Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Further results for ARX models in adaptive tracking

Bernard Bercu and Victor Vazquez

Abstract— **We introduce a new concept of strong controllability for ARX models in adaptive tracking. This new notion is related to the Schur complement of a suitable limiting matrix. It allows us to extend the previous convergence results associated with both least squares and weighted least squares algorithms. In particular, we show the almost sure convergence as well as the central limit theorem for this two algorithms.**

I. INTRODUCTION

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

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

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

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

One can obviously see that (1) can be rewritten as

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

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

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

We shall often make use of the regression matrix

$$S_n = \sum_{k=0}^{n} \Phi_k \Phi_k^t.$$

We assume that the driven noise (ε_n) is a martingale difference sequence adapted to the filtration $\mathbb{F} = (\mathcal{F}_n)$ associated with $\text{ARX}_d(p,q)$ process given by (1). In addition, 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 covariance matrix.

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

B. Bercu is with Université Bordeaux 1, Institut de Mathématiques de Bordeaux, UMR 5251, 351 cours de la libération, 33405 Talence cedex, France Bernard.Bercu@math.u-bordeaux1.fr

V. Vazquez is with Universidad Autonoma de Puebla, Facultad de Ciencias Fisico Matematicas, Avenida San Claudio y Rio Verde, 72570 Puebla, Mexico Victor.Vazquez@siu.buap.mx

A wide literature is available concerning the estimation of θ and on the optimality of the tracking control, see e.g. [1], [3], [4], [7], [8], [10], [12], [15], [16]. In the particular case $q = 0$, it was shown in [4] 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

$$L = \text{diag}\,(\Gamma, \cdots, \Gamma)\,.$$

Our purpose is to extend the previous results in [4], [5], [10], [11], [14] via the introduction of a new concept of strong controllability. Under the classical causality assumption on the polynomial B, this new notion allows us to prove that

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

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

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

Moreover, the matrices H and K are explicitly calculated. It is well-known [13] that $\det(\Lambda) = \det(L)\det(S)$ where the symmetric matrix $S = H - KL^{-1}K^t$ 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. We shall propose a suitable assumption under which S is positive definite. This assumption is really easy to understand and it can't be avoided. Then, we shall carry out a sharp analysis of the almost sure convergence for both least squares (LS) and weighted least squares (WLS) estimators of the unknown parameter θ. We also provide a central limit theorem and a law of iterated logarithm for these two estimators.

II. STRONG CONTROLLABILITY

In all the sequel, we shall make use of the well-known causality assumption on B.

Definition 1: We shall say that the matrix polynomial B is causal if for all $z \in \mathbb{C}$ with $|z| \leq 1$

$$(\text{A}_1) \qquad \det(B(z)) \neq 0.$$

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. 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 2: An $\text{ARX}_d(p, q)$ process is said to be strongly controllable if B is causal and Π is invertible,

$$(\text{A}_2) \qquad \det(\Pi) \neq 0.$$

Remark 1: 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 $\det(A_2 - B_1A_1) \neq 0$, if $p = 1$, $q = 2$ to $\det(A_1) \neq 0$, while if $p = q = 2$ to

$$\det \begin{pmatrix} A_1 & A_2 - B_1A_1 \\ A_2 - B_1A_1 & -B_1A_2 + (B_1^2 - B_2)A_1 \end{pmatrix} \neq 0.$$

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

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

Let H be the symmetric square matrix of order dq

$$H = \begin{pmatrix} H_1 & H_2 & \cdots & H_{q-1} & H_q \\ H_2^t & H_1 & H_2 & \cdots & H_{q-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ H_{q-1}^t & \cdots & H_2^t & H_1 & H_2 \\ H_q^t & H_{q-1}^t & \cdots & H_2^t & H_1 \end{pmatrix}. \tag{5}$$

For all $1 \leq i \leq p$, 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

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

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

$$L = \begin{pmatrix} \Gamma & 0 & \cdots & 0 & 0 \\ 0 & \Gamma & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \Gamma & 0 \\ 0 & 0 & \cdots & 0 & \Gamma \end{pmatrix} \tag{6}$$

and denote by Λ the symmetric square matrix of order δ

$$\Lambda = \begin{pmatrix} L & K^t \\ K & H \end{pmatrix}. \tag{7}$$

Lemma 1: Let S be the Schur complement of L in Λ

$$S = H - KL^{-1}K^t. \tag{8}$$

If (A_1) and (A_2) hold, S and Λ are invertible and Λ^{-1} is given by

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

III. ESTIMATION AND ADAPTIVE CONTROL

First of all, we focus our attention on the estimation of the parameter θ. We shall make use of the WLS estimator of Bercu and Duflo [2], [3], [9] given, for all $n \geq 0$, by

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n + a_nS_n^{-1}(a)\Phi_n\left(X_{n+1} - U_n - \widehat{\theta}_n^t\Phi_n\right)^t \tag{9}$$

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

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

we obtain the WLS estimator of Bercu and Duflo [2], [3]. Next, we are concern with the choice of the 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 [1] given, for all $n \geq 0$, by

$$U_n = x_{n+1} - \widehat{\theta}_n^t\Phi_n. \tag{10}$$

By substituting (10) into (2), we obtain the closed-loop system

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1} \tag{11}$$

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

$$\sum_{k=1}^{n} \| x_k \|^2 = o(n) \qquad \text{a.s.} \tag{12}$$

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

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

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

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

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

IV. MAIN RESULTS

We shall now present the results recently obtained by Bercu and Vazquez in [6]. The first one deals with the almost sure properties of the LS estimator.

Theorem 1: Assume that the $\text{ARX}_d(p, q)$ process is strongly controllable and that (ε_n) has finite conditional moment of order > 2. Then, we have

$$\lim_{n \to \infty} \frac{S_n}{n} = \Lambda \qquad \text{a.s.} \tag{13}$$

where the limiting matrix Λ is given by (7). In addition, the tracking is optimal

$$\| C_n - \Gamma_n \| = \mathcal{O}\left(\frac{\log n}{n}\right) \qquad \text{a.s.} \tag{14}$$

We can sharpen (14) by

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

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

$$\| \widehat{\theta}_n - \theta \|^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \qquad \text{a.s.} \tag{15}$$

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

Theorem 2: Assume that the $\text{ARX}_d(p, q)$ process 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, we have

$$\lim_{n \to \infty} (\log n)^{1+\gamma} \frac{S_n(a)}{n} = \Lambda \qquad \text{a.s.} \tag{16}$$

where the limiting matrix Λ is given by (7). In addition, the tracking is optimal

$$\| C_n - \Gamma_n \| = o\left(\frac{(\log n)^{1+\gamma}}{n}\right) \qquad \text{a.s.} \tag{17}$$

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

$$\| \widehat{\theta}_n - \theta \|^2 = \mathcal{O}\left(\frac{(\log n)^{1+\gamma}}{n}\right) \qquad \text{a.s.} \tag{18}$$

Remark 2: If the matrix polynomial B is causal but the $\text{ARX}_d(p, q)$ process is not strongly controllable, we only obtain for both LS and WLS estimators that

$$\lim_{n \to \infty} \Lambda(\widehat{\theta}_n - \theta) = 0 \qquad \text{a.s.}$$

Theorem 3: Assume that the $\text{ARX}_d(p, q)$ process 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 $2 < \beta < \alpha$

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

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

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Lambda^{-1} \otimes \Gamma) \tag{19}$$

where 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

$$\begin{aligned} &\limsup_{n \to \infty} \left(\frac{n}{2 \log \log n}\right)^{1/2} v^t(\widehat{\theta}_n - \theta)u \\ = \; & -\liminf_{n \to \infty} \left(\frac{n}{2 \log \log n}\right)^{1/2} v^t(\widehat{\theta}_n - \theta)u \\ = \; & \left(v^t \Lambda^{-1} v\right)^{1/2} \left(u^t \Gamma u\right)^{1/2} \qquad \text{a.s.} \end{aligned} \tag{20}$$

In particular, it implies that

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

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

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

V. SIMULATIONS

The goal of this section is illustrate our asymptotic results by simulations. We shall also show that our new concept of strong controllability can't be avoided. In order to keep this section brief, we shall only focus our attention on a strongly controllable $ARX_d(p, q)$ model in dimension $d = 2$ with $p = 1$ and $q = 1$. Our numerical simulations are based on $M = 500$ realizations of sample size $N = 1000$. For the sake of simplicity, the reference trajectory (x_n) is chosen to be identically zero and (ε_n) is a Gaussian white noise $\mathcal{N}(0, 1)$. Consider the $ARX_2(1, 1)$ model

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

where

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

First of all, it is easy to see that this $ARX_2(1,1)$ process is strongly controllable because $\det(A) = 2$. For all $k \geq 1$, we clearly have $P_k = -(-B)^{k-1}A$. Since the matrices A and B are diagonal, they commute which leads to

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

Consequently,

$$H = \frac{4}{21}\begin{pmatrix} 48 & 0 \\ 0 & 7 \end{pmatrix}.$$

Therefore, the limiting matrix Λ given by (7) is

$$\Lambda = \frac{1}{21}\begin{pmatrix} 21 & 0 & 0 & 0 \\ 0 & 21 & 0 & 0 \\ 0 & 0 & 192 & 0 \\ 0 & 0 & 0 & 28 \end{pmatrix}.$$

Figure 1 shows the almost sure convergence of the LS estimator $\widehat{\theta}_n$ to the four coordinates of θ which are different from zero. One can observe that $\widehat{\theta}_n$ performs very well in the estimation of θ.

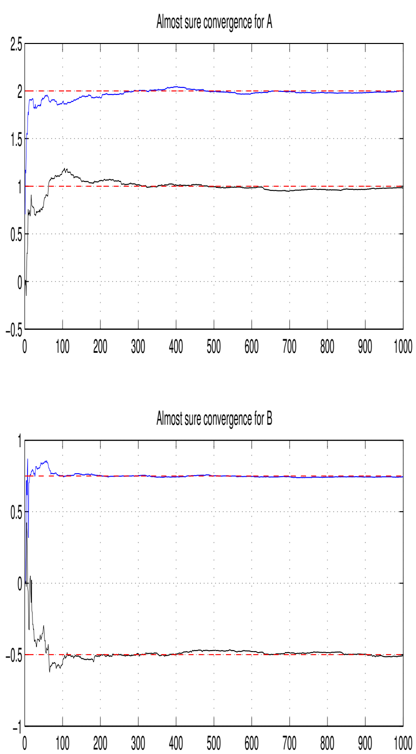

Fig. 1. Almost Sure Convergence

Figure 2 shows the central limit theorem 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.

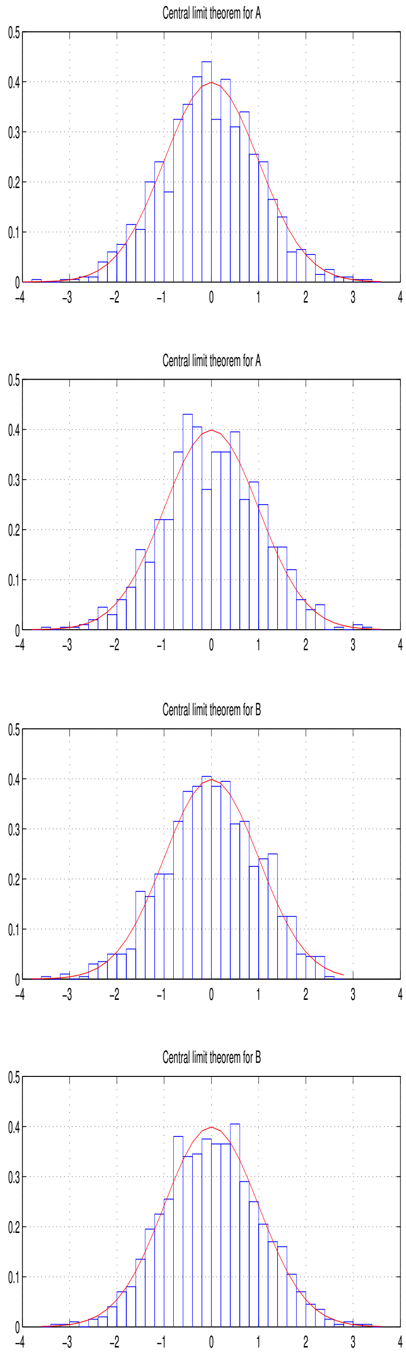

Fig. 2. Central Limit Theorem

Our goal is now to show that our strong controllability assumption can't be avoided. Consider the same $ARX_2(1,1)$ model with

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

We have only change the second diagonal term of the matrix A. As $\det(A) = 0$, this $ARX_2(1,1)$ process is not strongly controllable. In addition,

$$H = A^2(I_2 - B^2)^{-1} = \frac{64}{7}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

which leads to

$$\Lambda = \frac{1}{21}\begin{pmatrix} 21 & 0 & 0 & 0 \\ 0 & 21 & 0 & 0 \\ 0 & 0 & 192 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Consequently, only the matrix A and the first diagonal term of the matrix B are properly estimated as one can see in Figures 3 and 4.

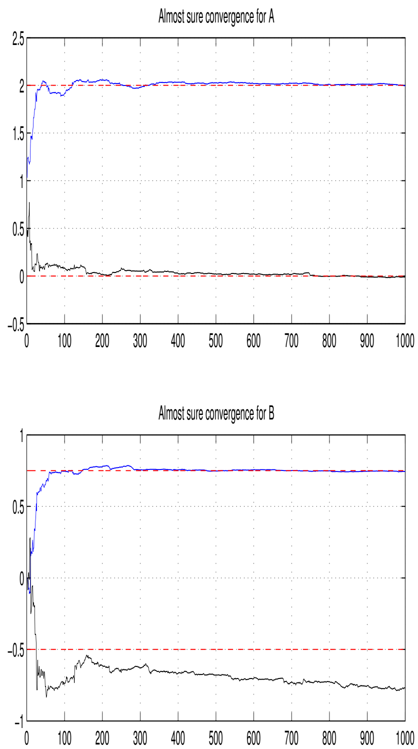

Fig. 3. Almost Sure Convergence

VI. CONCLUSION

Via our new concept of strong controllability, we have extended the analysis of the almost sure convergence for both LS and WLS estimators of the unknown parameter of $\text{ARX}_d(p,q)$ models. It enables us to provide a positive answer to a conjecture in [4] by establishing a CLT and a LIL for these two estimators. 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 is well-known that it is really difficult to investigate the almost sure asymptotic properties for both ELS and WLS estimators in the ARMAX framework [3], [10]. 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.

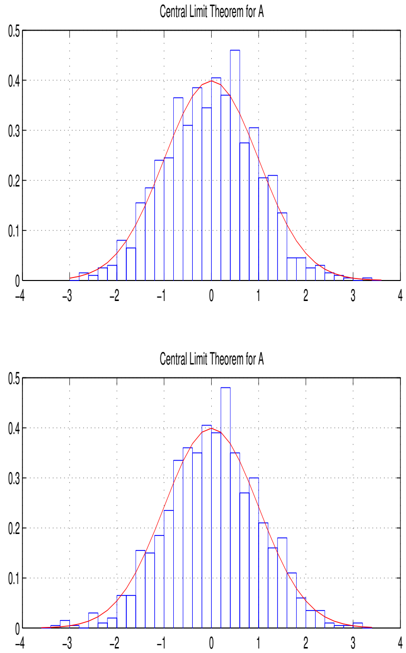

Fig. 4. Central Limit Theorem

REFERENCES

[1] K. J. Aström and B. Wittenmark, Adaptive Control, 2nd edition, Addison-Wesley, New York, 1995.
[2] B. Bercu and M. Duflo, Moindres carrés pondérés et poursuite, *Annals de l'Institut Henri Poincaré*, 28 (1992), pp. 403-430.
[3] B. Bercu, Weighted estimation and tracking for ARMAX models, *SIAM J. Control Optim.*, 33 (1995), pp. 89-106.
[4] B. Bercu, Central limit theorem and law of iterated logarithm for least squares algorithms in adaptive tracking, *SIAM J. Control Optim.*, 36 (1998), pp. 910-928.
[5] B. Bercu and B. Portier, Adaptive control of parametric nonlinear autoregressive models via a new martingale approach, *IEEE Trans. Automat. Control*, 47, (2002), pp. 1524-1528.
[6] B. Bercu and V. Vazquez, A new concept of strong controllability via the Schur complement in adaptive tracking, Submitted for publication, 2008.
[7] P. E. Caines, Linear Stochastic Systems, *John Wiley*, New York, 1988.
[8] H. F. Chen and L. Guo, Identification and Stochastic Adaptive Control, *Birkhäuser*, Boston, 1991.
[9] M. Duflo, Random Iterative Models, *Springer Verlag*, Berlin, 1997.
[10] L. Guo and H. F. Chen, The Aström Wittenmark self-tuning regulator revisited and ELS-based adaptive trackers, *IEEE Trans. Automat. Control*, 36 (1991), pp. 802-812.
[11] L. Guo, Further results on least squares based adaptive minimum variance control, *SIAM J. Control Optim.*, 32 (1994), pp. 187-212.
[12] L. Guo, Self convergence of weighted least squares with applications to stochastic adaptive control, *IEEE Trans. Automat. Control*, 41 (1996), pp. 79-89.
[13] R. A. Horn and C. R. Johnson, Matrix Analysis, *Cambridge University Press*, New York, 1990.
[14] A. Jankumas, Optimal adaptive control for estimation of parameters of ARX models, IEEE Trans. Automat. Control, 45 (2000), pp. 964-968.
[15] 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*, 31 (1986), pp. 898-906.
[16] C. Z. Wei, Adaptive prediction by least squares predictors in stochastic regression models with applications to time series, *Annals of Statistics*, 15 (1987), pp. 1667-1682.