SIAM J. CONTROL AND OPTIMIZATION
Vol. 33, No. 1, pp. 89–106, January 1995

004

WEIGHTED ESTIMATION AND TRACKING FOR ARMAX MODELS *

B. BERCU†

Abstract. For complex multivariate ARMAX models, the author studies the weighted least squares algorithm which offers, by the choice of suitable weightings, the advantages of both the extended least squares and the stochastic gradient algorithms. Concerning adaptive tracking problems, the strong consistency of the estimator and control optimality are both ensured. Almost sure rates of convergence are also provided.

Résumé. Pour les modèles ARMAX vectoriels complexes, on étudie l'algorithme des moindres carrés pondérés qui conjugue, par le choix de pondérations convenables, à la fois les avantages des algorithmes des moindres carrés généralisés et du gradient. Concernant les problèmes de poursuite adaptative, on assure la consistance de l'estimateur et l'optimalité du contrôle. On précise également les vitesses de convergence presque sûre.

Key words. linear systems, estimation, strong consistency, adaptive tracking

AMS subject classifications. 60G42, 93E12, 62M20, 93E20

1. Introduction. In the study of recursive identification and adaptive tracking for ARMAX linear systems, the major goal is to find a stochastic algorithm that ensures both strong consistency of the estimator and control optimality. On one hand, if we focus our attention on the strong consistency, we choose the extended least squares (ELS) algorithm [15], [19], [22], [23], [24]. On the other hand, if we are interested in adaptive tracking, we should use the stochastic gradient (SG) algorithm [8], [18]. Therefore, a natural question is: Can we find a stochastic algorithm that combines both advantages of the ELS for strong consistency and of the SG for adaptive tracking? A positive answer was recently given by Bercu and Duflo [4] when they proposed a new weighted least squares (WLS) algorithm. In this paper we complete their work, giving a solution to the twenty-year-old adaptive tracking problem proposed by Aström and Wittenmark [1] for ARMAX models.

The paper is organized as follows. In §2, we describe the WLS algorithm. The main difference from the ELS algorithm is the introduction of a random weighting sequence $a = (a_n)$. Section 3 is devoted to the crucial choice of $a = (a_n)$. The main results of the paper are given in §4. We can see that the WLS algorithm equals the performance of the ELS for the strong consistency and matches the best result of the SG for the adaptive tracking. More precisely, the relation (24) is similar to the one obtained by Lai and Wei [24], [25] or Chen and Guo [12] for the ELS estimator. Moreover, concerning the prediction errors sequence, the relation (26) is exactly the same as the one proved by Goodwin, Ramadge, and Caines [18] for the SG algorithm. Finally, in §§5 and 6, we solve, in a simple way, the adaptive tracking problem. We prove both strong consistency of the WLS estimator and control optimality. We also provide almost sure rates of convergence. Section 7 is devoted to a survey on earlier related works on adaptive tracking. Comparing our work with previous similar results, we show how the WLS algorithm is well suited for adaptive tracking problems. A short conclusion is given in §8. All technical proofs are collected in the Appendices.

Notations. In the following sections, for any matrix A, ${}^tA$ denotes the transpose of A, ${}^*A$ represents the Hermitian adjoint of A and we set $\|A\|^2 = \text{tr}(A^*A)$. Moreover, if A is a square matrix, $\text{tr}(A)$ denotes the trace of A, $\det(A)$ the determinant of A, and $\lambda_{\min}(A)$, $\lambda_{\max}(A)$ the minimum and the maximum eigenvalues of A, respectively. In addition, if A and B are two positive definite Hermitian matrices, then $A \leq B$ if $B - A$ is positive definite. Finally, for any positive integer d, I_d is the identity matrix of order d.

* Received by the editors August 31, 1992; accepted for publication (in revised form) August 5, 1993.

† Laboratoire de Statistiques, Batiment 425 Mathématiques, Université de Paris-Sud, 91 405 Orsay Cedex, France.

2. Weighted estimation. Let $(\Omega, \mathcal{A}, P)$ be a probability space with a filtration $\mathbf{F} = (\mathcal{F}_n)_{n\geq 0}$, where $\mathcal{F}_n$ is the σ-algebra generated by events occurring up to time n. We consider the following complex multivariate ARMAX model of order (p, q, r):

$$A(R)Y_n = B(R)U_n + C(R)\varepsilon_n, \tag{1}$$

where Y, U, and ε are the d_1-dimensional output, d_2-dimensional input, and d_1-dimensional driven noise, respectively. Set for the shift-back operator R,

$$A(R) = I_{d_1} - A_1 R - \cdots - A_p R^p, \tag{2}$$

$$B(R) = B_1 R + \cdots + B_q R^q, \tag{3}$$

$$C(R) = I_{d_1} + C_1 R + \cdots + C_r R^r, \tag{4}$$

where A_i, B_j, C_k are unknown matrices. Assume that the control $U = (U_n)$ and the noise $\varepsilon = (\varepsilon_n)$ are adapted to $\mathbf{F}$ and that ε is a martingale difference sequence with

$$\sup_{n\geq 0} E[\|\varepsilon_{n+1}\|^2 \mid \mathcal{F}_n] < \sigma^2 \quad \text{a.s.}, \tag{5}$$

where σ^2 is deterministic. The initial state ${}^t\Psi_0 = ({}^tY_0^p, {}^tU_0^q, {}^t\varepsilon_0^r)$ is $\mathcal{F}_0$ measurable and, for $n \geq 0$,

$${}^t\Psi_n = ({}^tY_n^p, {}^tU_n^q, {}^t\varepsilon_n^r), \tag{6}$$

where ${}^tY_n^p = ({}^tY_n, \ldots, {}^tY_{n-p+1})$, ${}^tU_n^q = ({}^tU_n, \ldots, {}^tU_{n-q+1})$ and ${}^t\varepsilon_n^r = ({}^t\varepsilon_n, \ldots, {}^t\varepsilon_{n-r+1})$. Let $\hat{\theta}_n$ be an estimator of θ where

$${}^*\theta = (A_1, \ldots, A_p, B_1, \ldots, B_q, C_1, \ldots, C_r). \tag{7}$$

The noise ε is predicted by the a posteriori error $\hat{\varepsilon}$ with $\hat{\varepsilon}_0^r = 0$ and, for $n \geq 0$,

$$\hat{\varepsilon}_{n+1} = Y_{n+1} - {}^*\hat{\theta}_{n+1}\Phi_n, \tag{8}$$

where

$${}^t\Phi_n = ({}^tY_n^p, {}^tU_n^q, {}^t\hat{\varepsilon}_n^r). \tag{9}$$

Let $a = (a_n)$ be a sequence of random variables adapted to $\mathbf{F}$, positive, nonincreasing, and ≤ 1. $a = (a_n)$ is called a weighting sequence. We propose, in order to estimate θ, the WLS estimator $\hat{\theta}_n$ introduced by Bercu and Duflo [4] and given, for $n \geq 0$, by

$$\hat{\theta}_{n+1} = \hat{\theta}_n + a_n S_n^{-1}(a)\Phi_n {}^*(Y_{n+1} - {}^*\hat{\theta}_n\Phi_n), \tag{10}$$

where the initial value $\hat{\theta}_0$ is arbitrarily chosen and, for $n \geq 0$,

$$S_n(a) = \sum_{k=0}^{n} a_k \Phi_k {}^*\Phi_k + S \tag{11}$$

with any positive definite Hermitian and deterministic matrix S. We also write, for such a matrix Q,

$$Q_n(a) = \sum_{k=0}^{n} a_k \Psi_k {}^*\Psi_k + Q. \tag{12}$$

The inverses of the matrices $S_n(a)$ and $Q_n(a)$ are recursively generated by the matrix inversion formula of Riccati. Denote by $s_n(a)$ and $q_n(a)$ the traces of $S_n(a)$ and $Q_n(a)$, respectively. We also make use of

$$s_n = \sum_{k=0}^{n} \|\Phi_k\|^2 + s, \tag{13}$$

where $s = \text{tr}(S)$, and of the prediction errors sequence $\pi = (\pi_n)$ defined by

$$\pi_n = {}^*\theta\Psi_n - {}^*\hat{\theta}_n\Phi_n. \tag{14}$$

3. Admissibility. Consider a weighting sequence $a = (a_n)$ and set, for $n \geq 0$,

$$f_n(a) = a_n\, {}^*\Phi_n S_n^{-1}(a)\Phi_n, \qquad \Delta = \sum_{n=0}^{\infty} a_n f_n(a). \tag{15}$$

a is said to be admissible if Δ is integrable. Let F be the family of continuous and nonincreasing functions f from $\mathbf{R}^+$ to $\mathbf{R}^+$ such that $xf(x)$ converges to 0 as x goes to infinity and

$$\int_c^{+\infty} f(x)dx < +\infty \tag{16}$$

for any constant $c > 0$. We have the classical inequality

$$f_n(a) \leq \inf\{1, \log(\det S_n(a)) - \log(\det S_{n-1}(a))\}. \tag{17}$$

Using (17), Bercu and Duflo [4] have shown that a weighting sequence $a = (a_n)$ such that $a_n = f(n)$ or $a_n = f(\log s_n)$ with $f \in F$ is admissible. More precisely, they have proved that Δ is always almost surely bounded.

Throughout the following, we always assume that the weighting sequence a is admissible.

Remark. It is important to see that the WLS algorithm does not include the ELS as a special case since the weighting sequence with general constant term equal to 1 is not admissible.

4. Strong consistency. We make use of the following traditional assumption of passivity:

(A_1) $C^{-1} - \frac{1}{2}I_{d_1}$ is strictly positive real.

THEOREM 1. *For the model and the* WLS *algorithm* (1) *to* (14), *assume that* (A_1) *is satisfied. Then we have*

$$E\left[\sup_{n\geq 0} \|S_n^{1/2}(a)(\hat{\theta}_{n+1} - \theta)\|^2\right] < +\infty, \tag{18}$$

$$E\left[\sum_{n=0}^{\infty} a_n \|{}^*(\hat{\theta}_{n+1} - \theta)\Phi_n\|^2\right] < +\infty, \tag{19}$$

$$E\left[\sum_{n=0}^{\infty} \lambda_{\min} S_{n-1}(a)\|\hat{\theta}_{n+1} - \hat{\theta}_n\|^2\right] < +\infty, \tag{20}$$

$$E\left[\sum_{n=0}^{\infty} a_n \|\Phi_n - \Psi_n\|^2\right] < +\infty, \tag{21}$$

$$E\left[\sum_{n=0}^{\infty} a_n(1-f_n(a))\|\pi_n\|^2\right] < +\infty. \tag{22}$$

Proof. The proof is given in Appendix A.

COROLLARY 1. If (A_1) *holds, the* WLS *estimator given by* (10) *is strongly consistent on*

$$I = \{\lim_{n\to+\infty} \lambda_{\min} S_n(a) = +\infty\} \tag{23}$$

and, on I, we have

$$\|\hat{\theta}_{n+1} - \theta\|^2 = O(\{\lambda_{\min} S_n(a)\}^{-1}) \quad a.s. \tag{24}$$

Remark. As $S_n(a) \geq S$, the WLS estimator is always almost surely bounded. By (24), we can conclude that the WLS algorithm behaves as well as the ELS for ARMAX parameter estimation [12], [24], [25].

COROLLARY 2. *If* (A_1) *holds, the prediction errors sequence satisfies*

$$E\left[\sum_{n=0}^{\infty} (a_n^{-1} + \|\Phi_n\|^2)^{-1}\|\pi_n\|^2\right] < +\infty. \tag{25}$$

More particularly, if, for $n \geq 0$, $s_n^{-1} \leq c a_n$ with c deterministic > 0, *then*

$$E\left[\sum_{n=0}^{\infty} \frac{\|\pi_n\|^2}{s_n}\right] < +\infty. \tag{26}$$

Finally, we also have

$$E\left[\sum_{n=0}^{\infty} \left(a_n^{-1} + \frac{\|\Phi_n\|^2}{\lambda_{\min} S_{n-1}(a)}\right)^{-1} \|\pi_n\|^2\right] < +\infty. \tag{27}$$

Remark. By (26), we can conclude that the WLS algorithm behaves as well as the SG for adaptive tracking [8], [18].

Proof. $\pi_n = {}^*\theta\Psi_n - {}^*\hat{\theta}_n\Phi_n$, so we can rewrite $\pi_n = \tau_{n+1} + {}^*(\hat{\theta}_{n+1} - \hat{\theta}_n)\Phi_n$ where $\tau_{n+1} = {}^*\theta\Psi_n - {}^*\hat{\theta}_{n+1}\Phi_n$. Then, by use of (19) together with (21),

$$E\left[\sum_{n=0}^{\infty} a_n\|\tau_{n+1}\|^2\right] < +\infty. \tag{28}$$

Hence, (25) and (27) immediately follow from (20) and (28). Moreover, if $s_n^{-1} \leq c a_n$ with c deterministic > 0, (20) and (28) imply (26). □

COROLLARY 3. *Assume that the driven noise ε satisfies the strong law of large numbers* (LN) *with, for $n \geq 0$,*

$$E[\varepsilon_{n+1}{}^*\varepsilon_{n+1} \mid \mathcal{F}_n] = \Gamma, \tag{29}$$

where Γ is a deterministic covariance matrix. To estimate Γ, we propose the two empirical estimators

- $\hat{\Gamma}_n = \frac{1}{n}\sum_{k=1}^{n} (Y_k - {}^*\hat{\theta}_{k-1}\Phi_{k-1})\,{}^*(Y_k - {}^*\hat{\theta}_{k-1}\Phi_{k-1})$,
- $\tilde{\Gamma}_n = \frac{1}{n}\sum_{k=1}^{n} \hat{\varepsilon}_k\,{}^*\hat{\varepsilon}_k$.

Suppose that (A_1) *is satisfied and that $s_n^{-1} = O(a_n)$. Then, on the set $\{s_n = O(n)$ and $s_n \to +\infty\}$, $\hat{\Gamma}_n$ and $\tilde{\Gamma}_n$ areboth strongly consistent estimators of Γ.*

Proof. The proof is obvious using (21) and (26) with Kronecker's lemma. □

5. Adaptive tracking. We still consider the model and the WLS algorithm (1) to (14). The goal of adaptive tracking is to find a control sequence $U = (U_n)$ that forces the output $Y = (Y_n)$ to follow a given reference trajectory $y = (y_n)$. We first use the traditional adaptive tracking control (ATC) introduced by Aström and Wittenmark [1] such that, for $n \geq 0$,

$$y_{n+1} = {}^*\hat{\theta}_n \Phi_n. \tag{30}$$

It is well known that the ATC is almost surely defined if the following assumption is satisfied:

(A_2) For $n \geq 0$, the $\mathcal{F}_n$ conditional distribution of ε_{n+1} is absolutely continuous with respect to the Lebesgue measure.

Remark. If (A_2) is satisfied, Caines [8] has shown how to solve the zero divisor problem for the ATC. We will see in the next section how to avoid this assumption.

Throughout the following, we assume that the driven noise ε has constant conditional covariance matrix Γ given by (29). We also make use of the two classical assumptions about ε:

(N_1) ε has finite conditional moment of order >2;

(N_2) ε is independent and identically distributed with mean 0 and covariance matrix Γ. Therefore, if (N_1) or (N_2) are fulfilled, ε satisfies LN, i.e., if

$$\Gamma_n = \frac{1}{n}\sum_{k=1}^{n} \varepsilon_k {}^*\varepsilon_k, \tag{31}$$

Γ_n converges to Γ almost surely. Finally, we need the following usual assumption of causality:

(A_3) $d_2 \leq d_1$ and the matrix B_1 is of full rank d_2. Moreover, if B_+ denotes the left inverse of B_1 and if $D(R) = B_+ R^{-1} B(R)$ for the shift-back operator R, then D is causal.

Throughout this section, we use a similar approach as that of Bercu and Duflo [4] in the ARX framework. Let $C = (C_n)$ be the average cost matrix sequence defined by

$$C_n = \frac{1}{n}\sum_{k=1}^{n} (Y_k - y_k)^*(Y_k - y_k). \tag{32}$$

The ATC is said to be optimal if C_n converges almost surely to Γ. If we use the ATC, we have from (30)

$$Y_{n+1} - y_{n+1} = \pi_n + \varepsilon_{n+1}. \tag{33}$$

Then, we can easily prove that

$$\|C_n - \Gamma_n\| = O\left(\frac{1}{n}\sum_{k=1}^{n} \|\pi_{k-1}\|^2\right) \quad \text{a.s.} \tag{34}$$

Therefore, in order to show the ATC optimality, we only have to prove that the prediction errors sequence satisfies

$$\sum_{k=1}^{n} \|\pi_k\|^2 = o(n) \quad \text{a.s.} \tag{35}$$

THEOREM 2. *For the model and the* WLS *algorithm* (1) *to* (14), *assume that* (A_1)–(A_3) *and* (N_1) *or* (N_2) *are satisfied. For the tracking trajectory* $y = (y_n)$, *suppose that, for* $n \geq 0$, y_{n+1} *is measurable with respect to* $\mathcal{F}_n$ *and*

$$\sum_{k=1}^{n} \|y_k\|^2 = O(n) \quad \textit{a.s.} \tag{36}$$

If $a_n = f(\log s_n)$ *with* $f \in F$ *and if* $s_n^{-1} = O(a_n)$, *then the* ATC *is optimal. Moreover, we have*

$$\sum_{n=1}^{\infty} \frac{1}{n} \|Y_n - y_n - \varepsilon_n\|^2 < +\infty \quad a.s. \tag{37}$$

Finally, $\hat{\Gamma}_n$ *and* $\tilde{\Gamma}_n$ *are both strongly consistent estimators of* Γ.

Remark. For the SG algorithm, the ATC optimality was established by Goodwin, Ramadge, and Caines [18]. Such a theorem was never proven for the ELS algorithm.

Proof. The proof is given in Appendix B.

We now give a useful excitation transfer (ET) lemma similar to the one established by Lai and Wei [25]. We begin by stating the following assumption of irreducibility which uses the same notation as (A_3):

(A_4) The matrix $B_+ B_q$ is regular and the polynomials of matrices $B_+ A$, $B_+ C$ and D are left coprime.

EXCITATION TRANSFER LEMMA. *Suppose that* (A_3) *and* (A_4) *are satisfied. Then we can find a constant* $M > 0$ *such that, for* $n \geq s$,

$$\lambda_{\min}\left(\sum_{k=0}^{n} \Psi_k{}^* \Psi_k\right) \geq M \lambda_{\min}\left(\sum_{k=s}^{n} H_{k+1}{}^* H_{k+1}\right) \quad a.s., \tag{38}$$

where ${}^t H_n = ({}^t Y_n^{p+s+1}, {}^t \varepsilon_n^{r+s+1})$ *and* $s = d_2(q-1)$.

Proof. A proof can be found in Lai and Wei [25] or Duflo [17].

THEOREM 3. *For the model and* WLS *algorithm* (1) *to* (14), *assume that* (A_1)–(A_4) *and* (N_1) *or* (N_2) *are satisfied. Assume that the covariance matrix* Γ *is regular. For the tracking trajectory* $y = (y_n)$, *suppose that* y_{n+1} *is* $\mathcal{F}_{n-p-s} \cap \mathcal{F}_{n-r-s}$*-measurable with* $\|y_n\|^2 = o(n)$ *and*

$$\sum_{k=1}^{n} \|y_k\|^2 = O(n) \quad a.s. \tag{39}$$

Moreover, suppose that y *is exciting with order* $p+s+1$, *i.e.,*

$$\liminf \lambda_{\min}\left(\frac{1}{n} \sum_{k=p+s}^{n} y_k^{p+s+1*} y_k^{p+s+1}\right) > 0 \quad a.s. \tag{40}$$

If $a_n = f(\log s_n)$ *with* $f \in F$ *and if* $s_n^{-1} = O(a_n^2)$, *then the* ATC *is optimal*

$$\|\Phi_n\|^2 = o(n), \quad f_n(a) = o(1) \quad a.s., \tag{41}$$

$$\|C_n - \Gamma_n\| = o\left(\frac{1}{n f(\log n)}\right) \quad a.s., \tag{42}$$

$$\sum_{n=1}^{\infty} f(\log n) \|Y_n - y_n - \varepsilon_n\|^2 < +\infty \quad a.s. \tag{43}$$

Moreover, the WLS *estimator* $\hat{\theta}_n$ *converges almost surely to* θ *and we obtain*

$$\|\hat{\theta}_{n+1} - \theta\|^2 = O\left(\frac{1}{n f(\log n)}\right) \quad a.s. \tag{44}$$

Finally, $\hat{\Gamma}_n$ and $\tilde{\Gamma}_n$ are both strongly consistent estimators of Γ and we find relations similar to (42) *with $\hat{\Gamma}_n$ or $\tilde{\Gamma}_n$ instead of C_n.*

Proof. The proof is given in Appendix C.

Remark. We can prove the ATC optimality and the strong consistency with a condition less restrictive than (40) for the tracking trajectory. More precisely, let $\lambda = (\lambda_n)$ be a deterministic positive sequence, increasing to infinity, such that (λ_n) has the same behavior as (λ_{n-1}) and $\lambda_n = O(n)$. Assume that y is λ-exciting with order $p + s + 1$, i.e.,

$$\liminf \lambda_{\min}\left(\frac{1}{\lambda_n}\sum_{k=p+s}^{n} y_k^{p+s+1*} y_k^{p+s+1}\right) > 0 \quad \text{a.s.} \tag{45}$$

Then, if $\lambda_n^{-1} = O(a_n^2)$, the ATC is optimal and the WLS estimator is strongly consistent with

$$\|\hat{\theta}_{n+1} - \theta\|^2 = O\left(\frac{1}{\lambda_n f(\log n)}\right) \quad \text{a.s.} \tag{46}$$

We next consider the continually disturbed control (CDC) introduced by Caines [6], [8] such that, for $n \geq 0$,

$$y_{n+1} + \xi_{n+1} = {}^*\hat{\theta}_n \Phi_n, \tag{47}$$

where ξ is a d_1-dimensional exogenous noise, adapted to $\mathbf{F}$, with mean 0 and covariance matrix Λ. The CDC is said to be residually optimal if C_n converges almost surely to $\Gamma + \Lambda$.

THEOREM 4. *For the model and* WLS *algorithm* (1) *to* (14), *take the same assumptions as in Theorem* 3 *except condition* (40) *for the tracking trajectory y. Moreover, for the exogenous noise ξ, assume that the* LN *is satisfied with Λ regular. In addition, assume that ξ is independent of ε, of y, and of the initial state Ψ_0. If $a_n = f(\log s_n)$ with $f \in F$ and if $s_n^{-1} = O(a_n^2)$, then the* CDC *excited by ξ is residually optimal*

$$\|\Phi_n\|^2 = o(n), \quad f_n(a) = o(1) \quad a.s., \tag{48}$$

$$\sum_{n=1}^{\infty} f(\log n)\|Y_n - y_n - \varepsilon_n - \xi_n\|^2 < +\infty \quad a.s. \tag{49}$$

Moreover, the WLS *estimator $\hat{\theta}_n$ converges almost surely to θ and we obtain*

$$\|\hat{\theta}_{n+1} - \theta\|^2 = O\left(\frac{1}{n f(\log n)}\right) \quad a.s. \tag{50}$$

Proof. The proof is similar to that of Theorem 3.

6. Modified adaptive tracking. We now use a similar approach as that of Guo and Chen [19], [15] in the ELS framework. Assume that (A_1) and (A_3) are satisfied. Without assumption (A_2), to avoid the zero divisor problem with the ATC, we propose a modified WLS estimator. Throughout the following, the major restriction is that the noise is supposed to satisfy (N_1). All the results of this section are also true, without modification, if we assume that (A_2) is satisfied.

Let $\hat{B}_n^1$ be the matrix component of $\hat{\theta}_n$ that estimates B_1. P_n and Q_n are the orthogonal matrices associated with the singular value decomposition of $\hat{B}_n^1$. The columns of P_n are eigenvectors of $\hat{B}_n^{1*}\hat{B}_n^1$ and the columns of Q_n are eigenvectors of ${}^*\hat{B}_n^1\hat{B}_n^1$ [20]. We set

$$\tilde{B}_n^1 = \begin{cases} \hat{B}_n^1 & \text{if } \lambda_{\min}({}^*\hat{B}_n^1\hat{B}_n^1) > 0, \\ \hat{B}_n^1 + \sqrt{\nu_n} P_n {}^*Q_n & \text{otherwise,} \end{cases} \tag{51}$$

for any positive, deterministic and summable sequence $\nu = (\nu_n)$. By (51), $\tilde{B}_n^1$ is clearly of full rank d_2. Denote by $\tilde{\theta}_n$ the modified WLS estimator of θ where $\hat{B}_n^1$ is replaced by $\tilde{B}_n^1$ in $\hat{\theta}_n$. It immediately follows from (51) that

$$\|\hat{\theta}_n - \tilde{\theta}_n\|^2 \le \nu_n \quad \text{a.s.} \tag{52}$$

Hence, since $\nu_n = o(1)$, the WLS algorithm is not modified for parameter estimation. We first consider the modified ATC such that, for $n \ge 0$,

$$y_{n+1} = {}^*\tilde{\theta}_n \Phi_n. \tag{53}$$

We clearly have from (53)

$$Y_{n+1} - y_{n+1} = \tilde{\pi}_n + \varepsilon_{n+1}, \tag{54}$$

where

$$\tilde{\pi}_n = \pi_n + (\hat{B}_n^1 - \tilde{B}_n^1)U_n. \tag{55}$$

Therefore, the modified ATC is optimal if

$$\sum_{k=1}^{n} \|\tilde{\pi}_k\|^2 = o(n) \quad \text{a.s} \tag{56}$$

THEOREM 5. *For the model and the* WLS *algorithm* (1) *to* (14), *assume that* (A_1), (A_3), *and* (N_1) *are satisfied. For a positive, nonincreasing and deterministic sequence* $\alpha = (\alpha_n)$ *such that* $\alpha_n = O(n)$, *assume that we have* $\|\varepsilon_n\|^2 = O(\alpha_n)$. *For the tracking trajectory* $y = (y_n)$, *suppose that, for* $n \ge 0$, y_{n+1} *is* $\mathcal{F}_n$*-measurable with* $\|y_n\|^2 = O(\alpha_n)$ *and*

$$\sum_{k=1}^{n} \|y_k\|^2 = O(n) \quad a.s. \tag{57}$$

If $a_n = f(\log s_n)$ *with* $f \in F$ *and if* $s_n^{-1} = O(a_n)$, *then the modified* ATC *is optimal. Moreover, consider the positive random sequence* $v = (v_n)$ *such that* $v_n = \alpha_n + a_n^{-1}$. *Then we also have*

$$\|\Phi_n\|^2 = O(v_{n+1}) \quad a.s., \tag{58}$$

$$\|C_n - \Gamma_n\| = o\left(\frac{v_n}{n}\right) \quad a.s., \tag{59}$$

$$\sum_{n=1}^{\infty} \frac{1}{v_n} \|Y_n - y_n - \varepsilon_n\|^2 < +\infty \quad a.s. \tag{60}$$

Finally, $\hat{\Gamma}_n$ *and* $\tilde{\Gamma}_n$ *are both strongly consistent estimators of* Γ *and we find relations similar to* (59) *with* $\hat{\Gamma}_n$ *or* $\tilde{\Gamma}_n$ *instead of* C_n.

Proof. The proof is given in Appendix D.

Concerning adaptive tracking, we now give the last but most important theorem of this paper. It ensures both strong consistency for the WLS estimator and modified continually disturbed control (CDC) optimality. We recall here that the following theorem is also true, without modification, if we assume that (A_2) is satisfied. Before stating it, we assume in (51) that the sequence $(n\nu_n)$ is summable.

Consider a deterministic positive sequence $\lambda = (\lambda_n)$, increasing to infinity, such that, for $n \geq 1$, $\lambda_n - \lambda_{n-1} \leq 1$, (λ_n) has the same behavior as (λ_{n-1}) and $\lambda_n = O(n)$. Let ξ be a d_1-dimensional exogenous noise adapted to **F** with mean 0 and covariance matrix Λ. Set, for $n \geq 1$, $\chi_n = \sqrt{\lambda_n - \lambda_{n-1}}\xi_n$. We use the CDC introduced by Bercu and Duflo [4] such that, for $n \geq 0$,

$$y_{n+1} + \chi_{n+1} =^* \tilde{\theta}_n \Phi_n. \tag{61}$$

It follows immediately from (61) that

$$Y_{n+1} - y_{n+1} - \chi_{n+1} = \tilde{\pi}_n + \varepsilon_{n+1}. \tag{62}$$

THEOREM 6. *For the model and the* WLS *algorithm* (1) *to* (14), *assume that* (A_1), (A_3), (A_4), *and* (N_1) *are satisfied and that* Γ *is regular. For a positive, nonincreasing and deterministic sequence* $\alpha = (\alpha_n)$ *such that* $\alpha_n = O(n)$, *assume that* $\|\varepsilon_n\|^2 = O(\alpha_n)$. *For the tracking trajectory* $y = (y_n)$, *suppose that* y_{n+1} *is* $\mathcal{F}_{n-p-s} \cap \mathcal{F}_{n-r-s}$*-measurable with* $\|y_n\|^2 = O(\alpha_n)$ *and*

$$\sum_{k=1}^{n} \|y_k\|^2 = O(n) \quad a.s. \tag{63}$$

Moreover, assume that the exogenous noise ξ *satisfies* (N_1) *with* Λ *regular. In addition, suppose that* ξ *is independent of* ε, *of the initial state* Ψ_0, *and of the tracking trajectory* y, *and that* $\|\xi_n\|^2 = O(\alpha_n)$. *Consider the positive random sequence* $v = (v_n)$ *such that* $v_n = \alpha_n + a_n^{-1}$. *Assume that* $v_n = o(\lambda_n)$ *and* $\lambda_n^{-1} = O(a_n^2)$. *If* $a_n = f(\log s_n)$ *with* $f \in F$ *and if* $s_n^{-1} = O(a_n^2)$, *then*

$$\|\Phi_n\|^2 = o(\lambda_n), \quad f_n(a) = o(1) \quad a.s., \tag{64}$$

$$\sum_{n=1}^{\infty} f(\log n)\|Y_n - y_n - \chi_n - \varepsilon_n\|^2 < +\infty \quad a.s., \tag{65}$$

$$\frac{1}{\lambda_n}\sum_{k=1}^{n}(Y_k - y_k - \varepsilon_k)^*(Y_k - y_k - \varepsilon_k) \to \Lambda \quad a.s., \tag{66}$$

$$\|C_n - \Gamma_n\| = O\left(\frac{\lambda_n}{n}\right) \quad a.s. \tag{67}$$

Therefore, if $\lambda_n = o(n)$, *the modified* CDC *excited by* χ *is optimal. Moreover, the* WLS *estimator* $\hat{\theta}_n$ *converges almost surely to* θ *and*

$$\|\hat{\theta}_{n+1} - \theta\|^2 = O\left(\frac{1}{\lambda_n f(\log n)}\right) \quad a.s. \tag{68}$$

Finally, $\hat{\Gamma}_n$ *and* $\tilde{\Gamma}_n$ *are both strongly consistent estimators of* Γ *and we find relations similar to* (67) *with* $\hat{\Gamma}_n$ *or* $\tilde{\Gamma}_n$ *instead of* C_n.

Proof. The proof is given in Appendix E.

Remark. If assumption (N_1) is satisfied with $\alpha > 2$, then, by use of the conditional Borel–Cantelli lemma, we can take $\alpha_n = n^\beta$ with $2\alpha^{-1} < \beta < 1$. We can also choose $a_n = (\log s_n)^{-1-\gamma}$ with $\gamma > 0$. Therefore, if we take $\lambda_n = n^\delta$ with $\beta < \delta < 1$, we obtain the convergence rates $n^{-\delta}(\log n)^{1+\gamma}$ for the strong consistency and $n^{\delta-1}$ for the optimality. In addition, if ε and ξ are Gaussian white noises, then, using again the Borel – Cantelli lemma, we can take $\alpha_n = \log n$. On one hand, if we focus our attention on the strong consistency, we can use the same choice as above and find the convergence rate $n^{-\delta}(\log n)^{1+\gamma}$. On the other hand, if we are interested in the optimality, we can take $\lambda_n = (\log n)^\delta$ with $2(1+\gamma) \leq \delta$ and we obtain the convergence rate $n^{-1}(\log n)^\delta$. One can realize that the attenuation $\lambda = (\lambda_n)$ plays a prominent part, reducing the role of the weighting sequence $a = (a_n)$.

7. Survey on adaptive tracking. We now give a short survey on earlier related works on adaptive tracking. We complete this section by comparing our work with previous similar results.

Concerning the SG algorithm, Goodwin, Ramadge, and Caines [18] proved global convergence and adaptive tracking control (ATC) optimality. In the scalar tracking problem, Becker, Kumar, and Wei [2] established convergence to a random multiple of the parameter to be estimated. If the tracking trajectory is sufficiently rich, Kumar and Praly [21] showed, in the scalar case, strong consistency and ATC optimality. Caines [6], [8] realized that, in order to enforce strong consistency, it is necessary to modify the ATC of Aström and Wittenmark [1] and he introduced the CDC. In the scalar case, Caines and Lafortune [7] obtained the first results of CDC optimality and persistent excitation. For the same purpose, Chen [9], [11] chose a weak hypothesis of excitation and using this assumption, Chen and Caines [10] proved, in the scalar case, strong consistency and CDC residual optimality. In a multidimensional framework and with a restrictive assumption on the noise, Chen and Guo [13] established both strong consistency and CDC optimality.

Concerning the ELS algorithm, Solo [27] gave, in the scalar case, a persistent excitation condition in order to guarantee strong consistency. Lai and Wei [23], [24] proposed a weaker excitation condition to obtain strong consistency. For bounded noise, they used a rather complicated control to obtain both strong consistency and CDC optimality. Under the same condition but in a multidimensional framework, Lai and Wei [25] showed strong consistency and gave an excitation transfer theorem useful in obtaining persistent excitation results. Analogously, in a multidimensional framework, Chen and Guo [12] gave conditions to obtain strong consistency. Then they used a rather complex control to prove both strong consistency and CDC optimality with almost sure rates of convergence. Chen and Zhang [14] established similar results in the multi-delay case. Recently Kumar [22] showed, for white Gaussian noise and in a regression framework, the existence of an almost sure limit for the least squares estimator, for almost all parameter values. Strong consistency and optimality results followed. Sin and Goodwin [26] introduced the modified least squares (MLS) algorithm and obtained results similar to those of Goodwin, Ramadge, and Caines [18]. Chen [9], [11] also introduced an algorithm similar to the MLS and in a multidimensional framework he proved strong consistency and CDC residual optimality.

Recently, Guo and Chen [19], [15] established the most important result concerning adaptive tracking for ARMAX models. They found a solution to the twenty-year-old adaptive tracking problem proposed by Aström and Wittenmark [1]. In a multidimensional framework, they proved both strong consistency of ELS estimator and CDC optimality. They also provided almost sure rates of convergence. The key idea was an over-estimation of the ARMAX regression vector norm.

With the WLS algorithm, we have also given a solution to the Aström and Wittenmark [1] adaptive tracking problem. We now compare our work to the results of Guo and Chen [19].

- One can remark that the WLS algorithm is similar to the ELS. The main difference is the easy introduction of a random weighting sequence $a = (a_n)$ in relation (10).
- To obtain strong consistency results, Guo and Chen [19] always required that the driven noise ε had finite conditional moment of order > 2. In §4, we showed how to avoid this assumption by the choice of an admissible weighting sequence. Moreover, it is easy to see via (24) that the WLS estimator performs as well as that of the ELS for ARMAX parameter estimation.
- Furthermore, to obtain adaptive tracking results, Guo and Chen [19] proposed a modified ELS estimator. One can realize that they established CDC optimality by use of a rather technical procedure. Our modification (51) is really simple. Moreover, via (26), we can easily prove the CDC optimality. In addition, our results are also true without modification if the continuity assumption (A_2) on the distribution of ε is satisfied. One can remark that such a result has not been proved by Guo and Chen [19] with pure ELS estimator.
- Finally, we have shown that the WLS algorithm is really easy to handle. We can choose the weighting sequence or the attenuation as we want to privilege the strong consistency or the optimality. One can also realize that our convergence rates are more precise. For example, suppose that we focus our attention on the control optimality. If ε is a Gaussian white noise, we can take the attenuation $\lambda_n = (\log n)^4$. Then, we obtain from (66) a convergence rate in power of $\log n$. It improves the result of Guo and Chen [19] as they founded a convergence rate in power of n.

8. Conclusion. Finally, as it was done for the ELS algorithm, we have shown that the WLS algorithm has rather attractive properties. Under classical assumptions, we have proved both strong consistency of the WLS estimator and CDC optimality. We have also established almost sure rates of convergence. We can easily guess that the weighted estimation can be used in many other frameworks. For instance, the adaptive tracking problems for linear ARMAX models with time varying parameters or for functional ARMAX models remain to be studied, following the choice of suitable weighting sequences.

Appendix A. We make use of the following two lemmas.

LEMMA 1. *Set* $f_n(a) = a_n {}^*\Phi_n S_n^{-1}(a)\Phi_n$ *and* $g_n(a) = a_n {}^*\Phi_n S_{n-1}^{-1}(a)\Phi_n$. *Then*

$$(1 - f_n(a)) = (1 + g_n(a))^{-1} \quad \textit{so} \quad 0 \le f_n(a) \le 1;$$

$$\hat{\varepsilon}_{n+1} = (1 - f_n(a))(\pi_n + \varepsilon_{n+1}), \qquad \hat{\theta}_{n+1} = \hat{\theta}_n + a_n S_{n-1}^{-1}(a)\Phi_n {}^*\hat{\varepsilon}_{n+1};$$

$$f_n(a) \le \inf\{1, \log(\det S_n(a)) - \log(\det S_{n-1}(a))\}.$$

LEMMA 2. *Assume that* ε *is a martingale difference sequence satisfying* (5). *For a vectorial random sequence* $\varphi = (\varphi_n)$ *adapted to* $\mathbf{F}$, *set*

$$M_{n+1} = \sum_{k=0}^{n*} \varphi_k \varepsilon_{k+1}.$$

Then we always have

$$E\left[\sup_{k \le n+1} |M_k|^2\right] = O\left(E\left[\sum_{k=0}^{n} \|\varphi_k\|^2\right]\right).$$

We can easily prove Lemma 1 using the same arguments as without the weighting sequence $a = (a_n)$. Lemma 2 can be established by the use of a stopping time argument together with the Kolmogorov's inequality [17]. It can also be proved via the Burkholder, Davis, and Gundy inequality [16], [28].

Proof of Theorem 1. For $n \geq 0$, set $\check{\theta}_n = \hat{\theta}_n - \theta, \check{\varepsilon}_n = \hat{\varepsilon}_n - \varepsilon_n$, and $v_n = \text{tr}({}^*\check{\theta}_n S_{n-1}(a)\check{\theta}_n)$, where $S_{-1}(a) = S$. By Lemma 1, we can find the following relation similar, without the weighting sequence $a = (a_n)$, to the well-known equality due to Caines [8], Chen [11], Duflo [17], Guo and Chen [11], [15], or Lai and Wei [25]:

$$v_{n+1} + P_{n+1} = v_0 + \sum_{k=0}^{n} a_k \|\alpha_{k+1}\|^2 + 2\sum_{k=0}^{n} a_n f_k(a)\|\varepsilon_{k+1}\|^2 + 2\text{Re}(M_{n+1}) - 2\text{Re}(L_{n+1}), \quad \text{(A.1)}$$

with $\alpha_n = -{}^*\check{\theta}_n \Phi_{n-1}$; $\beta_n = {}^*\check{\theta}_n \Phi_n + f_n(a)\pi_n$; and

- $P_{n+1} = \sum_{k=0}^{n} a_k f_k(a)(1 - f_k(a))\|\pi_k + \varepsilon_{k+1}\|^2$,
- $M_{n+1} = \sum_{k=0}^{n} a_k {}^*\beta_k \varepsilon_{k+1}$,
- $L_{n+1} = \sum_{k=0}^{n} a_k {}^*\alpha_{k+1} \check{\varepsilon}_{k+1}$.

Moreover, since (A_1) is satisfied, $\alpha_n = C(R)\hat{\varepsilon}_n$ and $a = (a_n)$ is positive and nonincreasing, we can find a positive constant l and an integrable random variable L such that

$$2\text{Re}(L_{n+1}) + L \geq (1 + l)\sum_{k=0}^{n} a_k \|\alpha_{k+1}\|^2. \quad \text{(A.2)}$$

In addition, it immediately follows from Lemma 2 that

$$E\left[\sup_{k \leq n+1} |M_k|^2\right] = O\left(E\left[\sum_{k=0}^{n} a_k^2 \|\beta_k\|^2\right]\right). \quad \text{(A.3)}$$

Therefore, recalling that $\beta_n = -\alpha_{n+1} - f_n(a)\varepsilon_{n+1}$, we obtain that either

$$E\left[\sup_{n \geq 1} |M_n|\right] < +\infty \quad \text{(A.4)}$$

or

$$E\left[\sup_{k \leq n+1} |M_k|\right] = o\left(E\left[\sum_{k=0}^{n} a_k(\|\alpha_{k+1}\|^2 + f_k(a)\|\varepsilon_{k+1}\|^2)\right]\right). \quad \text{(A.5)}$$

Finally, by (A.1) and (A.2), we find that

$$E\left[\sup_{k \leq n} v_{k+1}\right] = O\left(E\left[\sum_{k=0}^{n} a_k f_k(a)\|\varepsilon_{k+1}\|^2\right]\right). \quad \text{(A.6)}$$

Now, from (5) and (15) together with the monotone convergence theorem,

$$E\left[\sum_{n=0}^{\infty} a_n f_n(a)\|\varepsilon_{n+1}\|^2\right] < +\infty. \quad \text{(A.7)}$$

Then we obtain (18) from (A.6) and (A.7). Next, we also obtain (19) and (21) from (A.1), (A.2), (A.7) and the passivity assumption. It remains to show (20) and (22). Let $\pi_n = {}^*\theta\Psi_n - {}^*\hat{\theta}_n\Phi_n$

be the prediction error at time n. By use of Lemma 1, we have $(1 - f_n(a))\pi_n = \check{\varepsilon}_{n+1} + f_n(a)\varepsilon_{n+1}$. Hence, (21) and (A.7) imply

$$E\left[\sum_{n=0}^{\infty} a_n(1 - f_n(a))^2\|\pi_n\|^2\right] < +\infty. \tag{A.8}$$

Recalling (A.1), we also find that

$$E\left[\sup_{n\geq 1} P_n\right] < +\infty, \tag{A.9}$$

$$E\left[\sum_{n=0}^{\infty} a_n f_n(a)(1 - f_n(a))\|\pi_n\|^2\right] < +\infty, \tag{A.10}$$

and we clearly deduce (22) from (A.8) and (A.10). Furthermore, by the matrix inversion formula of Riccati, we obtain

$$^*\Phi_n S_n^{-2}(a)\Phi_n = (1 - f_n(a))^{2*}\Phi_n S_{n-1}^{-2}(a)\Phi_n. \tag{A.11}$$

So if we set $d = d_1 p + d_2 q + d_1 r$, we obtain, by Lemma 1,

$$a_n \lambda_{\min} S_{n-1}(a)^*\Phi_n S_n^{-2}(a)\Phi_n \leq d f_n(a)(1 - f_n(a)). \tag{A.12}$$

Finally, (10) and (14), together with (A.12) and (22), imply (20), completing the proof of Theorem 1. □

Appendix B.

Proof of Theorem 2. Denote by s_∞ and a_∞ the limits of the sequences $s = (s_n)$ and $a = (a_n)$, respectively. To use relation (26) together with Kronecker's lemma, we first have to show that $s_\infty = +\infty$. If we assume that $s_\infty < +\infty$, it follows from the assumption $s_n^{-1} = O(a_n)$ that necessarily $a_\infty > 0$. Hence, using (21), we have

$$\sum_{n=0}^{\infty} \|\Phi_n - \Psi_n\|^2 < \infty. \tag{B.1}$$

If we set $q_n = \sum_{k=0}^{n} \|\Psi_k\|^2 + s$, we can easily see that

$$q_n \leq 2\sum_{k=0}^{n} \|\Phi_k - \Psi_k\|^2 + 2s_n, \tag{B.2}$$

so $q_\infty < +\infty$ where q_∞ denotes the limit of the sequence $q = (q_n)$. But we also have from (6) that

$$q_n \geq \sum_{k=0}^{n} \|\varepsilon_k\|^2 \tag{B.3}$$

and as ε satisfies the LN, we get $n = O(q_n)$. Finally, we lead to a contradiction so that $s_\infty = +\infty$, $a_\infty = 0$. Moreover, we also have

$$s_n \leq 2\sum_{k=0}^{n} \|\Phi_k - \Psi_k\|^2 + 2q_n. \tag{B.4}$$

Then, by use of (21) together with Kronecker's lemma, we find that

$$\sum_{k=0}^{n} \|\Phi_k - \Psi_k\|^2 = o(a_n^{-1}) \tag{B.5}$$

and, as $a_n^{-1} = O(s_n)$, we obtain $s_n = O(q_n)$. In addition, from (6), we have

$$q_n = O\left(\sum_{k=0}^{n} \|Y_k\|^2 + \sum_{k=0}^{n} \|U_k\|^2 + \sum_{k=0}^{n} \|\varepsilon_k\|^2\right). \tag{B.6}$$

Hence, assumptions (N_1) or (N_2) imply that

$$q_n = O\left(n + \sum_{k=0}^{n} \|Y_k\|^2 + \sum_{k=0}^{n} \|U_k\|^2\right). \tag{B.7}$$

Recalling (1), we have $U_{n-1} = D^{-1}(R)B_+A(R)Y_n - D^{-1}(R)B_+C(R)\varepsilon_n$, where R is the shift-back operator. Then, using (B.7), we find that

$$q_{n-1} = O\left(n + \sum_{k=1}^{n} \|Y_k\|^2\right) \tag{B.8}$$

and so, as $s_n = O(q_n)$, we prove that

$$s_{n-1} = O\left(n + \sum_{k=1}^{n} \|Y_k\|^2\right). \tag{B.9}$$

By (26) and the assumption (36) for the tracking trajectory, we have

$$\sum_{k=1}^{n} \|Y_k\|^2 = o(s_{n-1}) + O(n). \tag{B.10}$$

Finally, using (B.9), we obtain that

$$\sum_{k=1}^{n} \|Y_k\|^2 = O(n), \qquad s_n = O(n). \tag{B.11}$$

From (26) together with Kronecker's lemma, we conclude that the ATC is optimal. Moreover, (26) immediately implies (37). We complete the proof using Corollary 3. □

Appendix C.

Proof of Theorem 3. Using the same ideas developed by Bercu and Duflo [4] in the ARX framework, we now prove the Theorem 3. We have already seen in Theorem 2 the ATC optimality with

$$\sum_{k=1}^{n} \|\pi_k\|^2 = o(n). \tag{C.1}$$

Set, for $n \geq 0$, ${}^tL_r = ({}^ty_n^{p+s+1} + {}^t\varepsilon_n^{p+s+1}, {}^t\varepsilon_n^{r+s+1})$. Using the notation of the ET Lemma, we obtain from (33),

$$\|H_{n+1} - L_{n+1}\|^2 = \sum_{k=1}^{p+s+1} \|\pi_{n-k+1}\|^2. \tag{C.2}$$

Then it follows from (C.1) and (C.2) that

$$\sum_{k=0}^{n} \|H_{k+1} - L_{k+1}\|^2 = o(n). \tag{C.3}$$

Moreover, as y is strongly exciting with order $p + s + 1$, we have

$$\liminf \lambda_{\min} \left(\frac{1}{n} \sum_{k=0}^{n} y_{k+1}^{p+s+1*} y_{k+1}^{p+s+1} \right) > 0. \tag{C.4}$$

Furthermore, Γ is regular. Hence, if we assume that y_{n+1} is $\mathcal{F}_{n-p-s} \cap \mathcal{F}_{n-r-s}$-measurable, we find, by use of a classical excitation transfer property proved by Duflo [17] or Lai and Wei [25], that

$$\liminf \lambda_{\min} \left(\frac{1}{n} \sum_{k=0}^{n} L_{k+1}{}^* L_{k+1} \right) > 0. \tag{C.5}$$

Then (C.3) together with (C.5) imply

$$\liminf \lambda_{\min} \left(\frac{1}{n} \sum_{k=0}^{n} H_{k+1}{}^* H_{k+1} \right) > 0. \tag{C.6}$$

Now, if we set

$$Q_n = \sum_{k=0}^{n} \Psi_k{}^* \Psi_k + Q, \qquad a_n Q_n \leq Q_n(a) \leq Q_n. \tag{C.7}$$

Hence, by use of the ET Lemma, we find that $n = O(\lambda_{\min} Q_n)$, which implies $na_n = O(\lambda_{\min} Q_n(a))$. In addition, we have already proved that $s_n = O(n)$. Consequently, from the assumption $(a_n s_n)^{-1} = O(a_n)$, we find that $\lambda_{\min} Q_n(a) \to +\infty$. Next, $q_n = O(n)$, so $\log(\lambda_{\max} Q_n(a)) = O(\log(n))$ and $\log(\lambda_{\max} Q_n(a)) = o(\lambda_{\min} Q_n(a))$. Therefore, via a well-known transfer property, we can conclude that $na_n = O(\lambda_{\min} S_n(a))$. Finally, the WLS estimator is strongly consistent and from (24) we obtain the convergence rate given in (44). Moreover, by use of (27), we also find that

$$\sum_{n=0}^{\infty} a_n \|\pi_n\|^2 < +\infty. \tag{C.8}$$

Hence, we obtain, from Kronecker's lemma,

$$\sum_{k=1}^{n} \|\pi_k\|^2 = o(a_n^{-1}). \tag{C.9}$$

Finally, from (34) and (C.9), we obtain the convergence rate given in (42). In addition, we immediately obtain (43) from (C.8). To complete the proof of Theorem 3, we now show that $\|\Phi_n\|^2 = o(n)$. It will clearly lead to $f_n(a) = o(1)_-$. From (C.8), we have $\|\pi_n\|^2 = o(n)$. Then, (33) together with the assumption $\|y_n\|^2 = o(n)$ imply $\|Y_{n+1}\|^2 = o(n)$. Recalling (1) and the causality assumption (A_3), we have

$$U_n = D^{-1}(R) B_+ A(R) Y_{n+1} - D^{-1}(R) B_+ C(R) \varepsilon_{n+1}, \tag{C.10}$$

where R is the shift-back operator. Hence, we can see that $\|U_n\|^2 = o(n)$. Finally $\|\Psi_n\|^2 = o(n)$ and from (21), $\|\Phi_n\|^2 = o(n)$, completing the proof of Theorem 3. □

Appendix D.

Proof of Theorem 5. From the causality assumption (A_3), Caines [8] or Guo and Chen [19] proved that we can find a positive constant $\lambda < 1$ such that

$$\|U_{n-1}\|^2 = O(F_n + \alpha_n), \tag{D.1}$$

where

$$F_n = \sum_{k=0}^{n} \lambda^{n-k} \|Y_k\|^2. \tag{D.2}$$

Furthermore, we have from (25) that

$$\|\pi_n\|^2 = o(a_n^{-1} + \|\Phi_n\|^2). \tag{D.3}$$

In addition, (54) and (55) together with (D.1) imply

$$\|Y_{n+1}\|^2 \leq O(\alpha_{n+1}) + O(\|\pi_n\|^2) + \nu_n O(F_{n+1}). \tag{D.4}$$

Therefore, it follows from (21) and (D.1) – (D.4) that

$$\|Y_{n+1}\|^2 \leq O(v_{n+1}) + o(F_{n+1}), \tag{D.5}$$

where $v_n = \alpha_n + a_n^{-1}$. Moreover, as $F_{n+1} = \lambda F_n + \|Y_{n+1}\|^2$, we obtain

$$F_{n+1} \leq \mu F_n + O(v_{n+1}) \tag{D.6}$$

for some positive constant $\mu < 1$. Finally $F_n = O(v_n)$ and we obtain that

$$\|\Phi_n\|^2 = O(v_{n+1}). \tag{D.7}$$

Recalling (55), we also have

$$\|\tilde{\pi}_n\|^2 \leq 2\|\pi_n\|^2 + 2\nu_n \|\Phi_n\|^2. \tag{D.8}$$

Hence, by use of (26) and (D.8), since $\nu = (\nu_n)$ is summable, we find that

$$\sum_{n=0}^{\infty} \frac{\|\tilde{\pi}_n\|^2}{s_n} < +\infty. \tag{D.9}$$

Therefore, as in the proof of Theorem 2, it follows from (D.9) that $s_n = O(n)$. Finally, (56) and (D.9) imply the modified ATC optimality. In addition, it immediately follows from (D.7) that $a_n^{-1} + \|\Phi_n\|^2 = O(v_{n+1})$. Then, we establish from (25) that

$$\sum_{n=0}^{\infty} \frac{\|\tilde{\pi}_n\|^2}{v_{n+1}} < +\infty. \tag{D.10}$$

Finally, (54) and (D.10) together with Kronecker's lemma imply (59) and (60), completing the proof of Theorem 5. □

Appendix E.

Proof of Theorem 6. We prove Theorem 6 using the same approach as Bercu and Duflo [4] in the ARX framework. The exogenous noise ξ satisfies assumption (N_1). Then, as $\lambda_n - \lambda_{n-1} \leq 1$, we obtain, by use of Chow's lemma,

$$\sum_{n=1}^{\infty} \lambda_n^{-1}(\lambda_n - \lambda_{n-1})(\xi_n {}^*\xi_n - \Lambda) < +\infty. \tag{E.1}$$

But $\chi_n = \sqrt{\lambda_n - \lambda_{n-1}}\xi_n$, so (E.1) implies immediately that

$$\sum_{n=1}^{\infty} \lambda_n^{-1}(\chi_n {}^*\chi_n - (\lambda_n - \lambda_{n-1})\Lambda) < +\infty. \tag{E.2}$$

Then, as $\lambda = (\lambda_n)$ increases to infinity, we obtain, by Kronecker's lemma,

$$\frac{1}{\lambda_n}\sum_{k=1}^{n} \chi_k {}^*\chi_k \to \Lambda. \tag{E.3}$$

Since Λ is regular, we immediately obtain from (E.3) that

$$\liminf \lambda_{\min}\left(\frac{1}{\lambda_n}\sum_{k=1}^{n} \chi_k {}^*\chi_k\right) > 0. \tag{E.4}$$

Moreover, as $\|\xi_n\|^2 = O(\alpha_n)$, we also have $\|\chi_n\|^2 = O(\alpha_n)$. Consequently, by use of (62) together with the proof of Theorem 5, we obtain the first relation of (64). Therefore, since $v_n = o(\lambda_n)$, we obtain, from (25),

$$\sum_{k=1}^{n} \|\tilde{\pi}_k\|^2 = o(\lambda_n). \tag{E.5}$$

Set, for $n \geq 0$, ${}^tL_n = ({}^ty_n^{p+s+1} + {}^t\varepsilon_n^{p+s+1} + {}^t\chi_n^{p+s+1},\ {}^t\varepsilon_n^{r+s+1})$. Using the notation of the ET Lemma, we obtain, from (62),

$$\|H_{n+1} - L_{n+1}\|^2 = \sum_{k=1}^{p+s+1} \|\tilde{\pi}_{n-k+1}\|^2. \tag{E.6}$$

Then, from (E.5) and (E.6), we obtain

$$\sum_{k=0}^{n} \|H_k - L_k\|^2 = o(\lambda_n). \tag{E.7}$$

In addition, by (E.4), we also have

$$\liminf \lambda_{\min}\left(\frac{1}{\lambda_n}\sum_{k=1}^{n} L_k {}^*L_k\right) > 0. \tag{E.8}$$

Finally, (E.7), (E.8), and the ET Lemma imply $\lambda_n = O(\lambda_{\min}Q_n)$. Therefore, as in the proof of Theorem 3, the assumption $(a_n\lambda_n)^{-1} = O(a_n)$ implies $\lambda_n a_n = O(\lambda_{\min}S_n(a))$. Hence, we clearly obtain the second relation of (64). Moreover (68) follows immediately from (24). Recalling (27), (55), and (64), as $v_n = o(\lambda_n)$ and the sequence $(n\nu_n)$ is summable, we also find that

$$\sum_{n=0}^{\infty} a_n\|\tilde{\pi}_n\|^2 < +\infty. \tag{E.9}$$

Then we clearly obtain (65) from (E.9). Finally, we obtain (66) and (67) from (62), (E.3), and (E.5), completing the proof of Theorem 6. □

Acknowledgment. The author expresses his gratitude to Prof. Marie Duflo. This work was completed thanks to her constant attention.

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. Henri Poincare, 27 (1991), pp. 425–443.
[4] B. Bercu and M. Duflo, *Moindres carrés pondérés et poursuite*, Ann. Inst. Henri Poincare, 28 (1992), pp. 403–430.
[5] B. Bercu, *Estimation pondérée et poursuite pour les modèles* ARMAX, Note au CRAS, Série 1, 314 (1992), pp. 403–406.
[6] P. E. Caines, *Stochastic adaptive control: Randomly varying parameters and continually disturbed controls*, in Control Science and Technology for the Progress of Society, H. Akashi, ed., Pergamon Press, Oxford, 1981, pp. 925–930.
[7] P. E. Caines and S. Lafortune, *Adaptive control with recursive identification for stochastic linear systems*, IEEE Trans. Automat. Control, AC-29 (1984), pp. 312–321.
[8] P. E. Caines, *Linear Stochastic Systems*, John Wiley, New York, 1988.
[9] H. F. Chen, *Recursive system identification and adaptive control by use of the modified least squares algorithm*, SIAM J. Control Optim., 22 (1984), pp. 758–776.
[10] H. F. Chen and P. E. Caines, *The strong consistency of the stochastic gradient algorithm of adaptive control*, IEEE Trans. Automat. Control, AC-30 (1985), pp. 189–192.
[11] H. F. Chen, *Recursive Estimation and Control for Stochastic Systems*, John Wiley, New York, 1985.
[12] 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.
[13] ———, *Asymptotically optimal adaptive control with consistent parameter estimates*, SIAM J. Control Optim., 25 (1987), pp. 558–575.
[14] H. F. Chen and J. F. Zhang, *Convergence rates in stochastic adaptive tracking*, Internat. J. Control, 49 (1989), pp. 1915–1935.
[15] H. F. Chen and L. Guo, *Identification and stochastic adaptive control*, Birkhäuser, Boston, 1991.
[16] Y. S. Chow and H. Teicher, *Probability theory: Independence interchangeability and martingales*, Springer-Verlag, Berlin, 1978.
[17] M. Duflo, *Méthodes récursives aléatoires*, Masson, Paris, 1990.
[18] G. C. Goodwin, P. J. Ramadge, and P. E. Caines, *Discrete time stochastic adaptive control*, SIAM J. Control Optim., 19 (1981), pp. 829–853.
[19] 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.
[20] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
[21] P. R. Kumar and L. Praly, *Self-tuning trackers*, SIAM J. Control Optim., 25 (1987), pp. 1053–1071.
[22] P. R. Kumar, *Convergence of adaptive control schemes using least squares parameter estimates*, IEEE Trans. Automat. Control, 35 (1990), pp. 416–424.
[23] T. L. Lai and C. Z. Wei, *Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems*, Ann. Statist., 10 (1982), pp. 154–166.
[24] 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.
[25] 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.
[26] K. S. Sin and G. C. Goodwin, *Stochastic adaptive control using a modified least squares algorithm*, Automatica, 18 (1982), pp. 315–321.
[27] V. Solo, *The convergence of* AML, IEEE Trans. Automat. Control, AC-24 (1979), pp. 958–962.
[28] C. Z. Wei, *Adaptive prediction by least squares predictors in stochastic regression models with applications to time series*, Ann. Statist., 15 (1987), pp. 1667–1682.