SIAM J. CONTROL OPTIM.
Vol. 36, No. 3, pp. 910–928, May 1998

CENTRAL LIMIT THEOREM AND LAW OF ITERATED LOGARITHM FOR LEAST SQUARES ALGORITHMS IN ADAPTIVE TRACKING*

BERNARD BERCU†

Abstract. In autoregressive adaptive tracking, we prove that the least squares and the weighted least squares algorithms possess the same asymptotic properties, sharing the same central limit theorem and the same law of iterated logarithm. We also obtain the same asymptotic behavior and show the limitations of these results in the autoregressive with moving average framework.

Key words. linear regression, least squares, central limit theorem, law of iterated logarithm

AMS subject classifications. 62J05, 93E24, 60F05, 60F15

PII. S0363012995294183

Notations. For any square matrix A, $\mathrm{tr}(A)$ is the trace of A and $\det(A)$ denotes the determinant of A. In addition, $\lambda_{min}A$ and $\lambda_{max}A$ are the minimum and the maximum eigenvalues of A, respectively. Finally, for any vectorial sequence $X=(X_n)$ and any integer $p\geq 1$, $X_n^p=(X_n^t,\dots,X_{n-p+1}^t)$.

1. Introduction.

Let $(\Omega,\mathcal{A},P)$ be a probability space endowed with a filtration $\mathbf{F}=(\mathcal{F}_n)_{n\geq 0}$, where $\mathcal{F}_n$ is the σ-algebra of the events occurring up to time n. Consider the controlled autoregressive with moving average (ARMA) model of order (p,r) given, for all $n\geq 0$, by

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

where X_n, U_n, and ε_n are, respectively, the d-dimensional system output, input, and driven noise and $\Psi_n=(X_n^p,\varepsilon_n^r)^t$. In order to estimate the unknown $\delta\times d$ matrix θ with $\delta=d\,(p+r)$, we use the weighted least squares (WLS) algorithm that satisfies, for all $n\geq 0$,

$$\hat{\theta}_{n+1}=\hat{\theta}_n+a_nS_n^{-1}(a)\Phi_n\left(X_{n+1}-U_n-\hat{\theta}_n^t\Phi_n\right)^t, \tag{2}$$

$$S_n(a)=\sum_{k=0}^{n}a_k\Phi_k\Phi_k^t+S, \tag{3}$$

$$\hat{\varepsilon}_{n+1}=X_{n+1}-U_n-\hat{\theta}_{n+1}^t\Phi_n,\qquad \Phi_n=(X_n^p,\hat{\varepsilon}_n^r)^t, \tag{4}$$

where the initial value $\hat{\theta}_0$ is arbitrarily chosen and S is a deterministic, symmetric, and positive definite matrix. We set

$$S_n=\sum_{k=0}^{n}\Phi_k\Phi_k^t+S,\qquad s_n=\mathrm{tr}(S_n). \tag{5}$$

The choice of the weighted sequence $a=(a_n)$ is crucial. If

$$a_n=1 \tag{6}$$

*Received by the editors November 3, 1995; accepted for publication (in revised form) March 5, 1997.
http://www.siam.org/journals/sicon/36-3/29418.html

†Laboratoire de Statistiques, Batiment 425 Mathématiques, Université de Paris-Sud, 91 405 Orsay Cedex, France (bernard.bercu@math.u-psud.fr).

we find again the extended least squares (ELS) algorithm. Otherwise, if

$$a_n = \left(\frac{1}{\log s_n}\right)^{1+\gamma} \tag{7}$$

with $\gamma > 0$, we obtain the WLS algorithm proposed by Duflo and Bercu [4], [5]. For these two algorithms, a wide literature concerning the strong consistency and the optimality in adaptive tracking is available (see, e.g., [4], [5], [6], [8], [9], [10], [11], [13], [15], [20], [26]). In these papers, it is always necessary to establish an excitation property for the regressive sequence $\Phi = (\Phi_n)$. To be more precise, for the strong consistency, one has to prove that

$$\lambda_{min} S_n \longrightarrow +\infty, \qquad \log \lambda_{max} S_n = o(\lambda_{min} S_n) \qquad \text{almost surely (a.s.)} \tag{8}$$

and, for the optimality, that $s_n = O(n)$ a.s. In fact, one always has to show that $n = O(\lambda_{min} S_n)$ and $\lambda_{max} S_n = O(n)$ a.s. In autoregressive (AR) adaptive tracking with $r = 0$, we improve the previous results showing the almost sure convergence

$$\frac{S_n}{n} \longrightarrow L_p, \tag{9}$$

$L_p = \operatorname{diag}(\Gamma, \dots, \Gamma)$, where Γ is the conditional covariance matrix of the driven noise. This convergence allows us to obtain a central limit theorem (CLT) and a law of iterated logarithm (LIL) for both LS and WLS algorithms. Since the WLS introduces less weight to the more recent information than the LS, one may expect that WLS may be inferior to LS in asymptotic properties. However, we prove that in the AR framework, the LS and WLS algorithms possess the same asymptotic properties, sharing the same CLT,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, L_p^{-1} \otimes \Gamma), \tag{10}$$

and the same LIL. In addition, we also obtain that the ELS and WLS algorithms have the same asymptotic behavior in the ARMA framework. Finally, there is no loss in asymptotic efficiency by using WLS, which has many other advantages [4], [5], [17], [22] over LS or ELS in adaptive control theory.

The paper is organized as follows. In section 2, we establish in the AR framework the same CLT and LIL for LS and WLS algorithms. In AR adaptive tracking, the limit matrix given in (9) is positive definite, while this is no longer true in ARMA adaptive tracking. In the ARMA framework, in order to obtain strong consistency results, it is necessary to introduce an excitation on the adaptive tracking control. In section 3, we prove that the effect of this excitation is to make the limit matrix in (9) positive definite. Therefore, for the ARMA models of orders one, we establish the same CLT and LIL for ELS and WLS algorithms. In section 4, we show by simulations the limitation of these last results if the ARMA orders are greater than one. A short conclusion is given in section 5. All technical proofs are collected in the Appendices.

2. AR adaptive tracking. We first consider the AR framework with $r = 0$. Let $x = (x_n)$ be a predictable reference trajectory, to track, step by step, by the observation $X = (X_n)$. To this end, we use the adaptive tracking control proposed by Aström and Wittenmark [1] given, for all $n \geq 0$, by

$$U_n = x_{n+1} - \hat{\theta}_n^t \Phi_n. \tag{11}$$

Relation (1) can then be rewritten as

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1}, \tag{12}$$

where $\pi_n = (\theta - \hat{\theta}_n)^t \Phi_n$. Throughout the following, we assume that the reference trajectory x satisfies

$$\sum_{k=1}^{n} \| x_k \|^2 = o(n) \qquad \text{a.s.} \tag{13}$$

We also assume that the driven noise $\varepsilon = (\varepsilon_n)$ is a martingale difference sequence with

$$E\,[\ \varepsilon_{n+1}\varepsilon_{n+1}^t \mid \mathcal{F}_n\ ] = \Gamma, \tag{14}$$

where Γ is a positive definite deterministic covariance matrix. Finally, we assume that ε satisfies the strong law of large numbers; i.e., if

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

Γ_n converges a.s. to Γ. This is the case if, for example, ε has finite conditional moment of order > 2 or ε is a white noise, i.e., if ε is independant and identically distributed with mean 0 and covariance matrix Γ. Let (C_n) be the average cost matrix sequence defined by

$$C_n = \frac{1}{n}\sum_{k=1}^{n} (X_k - x_k)(X_k - x_k)^t. \tag{16}$$

The adaptive tracking is said to be optimal if C_n converges a.s. to Γ. Let L_p be the block diagonal square matrix of order $\delta_p = dp$,

$$L_p = \operatorname{diag}(\Gamma, \ldots, \Gamma). \tag{17}$$

THEOREM 2.1. *Consider the AR framework with $r=0$. Assume that ε has finite conditional moment of order > 2. Then, for the LS algorithm, we have*

$$\frac{S_n}{n} \longrightarrow L_p \qquad \textit{a.s.} \tag{18}$$

In addition, the tracking is optimal:

$$\| C_n - \Gamma_n \| = O\left(\frac{\log n}{n}\right) \qquad \textit{a.s.} \tag{19}$$

We can be more precise in (19) *as follows*

$$\frac{1}{\log n}\sum_{k=1}^{n} (X_k - x_k - \varepsilon_k)(X_k - x_k - \varepsilon_k)^t \longrightarrow \delta_p \Gamma \qquad \textit{a.s.} \tag{20}$$

Finally, $\hat{\theta}_n$ is a strongly consistent estimator of θ:

$$\| \hat{\theta}_n - \theta \|^2 = O\left(\frac{\log n}{n}\right) \qquad \textit{a.s.} \tag{21}$$

Proof. The proof is given in Appendix A. □

THEOREM 2.2. *Consider the AR framework with* $r=0$. *Assume that either* ε *is a white noise or* ε *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*

$$(\log n)^{1+\gamma} \frac{S_n(a)}{n} \longrightarrow L_p \qquad a.s. \tag{22}$$

In addition, the tracking is optimal:

$$\| C_n - \Gamma_n \| = o\left(\frac{(\log n)^{1+\gamma}}{n}\right) \qquad a.s. \tag{23}$$

Finally, $\hat{\theta}_n$ *is a strongly consistent estimator of* θ*:*

$$\| \hat{\theta}_n - \theta \|^2 = O\left(\frac{(\log n)^{1+\gamma}}{n}\right) \qquad a.s. \tag{24}$$

Proof. The proof is given in Appendix B. □

Remark. Theorem 2.2 is similar to Theorem 2.1. On the one hand, it is not necessary to require a conditional moment of order > 2 for the noise ε. On the other hand, we note a loss in $(\log n)^\gamma$ in the rates of convergence.

THEOREM 2.3. *Consider the AR framework with* $r=0$. *Assume that* ε *has finite conditional moment of order* $\alpha > 2$ *and that* x *has the same regularity in norm as* ε*; i.e., for all* $2 < \beta < \alpha$,

$$\sum_{k=1}^{n} \| x_k \|^\beta = O(n) \qquad a.s. \tag{25}$$

Then, the LS and the WLS algorithms share the same CLT,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, L_p^{-1} \otimes \Gamma), \tag{26}$$

with $L_p^{-1} \otimes \Gamma = \mathrm{diag}(\Gamma^{-1} \otimes \Gamma, \ldots, \Gamma^{-1} \otimes \Gamma)$. *In addition, for any vectors* $u \in \mathbb{R}^d$ *and* $v \in \mathbb{R}^{dp}$, *they also share the same LIL,*

$$\begin{aligned} \limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} v^t(\hat{\theta}_n - \theta)u &= - \liminf_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} v^t(\hat{\theta}_n - \theta)u \\ &= (v^t L_p^{-1} v)^{1/2} (u^t \Gamma u)^{1/2} \qquad a.s. \end{aligned} \tag{27}$$

In particular,

$$\left(\frac{\lambda_{min}\Gamma}{\lambda_{max}\Gamma}\right) \le \limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right) \| \hat{\theta}_n - \theta \|^2 \le \left(\frac{\lambda_{max}\Gamma}{\lambda_{min}\Gamma}\right) \qquad a.s. \tag{28}$$

Proof. The proof is given in Appendix C. □

Remark. First, one can realize that (28) improves Theorem 3.1 of Guo [16] for the LS algorithm. Next, we can also prove that Theorem 2.3 holds for the cost matrix sequence (C_n). To be more precise, assume that ε satisfies the following CLT:

$$\sqrt{n}(\Gamma_n - \Gamma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Lambda), \tag{29}$$

where Λ is an appropriate deterministic covariance matrix. Then, by (19) or (23), it immediately follows that

$$\sqrt{n}(C_n - \Gamma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Lambda) \tag{30}$$

for both LS and WLS algorithms. Moreover, via (19) or (23), we can also obtain an LIL for the sequence (C_n). Finally, in AR adaptive tracking, we can avoid the restrictive assumption (13) on the reference trajectory x. Using the same approach developed in Appendix A, we only need to assume that x satisfies the strong law of large numbers

$$\frac{1}{n}\sum_{k=1}^{n} x_k x_k^t \longrightarrow \Delta \qquad \text{a.s.,} \tag{31}$$

where Δ is a deterministic covariance matrix. Then, we just have to replace Γ by $\Gamma + \Delta$ in relation (17).

3. ARMA adaptive tracking. We now consider the ARMA framework. We always use the adaptive tracking control given, for all $n \geq 0$, by

$$U_n = x_{n+1} - \hat{\theta}_n^t \Phi_n, \tag{32}$$

where the reference trajectory x satisfies (13). Relation (1) can be rewritten as

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1}, \tag{33}$$

where $\pi_n = \theta^t \Psi_n - \hat{\theta}_n^t \Phi_n$. Let L_r be the block diagonal square matrix of order $\delta_r = dr$,

$$L_r = \operatorname{diag}(\Gamma, \dots, \Gamma). \tag{34}$$

For $s = \inf\{p, r\}$, let K be the rectangular matrix of dimension $\delta_p \times \delta_r$ with all coefficients equal to 0 except its left superior block, which is the block diagonal square matrix of order ds, L_s. Finally, let L be the square matrix of order $\delta = \delta_p + \delta_r$:

$$L = \begin{pmatrix} L_p & K \\ K^t & L_r \end{pmatrix}. \tag{35}$$

Throughout the following, we make use of the traditional assumption of passivity: if C is the matrix polynomial associated with the moving average (MA) part of (1) and I_d is the identity matrix of order d,

$$C^{-1} - \frac{1}{2} I_d \tag{P}$$

is strictly positive real. In the ARMA framework, many results concerning the tracking optimality are available (see, e.g., [2], [5], [11], [15], [16]). It is also well known that we can't directly obtain strong consistency for both ELS or WLS algorithms (see, e.g., [5], [7], [11], [13], [17]). Nevertheless, we prove in the following lemma that convergences (18) and (22) still hold here, replacing L_p by L. This can lead to interesting asymptotic properties.

LEMMA 3.1. *For the ARMA model, assume that* (P) *is satisfied. Then, for the ELS algorithm, if ε has finite conditional moment of order > 2, we have*

$$\frac{S_n}{n} \longrightarrow L \qquad \textit{a.s.} \tag{36}$$

In addition, for the WLS algorithm with $a_n^{-1} = (\log s_n)^{1+\gamma}$, *where* $\gamma > 0$, *if* ε *is a white noise or if* ε *has finite conditional moment of order* > 2, *we have*

$$(\log n)^{1+\gamma} \frac{S_n(a)}{n} \longrightarrow L \quad a.s. \tag{37}$$

Proof. The proof is given in Appendix D. □

THEOREM 3.2. *For the ARMA model, assume that* (P) *is satisfied and consider the regulation problem with* $x = 0$. *Assume that* ε *has finite conditional moment of order* > 2. *For a positive, nonincreasing, and deterministic sequence* (α_n) *such that* $\alpha_n = O(n)$, *assume that* $\| \varepsilon_n \|^2 = O(\alpha_n)$. *Let* (λ_n) *be a positive, nonincreasing, and deterministic sequence such that* $n^c \alpha_n = O(\lambda_n)$, $n^{1+c}\alpha_n = O(\lambda_n^2)$ *with* $0 < c < 1$ *for the ELS algorithm, and* $c = 0$ *for the WLS algorithm. Finally, assume that*

$$\| \Gamma_n - \Gamma \| = o\left(\frac{\lambda_n}{n}\right) \quad a.s. \tag{38}$$

Then, for both ELS and WLS algorithms, the tracking is optimal:

$$\| C_n - \Gamma \| = o\left(\frac{\lambda_n}{n}\right) \quad a.s. \tag{39}$$

Moreover, we also have

$$\left\| \frac{S_n}{n} - L \right\| = o\left(\frac{\lambda_n}{n}\right) \quad a.s. \tag{40}$$

Finally, on the one hand, it results for the ELS estimator that

$$\| L^{1/2}(\hat{\theta}_n - \theta) \|^2 = o\left(\lambda_n \frac{\log n}{n}\right) \quad a.s. \tag{41}$$

On the other hand, we have for the WLS estimator that

$$\| L^{1/2}(\hat{\theta}_n - \theta) \|^2 = o\left(\frac{\lambda_n}{n}\right) \quad a.s. \tag{42}$$

Remark. If ε has finite conditional moment of order $\alpha > 2$, we can take by Chow's lemma (see, e.g., Corollary 2.8.5 of Stout [23] or [12]) the sequence (λ_n) such that

$$\sum_{k=1}^{\infty} \left(\frac{1}{\lambda_n}\right)^{\alpha/2} < +\infty. \tag{43}$$

We can choose, for example, $\lambda_n = n^\beta$ with $2\alpha^{-1} < \beta < 1$. One can realize that (41) improves Theorem 3.2 (i) of Guo [16].

Proof. By Theorem 1 of Guo and Chen [15] and Theorem 5 of Bercu [5] on the prediction errors sequence (π_n), respectively, we have

$$\sum_{k=0}^{n} \| \pi_k \|^2 = o(n^c \alpha_n) \quad \text{a.s.} \tag{44}$$

with $c > 0$ for the ELS algorithm and

$$\sum_{k=0}^{n} \| \pi_k \|^2 = o(a_n^{-1} + \alpha_n) \quad \text{a.s.} \tag{45}$$

for the WLS algorithm. Then, for these two algorithms, we find that

$$\sum_{k=0}^{n} \| \pi_k \|^2 = o(\lambda_n) \qquad \text{a.s.} \tag{46}$$

By (33), we also have

$$\| C_n - \Gamma_n \| = O\left(\frac{1}{n}\sum_{k=1}^{n} \| \pi_{k-1} \|^2\right) \qquad \text{a.s.,} \tag{47}$$

and we immediately obtain relation (39). Therefore (33), (44), and (45), together with the second assumption on the sequence $\lambda = (\lambda_n)$, imply (40). Finally, for the ELS estimator, by Theorem 1 of Lai and Wei [19], we have $\| \hat{\theta}_{n+1} - \theta \|^2 = O(\log n)$ a.s. Moreover, by Theorem 1 of Bercu [5], the WLS estimator is always a.s. bounded, $\| \hat{\theta}_{n+1} - \theta \|^2 = O(1)$. Therefore, (40) clearly implies (41) and (42), completing the proof of Theorem 3.2. ◻

In order to obtain strong consistency for ELS and WLS algorithms, we are brought to introduce an excitation on the adaptive tracking control. As one can see below, the effect of this excitation is to make the limit matrix in Lemma 3.1 positive definite. First, we use the continually disturbed control given, for all $n \geq 0$, by

$$U_n = x_{n+1} - \hat{\theta}_n^t \Phi_n + \xi_{n+1}, \tag{48}$$

where the reference trajectory x satisfies (13) and ξ is an exogenous noise of dimension d, adapted to $\mathbf{F}$, with mean 0 and positive definite covariance matrix Λ. In addition, we assume that ξ is independent of ε, of x, and of the initial state of the system. Let

$$\Delta_n = \frac{1}{n}\sum_{k=1}^{n} (\varepsilon_k + \xi_k)(\varepsilon_k + \xi_k)^t. \tag{49}$$

Assume that ξ satisfies the strong law of large numbers, so Δ_n converges a.s. to $\Gamma + \Lambda$. Relation (1) can be rewritten as

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1} + \xi_{n+1}. \tag{50}$$

The adaptive tracking is said to be residually optimal if C_n converges a.s. to $\Gamma + \Lambda$. Let H be the square matrix of order $\delta = \delta_p + \delta_r$,

$$H = \begin{pmatrix} H_p & K \\ K^t & L_r \end{pmatrix}, \tag{51}$$

where H_p is the block diagonal square matrix of order $\delta_p = dp$:

$$H_p = \operatorname{diag}(\Gamma + \Lambda, \ldots, \Gamma + \Lambda). \tag{52}$$

THEOREM 3.3. *For the ARMA model, assume that* (P) *is satisfied. Assume that* ε *has finite conditional moment of order* > 2. *Then, for the ELS algorithm, we have*

$$\frac{S_n}{n} \longrightarrow H \qquad \textit{a.s.} \tag{53}$$

In addition, the tracking is residually optimal:

$$\| C_n - \Delta_n \| = O\left(\frac{\log n}{n}\right) \qquad \textit{a.s.} \tag{54}$$

Finally, $\hat{\theta}_n$ is a strongly consistent estimator of θ:

$$\| \hat{\theta}_n - \theta \|^2 = O\left(\frac{\log n}{n}\right) \qquad a.s. \tag{55}$$

Proof. The proof is given in Appendix D. □

THEOREM 3.4. *For the ARMA model, assume that* (P) *is satisfied. Assume that either ε is a white noise or ε 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*

$$(\log n)^{1+\gamma} \frac{S_n(a)}{n} \longrightarrow H \qquad a.s. \tag{56}$$

In addition, the tracking is residually optimal:

$$\| C_n - \Delta_n \| = o\left(\frac{(\log n)^{1+\gamma}}{n}\right) \qquad a.s. \tag{57}$$

Finally, $\hat{\theta}_n$ is a strongly consistent estimator of θ:

$$\| \hat{\theta}_n - \theta \|^2 = O\left(\frac{(\log n)^{1+\gamma}}{n}\right) \qquad a.s. \tag{58}$$

Proof. The proof is given in Appendix D. □

Remark. We note that Theorems 3.3 and 3.4 are similar to Theorems 2.1 and 2.2. In addition, it is easy to see that the matrix H is positive definite. In fact, if $p \leq r$, then $\det H = (\det \Gamma)^r (\det \Lambda)^p$, and if $p > r$, then $\det H = (\det \Gamma)^r (\det \Lambda)^r (\det(\Gamma + \Lambda))^{p-r}$.

THEOREM 3.5. *For the ARMA model, assume that* (P) *is satisfied, with p and r equal to* 1. *Assume that ε and ξ have finite conditional moments of order $\alpha > 2$. On the one hand, assume that x satisfies* (25) *and*

$$\sum_{k=1}^{n} \| x_k \|^2 = o\left(\frac{n}{\log n}\right) \qquad a.s. \tag{59}$$

for the ELS algorithm. On the other hand, assume that x satisfies (25) *and*

$$\sum_{k=1}^{n} \| x_k \|^2 = o\left(\frac{n}{(\log n)^{2+2\gamma}}\right) \qquad a.s. \tag{60}$$

for the WLS algorithm with $a_n^{-1} = (\log s_n)^{1+\gamma}$, where $\gamma > 0$. Then, the ELS and the WLS algorithms share the same CLT,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, H^{-1} \otimes \Gamma). \tag{61}$$

For any vectors $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^\delta$, they also share the same LIL,

$$\begin{aligned} \limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} v^t(\hat{\theta}_n - \theta)u &= -\liminf_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} v^t(\hat{\theta}_n - \theta)u \\ &= (v^t H^{-1} v)^{1/2} (u^t \Gamma u)^{1/2} \qquad a.s. \end{aligned} \tag{62}$$

In particular,

$$\left(\frac{\lambda_{min}\Gamma}{\lambda_{max}H}\right) \leq \limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right) \| \hat{\theta}_n - \theta \|^2 \leq \left(\frac{\lambda_{max}\Gamma}{\lambda_{min}H}\right) \qquad a.s. \tag{63}$$

Proof. The proof is given in Appendix E. □

4. Simulations. The goal of this section is to show that Theorem 3.5 is no longer true if the orders p or r are greater than 1. From relations (1) and (2), we have

$$S_{n-1}(a)(\hat{\theta}_n - \theta) = M_n(a) - R_{n-1}(a)\theta, \tag{64}$$

$$M_n(a) = M_0 + \sum_{k=1}^{n} a_{k-1}\Phi_{k-1}\varepsilon_k^t, \qquad R_n(a) = \sum_{k=0}^{n} a_k\Phi_k(\Phi_k - \Psi_k)^t \tag{65}$$

with $M_0 = S(\hat{\theta}_0 - \theta)$. By Lemmas C.1 or C.2 in Appendix C, we know how to deal with $M_n(a)$. The remainder $R_n(a)$ is much more complicated to study. This remainder vanishes in the AR framework. Consequently, we can easily establish CLT and LIL as in Theorem 2.3. In order to obtain similar results in the ARMA framework, we have to prove that the remainder $R_n(a)$ can be neglected. This was done with p and r equal to 1 in Theorem 3.5. Unfortunately, if p or r is greater than 1, $R_n(a)$ plays a prominent part and is really very complicated to study. We shall now show it by simulations for the ELS algorithm. Consider the following two models:

$$X_{n+1} = \frac{5}{4}X_n + \frac{1}{2}X_{n-1} + U_n + \frac{3}{4}\varepsilon_n + \varepsilon_{n+1}, \tag{I}$$

$$X_{n+1} = \frac{5}{4}X_n + U_n + \frac{3}{4}\varepsilon_n + \frac{1}{4}\varepsilon_{n-1} + \varepsilon_{n+1}, \tag{II}$$

where ε is a Gaussian white noise $N(0,1)$. For simplicity, we study the regulation problem taking the reference trajectory $x = 0$. Therefore, we use the continually disturbed control

$$U_n = -\hat{\theta}_n^t\Phi_n + \xi_{n+1}, \tag{66}$$

where ξ is an exogenous Gaussian white noise $N(0,1)$. We base our simulations on $M =$ 500 realizations of sample size $N = 10000$. In order to keep this section brief, we focus our attention on the behavior of the statistic

$$Z_N = \sqrt{N}H^{1/2}(\hat{\theta}_N - \theta), \tag{67}$$

where the matrix H is for models (I) and (II), respectively:

$$\begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{68}$$

We expect at least that each component of Z_N has $N(0,1)$ distribution. Figure 1 represents the three coordinates of Z_N in model (I). One can realize that the second coordinate is not $N(0,1)$. Figure 2 represents the three coordinates of Z_N in model (II). One can realize that the third coordinate is not $N(0,1)$. Next, if we consider an ARMA model of orders $p=2$ and $r=2$, we can also see that the second and the fourth coordinates of Z_N are not $N(0,1)$. We can conclude that if p or r is greater than 1, $R_n(a)$ plays a prominent part which can't be neglected. It would be very nice to clarify the behavior of $R_n(a)$ in ARMA adaptive tracking.

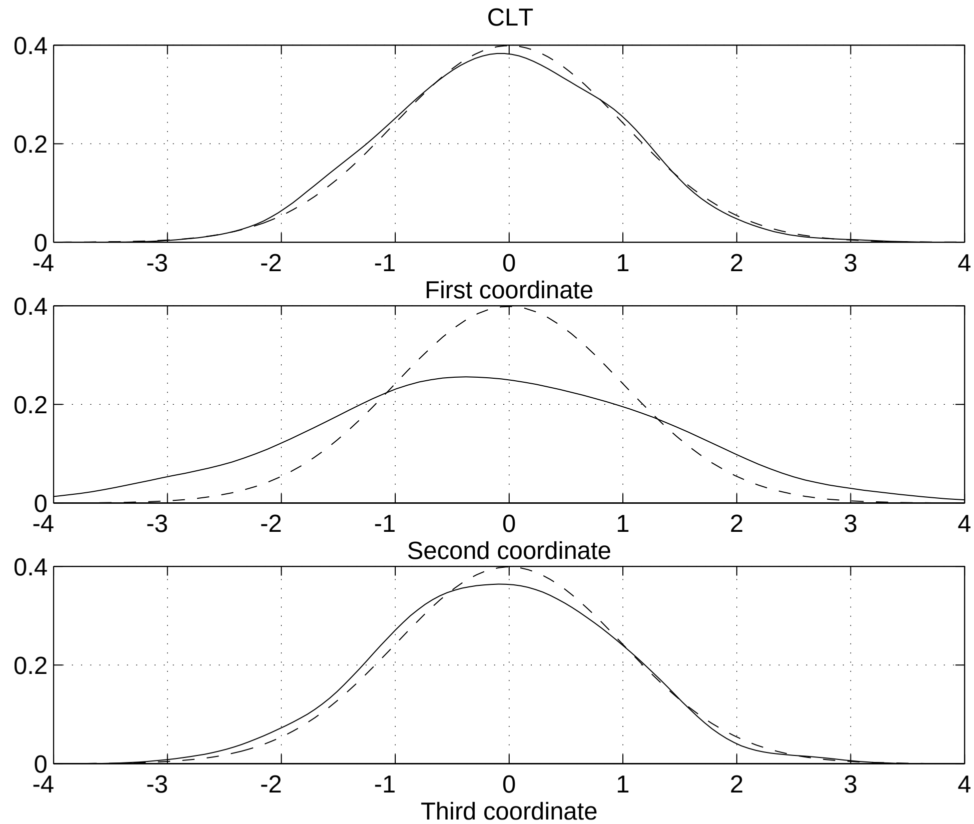

Fig. 1.

5. Conclusion. In AR adaptive tracking, we have proved that the LS and the WLS algorithms share the same CLT and LIL. We have also extended and shown the limitations of these results in the ARMA framework. One can ask the natural question: Why make use of the WLS algorithm?

• First, we have seen in this paper that WLS performs as well as ELS for parameter estimation when the system is persistently excited. There is no loss in asymptotic efficiency by using the WLS algorithm.

• Next, as it was shown in [5], the WLS algorithm is more convenient than the ELS in the analysis of autoregressive with moving average and exogenous control (ARMAX) adaptive tracking thanks to the behavior of the prediction errors sequence. The convergence rates proved for the tracking optimality are in general better for the WLS [5] than for the ELS [15].

• In the ARMAX framework, the leading matrix associated with the control is usually called the high frequency gain. For ARX models with known or unknown high frequency gain, strong consistency and tracking optimality results have been established in [16]. It is reasonable to conjecture that CLT and LIL could also be proved for ARX models with known high frequency gain. However, it would be extremely difficult in the general case.

• Finally, Guo [17] has recently proved the almost sure self-convergence of the WLS algorithm. This property can lead to various applications in adaptive control theory such as adaptive pole-placement and LQG problems [17], [22] for ARMAX models.

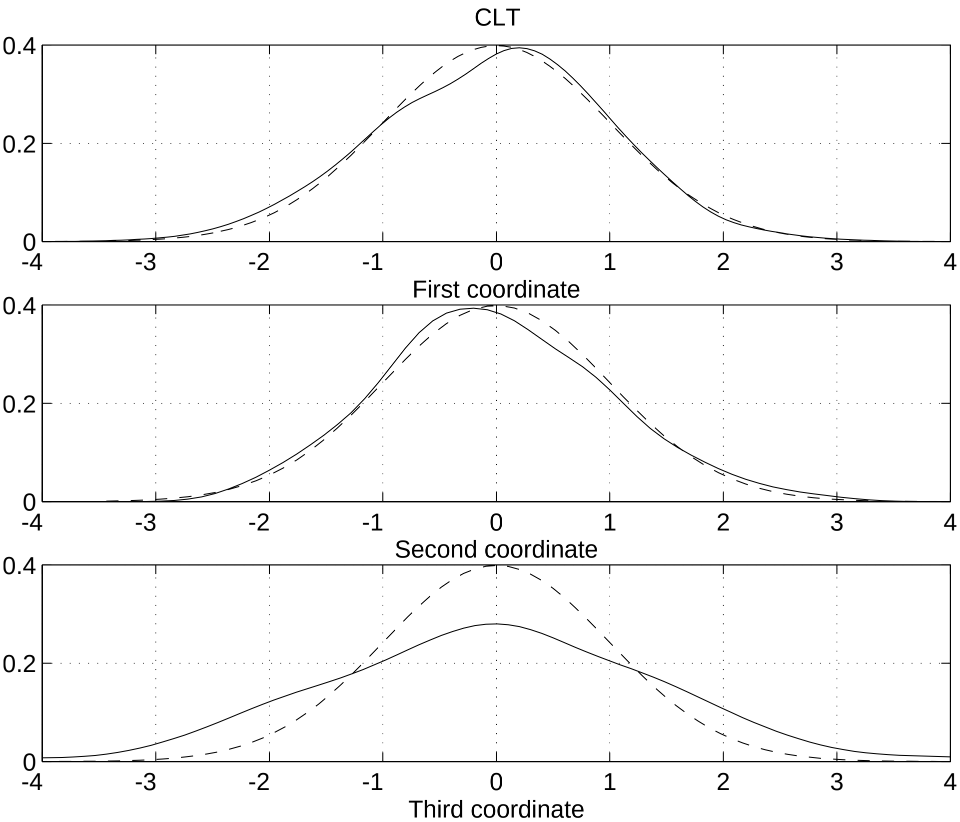

FIG. 2.

Appendix A.

Proof of Theorem 2.1. By the strong law of large numbers and relation (12), we easily prove that $n = O(s_n)$. By Lemma 1 of Guo and Chen [15] or Theorem 1 of Bercu [3] on the prediction errors sequence (π_n), we have

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

where $f_n = \Phi_n^t S_n^{-1} \Phi_n$. If ε has finite conditional moment of order $\alpha > 2$, using the same approach as Chen and Guo [11], [15], we can show by (A.1) that $\parallel \Phi_n \parallel^2 = O(s_n^\beta)$ with $2\alpha^{-1} < \beta < 1$. We also find by (A.1) and (12) that

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

$$\sum_{k=1}^{n} \parallel X_{k+1} \parallel^2 = o(s_n^\beta \log s_n) + O(n) \qquad \text{a.s.} \tag{A.3}$$

Finally, we obtain that $s_n = o(s_n) + O(n)$, so $s_n = O(n)$. Consequently, we prove the tracking optimality, as by (12) and (A.2),

$$\| C_n - \Gamma_n \| = O\Big(\frac{1}{n}\sum_{k=1}^{n} \| \pi_{k-1} \|^2\Big) \qquad \text{a.s.,} \tag{A.4}$$

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

We still have to establish the convergence rate given in (19). As the reference trajectory x satisfies (13), we have already proven the almost sure convergence

$$\frac{1}{n}\sum_{k=0}^{n} X_k X_k^t \longrightarrow \Gamma. \tag{A.6}$$

Recalling (12), we have for $1 \leq i \leq p-1$ that

$$\sum_{k=1}^{n} X_k X_{k-i}^t = \sum_{k=1}^{n} (\pi_{k-1} + x_k) X_{k-i}^t + \sum_{k=1}^{n} \varepsilon_k X_{k-i}^t. \tag{A.7}$$

The right-hand side of (A.7) is a regressive sequence. Therefore, we have a.s.

$$\left\| \sum_{k=1}^{n} X_k X_{k-i}^t \right\| \leq \sum_{k=1}^{n} \| \pi_{k-1} + x_k \| \| X_{k-i} \| + o\left(\sum_{k=1}^{n} \| X_{k-i} \|^2\right). \tag{A.8}$$

We prove, by (13) and (A.5), together with the Cauchy–Schwarz inequality, that

$$\sum_{k=1}^{n} X_k X_{k-i}^t = o(n) \qquad \text{a.s.,} \tag{A.9}$$

which implies the convergence (18). As the matrix L_p is positive definite, it clearly follows that $n = O(\lambda_{min} S_n)$, $\| \Phi_n \|^2 = o(n)$, and f_n tends a.s. towards 0. Then, by (A.1), we find that

$$\sum_{k=1}^{n} \| \pi_k \|^2 = O(\log n) \qquad \text{a.s.,} \tag{A.10}$$

and consequently, we obtain the relation (19). By a well-known result established in Theorem 1 of Lai and Wei [19], [20] for the LS estimator, we also have

$$\| \hat{\theta}_{n+1} - \theta \|^2 = O\left(\frac{\log s_n}{\lambda_{min} S_n}\right) \qquad \text{a.s.,} \tag{A.11}$$

which implies (21). Moreover, if $\breve{\theta}_n = \hat{\theta}_n - \theta$, we immediately deduce from (28) that

$$\| S_{n-1}^{1/2} \breve{\theta}_n \|^2 = o(\log n) \qquad \text{a.s.} \tag{A.12}$$

By Duflo, Senoussi, and Touati [14, p. 560], we also have the almost sure convergence

$$\frac{1}{\log n}\left(\breve{\theta}_n^t S_{n-1} \breve{\theta}_n + \sum_{k=0}^{n-1} (1 - f_k) \pi_k \pi_k^t\right) \longrightarrow \delta_p \Gamma. \tag{A.13}$$

Finally, (A.12) and (A.13) imply (20), completing the proof of Theorem 2.1. □

Appendix B.

Proof of Theorem 2.2. By Theorem 1 of Bercu and Duflo [4], [5] on the prediction errors sequence (π_n), we have

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

where $f_n(a) = a_n \Phi_n^t S_n^{-1}(a) \Phi_n$. Then, as $a_n^{-1} = O(s_n)$, we find by (B.1) together with Kronecker's lemma that

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

Contrary to the LS algorithm, we can easily prove that $s_n = O(n)$. In fact, (B.2) and (12) immediately imply

$$\sum_{k=1}^{n} \parallel X_{k+1} \parallel^2 = o(s_n) + O(n) \qquad \text{a.s.} \tag{B.3}$$

Therefore, $s_n = o(s_n) + O(n)$, so $s_n = O(n)$. Finally, we have established the tracking optimality. In Appendix A, we have also shown that $n^{-1} S_n$ converges a.s. to L_p. Consequently, as the weighting sequence $a = (a_n)$ is nonincreasing, it results that $a_n S_n \leq S_n(a)$ so $n a_n = O(\lambda_{min} S_n(a))$ and $f_n(a)$ tends a.s. towards 0. We can conclude by (B.1) that

$$\sum_{k=1}^{n} \parallel \pi_k \parallel^2 = o(a_n^{-1}) \qquad \text{a.s.,} \tag{B.4}$$

which implies relation (23) as s_n has the same order as n, so a_n^{-1} is a.s. equivalent to $(\log n)^{1+\gamma}$. We can also deduce (24), as by Theorem 1 of Bercu and Duflo [4], [5],

$$\parallel \hat{\theta}_{n+1} - \theta \parallel^2 = O\left(\frac{1}{\lambda_{min} S_n(a)} \right) \qquad \text{a.s.} \tag{B.5}$$

Now, we have

$$S_n(a) = a_{n+1} S_n + \sum_{k=1}^{n} b_k \frac{S_k}{k} + R \tag{B.6}$$

with $b_n = n(a_n - a_{n+1})$ and $R = S_0(a) - a_1 S_0$. In addition,

$$\sum_{k=1}^{n} b_k = \sum_{k=1}^{n} a_k - n a_{n+1}. \tag{B.7}$$

Next, as a_n^{-1} is a.s. equivalent to $(\log n)^{1+\gamma}$,

$$\sum_{k=1}^{n} b_k \sim (1+\gamma) \frac{n a_n}{\log n}, \qquad \sum_{k=1}^{n} b_k = o(n a_n) \qquad \text{a.s.} \tag{B.8}$$

Finally, (B.6), together with Toeplitz's lemma, imply the convergence (22), completing the proof of Theorem 2.2. □

Appendix C.

Proof of Theorem 2.3. In order to prove Theorem 2.3, we need the two following lemmas on regressive sequences. They result from the CLT on triangular arrays [18], [21], [25] and from the LIL on martingales [14], [23], [24]. Let $\varepsilon = (\varepsilon_n)$ be a d-dimensional noise, adapted to $\mathbf{F}$, which satisfies (14) where Γ is a deterministic covariance matrix. Let $\varphi = (\varphi_n)$ be a δ-dimensional sequence of random vectors, adapted to $\mathbf{F}$. Set, for $n \geq 0$,

$$M_n = M_0 + \sum_{k=1}^{n} \varphi_{k-1}\varepsilon_k^t, \qquad S_n = \sum_{k=0}^{n} \varphi_k \varphi_k^t + S.$$

LEMMA C.1. *Let (c_n) be a deterministic real sequence increasing to infinity. Assume that, for all $\epsilon > 0$,*

$$(H_1) \qquad c_n^{-1} S_{n-1} \xrightarrow{\mathcal{P}} L,$$

$$(H_2) \qquad c_n^{-1} \sum_{k=1}^{n} E\left[\| \Delta M_k \|^2 \mathbf{1}_{\left\{\| \Delta M_k \| \geq \epsilon \sqrt{c_n}\right\}} \mid \mathcal{F}_{k-1} \right] \xrightarrow{\mathcal{P}} 0,$$

where $\Delta M_n = M_n - M_{n-1}$. Then, $c_n^{-1} M_n$ tends a.s. towards 0 and

$$\frac{1}{\sqrt{c_n}} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, L \otimes \Gamma).$$

In addition, if L is positive definite, we have the CLT

$$\sqrt{c_n} S_{n-1}^{-1} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, L^{-1} \otimes \Gamma).$$

Remark. Assume that ε has finite conditional moment of order > 2. Then, Lindeberg's condition (H$_2$) is satisfied if $\| \varphi_n \|^2 = o(c_n)$ a.s.

LEMMA C.2. *Let (c_n) be a deterministic real sequence increasing to infinity. Assume that the noise ε has finite conditional moment of order $\alpha > 2$. Also assume that*

$$(H_3) \qquad c_n^{-1} S_{n-1} \longrightarrow L \quad a.s.,$$

$$(H_4) \qquad \sum_{n=1}^{\infty} \left(\frac{\| \varphi_n \|}{\sqrt{c_n}} \right)^{\beta} < +\infty \quad a.s.,$$

with $2 < \beta \leq \alpha$. Then, for any vector $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^\delta$ such that $v^t L v > 0$, we have

$$\limsup_{n\to\infty} \left(\frac{1}{2c_n \log\log c_n} \right)^{1/2} v^t M_n u = -\liminf_{n\to\infty} \left(\frac{1}{2c_n \log\log c_n} \right)^{1/2} v^t M_n u$$
$$= (v^t L v)^{1/2} (u^t \Gamma u)^{1/2} \quad a.s.$$

In addition, if L is positive definite, we have the LIL

$$\limsup_{n\to\infty} \left(\frac{c_n}{2\log\log c_n} \right)^{1/2} v^t S_{n-1}^{-1} M_n u = -\liminf_{n\to\infty} \left(\frac{c_n}{2\log\log c_n} \right)^{1/2} v^t S_{n-1}^{-1} M_n u$$
$$= (v^t L^{-1} v)^{1/2} (u^t \Gamma u)^{1/2} \quad a.s.$$

Theorem 2.3 is a direct application of Lemmas C.1 and C.2. By the relations (1) and (2), we have

$$\hat{\theta}_n - \theta = S_{n-1}^{-1}(a)M_n(a), \tag{C.1}$$

$$M_n(a) = M_0 + \sum_{k=1}^{n} a_{k-1}\Phi_{k-1}\varepsilon_k^t \tag{C.2}$$

with $M_0 = S(\hat{\theta}_0 - \theta)$. On the one hand, for the LS algorithm, we choose $\varphi_n = \Phi_n$ and $c_n = n$. On the other hand, for the WLS algorithm, we take $\varphi_n = a_n\Phi_n$ and $c_n = n/(\log n)^{2+2\gamma}$. First, for the LS algorithm, (26) can be clearly deduced via Lemma C.1 together with (18) and equation (C.1). In addition, if ε has finite conditional moment of order $\alpha > 2$, for all $2 < \beta < \alpha$, we have by Chow's lemma (see, e.g., Corollary 2.8.5 of Stout [23]) that

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

Since the reference trajectory x satisfies (25), we show by (A.10) that

$$\sum_{k=1}^{n} \| X_k \|^{\beta} = O(n), \qquad \sum_{k=1}^{n} \| \Phi_k \|^{\beta} = O(n) \quad \text{a.s.} \tag{C.4}$$

Therefore, as $\beta > 2$, (C.4) implies

$$\sum_{n=1}^{\infty} \left(\frac{\| \Phi_n \|}{\sqrt{n}} \right)^{\beta} < +\infty \qquad \text{a.s.} \tag{C.5}$$

Finally, we find (27) via Lemma C.2 together with (18) and equation (C.1). Next, for the WLS algorithm, set

$$Q_n(a) = \sum_{k=0}^{n} a_k^2 \Phi_k \Phi_k^t + S. \tag{C.6}$$

As in (22), we prove that $c_n^{-1}Q_n(a)$ converges a.s. to L_p. Hence, (22) and equation (C.1) clearly imply (26). In addition, if ε has finite conditional moment of order $\alpha > 2$, for all $2 < \beta < \alpha$, we have by Chow's lemma [23], together with (B.4),

$$\sum_{k=1}^{n} (a_k \| \Phi_k \|)^{\beta} = O(n) \qquad \text{a.s.} \tag{C.7}$$

Then, it follows from (C.7) that

$$\sum_{n=1}^{\infty} \left(\frac{a_n \| \Phi_n \|}{\sqrt{c_n}} \right)^{\beta} < +\infty \qquad \text{a.s.} \tag{C.8}$$

Finally, we prove (27) via Lemma C.2 together with (22) and equation (C.1), completing the proof of Theorem 2.3. ∎

Appendix D.

Proof of Theorems 3.3 and 3.4. First, we prove Lemma 3.1 for both ELS and WLS algorithms. On the one hand, relation (A.1) holds for the ELS algorithm in the ARMAX framework [3], [11], [15]. In addition, by Theorem 1 of Lai and Wei [19], [20], we also have

$$\sum_{k=1}^{n} \| \Phi_k - \Psi_k \|^2 = O(\log s_n) \qquad \text{a.s.} \tag{D.1}$$

If ε has finite conditional moment of order $\alpha > 2$, we can show as in Appendix A that $\| \Phi_n \|^2 = O(s_n^\beta)$ with $2\alpha^{-1} < \beta < 1$. Hence, we find by (A.1) and (33) that

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

$$\sum_{k=1}^{n} \| X_{k+1} \|^2 = o(s_n^\beta \log s_n) + O(n) \qquad \text{a.s.} \tag{D.3}$$

Finally, (D.1) together with (D.3) imply that $s_n = o(s_n) + O(n)$, so $s_n = O(n)$. Consequently, we find by (D.2) that

$$\sum_{k=1}^{n} \| \pi_k \|^2 = o(n) \qquad \text{a.s.} \tag{D.4}$$

We now recall that the reference trajectory x satisfies (13). Therefore, exactly as in Appendix A, (33), (D.1), and (D.4) imply the convergence (36) for the ELS algorithm. On the other hand, concerning the WLS algorithm, relation (B.1) holds in the ARMAX framework [4], [5]. In addition, we also have, by Theorem 1 of Bercu [5],

$$\sum_{n=1}^{\infty} a_n \| \Phi_n - \Psi_n \|^2 < +\infty \qquad \text{a.s.} \tag{D.5}$$

As $a_n^{-1} = (\log s_n)^{1+\gamma}$ with $\gamma > 0$, we find by (33), (B.2), and (D.5) together with Kronecker's lemma that $s_n = o(s_n) + O(n)$, so $s_n = O(n)$. Consequently, we immediately obtain by (B.2) that

$$\sum_{k=1}^{n} \| \pi_k \|^2 = o(n) \qquad \text{a.s.} \tag{D.6}$$

Therefore, (33), (D.5), and (D.6) imply the convergence (36) for the WLS algorithm. Finally, via (B.6)–(B.8), we also find the convergence (37) for the WLS algorithm, completing the proof of Lemma 3.1. We now prove Theorems 3.3 and 3.4. We can easily switch to the continually disturbed tracking situation. Indeed, as ξ is an exogenous noise that satisfies the strong law of large numbers, we prove by (50) the convergences (53) and (56) exactly as in Lemma 3.1. Furthermore, since the matrix H is positive definite, we find for the ELS algorithm that $n = O(\lambda_{min} S_n)$, and for the WLS algorithm, that $n a_n = O(\lambda_{min} S_n(a))$. Finally, as the relations (A.1), (A.11) and (B.1), (B.5) hold in the ARMAX framework, Theorems 3.3 and 3.4 are established. □

Appendix E.

Proof of Theorem 3.5. We finally prove Theorem 3.5 for both ELS and WLS algorithms. On the one hand, for the ELS algorithm, by (1) and (2), we have

$$S_{n-1}(\hat{\theta}_n - \theta) = M_n - R_{n-1}\theta, \tag{E.1}$$

$$M_n = M_0 + \sum_{k=1}^{n} \Phi_{k-1}\varepsilon_k^t, \qquad R_n = \sum_{k=0}^{n} \Phi_k(\Phi_k - \Psi_k)^t \tag{E.2}$$

with $M_0 = S(\hat{\theta}_0 - \theta)$. In order to study the remainder R_n, it is enough by (D.1) to work on

$$P_n = \sum_{k=0}^{n} X_k \breve{\varepsilon}_k^t, \qquad Q_n = \sum_{k=0}^{n} \varepsilon_k \breve{\varepsilon}_k^t, \tag{E.3}$$

where $\breve{\varepsilon}_n = \hat{\varepsilon}_n - \varepsilon_n$. The first equality of (4) can be rewritten as

$$\breve{\varepsilon}_{n+1} = (1 - f_n)\pi_n - f_n \varepsilon_{n+1} \tag{E.4}$$

with $f_n = \Phi_n^t S_n^{-1} \Phi_n$. By (A.1) and (E.4) together with Chow's lemma [23], we have the almost sure convergence

$$\frac{1}{\log n} \sum_{k=0}^{n} \varepsilon_k \breve{\varepsilon}_k^t \longrightarrow -\delta\Gamma. \tag{E.5}$$

Therefore, we immediately obtain $Q_n = o(\sqrt{n})$ a.s. In addition, by (A.1), we also have

$$\left\| \sum_{k=1}^{n} \pi_{k-1} \breve{\varepsilon}_k^t \right\| = O(\log n) \qquad \text{a.s.} \tag{E.6}$$

Finally, as the trajectory x satisfies relation (59), we can conclude by (50), (A.1), and the Cauchy–Schwarz inequality that $P_n = o(\sqrt{n})$, so $R_n = o(\sqrt{n})$ a.s. Lemmas C.1 and C.2 together with (53) lead to (61) and (62) for the ELS algorithm. On the other hand, for the WLS algorithm, by (1) and (2), we have

$$S_{n-1}(a)(\hat{\theta}_n - \theta) = M_n(a) - R_{n-1}(a)\theta, \tag{E.7}$$

$$M_n(a) = M_0 + \sum_{k=1}^{n} a_{k-1}\Phi_{k-1}\varepsilon_k^t, \qquad R_n(a) = \sum_{k=0}^{n} a_k \Phi_k(\Phi_k - \Psi_k)^t \tag{E.8}$$

with $M_0 = S(\hat{\theta}_0 - \theta)$. Set

$$P_n(a) = \sum_{k=0}^{n} a_k X_k \breve{\varepsilon}_k^t, \qquad Q_n(a) = \sum_{k=0}^{n} a_k \varepsilon_k \breve{\varepsilon}_k^t. \tag{E.9}$$

The first equality of (4) can be rewritten as

$$\breve{\varepsilon}_{n+1} = (1 - f_n(a))\pi_n - f_n(a)\varepsilon_{n+1} \tag{E.10}$$

with $f_n(a) = a_n \Phi_n^t S_n^{-1}(a)\Phi_n$. The main property of the weighted sequence $a = (a_n)$ is that

$$\sum_{n=1}^{\infty} a_n f_n(a) < +\infty \qquad \text{a.s.} \tag{E.11}$$

Therefore, as $a=(a_n)$ is nonincreasing, we find by (B.1), (E.10), and (E.11) that

$$\left\|\sum_{k=0}^{n} a_k \varepsilon_k \check{\varepsilon}_k^t\right\| = o((\log n)^{1+\gamma}) \qquad \text{a.s.,} \tag{E.12}$$

so $Q_n(a)=o(\sqrt{c_n})$ a.s. with $c_n = n/(\log n)^{2+2\gamma}$. In addition, by (B.1) and (D.5), as $f_n(a)$ tends a.s. towards 0,

$$\left\|\sum_{k=1}^{n} \pi_{k-1} \check{\varepsilon}_k^t\right\| = O(1) \qquad \text{a.s.} \tag{E.13}$$

Finally, as the trajectory x satisfies relation (60), we can conclude by (50), (B.1), and the Cauchy–Schwarz inequality that $P_n(a) = o(\sqrt{c_n})$, so $R_n(a) = o(\sqrt{c_n})$ a.s. Lemmas C.1 and C.2, together with (56), lead to (61) and (62) for the WLS algorithm, completing the proof of Theorem 3.5. □

REFERENCES

[1] K. J. Aström and B. Wittenmark, *On self tuning regulators*, Automatica, 9 (1973), pp. 185–199.

[2] A. H. Becker, P. R. Kumar, and C. Z. Wei, *Adaptive control with the stochastic approximation algorithm: Geometry and convergence*, IEEE Trans. Automat. Control, AC-30 (1985), pp. 330–338.

[3] B. Bercu, *Sur l'estimateur des moindres carrés généralisé d'un modèle ARMAX. Application à l'identification des modèles ARMA*, Ann. Inst. H. Poincaré, 27 (1991), pp. 425–443.

[4] B. Bercu and M. Duflo, *Moindres carrés pondérés et poursuite*, Ann. Inst. Henri Poincaré, 28 (1992), pp. 403–430.

[5] B. Bercu, *Weighted estimation and tracking for ARMAX models*, SIAM J. Control Optim., 33 (1995), pp. 89–106.

[6] B. Bercu, *Théorème de limite centrale et loi du logarithme itéré pour les algorithmes des moindres carrés en poursuite adaptative*, Note au C. R. Acad. Sci. Paris Sér. I, 320 (1995), pp. 493–496.

[7] P. E. Caines, *Linear Stochastic Systems*, John Wiley, New York, 1988.

[8] H. F. Chen, *Recursive Estimation and Control for Stochastic Systems*, John Wiley, New York, 1985.

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

[10] H. F. Chen and J. F. Zhang, *Convergence rates in stochastic adaptive tracking*, Internat. J. Control, 49 (1989), pp. 1915–1935.

[11] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*, Birkhäuser, Boston, 1991.

[12] Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability and Martingales*, Springer-Verlag, Berlin, 1978.

[13] M. Duflo, *Random Iterative Methods*, Springer-Verlag, Berlin, 1996.

[14] M. Duflo, R. Senoussi, and A. Touati, *Sur la loi des grands nombres pour les martingales vectorielles et l'estimateur des moindres carrés d'un modéle de régression*, Ann. Inst. H. Poincaré, 26 (1990), pp. 549–566.

[15] 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.

[16] L. Guo, *Further results on least squares based adaptive minimum variance control*, SIAM J. Control Optim., 32 (1994). pp. 187–212.

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

[18] D. Hall and C. Heyde, *Martingale Limit Theory and Its Applications*, Academic Press, New York, 1980.

[19] 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, AC-31 (1986), pp. 898–906.

[20] T. L. Lai and C. Z. Wei, *On the concept of excitation in least squares identification and adaptive control*, Stochastics, 16 (1986), pp. 227–254.
[21] R. S. Liptser and A. N. Shiryaev, *A functional central limit theorem for semimartingales*, Theory Probab. Appl., 25 (1980), pp. 667–688.
[22] K. Nassiri-Toussi and W. Ren, *Indirect adaptive pole-placement control of MIMO stochastic systems: Self-tuning results*, IEEE Trans. Automat. Control, 42 (1997), pp. 38–52.
[23] W. F. Stout, *Almost Sure Convergence*, Academic Press, New York, 1974.
[24] W. F. Stout, *A martingale analogue of Kolmogorov's law of the iterated logarithm*, Z. Wahrscheinlichkeitstheorie, 15 (1970), pp. 279–290.
[25] A. Touati, *On the functional convergence in distribution of sequences of semimartingales to a mixture of brownian motions*, Theory Probab. Appl., 36 (1991), pp. 752–771.
[26] A. Touati, *Vitesse de convergence en loi de l'estimateur des moindres carrés d'un modèle autoregressif (cas mixte)*, Ann. Inst. H. Poincaré, 32 (1996), pp. 211–230.