ON THE USEFULNESS OF PERSISTENT EXCITATION IN ARX ADAPTIVE TRACKING

BERNARD BERCU AND VICTOR VAZQUEZ

Abstract. The usefulness of persistent excitation is well-known in the control community. Thanks to a persistently excited adaptive tracking control, we show that it is possible to avoid the strong controllability assumption recently proposed in the multidimensional ARX framework. We establish the almost sure convergence for both least squares and weighted least squares estimators of the unknown parameters. A central limit theorem and a law of iterated logarithm are also provided. All this asymptotical analysis is related to the Schur complement of a suitable limiting matrix.

1. INTRODUCTION

The concept of persistent excitation is well-known in the control community. Since the pioneers works of Anderson [3] and Moore [24], this concept has been successfullly used in a large variety of fields of application going from economics [2], [13], to adaptive or learning control [12], [16], [22], [23], or mechanical engineering and robotics [1], [15], and [20]. In this paper, we use a persistently excited adaptive tracking control in the multidimensional ARX framework. It allows us to avoid the strong controllability assumption recently proposed by Bercu and Vazquez [8], [9]. More precisely, we shall establish the almost sure convergence for both least squares (LS) and weighted least squares (WLS) estimators of the unknown parameters of ARX model. The asymptotic normality as well as a law of iterated logarithm are also provided. 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

$$A(R)X_{n+1} = B(R)U_n + \varepsilon_{n+1} \tag{1.1}$$

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

$$\begin{aligned} A(z) &= I_d - A_1 z - \cdots - A_p z^p, \\ B(z) &= I_d + B_1 z + \cdots + B_q z^q, \end{aligned}$$

2000 *Mathematics Subject Classification.* Primary: 62G05 Secondary: 93C40, 15A09, 60F05, 60F15.

Key words and phrases. Estimation, adaptive control, persistent excitation, Schur complement, central limit theorem, law of iterated logarithm.

This work has been supported by INRIA, by CONACYT, and by the ECOS Scientific Cooperation Programme.

where A_i and B_j are unknown square matrices of order d and I_d is the identity matrix. Relation (1.1) may be rewritten in the compact form

$$X_{n+1} = \theta^t \Phi_n + U_n + \varepsilon_{n+1} \tag{1.2}$$

where the regression vector $\Phi_n = \left(X_n^p, U_{n-1}^q\right)^t$ with

$$\begin{aligned} X_n^p &= (X_n^t, \ldots, X_{n-p+1}^t), \\ U_n^q &= (U_n^t, \ldots, U_{n-q+1}^t), \end{aligned}$$

and the unknown parameter θ is given by

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

In all the sequel, we shall assume that the driven noise (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. Moreover, 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. In addition, we suppose that the driven noise (ε_n) satisfies the strong law of large numbers i.e. if

$$\Gamma_n = \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k \varepsilon_k^t, \tag{1.3}$$

then the sequence (Γ_n) converges to Γ a.s. 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.

The paper is organized as follows. Section 2 deals with the parameter estimation and the persistently excited adaptive tracking control. Section 3 is devoted to the introduction of the Schur complement approach together with some linear algebra calculations. In Section 4, we propose some usefull almost sure convergence properties together with a central limit theorem (CLT) and a law of iterated logarithm (LIL) for both LS and WLS estimators. Some numerical simulations are also provided in Section 5. Finally, a short conclusion is given in Section 6.

2. Estimation and Adaptive control

In the ARX tracking framework, we must deal with two objectives simultaneously. On the one hand, it is necessary to estimate the unknown parameter θ. On the other hand, the output (X_n) has to track, step by step, a predictable reference trajectory (x_n). First, we focus our attention on the estimation of the parameter θ. We shall make use of the WLS algorithm which satisfies, for all $n \geq 0$,

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n + a_n S_n^{-1}(a)\Phi_n \left(X_{n+1} - U_n - \widehat{\theta}_n^t \Phi_n\right)^t \tag{2.1}$$

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 the useless invertibility assumption. The choice of the weighted sequence (a_n) is crucial. If

$$a_n = 1$$

we find the standard LS estimator, while if $\gamma > 0$,

$$a_n = \Big(\frac{1}{\log s_n}\Big)^{1+\gamma} \quad \text{with} \quad s_n = \sum_{k=0}^{n} \| \Phi_k \|^2,$$

we obtain the WLS estimator introduced by Bercu and Duflo [5], [6]. Next, we are concern with the choice of the adaptive control sequence (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 a predictable reference trajectory (x_n). We propose to make use of the persistently excited adaptive tracking control given, for all $n \geq 0$, by

$$U_n = x_{n+1} - \widehat{\theta}_n^{\,t} \Phi_n + \xi_{n+1} \tag{2.2}$$

where (ξ_n) is an exogenous noise of dimension d, adapted to $\mathbb{F}$, with mean 0 and positive definite covariance matrix Δ. In addition, we assume that (ξ_n) is independent of (ε_n), of (x_n), and of the initial state of the system. Moreover, we suppose that (ξ_n) satisfies the strong law of large numbers. Consequently, if

$$\Delta_n = \frac{1}{n}\sum_{k=1}^{n}(\varepsilon_k + \xi_k)(\varepsilon_k + \xi_k)^t, \tag{2.3}$$

then the sequence (Δ_n) converges to $\Gamma + \Delta$ a.s. By substituting (2.2) into (1.2), we obtain the closed-loop system

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1} + \xi_{n+1} \tag{2.4}$$

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

$$\sum_{k=1}^{n} \| x_k \|^2 = o(n) \qquad \text{a.s.} \tag{2.5}$$

Finally, let (C_n) be the average cost matrix sequence defined by

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

The tracking is said to be residually optimal if (C_n) converges to $\Gamma + \Delta$ a.s.

3. On the Schur Complement

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 $|z| \leq 1$

$$\det(B(z)) \neq 0. \tag{3.1}$$

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. [14] page 155). Hence, for all $z \in \mathbb{C}$ with $|z| \leq r$, we have

$$B^{-1}(z) = \sum_{k=0}^{\infty} D_k z^k. \tag{3.2}$$

where all the matrices D_k can be explicitly calculated via the recursive equations $D_0 = I_d$ and, for all $k \geq 1$

$$D_k = -\sum_{j=0}^{k-1} D_j B_{k-j} \quad \text{if} \quad k \leq q, \tag{3.3}$$

$$D_k = -\sum_{j=1}^{q} D_{k-j} B_j \quad \text{if} \quad k > q. \tag{3.4}$$

In a similar way, for all $z \in \mathbb{C}$ such that $|z| \leq r$, we shall denote

$$P(z) = B^{-1}(z)(A(z) - I_d) = \sum_{k=1}^{\infty} P_k z^k. \tag{3.5}$$

All the matrices P_k may be explicitly calculated as functions of the matrices A_i and B_j. As a matter of fact, for all $k \geq 1$

$$P_k = -\sum_{j=0}^{k-1} D_j A_{k-j} \quad \text{if} \quad k \leq p, \tag{3.6}$$

$$P_k = -\sum_{j=1}^{p} D_{k-j} A_j \quad \text{if} \quad k > p. \tag{3.7}$$

For all $1 \leq i \leq q$, denote by H_i be the square matrix of order d

$$H_i = \sum_{k=i}^{\infty} P_k \Gamma P_{k-i+1}^t + \sum_{k=i-1}^{\infty} Q_k \Delta Q_{k-i+1}^t.$$

where, for all $k \geq 0$, $Q_k = D_k + P_k$ with $Q_0 = I_d$. In addition, let H be the symmetric square matrix of order dq

$$H = \begin{pmatrix} H_1 & H_2 & \cdots & H_{q-1} & H_q \\ H_2^t & H_1 & H_2 & \cdots & H_{q-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ H_{q-1}^t & \cdots & H_2^t & H_1 & H_2 \\ H_q^t & H_{q-1}^t & \cdots & H_2^t & H_1 \end{pmatrix}. \tag{3.8}$$

For all $0 \leq i \leq p-1$, let $K_i = P_i\Gamma + Q_i\Delta$ with $K_0 = \Delta$ and denote by K the rectangular matrix of dimension $dq \times dp$ given, if $p \geq q$, by

$$K = \begin{pmatrix} K_0 & K_1 & K_2 & \cdots & \cdots & K_{p-2} & K_{p-1} \\ 0 & K_0 & K_1 & \cdots & \cdots & K_{p-3} & K_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & K_0 & K_1 & K_2 & \cdots & K_{p-q+1} \\ 0 & \cdots & \cdots & K_0 & K_1 & \cdots & K_{p-q} \end{pmatrix}$$

while, if $p \leq q$, by

$$K = \begin{pmatrix} K_0 & K_1 & \cdots & K_{p-2} & K_{p-1} \\ 0 & K_0 & K_1 & \cdots & K_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & K_0 & K_1 \\ 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

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

$$L = \begin{pmatrix} \Gamma+\Delta & 0 & \cdots & 0 & 0 \\ 0 & \Gamma+\Delta & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \Gamma+\Delta & 0 \\ 0 & 0 & \cdots & 0 & \Gamma+\Delta \end{pmatrix}. \tag{3.9}$$

Denote by Λ the symmetric square matrix of order δ

$$\Lambda = \begin{pmatrix} L & K^t \\ K & H \end{pmatrix}. \tag{3.10}$$

This lemma is the keystone of all our asymptotic results.

Lemma 3.1. Let S be the Schur complement of L in Λ

$$S = H - KL^{-1}K^t. \tag{3.11}$$

If B is causal, then S and Λ are invertible and

$$\Lambda^{-1} = \begin{pmatrix} L^{-1}+L^{-1}K^tS^{-1}KL^{-1} & -L^{-1}K^tS^{-1} \\ -S^{-1}KL^{-1} & S^{-1} \end{pmatrix}. \tag{3.12}$$

Proof. The proof is given in Appendix A. □

Remark 3.1. One can see the usefulness of persistent excitation in ARX tracking. As we make use of a persistently excited adaptive tracking control given, it is possible to get ride of the strong controllability assumption recently proposed by Bercu and Vazquez [8], [9]. On the other hand, we will see in the next section that the tracking is not optimal but it is residually optimal. It is necessary to make a compromise between estimation and tracking optimality.

4. MAIN RESULTS

Our first result concerns to the a.s. asymptotic properties of the LS estimator.

Theorem 4.1. *Assume that B is causal and that (ε_n) has finite conditional moment of order > 2. Then, for the LS estimator, we have*

$$\lim_{n\to\infty} \frac{S_n}{n} = \Lambda \qquad a.s. \tag{4.1}$$

where the limiting matrix Λ is given by (3.10). In addition, the tracking is residually optimal

$$\| C_n - \Delta_n \| = \mathcal{O}\left(\frac{\log n}{n}\right) \qquad a.s. \tag{4.2}$$

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

$$\| \widehat{\theta}_n - \theta \|^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \qquad a.s. \tag{4.3}$$

Proof. The proof is given in Appendix B. □

Our second result is related to the almost sure properties of the WLS estimator.

Theorem 4.2. *Assume that B is causal. 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 estimator, we have*

$$\lim_{n\to\infty} (\log n)^{1+\gamma} \frac{S_n(a)}{n} = \Lambda \qquad a.s. \tag{4.4}$$

where the limiting matrix Λ is given by (3.10). In addition, the tracking is residually optimal

$$\| C_n - \Delta_n \| = o\left(\frac{(\log n)^{1+\gamma}}{n}\right) \qquad a.s. \tag{4.5}$$

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

$$\| \widehat{\theta}_n - \theta \|^2 = \mathcal{O}\left(\frac{(\log n)^{1+\gamma}}{n}\right) \qquad a.s. \tag{4.6}$$

Proof. The proof is given in Appendix C. □

Finally, we present the CLT and the LIL for both LS and WLS estimators.

Theorem 4.3. *Assume that B is causal and that (ε_n) and (ξ_n) have both finite conditional moments of order $\alpha > 2$. In addition, suppose that (x_n) satisfies for some $2 < \beta < \alpha$*

$$\sum_{k=1}^{n} \| x_k \|^\beta = \mathcal{O}(n) \qquad a.s. \tag{4.7}$$

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

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Lambda^{-1} \otimes \Gamma) \tag{4.8}$$

where the inverse matrix Λ^{-1} is given by (3.12) and the symbol $\otimes$ stands for the matrix Kronecker product. In addition, for any vectors $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^\delta$, they also share the same law of iterated logarithm

$$\begin{aligned}\limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} v^t(\widehat{\theta}_n - \theta)u &= -\liminf_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} v^t(\widehat{\theta}_n - \theta)u \\ &= \left(v^t\Lambda^{-1}v\right)^{1/2}\left(u^t\Gamma u\right)^{1/2} \quad a.s.\end{aligned} \tag{4.9}$$

In particular,

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

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

Proof. The proof is given in Appendix D. □

5. NUMERICAL SIMULATIONS

The goal of this section is to illustrate via some numerical experiments the main results of this paper. In order to keep this section brief, we consider a causal $ARX_d(p,q)$ model in dimension $d = 2$ with $p = 1$ and $q = 1$. Moreover, the reference trajectory (x_n) is chosen to be identically zero and the driven and exogenous noises (ε_n) and (ξ_n) are Gaussian $\mathcal{N}(0,1)$ white noises. Finally our numerical simulations are based on $M = 500$ realizations of sample size $N = 1000$. Consider the $ARX_2(1,1)$ model

$$X_{n+1} = AX_n + U_n + BU_{n-1} + \varepsilon_{n+1}$$

where

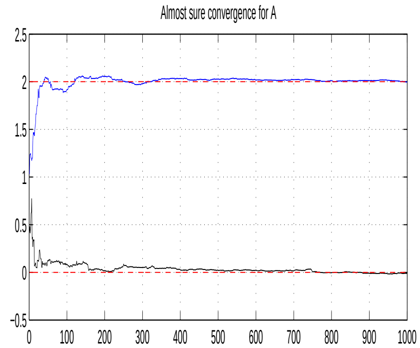
$$A = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad B = \frac{1}{4}\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}.$$

First of all, it is easy to see that this $ARX_2(1,1)$ process is not strongly controllable [8], [9], because $\det(A) = 0$. Consequently, if we use an adaptive tracking control (U_n) without persistent excitation (ξ_n), then only the matrix A and the first diagonal term of the matrix B can be properly estimated as one can see in Figure 1.
Next, we make use of the persistently excited adaptive tracking control given by

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

For all $k \geq 1$, we have $D_k = (-B)^k$ and $P_k = -(-B)^{k-1}A$ which clearly implies that

$$Q_k = -(-B)^{k-1}(A+B).$$



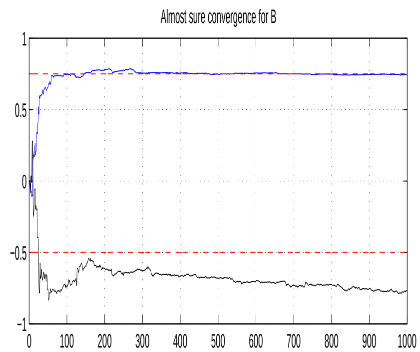


FIGURE 1. Almost Sure Convergence

Since the matrices A and B are both diagonal, we find that

$$\begin{aligned} H &= \sum_{k=1}^{\infty} P_k^2 + \sum_{k=0}^{\infty} Q_k^2, \\ &= I_2 + \sum_{k=1}^{\infty} B^{k-1} A^2 B^{k-1} + \sum_{k=1}^{\infty} B^{k-1}(A+B)^2 B^{k-1}, \\ &= I_2 + (A^2 + (A+B)^2) \sum_{k=0}^{\infty} B^{2k}, \\ &= I_2 + (A^2 + (A+B)^2)(I_2 - B^2)^{-1}. \end{aligned}$$

Consequently, we obtain that

$$H = \frac{1}{21} \begin{pmatrix} 576 & 0 \\ 0 & 28 \end{pmatrix}.$$

Therefore, the limiting matrix Λ given by (3.10) is

$$\Lambda = \frac{1}{21} \begin{pmatrix} 42 & 0 & 21 & 0 \\ 0 & 42 & 0 & 21 \\ 21 & 0 & 576 & 0 \\ 0 & 21 & 0 & 28 \end{pmatrix}.$$

It is not hard to see that $\det(\Lambda) = 89.7619$. One can observe in Figure 2 the almost sure convergence of the LS estimator $\widehat{\theta}_n$ to the four diagonal coordinates of θ. One can conclude that $\widehat{\theta}_n$ performs very well in the estimation of θ.

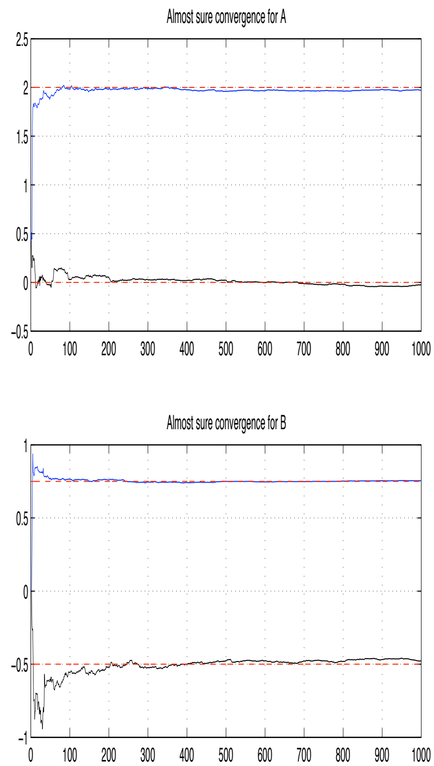


FIGURE 2. Almost Sure Convergence

Figure 3 shows the CLT for the four coordinates of

$$Z_N = \sqrt{N}\Lambda^{1/2}(\widehat{\theta}_N - \theta).$$

One can realize that each component of Z_N has $\mathcal{N}(0,1)$ distribution as expected.

6. CONCLUSION

Via the use of a persistently excited adaptive tracking control, we have shown that it was possible to get ride of the strong controllability assumption recently proposed by Bercu and Vazquez [8], [9]. We have established the almost sure convergence for the LS and WLS estimators in the multidimensional ARX framework. In addition, we have shown the residual optimality of the adaptive tracking. Moreover, both LS and the WLS estimators share the same CLT and LIL. We hope that similar analysis could be extended to the ARMAX framework.

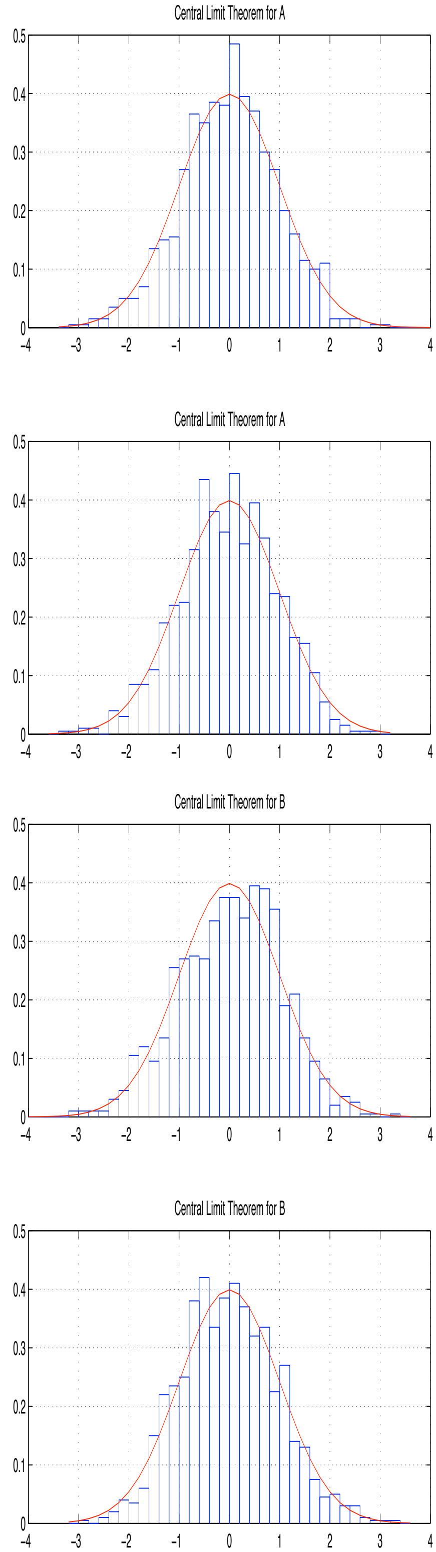


FIGURE 3. Central Limit Theorem

Appendix A

PROOF OF LEMMA 3.1

Let $\mathcal{A}$ and $\mathcal{B}$ be the infinite-dimensional diagonal square matrices given by

$$\mathcal{A} = \begin{pmatrix} \Gamma & 0 & \cdots & \cdots & \cdots \\ 0 & \Gamma & 0 & \cdots & \cdots \\ \cdots & 0 & \Gamma & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

$$\mathcal{B} = \begin{pmatrix} \Delta & 0 & \cdots & \cdots & \cdots \\ 0 & \Delta & 0 & \cdots & \cdots \\ \cdots & 0 & \Delta & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

Moreover, denote by $\mathcal{P}$ and $\mathcal{Q}$ the infinite-dimensional rectangular matrices with dq rows and an infinite number of columns, respectively given, if $p \geq q$, by

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

$$\mathcal{Q} = \begin{pmatrix} Q_p & Q_{p+1} & \cdots & Q_k & Q_{k+1} & \cdots \\ Q_{p-1} & Q_p & \cdots & Q_{k-1} & Q_k & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Q_{p-q+2} & Q_{p-q+3} & \cdots & Q_{k-q+2} & Q_{k-q+3} & \cdots \\ Q_{p-q+1} & Q_{p-q+2} & \cdots & Q_{k-q+1} & Q_{k-q+2} & \cdots \end{pmatrix},$$

while, if $p \leq q$, by

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

$$\mathcal{Q} = \begin{pmatrix} Q_p & Q_{p+1} & \cdots & \cdots & Q_k & Q_{k+1} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Q_1 & Q_2 & \cdots & \cdots & Q_{k-p+1} & Q_{k-p+2} & \cdots \\ Q_0 & Q_1 & Q_2 & \cdots & Q_{k-p} & Q_{k-p+1} & \cdots \\ 0 & Q_0 & Q_1 & \cdots & Q_{k-p-1} & Q_{k-p} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & Q_0 & Q_1 & \cdots \end{pmatrix}.$$

Furthermore, let $\Sigma = \Delta - \Delta(\Gamma + \Delta)^{-1}\Delta$ and denote by $\mathcal{C}$ the block diagonal matrix of order dp

$$\mathcal{C} = \begin{pmatrix} \Sigma & 0 & \cdots & 0 & 0 \\ 0 & \Sigma & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \Sigma & 0 \\ 0 & 0 & \cdots & 0 & \Sigma \end{pmatrix}.$$

One can observe that Σ is a positive definite matrix. Finally, if $p \geq q$, denote by V the matrix with dq rows and dp columns given by

$$V = \begin{pmatrix} D_0 & D_1 & D_2 & \cdots & \cdots & D_{p-2} & D_{p-1} \\ 0 & D_0 & D_1 & \cdots & \cdots & D_{p-3} & D_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & D_0 & D_1 & D_2 & \cdots & D_{p-q+1} \\ 0 & \cdots & \cdots & D_0 & D_1 & \cdots & D_{p-q} \end{pmatrix},$$

while, if $p \leq q$, the upper triangular square matrix of order dp given by

$$V = \begin{pmatrix} D_0 & D_1 & \cdots & D_{p-2} & D_{p-1} \\ 0 & D_0 & D_1 & \cdots & D_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & D_0 \end{pmatrix}.$$

On the one hand, if $p \geq q$, we can deduce from (3.11) after some straightforward, although rather lengthy, linear algebra calculations that

$$S = \mathcal{P}\mathcal{A}\mathcal{P}^t + \mathcal{Q}\mathcal{B}\mathcal{Q}^t + V\mathcal{C}V^t. \tag{A.1}$$

We shall focus our attention on the last term in (A.1). Since the matrix $\mathcal{C}$ is positive definite, it immediately follows that $V\mathcal{C}V^t$ is also positive definite. Consequently, the Schur complement S is invertible. On the other hand, if $p \leq q$, we can see from (3.11) that

$$S = \mathcal{P}\mathcal{A}\mathcal{P}^t + \mathcal{Q}\mathcal{B}\mathcal{Q}^t + R.$$

where R is the symmetric square matrix of order dq

$$R = \begin{pmatrix} V\mathcal{C}V^t & \mathcal{O} \\ \mathcal{O}^t & W \end{pmatrix}$$

where $\mathcal{O}$ stands for the zeros matrix of order $dp \times d(q-p)$ and W is the block diagonal matrix of order $d(q-p)$

$$W = \begin{pmatrix} \Delta & 0 & \cdots & 0 & 0 \\ 0 & \Delta & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \Delta & 0 \\ 0 & 0 & \cdots & 0 & \Delta \end{pmatrix}.$$

Taking into account the fact that $V\mathcal{C}V^t$ and W are both positive definite matrices, we obtain that R is also positive definite which implies that S is invertible. Finally, we infer from (3.10) that

$$\det(\Lambda) = \det(L)\det(S) = \det(\Gamma + \Delta)^p \det(S). \tag{A.2}$$

Consequently, we deduce from (A.2) that Λ is invertible and formula (3.12) can be found in [19] page 18, which completes the proof of Lemma 3.1. □

Appendix B

PROOF OF THEOREM 4.1

In order to prove Theorem 4.1, we shall make use of the same approach than Bercu [7] or Guo and Chen [17]. First of all, we recall that for all $n \geq 0$

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

It follows from (B.1) together with the strong law of large numbers for martingales (see e.g. Corollary 1.3.25 of [14]) that $n = \mathcal{O}(s_n)$ a.s. Moreover, by Theorem 1 of [7] or Lemma 1 of [17], we have

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

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 on the matrix polynomial B together with (B.2) that $\parallel \Phi_n \parallel^2 = \mathcal{O}(s_n^\beta)$ a.s. for all $2\alpha^{-1} < \beta < 1$. In addition, let $g_n = \Phi_n^t S_{n-1}^{-1} \Phi_n$ and $\delta_n = \mathrm{tr}(S_{n-1}^{-1} - S_n^{-1})$. It is well-known that

$$(1 - f_n)(1 + g_n) = 1$$

and (δ_n) tends to zero a.s. Consequently, as

$$1 + g_n \leq 2 + \delta_n \parallel \Phi_n \parallel^2,$$

we infer from from (B.2) that

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

Therefore, we obtain from (2.5), (B.1) and (B.3) that

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

Furthermore, as B is causal, we find from relation (1.1) that

$$U_n = B^{-1}(R)A(R)X_{n+1} - B^{-1}(R)\varepsilon_{n+1} \tag{B.5}$$

which implies by (B.4) that

$$\sum_{k=1}^{n} \parallel U_k \parallel^2 = o(s_n^\beta \log s_n) + \mathcal{O}(n) \quad \text{a.s.} \tag{B.6}$$

It remains to put together the two contributions (B.4) and (B.6) 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

$$\sum_{k=1}^{n} \| \pi_k \|^2 = o(n) \quad \text{a.s.} \tag{B.7}$$

Consequently, we obtain from (2.5), (B.1), (B.7) and the strong law of large numbers for martingales (see e.g. Theorem 4.3.16 of [14]) that

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

and, for all $1 \leq i \leq p-1$,

$$\sum_{k=0}^{n} X_k X_{k-i}^t = o(n) \quad \text{a.s.}$$

which implies that

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

where L is given by (3.9). Furthermore, it follows from (1.1), (B.1) and (B.5) that for all $n \geq 0$

$$\begin{aligned} U_n &= B^{-1}(R)A(R)X_{n+1} - B^{-1}(R)\varepsilon_{n+1}, \\ &= V_n + W_{n+1} + Z_{n+1}, \end{aligned}$$

where

$$\begin{aligned} V_n &= B^{-1}(R)A(R)(\pi_n + x_{n+1}), \\ W_{n+1} &= P(R)\varepsilon_{n+1}, \\ Z_{n+1} &= B^{-1}(R)A(R)\xi_{n+1}. \end{aligned}$$

Consequently, we deduce from the Cauchy-Schwarz inequality together with (2.5), (B.7), and the strong law of large numbers for martingales (see e.g. Theorem 4.3.16 of [14]) 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

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} U_k^q (U_k^q)^t = H \quad \text{a.s.} \tag{B.9}$$

where H is given by (3.8). Via the same lines, we also find that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k^p (U_{k-1}^q)^t = K^t \quad \text{a.s.} \tag{B.10}$$

Therefore, it follows from the conjunction of (B.8), (B.9) and (B.10) that

$$\lim_{n\to\infty} \frac{S_n}{n} = \Lambda \quad \text{a.s.} \tag{B.11}$$

where the limiting matrix Λ is given by (3.10). Thanks to Lemma 3.1, the matrix Λ is invertible. This is the key point for the rest of the proof. On the one hand, it follows from (B.11) 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

$$\sum_{k=1}^{n} \| \pi_k \|^2 = \mathcal{O}(\log n) \quad \text{a.s.} \tag{B.12}$$

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

$$\| C_n - \Delta_n \| = \mathcal{O}\left(\frac{1}{n} \sum_{k=1}^{n} \| \pi_{k-1} \|^2 \right) \quad \text{a.s.} \tag{B.13}$$

Consequently, we immediately obtain the tracking residual optimality (4.2) from (B.12) and (B.13). Furthermore, by a well-known result of Lai and Wei [21] on the LS estimator, we also have

$$\| \widehat{\theta}_{n+1} - \theta \|^2 = \mathcal{O}\left(\frac{\log \lambda_{max} S_n}{\lambda_{min} S_n} \right) \quad \text{a.s.} \tag{B.14}$$

Hence (4.3) clearly follows from (B.11) and (B.14), which completes the proof of Theorem 4.1. □

Appendix C

PROOF OF THEOREM 4.2

By Theorem 1 of [6], we have

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

where the coefficient $f_n(a) = a_n \Phi_n^t S_n^{-1}(a) \Phi_n$. Then, as the weighted sequence (a_n) is given by

$$a_n = \left(\frac{1}{\log s_n} \right)^{1+\gamma}$$

with $\gamma > 0$, we clearly have $a_n^{-1} = \mathcal{O}(s_n)$ a.s. Hence, it follows from (C.1) together with Kronecker's Lemma given e.g. by Lemma 1.3.14 of [14] that

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

Therefore, we obtain from (2.5), (B.1), (C.2) and the strong law of large numbers for martingales (see e.g. Theorem 4.3.16 of [14]) that

$$\sum_{k=1}^{n} \| X_{k+1} \|^2 = o(s_n) + \mathcal{O}(n) \quad \text{a.s.} \tag{C.3}$$

In addition, we also deduce from the causality assumption on the matrix polynomial B that

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

Consequently, we immediately infer from (C.3) and (C.4) that $s_n = o(s_n) + \mathcal{O}(n)$ so $s_n = \mathcal{O}(n)$ a.s. Hence, (C.2) implies that

$$\sum_{k=1}^{n} \| \pi_k \|^2 = o(n) \quad \text{a.s.} \tag{C.5}$$

Proceeding exactly as in Appendix A, we find from (C.5) that

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

Via an Abel transform, it ensures that

$$\lim_{n\to\infty} (\log n)^{1+\gamma} \frac{S_n(a)}{n} = \Lambda \quad \text{a.s.} \tag{C.6}$$

We obviously have from (C.6) that $f_n(a)$ tends to zero a.s. Consequently, we obtain from (C.1) and Kronecker's Lemma that

$$\sum_{k=1}^{n} \| \pi_k \|^2 = o((\log s_n)^{1+\gamma}) \quad \text{a.s.} \tag{C.7}$$

Then, (4.5) clearly follows from (B.13) and (C.7). Finally, by Theorem 1 of [6]

$$\| \widehat{\theta}_{n+1} - \theta \|^2 = \mathcal{O}\left(\frac{1}{\lambda_{min} S_n(a)} \right) \quad \text{a.s.} \tag{C.8}$$

Hence, we obtain (4.6) from (C.6) and (C.8), which completes the proof of Theorem 4.2. □

Appendix D

PROOF OF THEOREM 4.3

First of all, it follows from (1.2) and (2.1) that for all $n \geq 1$

$$\widehat{\theta}_n - \theta = S_{n-1}^{-1}(a) M_n(a) \tag{D.1}$$

where

$$M_n(a) = \widehat{\theta}_0 - \theta + \sum_{k=1}^{n} a_{k-1} \Phi_{k-1} \varepsilon_k^t. \tag{D.2}$$

We now make use of the CLT for multivariate martingales given e.g. by Lemma C.1 of [7], see also [14]. On the one hand, for the LS algorithm, we clearly deduce (4.8) from convergence (4.1) and decomposition (D.1). On the other hand, for the WLS algorithm, we also infer (4.8) from convergence (4.4) and (D.1). Next, we make use of the LIL for multivariate martingales given e.g. by Lemma C.2 of [7], see also [14], [25]. For the LS algorithm, since (ε_n) has finite conditional moment of order

$\alpha > 2$, we obtain from Chow's Lemma given e.g. by Corollary 2.8.5 of [25] that for all $2 < \beta < \alpha$

$$\sum_{k=1}^{n} \| \varepsilon_k \|^{\beta} = \mathcal{O}(n) \quad \text{a.s.} \tag{D.3}$$

The exogenous noise (ξ_n) shares the same regularity in norm than (ε_n) which means that for all $2 < \beta < \alpha$

$$\sum_{k=1}^{n} \| \xi_k \|^{\beta} = \mathcal{O}(n) \quad \text{a.s.} \tag{D.4}$$

Consequently, as the reference trajectory (x_n) satisfies (4.7), we deduce from (B.1) together with (B.12), (D.3) and (D.4) that for some $2 < \beta < \alpha$

$$\sum_{k=1}^{n} \| X_k \|^{\beta} = \mathcal{O}(n) \quad \text{a.s.} \tag{D.5}$$

Furthermore, it follows from (B.5) and (D.5) that

$$\sum_{k=1}^{n} \| U_k \|^{\beta} = \mathcal{O}(n) \quad \text{a.s.} \tag{D.6}$$

Hence, we clearly obtain from (D.5) and (D.6) that

$$\sum_{k=1}^{n} \| \Phi_k \|^{\beta} = \mathcal{O}(n) \quad \text{a.s.} \tag{D.7}$$

Therefore, as $\beta > 2$, (D.7) immediately implies that

$$\sum_{n=1}^{\infty} \left(\frac{\| \Phi_n \|}{\sqrt{n}} \right)^{\beta} < +\infty \quad \text{a.s.}$$

Finally, Lemma C.2 of [7] together with convergence (4.1) and (D.1) lead to (4.9). The proof for the WLS algorithm is left to the reader because it follows essentially the same arguments than the proof for the LS algorithm. It is only necessary to add the weighted sequence (a_n) and to make use of convergence (4.4). □

References

[1] A. A. Abouelsoud, M. F. Hassan and M. A. Sultan, An adaptive observer for robots with persistent excitation, International Journal of Systems Science, Vol. 27, pp. 809-820, 1996.

[2] E. Aggelogiannaki, P. Doganis and H. Sarimveis, An adaptive model predictive control configuration for production-inventory systems, Internationa Journal of Production Economics, Vol. 114, pp. 165-178, 2008,

[3] B. D. O. Anderson, Exponential convergence and persistent excitation, 21th IEEE Conference on Decision and Control, 1982.

[4] K. J. Aström and B. Wittenmark, Adaptive Control, 2nd edition, Addison-Wesley, New York, 1995.

[5] B. Bercu and M. Duflo, Moindres carrés pondérés et poursuite, Annals de l'Institut Henri Poincaré, Vol. 28, pp. 403-430, 1992.

[6] B. Bercu, Weighted estimation and tracking for ARMAX models, SIAM J. Control Optim., Vol. 33, pp. 89-106, 1995.
[7] B. Bercu, Central limit theorem and law of iterated logarithm for least squares algorithms in adaptive tracking, SIAM J. Control Optim., Vol. 36, pp. 910-928, 1998.
[8] B. Bercu and V. Vázquez, Further results for ARX models in adaptive tracking, 47th IEEE Conference on Decision and Control, 2008.
[9] B. Bercu and V. Vázquez, A new concept of strong controllability via the Schur complement in adapative tracking, submitted to Automatica, 2008.
[10] P. E. Caines, Linear Stochastic Systems, John Wiley, New York, 1988.
[11] H. F. Chen and L. Guo, Identification and Stochastic Adaptive Control, Birkhäuser, Boston, 1991.
[12] C. Chengyu, N. Hovakimyan and W. Jiang, Intelligent excitation for adaptive control with unknown parameters in reference input, IEEE Trans. Automat. Control, Vol. 52, pp. 1525-1532, 2007.
[13] L. Corrado and S. Holly, The linearisation and optimal control of large non-linear rational expectations models by persistent excitation source, Computational Economics, Vol. 28, pp. 139-153, 2006.
[14] M. Duflo, Random Iterative Models, Springer Verlag, Berlin, 1997.
[15] H. Fang, R. X. Fan, B. Thuilot and P. Marinet, Lateral control for autonomus vehicles with sliding angle reconstruction, Chinese Control and Decision Conference, 2008.
[16] A. Farrell, Persistent excitation conditions in passive learning control, Automatica, Vol. 33, pp. 699-703, 1997.
[17] L. Guo and H. F. Chen, The Aström Wittenmark self-tuning regulator revisited and ELS-based adaptive trackers, IEEE Trans. Automat. Control, Vol. 36, pp. 802-812, 1991.
[18] L. Guo, Self convergence of weighted least squares with applications to stochastic adaptive control, IEEE Trans. Automat. Control, Vol. 41, pp. 79-89, 1996.
[19] R. A. Horn and C. R. Johnson, Matrix Analysis, *Cambridge University Press*, New York, 1990.
[20] J. T. Huang, Persistent excitation in a shunt DC motor under adaptive control, Asian Journal of Control, Vol. 9, pp. 37-44., 2008.
[21] T. L. Lai and C. Z. Wei, Extended least squares and their applications to adaptive control and prediction in linear systems, IEEE Trans. Automat. Control, Vol. 31, pp. 898-906, 1986.
[22] D. Levanoy, On persistent excitation conditions for the filtering of convergent semimartingales, 43th IEEE Conference on Decicion and Control, 2004.
[23] D. Levanoy and P. E. Caines, On persistent excitation for linear systems with stochastic coefficients, SIAM J. Control Optim., Vol. 40, pp. 882-897, 2001.
[24] J. B. Moore, Persistency of excitation in extended least squares, IEEE Trans. Automat. Control, Vol. 28, pp. 60-68, 1983.
[25] W. F. Stout, Almost sure convergence, Academic Press, New York, 1974.
[26] C. Z. Wei, Adaptive prediction by least squares predictors in stochastic regression models with applications to time series, Annals of Statistics, Vol. 15, pp. 1667-1682, 1987.

E-mail address: `Bernard.Bercu@math.u-bordeaux1.fr`
E-mail address: `Victor.Vazquez@siu.buap.mx`

Université Bordeaux 1, Institut de Mathématiques de Bordeaux, UMR 5251, 351 cours de la libération, 33405 Talence cedex, France.

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