are given in Appendix B. □

Remark 9. Some numerical simulations illustrating the asymptotic results of Section 4 can be found in Bercu and Vazquez (2008).

5. Conclusion

Via our new concept of strong controllability, we have extended the analysis of the almost sure convergence for both LS and WLS algorithms in the multidimensional ARX framework. It enables us to provide a positive answer to a conjecture in Bercu (1998) by establishing a CLT and a LIL for these two stochastic algorithms. In our approach, the leading matrix associated with the matrix polynomial B, commonly called the high frequency gain, was supposed to be known and it was chosen as the identity matrix I_d. It would be a great challenge for the control community to carry out similar analysis with unknown high frequency gain and to extend it to ARMAX models.

Acknowledgements

The authors would like to thanks Jean Fresnel for very stimulating conversations on linear Algebra and Schur complement. Many thanks are also due to Han-Fu Chen, Michel Fliess, Lei Guo, and Alexander Nazin for fruitful discussions on controllability.

Appendix A. Proof of the keystone Lemma 3

Let Σ be the infinite-dimensional diagonal square matrix

$$\Sigma = \mathrm{diag}(\Gamma, \ldots, \Gamma, \ldots).$$

Moreover, denote by T the infinite-dimensional rectangular matrix with dq rows and an infinite number of columns given, if $p \geq q$, by

$$T = \begin{pmatrix} P_p & P_{p+1} & \cdots & P_k & P_{k+1} & \cdots \\ P_{p-1} & P_p & \cdots & P_{k-1} & P_k & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ P_{p-q+2} & P_{p-q+3} & \cdots & P_{k-q+2} & P_{k-q+3} & \cdots \\ P_{p-q+1} & P_{p-q+2} & \cdots & P_{k-q+1} & P_{k-q+2} & \cdots \end{pmatrix},$$

while, if $p \leq q$, by

$$T = \begin{pmatrix} P_p & P_{p+1} & \cdots & \cdots & P_k & P_{k+1} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ P_1 & P_2 & \cdots & \cdots & P_{k-p+1} & P_{k-p+2} & \cdots \\ 0 & P_1 & P_2 & \cdots & P_{k-p} & P_{k-p+1} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & P_1 & P_2 & \cdots \end{pmatrix}.$$

After some straightforward, although rather lengthy, linear algebra calculations, it is possible to deduce from (9) that

$$S = T \Sigma T^t. \tag{A.1}$$

It clearly follows from (A.1) that $\ker(S) = \ker(T^t)$. As a matter of fact, assume that $v \in \mathbb{R}^{dq}$ belongs to $\ker(T^t)$. Then, $T^t v = 0$, $Sv = 0$ so $\ker(T^t) \subset \ker(S)$. On the other hand, assume that $v \in \mathbb{R}^{dq}$ belongs to $\ker(S)$. Since $Sv = 0$, we clearly have $v^t S v = 0$, $v^t T \Sigma T^t v = 0$. However, the matrix Γ is positive definite. Consequently, $T^t v = 0$ and $\ker(S) \subset \ker(T^t)$. Moreover, it follows from the well-known rank theorem that

$$dq = \dim(\ker(S)) + \operatorname{rank}(S). \tag{A.2}$$

As soon as $\ker(S) = \{0\}$, $\dim(\ker(S)) = 0$ and we obtain from (A.2) that S is of full rank dq which means that S is invertible. Furthermore, the left hand side square matrix of order dq of the infinite-dimensional matrix T is precisely Π. Consequently, if Π is invertible, Π is of full rank dq, $\ker(\Pi) = \ker(\Pi^t) = \{0\}$ and the left null space of T reduces to the null vector of $\mathbb{R}^{dq}$. Hence, if Π is invertible, we deduce from (A.2) that S is also invertible. Finally, as

$$\det(\Lambda) = \det(L)\det(S) = \det(\Gamma)^p \det(S), \tag{A.3}$$

we obtain from (A.3) that Λ is invertible and (10) follows from Horn and Johnson (1990, page 18), completing the proof of Lemma 3. □

Appendix B. Proofs of Theorems 5, 6 and 8

In order to prove Theorem 5, we shall make use of the same approach than Bercu (1998) or Guo and Chen (1991). First of all, we recall that for all $n \geq 0$

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1}. \tag{B.1}$$

In addition, let $s_n = \sum_{k=0}^n \|\Phi_k\|^2$. It follows from (B.1) and the strong law of large numbers for martingales (see e.g. Corollary 1.3.25 of Duflo, 1997) that $n = \mathcal{O}(s_n)$ a.s. Moreover, by Lemma 1 of Guo and Chen (1991), we have

$$\sum_{k=1}^n (1 - f_k)\|\pi_k\|^2 = \mathcal{O}(\log s_n) \quad \text{a.s.} \tag{B.2}$$

where $f_n = \Phi_n^t S_n^{-1} \Phi_n$. Hence, if (ε_n) has finite conditional moment of order $\alpha > 2$, we can show by the causality assumption (A_1) on the matrix polynomial B together with (B.2) that $\|\Phi_n\|^2 = \mathcal{O}(s_n^\beta)$ a.s. for all $2\alpha^{-1} < \beta < 1$. In addition, let $g_n = \Phi_n^t S_{n-1}^{-1} \Phi_n$ and $\delta_n = \operatorname{tr}(S_{n-1}^{-1} - S_n^{-1})$. It follows from Proposition 4.2.12 of Duflo (1997) that $(1 - f_n)(1 + g_n) = 1$. Moreover, (δ_n) tends to zero a.s. Consequently, as $1 + g_n \leq 2 + \delta_n \|\Phi_n\|^2$, we can deduce from (B.2) that

$$\sum_{k=1}^n \|\pi_k\|^2 = o(s_n^\beta \log s_n) \quad \text{a.s.} \tag{B.3}$$

Therefore, we obtain from (14), (B.1) and (B.3) that

$$\sum_{k=1}^n \|X_{k+1}\|^2 = o(s_n^\beta \log s_n) + \mathcal{O}(n) \quad \text{a.s.} \tag{B.4}$$

Furthermore, we have from assumption (A_1) that the control $U_n = B^{-1}(R)A(R)X_{n+1} - B^{-1}(R)\varepsilon_{n+1}$. It implies by (B.4) that

$$\sum_{k=1}^n \|U_k\|^2 = o(s_n^\beta \log s_n) + \mathcal{O}(n) \quad \text{a.s.} \tag{B.5}$$

It remains to put together the two contributions (B.4) and (B.5) to deduce that $s_n = o(s_n) + \mathcal{O}(n)$ a.s. leading to $s_n = \mathcal{O}(n)$ a.s. Hence, it follows from (B.3) that $\sum_{k=1}^n \|\pi_k\|^2 = o(n)$ a.s. Consequently, we obtain from (14) and (B.1) that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n X_k X_k^t = \Gamma \quad \text{a.s.}$$

and, for all $1 \leq i \leq p-1$, $\sum_{k=0}^n X_k X_{k-i}^t = o(n)$ a.s. so

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n (X_k^p)^t X_k^p = L \quad \text{a.s.} \tag{B.6}$$

where $X_n^p = (X_n^t, \dots, X_{n-p+1}^t)$ and the matrix L is given by (7). Furthermore, we already saw that the control $U_n = V_n + W_{n+1}$ where the first term $V_n = B^{-1}(R)A(R)(\pi_n + x_{n+1})$ and the second term $W_{n+1} = B^{-1}(R)(A(R) - I_d)\varepsilon_{n+1}$. On the one hand, we obtain from (14) that

$$\sum_{k=1}^n \|V_k\|^2 = o(n) \quad \text{a.s.}$$

Thus, it follows from the Cauchy–Schwarz inequality that for all $1 \leq i \leq q$, $\sum_{k=1}^n V_k V_{k-i+1}^t = o(n)$ a.s. as well as $\sum_{k=1}^n V_k W_{k-i+2}^t = o(n)$ a.s. On the other hand, we deduce from the strong law of large numbers for martingales (see e.g. Theorem 4.3.16 of Duflo, 1997) that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \varepsilon_k \varepsilon_k^t = \Gamma \quad \text{a.s.}$$

while, for all $i, j \geq 0$ with $i \neq j$, $\sum_{k=1}^n \varepsilon_{k-i}\varepsilon_{k-j}^t = o(n)$ a.s. Consequently, as $W_n = P(R)\varepsilon_n$, we obtain that for all $1 \leq i \leq q$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n U_k U_{k-i+1}^t = H_i \quad \text{a.s.}$$

which ensures that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n (U_k^q)^t U_k^q = H \quad \text{a.s.} \tag{B.7}$$

where $U_n^q = (U_n^t, \dots, U_{n-q+1}^t)$ and the matrix H is given by (6). Via the same lines, we also find that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n (U_{k-1}^q)^t X_k^p = K \quad \text{a.s.} \tag{B.8}$$

Therefore, it follows from the conjunction of (B.6)–(B.8) that

$$\lim_{n\to\infty} \frac{S_n}{n} = \Lambda \quad \text{a.s.} \tag{B.9}$$

where the limiting matrix Λ is given by (8). Hereafter, we recall that the $\text{ARX}_d(p, q)$ model is strongly controllable. Thanks to

Lemma 3, the matrix Λ is invertible and Λ^{-1}, given by (10), may be explicitly calculated. This is the key point for the rest of the proof. On the one hand, it follows from (B.9) that $n = \mathcal{O}(\lambda_{\min}(S_n))$, $\|\Phi_n\|^2 = o(n)$ a.s. which implies that f_n tends to zero a.s. Hence, by (B.2), we find that

$$\sum_{k=1}^{n} \|\pi_k\|^2 = \mathcal{O}(\log n) \quad \text{a.s.} \tag{B.10}$$

On the other hand, we obviously have from (B.1)

$$\|C_n - \Gamma_n\| = \mathcal{O}\left(\frac{1}{n}\sum_{k=1}^{n}\|\pi_{k-1}\|^2\right) \quad \text{a.s.} \tag{B.11}$$

Consequently, we immediately obtain the tracking optimality (17) from (B.10) and (B.11). Furthermore, by a well-known result of Lai and Wei (1986) on the LS estimator, we also have

$$\|\widehat{\theta}_{n+1} - \theta\|^2 = \mathcal{O}\left(\frac{\log \lambda_{\max}(S_n)}{\lambda_{\min}(S_n)}\right) \quad \text{a.s.} \tag{B.12}$$

Hence (18) clearly follows from (B.9) and (B.12). Moreover, we infer from Lemma 1 of Wei (1987) together with (B.9) that $(\widehat{\theta}_{n+1} - \theta)^t S_n(\widehat{\theta}_{n+1} - \theta) = o(\log n)$ a.s. Therefore, it follows from Theorem 4.3.16 part 4 of Duflo (1997) that

$$\lim_{n\to\infty} \frac{1}{\log d_n}\sum_{k=0}^{n}(1 - f_k)\pi_k\pi_k^t = \Gamma \quad \text{a.s.} \tag{B.13}$$

where $d_n = \det(S_n)$. In addition, if $\delta = d(p+q)$, we deduce from (B.9) that

$$\lim_{n\to\infty} \frac{d_n}{n^\delta} = \det \Lambda \quad \text{a.s.} \tag{B.14}$$

Finally, (B.13) and (B.14) imply that

$$\lim_{n\to\infty} \frac{1}{\log n}\sum_{k=0}^{n}\pi_k\pi_k^t = \delta\Gamma \quad \text{a.s.}$$

which achieves the proof of Theorem 5. We now carry out the proof of Theorem 6. By Theorem 1 of Bercu (1995), we have

$$\sum_{n=1}^{\infty} a_n(1 - f_n(a))\|\pi_n\|^2 < +\infty \quad \text{a.s.} \tag{B.15}$$

where $f_n(a) = a_n\Phi_n^t S_n^{-1}(a)\Phi_n$. Then, as $a_n^{-1} = (\log(s_n))^{1+\gamma}$ with $\gamma > 0$, we clearly have $a_n^{-1} = \mathcal{O}(s_n)$ a.s. Hence, it follows from (B.15) together with Kronecker's Lemma given e.g. by Lemma 1.3.14 of Duflo (1997) that

$$\sum_{k=1}^{n} \|\pi_k\|^2 = o(s_n) \quad \text{a.s.} \tag{B.16}$$

Therefore, we obtain from (14), (B.1) and (B.16) that

$$\sum_{k=1}^{n} \|X_{k+1}\|^2 = o(s_n) + \mathcal{O}(n) \quad \text{a.s.} \tag{B.17}$$

In addition, we also deduce from assumption (A_1) that

$$\sum_{k=1}^{n} \|U_k\|^2 = o(s_n) + \mathcal{O}(n) \quad \text{a.s.} \tag{B.18}$$

Consequently, we infer from (B.17) and (B.18) that $s_n = o(s_n) + \mathcal{O}(n)$ so $s_n = \mathcal{O}(n)$ a.s. Hence, (B.16) implies that $\sum_{k=1}^{n}\|\pi_k\|^2 = o(n)$ a.s. Proceeding exactly as in the proof of Theorem 5, we find that

$$\lim_{n\to\infty} \frac{S_n}{n} = \Lambda \quad \text{a.s.}$$

Via an Abel transform, it ensures that

$$\lim_{n\to\infty} (\log n)^{1+\gamma}\frac{S_n(a)}{n} = \Lambda \quad \text{a.s.} \tag{B.19}$$

We obviously have from (B.19) that $f_n(a)$ tends to zero a.s. Consequently, we obtain from (B.15) and Kronecker's Lemma that $\sum_{k=1}^{n}\|\pi_k\|^2 = o((\log s_n)^{1+\gamma})$ a.s. Then, (20) clearly follows from (B.11). Finally, by Theorem 1 of Bercu (1995)

$$\|\widehat{\theta}_{n+1} - \theta\|^2 = \mathcal{O}\left(\frac{1}{\lambda_{\min}(S_n(a))}\right) \quad \text{a.s.} \tag{B.20}$$

Hence, we obtain (21) from (B.19) and (B.20), which completes the proof of Theorem 6. Finally, the proof of Theorem 8 is left to the reader as it follows essentially the same lines as those in Appendix C of Bercu (1998) in the controlled $\mathrm{AR}_d(p)$ framework.

References

Aström, K. J., & Wittenmark, B. (1995). *Adaptive control* (2nd ed.). New York: Addison-Wesley.

Baraldi, P., Manginelli, A. A., Maieron, M., Liberati, D., & Porro, C. A. (2007). An ARX model-based approach to trial by trial identification of fMRI-BOLD responses. *NeuroImage*, *37*, 189–201.

Bercu, B., & Duflo, M. (1992). Moindres carrés pondérés et poursuite. *Annals de l'Institut Henri Poincaré*, *28*, 403–430.

Bercu, B. (1995). Weighted estimation and tracking for ARMAX models. *SIAM Journal of Control and Optimization*, *33*, 89–106.

Bercu, B. (1998). Central limit theorem and law of iterated logarithm for least squares algorithms in adaptive tracking. *SIAM Journal of Control and Optimization*, *36*, 910–928.

Bercu, B., & Portier, B. (2002). Adaptive control of parametric nonlinear autoregressive models via a new martingale approach. *IEEE Transactions on Automatic Control*, *47*, 1524–1528.

Bercu, B., & Portier, B. (2008). Kernel density estimation and goodness-of-fit test in adaptive tracking. *SIAM Journal of Control and Optimization*, *47*, 2440–2457.

Bercu, B., & Vazquez, V. (2008). Further results for ARX models in adaptive tracking. In *47th IEEE conference on decision and control*.

Caines, P. E. (1988). *Linear stochastic systems*. New York: John Wiley.

Chang, H., & Tzenog, P. K. (2008). Analysis of the dynamic characteristics of pressure sensors using ARX system identification. *Sensors and Actuators A: Physical*, *141*, 367–375.

Chen, H. F., & Guo, L. (1991). *Identification and stochastic adaptive control*. Boston: Birkhäuser.

Duflo, M. (1997). *Random iterative models*. Berlin: Springer Verlag.

Fresnel, J. (2001). *Anneaux*. Paris: Hermann.

Fukata, K., Washio, T., & Motoda, H. (2006). A method to search ARX model orders and its application to sales dynamics analysis. In *ICDM workshops* (pp. 590–595).

Gelfand, I. M., Kapranov, M. M., & Zelevinsky, A. V. (1994). *Discriminants, resultants, and multidimensional determinants*. Boston: Birkhäuser.

Guo, L., & Chen, H. F. (1991). The Aström Wittenmark self-tuning regulator revisited and ELS-based adaptive trackers. *IEEE Trans. Automat. Control*, *36*, 802–812.

Guo, L. (1994). Further results on least squares based adaptive minimum variance control. *SIAM Journal of Control and Optimization*, *32*, 187–212.

Guo, L. (1995). Convergence and logarithm laws of self-tuning regulators. *Automatica*, *31*, 435–450.

Guo, L. (1996). Self convergence of weighted least squares with applications to stochastic adaptive control. *IEEE Trans. Automat. Control*, *41*, 79–89.

Horn, R. A., & Johnson, C. R. (1990). *Matrix analysis*. New York: Cambridge University Press.

Huang, K. Y., & Jane, C.-J. (2009). A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories. *Expert Systems with Applications*, *36*, 5387–5392.

Jankumas, A. (2000). Optimal adaptive control for estimation of parameters of ARX models. *IEEE Transactions on Automatic Control*, *45*, 964–968.

Lai, T. L., & Wei, C. Z. (1986). Extended least squares and their applications to adaptive control and prediction in linear systems. *IEEE Transactions on Automatic Control*, *31*, 898–906.

Li, T., & Zhang, J.-F. (2008). Decentralized tracking-type games for multi-agent systems with coupled ARX models: asymptotic Nash equilibria. *Automatica*, *44*, 713–725.

Liu, Y., Birch, A. A., & Allen, R. (2003). Dynamic cerebral autoregulation assessment using an ARX model: comparative study using step response and phase shift analysis. *Medical Engineering and Physics*, *25*, 647–653.

Nazin, A. V. (1993). An asymptotically efficient algorithm for the adaptive control of a multidimensional linear plant. *Automation and Remote Control*, *54*, 1115–1127.

Vidal, R. (2008). Recursive identification of switched ARX systems. *Automatica*, *44*, 2274–2287.

Wei, C. Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Annals of Statistics*, *15*, 1667–1682.

Bernard Bercu was born in Paris, France, in 1965. He received the Ph.D. degree in Mathematics in 1992 and the HDR in Mathematics in 2000, both from the University of Paris-Sud, Orsay, France. He was Professor of Mathematics at the University of Toulouse and he is currently Professor of Mathematics at the University of Bordeaux. His research interests lie in both Probability and Statistics, including stochastic algorithms and martingales, estimation theory and adaptive control, bifurcation and branching processes, large deviations theory and concentration inequalities with Statistical applications.

Victor Vazquez was born in Puebla, México, in 1982. He received the Ph.D. degree in Mathematics in 2010 from the Benemérita Universidad Autónoma de Puebla, México. He is working on stochastic algorithms and martingales, more precisely on estimation and adaptive control for ARX processes. He is currently a full time Professor of Mathematics at the Benemérita Universidad Autónoma de Puebla.