



UNIVERSITÉ BORDEAUX I

INSTITUTE OF MATHEMATICS

**ASYMPTOTICAL RESULTS FOR MODELS ARX IN
ADAPTIVE TRACKING**

By

Víctor Hugo Vázquez Guevara.

THESIS

Submitted in partial fulfillment of the requirements for the
Doctor in Science (Mathematics) (BUAP) and the Doctor
in Applied Mathematics and Scientific Calculus (UB-I)
degrees.

Directors of Thesis:

Dr. Bernard Bercu and Dr. Raúl Montes de Oca.

Puebla, junio 10 de 2010

Acknowledgement

I am heartily thankful to my supervisors, Dr. Bernard Bercu and Dr. Raúl Montes de Oca whose encouragement, supervision and support from the preliminary to the concluding level enabled me to develop an understanding of the subject.

I also want to thank to the members of the jury: Dr. Onésimo Hernández Lerma, Dr. Miguel Antonio Jiménez Pozo, Dr. Rolando Cavazos Cadena, Dr. Fabrice Gamboa, Dr. Jean Claude Fort, Dr. Hugo Adán Cruz and Dr. Francois Dupuy.

Finally, I want to thank to the CoNaCyT, the Mexican National Council for Science and Technology, to the INRIA, the French National Institute for Research in Computer Science and Control for the financial support offered during my studies of doctorate. In addition it is possible to mention that this thesis was realised inside the Cooperation Programme "Problemas de cambio estructural y de control adaptado para modelos de regresión" (M06-M01), SEP-CONACYT-ANUIES-ECOS Nord.

Summary

This thesis is devoted to asymptotical results for ARX models in adaptive tracking. It is divided into four parts.

The first part is a short introduction on ARMAX models together with a state of the art on the main results in the literature on adaptive tracking.

The second part deals with a new concept of strong controllability for ARX models in adaptive tracking. This new notion allows us to extend the previous convergence results. We prove the almost sure convergence for both least squares and weighted least squares algorithms. We also establish a central limit theorem and a law of iterated logarithm for these two algorithms.

The third part is dedicated to ARX models that are not strongly controllable. Thanks to a persistently excited adaptive tracking control, we show that it is possible to get rid of the strong controllability assumption.

The fourth part deals with the asymptotic behaviour of the Durbin-Watson statistic for ARX models in adaptive tracking via a martingale approach.

Résumé

Cette thèse est consacrée aux résultats asymptotiques pour les modèles ARX en poursuite adaptative. Elle est constituée de quatre parties.

La première partie est une brève introduction sur les modèles ARMAX et un état de l'art des principaux résultats de la littérature en poursuite adaptative.

La seconde partie porte sur l'introduction d'un nouveau concept de contrôlabilité forte pour les modèles ARX en poursuite adaptative. Il permet de généraliser les résultats antérieurs. On montre la convergence presque sûre des algorithmes des moindres carrés ordinaires et pondérés. On établit également le théorème de la limite centrale ainsi que la loi du logarithme itéré pour ces deux algorithmes.

La troisième partie est dédiée aux modèles ARX qui ne sont pas fortement contrôlables. On montre que, via un contrôle de poursuite excité, il est possible de s'affranchir de l'hypothèse de forte contrôlabilité.

La quatrième partie est consacrée au comportement asymptotique de la statistique de Durbin-Watson pour les modèles ARX en poursuite adaptative via des arguments martingales.

Contents

Acknowledgement	i
1 Introduction	1
1.1 The <i>ARX</i> models	2
1.2 The Noise	9
1.3 Estimation Algorithms	9
1.4 Tracking Control	12
1.5 Reference Trajectory	13
1.6 Review	14
2 Strongly Controllable ARX Models	17
2.1 Strong Controllability	17
2.2 Main Results	26
2.3 Numerical Experiments	43
3 Non Strong Controllable Models	50
3.1 Persistently Excited Tracking	50
3.2 Main Results	56
3.3 Note on the Reference Trajectory	64
3.4 Numerical Experiments	76
4 Work in Progress on the Durbin Watson Statistic	90
4.1 Preliminaries	90

4.2 The DW Statistic in the $AR_1(p)$ Framework	94
Conclusions	114
Further Work	115
Matlab Codes	116
Bibliography	139

Chapter 1

Introduction

The multidimensional Auto Regressive models with eXogenous input (ARX) are a versatile and useful tool in many areas of applied mathematics, such as Video Segmentation [59], Financial Mathematics [29],[38], [44], [64], Population Dynamics [48], Robotics [15], Medical Physics [39], Electronics [19], Mechatronics [35], Combustion Engines [53], Internet [56] and Neurosciences [5], [49]. On the other hand, adaptive tracking plays a crucial role in a wide range of application areas such as robotics [43] and Aeronautics [65]. Due to the amplitude of application areas of the ARX models it is important for us to investigate these models in order to obtain certain useful results around the tracking trajectories and the estimation of parameters.

In this introductory chapter, the ARX models with unknown parameter are presented together with three estimation algorithms and the notion of control in adaptive tracking. At the end of this Chapter there is included a brief review of the most significant previous papers, in our opinion, with a discussion of the main results of this Thesis.

1.1 The *ARX* models

Of course we could start this Thesis with a definition or with some complicated discussion of an application problem. However we prefer to open it with a "trivial" example in order to introduce the ARX models which are the main topic of this work. Let's consider a boxer who is going to fight in 90 days from today. In addition, the fighter has to increase his weight from Super welterweight (69.9 kg) to Middleweight (72.6 kg). To this end, the trainer will measure his weight every day during the following three months and he will take some decisions to approximate the boxer's weight as closely as possible to the picture shown in Figure 1.1. Once the trainer has done the measurement, he can modify the rhythm of the training sessions and reduce or increase the calories that the trainee eats. It's plausible to think that there exist some dependence in the weight measured on day " n ", let's say X_n , and the weight of the last " p " days before ($X_{n-1}, X_{n-2}, \dots, X_{n-p}$). In addition, the decisions of the trainer related to how hard the training sessions are and the regulation of the food have some impact on the weight of each following day; i.e., once the trainer knows the weight X_n of the fighter he takes some decisions that will have some effect, let's say U_n , on the weight of the next day: X_{n+1} . Moreover, we can say that the weight on day " n " not only depends on the decisions taken the day before, but it depends on the decisions of the last " q " days ($U_n, U_{n-1}, \dots, U_{n-q+1}$). However, large values of " p " and " q " may not be a good idea, because the decisions taken on the first two days of training may not have an important effect on the weight of day 30. Then, we can use, for example $p = 3$, $q = 2$ (just for an illustration) and the following form for indicating the relationship between the decisions of the past two days and the weight of the last three days:

$$X_{n+1} = a_1 X_n + a_2 X_{n-1} + a_3 X_{n-2} + a_4 U_n + a_5 U_{n-1} + \varepsilon_{n+1}. \quad (1.1.1)$$

The term ε_{n+1} is added because of the errors committed in the measurement. Of course this is not the only way of writing the expression we are aware of,

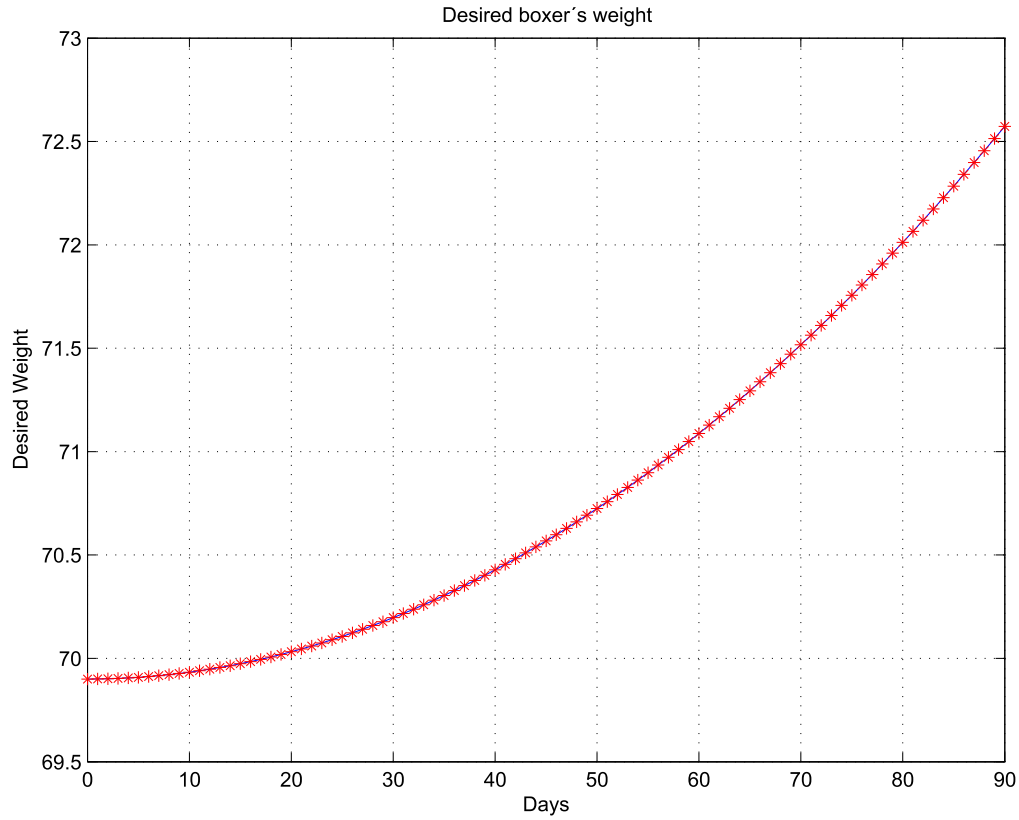


Figure 1.1: The desired boxer's weight per day

but this is the most related one with the main topic of this Thesis. It is interesting for us to know the values of a_1, a_2, a_3, a_4 and a_5 (at least an approximation) and we would like to obtain some expression for writing U_n in order to make X_n has the same shape as in Figure 1.1.

The models with a form similar to the one given in equation (1.1.1) are known as ARX models, and as we said before, this kind of models are the main subject of this Thesis.

Before describing in a formal way what an *ARX* model is, let us introduce

the linear regression models in order to define the *ARX* models as a particular case of them.

In all the rest of this Thesis, we shall assume that we work in a probability space (Ω, A, P) with a filtration $\mathcal{F} = (\mathcal{F}_n)_{n \geq 0}$, where $\mathcal{F}_n$ is the σ -algebra of the events occurring up to time n .

Definition 1.1.1 : A linear regression model with parameter θ is given by

$$X_n = \theta^t \phi_{n-1} + \varepsilon_n \quad (1.1.2)$$

where $\phi = (\phi_n)_{n \geq 0}$ is a sequence of random vectors of dimension δ which is adapted to $\mathcal{F}$ which means that for all $n \geq 0$ ϕ_n is $\mathcal{F}_n$ measurable in addition, ϕ_n is observable at time n , θ is a $\delta \times d$ unknown matrix and $\varepsilon = (\varepsilon_n)_{n \geq 0}$ is a d -dimensional driven noise adapted to $\mathcal{F}$.

We can make several choices on the matrix θ and the sequence ϕ ; among those which are interesting for the purpose of this Thesis we have:

1. AR models. If we choose

$$\theta^t = (A_1, \dots, A_p) \quad \text{and} \quad (1.1.3)$$

$$\phi_{n-1}^t = (X_{n-1}, \dots, X_{n-p}), \quad (1.1.4)$$

where $A_1, \dots, A_p$ are square matrices of order d and for all $n \geq 0$, ϕ_n is a vector of size $\delta = pd$, model (1.1.2) with (1.1.3) and (1.1.4) is called an Autoregressive Model of order p and is denoted by the symbol $AR_d(p)$. In addition the corresponding linear regressive model in its extended form is

$$X_n = A_1 X_{n-1} + A_2 X_{n-2} + \dots + A_p X_{n-p} + \varepsilon_n. \quad (1.1.5)$$

The $AR_d(p)$ models may be associated with the matrix valued polynomial, defined for $z \in \mathbb{C}$ by

$$A(z) = I_d - A_1 z - A_2 z^2 - \dots - A_p z^p. \quad (1.1.6)$$

Then, for all $n \geq 0$

$$A(R)X_n = \varepsilon_n \quad (1.1.7)$$

where R stands for the shift back operator defined, for any sequence $V = (V_n)_{n \geq 0}$ by

$$R(V_n) = V_{n-1}. \quad (1.1.8)$$

2. ARMA models. If we choose

$$\theta^t = (A_1, \dots, A_p, C_1, \dots, C_r) \quad \text{and} \quad (1.1.9)$$

$$\phi_{n-1}^t = (X_{n-1}^{(p)}, \varepsilon_{n-1}^{(r)}) = (X_{n-1}, \dots, X_{n-p}, \varepsilon_{n-1}, \dots, \varepsilon_{n-r}), \quad (1.1.10)$$

where $A_1, \dots, A_p, C_1, \dots, C_r$ are square matrices of order d and for all $n \geq 0$, ϕ_n is a vector of size $\delta = (p+r)d$, model (1.1.2) with (1.1.9) and (1.1.10) is called an Autoregressive Moving Average Model of order (p, r) and is denoted by the symbol $ARMA_d(p, r)$. In addition, the corresponding linear regression model in its extended form is

$$X_n = A_1 X_{n-1} + \dots + A_p X_{n-p} + C_1 \varepsilon_{n-1} + \dots + C_r \varepsilon_{n-r} + \varepsilon_n. \quad (1.1.11)$$

As in the case of the $AR_d(p)$ models, we can associate a couple of matrix valued polynomials with the $ARMA_d(p, r)$ models:

$$\begin{aligned} A(z) &= I_d - A_1 z - A_2 z^2 - \dots - A_p z^p, \\ C(z) &= I_d + C_1 z + C_2 z^2 + \dots + C_r z^r. \end{aligned} \quad (1.1.12)$$

Then for all $n \geq 0$

$$A(R)X_n = C(R)\varepsilon_n. \quad (1.1.13)$$

3. ARX models. Sometimes the regression vector ϕ_n depends on a so-called exogenous sequence $U = (U_n)$, which may be chosen at any time as a function of the previous observations; this may be a deterministic or a random sequence, defined beforehand before any observations. It is often a control, chosen in order to adjust the system as best as possible, for example in the tracking problem, around the position of a given trajectory. Then if

$$\theta^t = (A_1, \dots, A_p, B_1, \dots, B_q) \quad \text{and} \quad (1.1.14)$$

$$\phi_{n-1}^t = (X_{n-1}^{(p)}, U_{n-1}^{(q)}) = (X_{n-1}, \dots, X_{n-p}, U_{n-1}, \dots, U_{n-q}), \quad (1.1.15)$$

where $A_1, \dots, A_p, B_1, \dots, B_q$ are square matrices of order d and for all $n \geq 0$, ϕ_n is a vector of size $\delta = d(p+q)$, model (1.1.2) with (1.1.14) and (1.1.15) is called an Autoregressive with Exogenous Input Model of order (p, q) and is denoted by the symbol $ARX_d(p, q)$. Exactly as in the case of $AR_d(p)$ and $ARMA_d(p, r)$ models, we have the extended expression:

$$X_n = A_1 X_{n-1} + \dots + A_p X_{n-p} + B_1 U_{n-1} + \dots + B_q U_{n-q} + \varepsilon_n \quad (1.1.16)$$

and a couple of matrix valued polynomials related with these models

$$\begin{aligned} A(z) &= I_d - A_1 z - A_2 z^2 - \dots - A_p z^p \\ B(z) &= B_1 + B_2 z + B_3 z^2 + \dots + B_q z^q \end{aligned} \quad (1.1.17)$$

which lead us to the equation

$$A(R)X_n = B(R)U_{n-1} + \varepsilon_n. \quad (1.1.18)$$

4. ARMAX models. Finally, if

$$\begin{aligned} \theta^t &= (A_1, \dots, A_p, B_1, \dots, B_q, C_1, \dots, C_r) \quad \text{and} \\ \phi_n^t &= (X_{n-1}^{(p)}, U_{n-1}^{(q)}, \varepsilon_{n-1}^{(r)}), \end{aligned}$$

then the corresponding model is called an Autoregressive with Exogenous Input and Moving Average model or $ARMAX_d(p, q, r)$ for short. The extended form for the $ARMAX_d(p, q, r)$ is too long, that is why we only present the matrix valued polynomials associated with these models and the corresponding equations:

$$A(z) = I_d - A_1 z - A_2 z^2 - \dots - A_p z^p \quad (1.1.19)$$

$$B(z) = B_1 + B_2 z + B_3 z^2 + \dots + B_q z^q$$

$$C(z) = I_d + C_1 z + C_2 z^2 + \dots + C_r z^r$$

$$(1.1.20)$$

which implies the compact equation

$$A(R)X_n = B(R)U_{n-1} + C(R)\varepsilon_n. \quad (1.1.21)$$

We can see that an $ARMAX_d(p, q, r)$ model with $r = 0$ is essentially an $ARX_d(p, q)$ model. We also observe that an $ARX_d(p, q)$ with $q = 0$ is a simple $AR_d(p)$ model. Finally, an $ARMAX_d(p, q, r)$ model with $q = 0$ is an $ARMA_d(p, r)$ model. The main purpose of this Thesis deals with the $ARX_d(p, q)$ models but in our approach we assume that the leading matrix B_1 , commonly called the high frequency gain is known and equal to the identity matrix I_d of order d , but we want to keep the notation $ARX_d(p, q)$ for these models which can be rewritten as

$$X_n = \theta^t \phi_{n-1} + U_{n-1} + \varepsilon_n \quad (1.1.22)$$

where

$$\theta^t = (A_1, A_2, \dots, A_p, B_1, \dots, B_q) \quad \text{and} \quad (1.1.23)$$

$$\phi_{n-1} = \left(X_{n-1}^{(p)}, U_{n-2}^{(q)} \right)^t. \quad (1.1.24)$$

We can obtain the corresponding matrix polynomials associated with this new form of the $ARX_d(p, q)$ models:

$$\begin{aligned} A(z) &= I_d - A_1 z - A_2 z^2 - \dots - A_p z^p \\ B(z) &= I_d + B_1 z + B_2 z^2 + \dots + B_q z^q. \end{aligned} \quad (1.1.25)$$

In all the rest of this Thesis, we will only refer to this kind of $ARX_d(p, q)$ models with unknown parameter θ given by (1.1.23), with the additional objective of forcing the process $X = (X_n)_{n \geq 0}$ to be as close as possible in every step to a given reference trajectory $x = (x_n)_{n \geq 0}$, which is assumed to be predictable (x_n is $\mathcal{F}_{n-1}$ measurable), via the effect offered by the exogenous variables $U = (U_n)_{n \geq 0}$. Given that the matrices $A_1, \dots, A_p, B_1, \dots, B_q$ are not known, we need to estimate them at the same time as the exogenous variables control the dynamic of the process $X = (X_n)_{n \geq 0}$.

In addition, we shall make use of the causality assumption also known as the minimum phase condition on the matrix polynomial $B(z)$ [20]. More precisely, we assume that for all $z \in \mathbb{C}$, with $|z| \leq 1$

$$\det(B(z)) \neq 0. \quad (\text{A.1})$$

In other words, the polynomial $\det(B(z))$ only has zeros with modulus bigger than one. Consequently, if $r > 1$ is strictly less than the smallest modulus of the zeros of $\det(B(z))$, then $B(z)$ is invertible in the ball with center zero and radius r and $B^{-1}(z)$ is a holomorphic function [20]. For all $z \in \mathbb{C}$ such that $|z| \leq r$, we shall denote

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

the matrix valued polynomial such that $B^{-1}(z)B(z) = I_d$. We clearly have $D_0 = I_d$. In addition, the expressions of the matrices $D_1, D_2, \dots$ are given by

$$D_k = - \sum_{j=0}^{k-1} D_j B_{k-j}, \quad k \leq q \quad (1.1.26)$$

$$D_k = - \sum_{j=1}^q D_{k-j} B_j, \quad k > q. \quad (1.1.27)$$

In a similar way, we define a matrix sequence which will be useful for the development of the next two Chapters. For all $z \in \mathbb{C}$ such that $|z| \leq r$, we shall denote

$$P(z) = B^{-1}(z)(A(z) - I_d) = \sum_{k=1}^{\infty} P_k z^k. \quad (1.1.28)$$

We can find expressions for all matrices P_k as functions of the matrices A_i and B_j with $i = 1, 2, \dots, p$ and $j = 1, 2, \dots, q$, but such expressions may be simplified if we write matrices P_k as functions of the matrices A_i and D_j :

$$P_k = - \sum_{j=0}^{k-1} D_j A_{k-j}, \quad k \leq p \quad (1.1.29)$$

$$P_k = - \sum_{j=1}^p D_{k-j} A_j, \quad k > p. \quad (1.1.30)$$

1.2 The Noise

In the following two Chapters, we will assume that the driven noise $\varepsilon = (\varepsilon_n)_{n \geq 0}$ that is present in equation (1.1.22) satisfies for all $n \geq 0$ $\mathbb{E}[\varepsilon_{n+1} | \mathcal{F}_n] = 0$ and $\mathbb{E}[\varepsilon_{n+1} \varepsilon_{n+1}^t | \mathcal{F}_n] = \Gamma$ a.s. where Γ is a positive definite deterministic covariance matrix.

We also assume that the driven noise ε satisfies the strong law of large numbers; i.e. if

$$\Gamma_n = \frac{1}{n} \sum_{k=1}^n \varepsilon_k \varepsilon_k^t, \quad (1.2.1)$$

then Γ_n converges a.s. to Γ . That is the case if, for example:

1. (ε_n) is a white noise; i.e. a sequence of independent and identically distributed random variables.
2. (ε_n) has a finite conditional moment of order greater than 2.

1.3 Estimation Algorithms

In order to solve the problem of the estimation of the parameters given by (1.1.23) we will consider the Weighted Least Squares algorithm introduced by Bercu and Duflo [8], [13] (from now on WLS for short) of which the Least Squares (LS) algorithm is a particular case. This is the reason why we

will only develop the most general case. For this purpose we define, for an $ARX_d(p, q)$ model, the following two sequences, for $n \geq 0$,

$$S_n(a) = \sum_{k=0}^n a_k \phi_k \phi_k^t + I_\delta \quad \text{and} \quad (1.3.1)$$

$$s_n = \sum_{k=0}^n \|\phi_k\|^2 \quad (1.3.2)$$

where ϕ_n is given by (1.1.24) and the identity matrix I_δ with $\delta = d(p + q)$ is added in order to avoid useless invertibility assumption. The sequence $a = (a_n)_{n \geq 1}$ is called the weighted sequence and this sequence usually has the form, for $n \geq 0$

$$a_n = \frac{1}{g(\ln s_n)}$$

where g is an increasing function from $\mathbb{R}^+$ to itself such that

$$\int_1^\infty \frac{1}{g(t)} dt < +\infty \quad (1.3.3)$$

and $g(\ln t) = \mathcal{O}(t)$. In this work, we are concerned with the weighted sequence of Bercu and Duflo given by $g(t) = t^{1+\gamma}$ with $\gamma > 0$, [8] [13]. Consequently we have

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

The LS algorithm is obtained if we choose the weighted sequence to be constant and equal to one i.e.

$$a_n = 1.$$

One natural method consists in finding a $\delta \times d$ matrix in every step, let us say $\hat{\theta}_n$ for $n \geq 0$, that minimizes the weighted mean square distance between X and its prediction up to time n :

$$\hat{\theta}_n = \min_{\theta} \arg \sum_{k=1}^n a_{k-1} \|X_k - U_{k-1} - \theta^t \phi_{k-1}\|^2.$$

It is not hard to see that

$$\hat{\theta}_n = S_{n-1}^{-1}(a) \sum_{k=1}^n a_{k-1} \phi_{k-1} (X_k - U_{k-1})^t. \quad (1.3.4)$$

Furthermore, we have

$$\begin{aligned} S_n(a) \hat{\theta}_{n+1} &= \sum_{k=1}^{n+1} a_{k-1} \phi_{k-1} (X_k - U_{k-1})^t \\ &= \sum_{k=1}^n a_{k-1} \phi_{k-1} (X_k - U_{k-1})^t + a_n \phi_n (X_{n+1} - U_n)^t \\ &= (S_n(a) - a_n \phi_n \phi_n^t) \hat{\theta}_n + a_n \phi_n (X_{n+1} - U_n)^t \end{aligned}$$

which implies that

$$\hat{\theta}_{n+1} = \hat{\theta}_n + a_n S_n^{-1}(a) \phi_n (X_{n+1} - U_n - \hat{\theta}_n^t \phi_n)^t.$$

Then, we are in position to give the following definitions:

Definition 1.3.1 : *The Bercu and Dufo WLS algorithm is given by the recursion:*

$$\hat{\theta}_{n+1} = \hat{\theta}_n + a_n S_n^{-1}(a) \phi_n (X_{n+1} - U_n - \hat{\theta}_n^t \phi_n)^t \quad (1.3.5)$$

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

With the choice $a_n = 1$, for all $n \geq 0$, we are in presence of the LS algorithm.

Definition 1.3.2 : *The LS algorithm is defined recursively by:*

$$\hat{\theta}_{n+1} = \hat{\theta}_n + S_n^{-1} \phi_n (X_{n+1} - U_n - \hat{\theta}_n^t \phi_n)^t$$

with $S_n = \sum_{k=0}^n \phi_k \phi_k^t + I_\delta$.

In addition, there exists another estimator for θ which is well suited for tracking problems given by the Gradient algorithm. In order to obtain it, in the recursive definition of the least squares algorithm, we replace the matrix S_n by its trace s_n :

Definition 1.3.3 : *A stochastic Gradient algorithm in its recursive form is given for all $n \geq 0$ by*

$$\hat{\theta}_{n+1} = \hat{\theta}_n + s_n^{-1} \phi_n \left(X_{n+1} - U_n - \hat{\theta}_n^t \phi_n \right)^t.$$

In this Thesis we will focus on the Least Squares and the Weighted Least Squares algorithms but we will not develop any analysis for the Gradient algorithm since it is not in general consistent [20].

1.4 Tracking Control

In the $ARX_d(p, q)$ framework previously described with anteriority, the exogenous sequence $U = (U_n)_{n \geq 0}$ is the responsible of the regulation of the process $X = (X_n)_{n \geq 0}$ by forcing X_n to track step by step the predictable reference trajectory $x = (x_n)_{n \geq 0}$. The following lines show in an intuitive way the reason why we prefer to use the Åström and Wittenmark adaptive control. From equation (1.1.22) it follows that:

$$X_{n+1} - x_{n+1} = \theta^t \phi_n + U_n - x_{n+1} + \varepsilon_{n+1}. \quad (1.4.1)$$

If in fact for every $n \geq 0$ we have that X_{n+1} is very close to x_{n+1} and we rule out the noise ε_{n+1} , then

$$0_d \approx \theta^t \phi_n + U_n - x_{n+1} \quad (1.4.2)$$

where 0_d stands for the zero vector of dimension d . Ideally it should be possible to choose U_n such that

$$U_n = x_{n+1} - \theta^t \phi_n. \quad (1.4.3)$$

However the parameter θ is not known, so the estimator $\hat{\theta}_n$ given by the LS or the WLS algorithm is used instead of θ itself. The adaptive tracking control proposed by Aström and Wittenmark [4] is given, for all $n \geq 0$, by

$$U_n = x_{n+1} - \hat{\theta}_n^t \phi_n. \quad (1.4.4)$$

By substituting (1.4.4) in (1.1.22), we obtain the closed-loop system

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1} \quad (1.4.5)$$

with the prediction error $\pi_n = (\theta - \hat{\theta}_n)^t \phi_n$.

1.5 Reference Trajectory

The reference trajectory $x = (x_n)_{n \geq 0}$ shall satisfy the following properties:

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n x_k x_k^t = \nabla \quad \text{a.s.} \quad (1.5.1)$$

where ∇ is a positive definite symmetric matrix. In addition, for all $i \geq 1$,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n x_{k+i} x_k^t = \nabla_i \quad \text{a.s.} \quad (1.5.2)$$

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n x_{k+i} \varepsilon_k^t = 0 \quad \text{a.s.} \quad (1.5.3)$$

If for example, the reference trajectory is such that

$$\sum_{k=1}^n \|x_k\|^2 = o(n) \quad \text{a.s.} \quad (1.5.4)$$

we have that $\nabla = \nabla_i = 0$. However, it could be also interesting to investigate the case in which $\nabla_i = 0$ with ∇ positive.

1.6 Review

The following lines show a brief description of some aspects of previous papers concerning the *ARX* models in different frameworks around the estimation and the adaptive control. In a chronological order, we have:

1. Bercu [6] in the ARMAX framework proved the strong consistency of the extended least squares estimator for a recursive AML scheme under the hypothesis of passivity; in addition, he worked on the prediction errors related to such a model.

2. Chen and Guo [31] worked on the stability and optimality of the Aström and Wittenmark self-tuning regulator.

3. Bercu and Duflo [13] defined a weighted least squares algorithm that improves the least squares estimator, in the sense that the prediction errors have more interesting asymptotic properties.

4. Bercu [8] worked on ARMAX models, showing that the weighted least squares algorithm has the same advantages of both extended least squares and stochastic gradient algorithms, if suitable weightings are chosen. Concerning adaptive tracking problems, the strong consistency of the estimator and control optimality are both ensured.

5. Duflo [20] developed, among many other important topics, a very large set of tools around ARMAX models including the gradient algorithm and the weighted least squares algorithm, the adaptive tracking problem, the theory of martingales and vector martingales, etc.

6. Bercu [7] proved a central limit theorem and a law of iterated logarithm for both least squares and weighted least squares algorithms, together with the optimality of the tracking for AR models. In the ARMA framework, he obtained for the extended and weighted least squares algorithms the optimality of the tracking. In addition, for the same algorithms, he obtained

via a persistently excited adaptive tracking control a central limit theorem and a law of iterated logarithm.

We also present the description of some papers concerning to the Durbin Watson statistic which will be the main topic of the last Chapter of this work:

1. J. Durbin and G.S. Watson [22] [23] introduce their statistic for detecting first order serial correlations in classical linear regression models and tabulated bounds for its distribution, in addition they obtained critical values and described their use in practice.

2. Nerlove [51] and Sapir [54] exposed the bias of the Durbin Watson statistic in autoregressive schemes.

3. J. Durbin [21] obtained a procedure for testing the serial independence of the disturbances in a regression models where some of the regressors are lagged dependent variables.

4. Srivastava [57] established the asymptotic distribution of the Durbin Watson statistic without any condition on the design matrix in classical regression models.

5. Phillips and Loretan [52] via a von Neumann ratio approach proved the limit distribution of the Durbin Watson statistic under infinite variance errors.

In this Thesis we will extend the results obtained by Bercu [7] for the multidimensional $AR_d(p)$ models in adaptive tracking. In the second Chapter of this work, a concept of Strong Controllability is introduced together with its relationship with the usual notion of controllability. This concept will allow us to prove the invertibility of a suitable limit matrix via the analysis of the Schur complement of one of its blocks. Such property is needed in order to obtain the almost sure convergence of the least squares and the weighted least squares algorithms, the optimality of the tracking (in some sense, which will be explained later) as a consequence of the use of the Aström and Wittenmark

adaptive tracking control. Finally it will be proved that both least squares algorithms share the same central limit theorem (CLT) and the same law of iterated logarithm (LIL).

In the third Chapter, via a persistent excitation in the adaptive tracking control of Aström and Wittenmark, we will get rid of the Strong Controllability assumption on the $ARX_d(p, q)$ models and we will establish similar results to those obtained in the second Chapter; i.e., we will prove for any $ARX_d(p, q)$ model (not necessarily Strongly Controllable) the almost sure convergence of both LS and WLS algorithms, the residual optimality of the tracking and the shared CLT and LIL for both estimation algorithms.

In the final Chapter we introduce the Durbin Watson statistic and a series of numerical experiments that suggest some asymptotical results for the $AR_1(p)$ and $ARX_1(p, q)$ models when a serial correlated noise is considered.

Chapter 2

Strongly Controllable ARX Models

In this Chapter we introduce our concept of Strong Controllability for the $ARX_d(p, q)$ models. This concept will allow us to prove asymptotical results for both WLS and LS algorithms when the reference trajectory satisfies

$$\sum_{k=1}^n \|x_k\|^2 = o(n) \quad \text{a.s.} \quad (2.0.1)$$

It is possible to mention that these asymptotical results are associated to the analysis of the Schur complement of a suit able limiting matrix related with the sequence $(S_n)_{n \geq 0}$ and $(S_n(a))_{n \geq 0}$. The Strong Controllability allows us to extend the previous converge results in the literature. Finally, via some numerical experiments, we show that this concept can not be avoided in $ARX_d(p, q)$ adaptive tracking.

2.1 Strong Controllability

In this section, we will introduce the concept of Strong Controllability for models $ARX_d(p, q)$ which offers sufficient conditions to prove the asymptot-

ical results of the following section. In order to introduce our concept of Strong Controllability, we must define the square matrix of order $d \times q$ given, if $p \geq q$, by

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

and, if $p \leq q$, by

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

where the matrices $P_1, P_2, \dots$ are given in the relation (1.1.29) and (1.1.30). Now, we are in position to define the concept that will lead us to be able to prove the asymptotical results we are concerned with for the LS and WLS algorithms, when the reference trajectory satisfies condition (2.0.1).

Definition 2.1.1 : An $ARX_d(p, q)$ model is said to be strongly controllable if the corresponding matrix valued polynomial $B(z)$ is causal and

$$\det(\Pi) \neq 0. \quad (\text{A.2})$$

Remark 1 The concept of strong controllability is not really restrictive. For example, if $p = q = 1$, assumption (A_2) reduces to $\det(A_1) \neq 0$; if $p = 2$ and $q = 1$, to $\det(A_2 - B_1 A_1) \neq 0$; if $p = 1$ and $q = 2$, to $\det(A_1) \neq 0$; while if $p = q = 2$, to

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

Remark 2 One can observe that our Strong Controllability notion is closely related to the usual concept of controllability or, in other words, to the coprimeness of the matrix polynomials $A - I_d$ and B . As a matter of fact, the resultant of the polynomials $A - I_d$ and B is given by

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

This determinant involves q rows with the matrices $A_1, \dots, A_p$ and p rows with the matrices $B_1, \dots, B_q$. It is not hard to see that ([28], [30]) $\text{Res}(A - I_d, B) = \det(\text{SYLV})$ where SYLV is the Sylvester matrix

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

For all $z \in \mathbb{C}$ such that $|z| \leq r$, let us define an analogue polynomial to $P(z)$:

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

where the matrices G_k are similar to the matrices P_k and may be explicitly calculated as functions of the matrices which form the parameter θ via the following recurrent form:

$$G_k = - \sum_{j=0}^{k-1} A_{k-j} D_j, \quad k \leq p \quad (2.1.1)$$

$$G_k = - \sum_{j=1}^p A_j D_{k-j}, \quad k > p \quad (2.1.2)$$

with matrices D_k , $k \geq 0$ given by (1.1.26) and (1.1.27).

Let Ψ be the symmetric square matrix of order δ :

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

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

$$\Pi = \begin{pmatrix} G_p & G_{p+1} & \cdots & G_{p+q-2} & G_{p+q-1} \\ G_{p-1} & G_p & G_{p+1} & \cdots & G_{p+q-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ G_{p-q+2} & \cdots & G_{p-1} & G_p & G_{p+1} \\ G_{p-q+1} & G_{p-q+2} & \cdots & G_{p-1} & G_p \end{pmatrix}$$

and, if $p \leq q$, by

$$\Psi = \begin{pmatrix} G_p & G_{p+1} & \cdots & \cdots & G_{p+q-2} & G_{p+q-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ G_1 & G_2 & \cdots & \cdots & G_{q-1} & G_q \\ 0 & G_1 & G_2 & \cdots & G_{q-2} & G_{q-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & Q_1 & \cdots & Q_p \end{pmatrix}.$$

In addition, denote by E the rectangular matrix of dimension $dq \times dp$ given,

if $p \geq q$, by

$$E = \begin{pmatrix} 0 & G_1 & G_2 & \cdots & \cdots & G_{p-2} & G_{p-1} \\ 0 & 0 & G_1 & \cdots & \cdots & G_{p-3} & G_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & G_1 & G_2 & \cdots & G_{p-q+1} \\ 0 & \cdots & \cdots & 0 & G_1 & \cdots & G_{p-q} \end{pmatrix}$$

and, if $p \leq q$, by

$$E = \begin{pmatrix} 0 & G_1 & \cdots & G_{p-2} & G_{p-1} \\ 0 & 0 & G_1 & \cdots & G_{p-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & G_1 \\ 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

We can easily see that the product

$$SYLV\Psi = \begin{pmatrix} I_{dp} & O^t \\ E & \Pi \end{pmatrix} \quad (2.1.3)$$

where O stands for the rectangular zero matrix of dimension $dq \times dp$. As $\det(\Psi) = 1$, we deduce from (2.1.3) that

$$\text{Res}(A - I_d, B) = \det(SYLV) = \det(SYLV\Psi) = \det(\Pi).$$

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

$$\det(\Pi) \neq 0 \iff A - I_d \text{ and } B \text{ are coprime} \quad (2.1.4)$$

which corresponds to the usual notion of controllability. Otherwise, if $d \geq 2$, it is easy to provide a counterexample for which the equivalence (2.1.4) fails. For example, one can take $d = 2$, and $p = q = 1$ with

$$A_1 = B_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

The corresponding $ARX_2(1,1)$ model is not Strongly Controllable according to Definition 2.1.1, but the matrix polynomials $A - I_d$ and B are clearly coprime.

To be able to develop the asymptotical analysis for the LS and WLS algorithms, we need a Lemma concerning to the invertibility of the suitable matrix of order δ :

$$\Lambda = \begin{pmatrix} L & K^t \\ K & H \end{pmatrix}, \quad (2.1.5)$$

where for all $1 \leq i \leq q$, H_i is the square matrix of order d :

$$H_i = \sum_{k=i}^{\infty} P_k \Gamma P_{k-i+1}^t. \quad (2.1.6)$$

In addition, let H be the symmetric square matrix of order dq :

$$H = \begin{pmatrix} H_1 & H_2 & \cdots & H_{q-1} & H_q \\ H_2^t & H_1 & H_2 & \cdots & H_{q-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ H_{q-1}^t & \cdots & H_2^t & H_1 & H_2 \\ H_q^t & H_{q-1}^t & \cdots & H_2^t & H_1 \end{pmatrix}. \quad (2.1.7)$$

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

$$K = \begin{pmatrix} 0 & K_1 & K_2 & \cdots & \cdots & K_{p-2} & K_{p-1} \\ 0 & 0 & K_1 & \cdots & \cdots & K_{p-3} & K_{p-2} \\ 0 & 0 & 0 & \cdots & \cdots & K_{p-4} & K_{p-3} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & K_1 & K_2 & \cdots & K_{p-q+1} \\ 0 & \cdots & 0 & 0 & K_1 & \cdots & K_{p-q} \end{pmatrix} \quad (2.1.8)$$

and, if $p \leq q$, by

$$\tilde{K} = \begin{pmatrix} 0 & K_1 & K_2 & \cdots & \cdots & K_{p-2} & K_{p-1} \\ 0 & 0 & K_1 & \cdots & \cdots & K_{p-3} & K_{p-2} \\ 0 & 0 & 0 & \cdots & \cdots & K_{p-4} & K_{p-3} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \cdots & 0 & K_1 \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix}. \quad (2.1.9)$$

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

$$L = \begin{pmatrix} \Gamma & 0 & \cdots & \cdots & 0 \\ 0 & \Gamma & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ & \cdots & 0 & \Gamma & 0 \\ 0 & \cdots & \cdots & 0 & \Gamma \end{pmatrix}. \quad (2.1.10)$$

The relevance of the matrix Λ and of the following Lemma which is the keystone of the asymptotic results of the next section will be more evident during the proofs of the Main Results section of this Chapter.

Lemma 2.1.2 *Let S be the Schur complement of L in Λ :*

$$S = H - KL^{-1}K^t. \quad (2.1.11)$$

If (A_1) and (A_2) hold and the reference trajectory satisfies (2.0.1), then S and Λ are invertible and

$$\Lambda^{-1} = \begin{pmatrix} L^{-1} + L^{-1}K^tS^{-1}KL^{-1} & -L^{-1}K^tS^{-1} \\ -S^{-1}KL^{-1} & S^{-1} \end{pmatrix}. \quad (2.1.12)$$

Proof. Let Σ be the infinite-dimensional diagonal square matrix:

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

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

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

and, if $p \leq q$, by

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

We can find that, for $i, j = 1, 2, \dots, q$, and $j \geq i$,

$$H = [h_{ij}]$$

with

$$h_{ij} = \sum_{k=j-i+1}^{\infty} P_k \Gamma P_{k-j+i}^t \quad (2.1.13)$$

and $h_{ji} = h_{ij}^t$. In a similar way, we have that

$$KL^{-1}K^t = [r_{ij}]$$

with

$$r_{ij} = \sum_{k=j-i+1}^{p-i} P_k \Gamma P_{k-j+i}^t \quad (2.1.14)$$

and $r_{ji} = r_{ij}^t$. Thus we have from (2.1.13) and (2.1.14) that

$$S = [s_{ij}]$$

where

$$s_{ij} = \sum_{k=p-i+1}^{\infty} P_k \Gamma P_{k-j+i}^t \quad (2.1.15)$$

and $s_{ij} = s_{ij}^t$. We can also find that

$$T \Sigma T^t = [v_{ij}]$$

with

$$v_{ij} = \sum_{k=p-i+1}^{\infty} P_k \Gamma P_{k-j+i}^t \quad (2.1.16)$$

and $v_{ji} = v_{ij}^t$. Finally we have, from (2.1.15) and (2.1.16) that

$$S = T \Sigma T^t. \quad (2.1.17)$$

It clearly follows from this suitable decomposition that

$$\ker(S) = \ker(T^t). \quad (2.1.18)$$

As a matter of fact, assume that $v \in \mathbb{R}^{dq}$ belongs to $\ker(T^t)$. Then, $T^t v = 0$, $Sv = 0$ which leads to $\ker(T^t) \subset \ker(S)$. On the other hand, assume that $v \in \mathbb{R}^{dq}$ belongs to $\ker(S)$. Since $Sv = 0$, we clearly have $v^t Sv = 0$, $v^t T \Sigma T^t v = 0$. However, the matrix Γ is positive definite. Consequently, $T^t v = 0$ and $\ker(S) \subset \ker(T^t)$, which implies (2.1.18). Moreover, it follows from the well-known rank theorem that

$$dq = \dim(\ker(S)) + \text{rank}(S). \quad (2.1.19)$$

As soon as $\ker(S) = \{0\}$, $\dim(\ker(S)) = 0$ and we obtain from (2.1.19) that S is of full rank dq which means that S is invertible. Furthermore, the left hand side square matrix of order dq of the infinite-dimensional matrix T is precisely Π . Consequently, if Π is invertible, Π is of full rank dq , $\ker(\Pi) = \ker(\Pi^t) = \{0\}$ and the left null space of T reduces to the null vector of $\mathbb{R}^{dq}$. Hence, if Π is invertible, we deduce from (2.1.18) and (2.1.19) that S is also invertible. Finally, as

$$\det(\Lambda) = \det(L) \det(S) = \det(\Gamma)^p \det(S), \quad (2.1.20)$$

we obtain from (2.1.20) that Λ is invertible and formula (2.1.12) follows from [36] page 18, which completes the proof ■

2.2 Main Results

In this section, we will establish the extensions of some of the results on the $AR_d(p)$ models that appear in [7], in particular those which concern to the almost sure convergence of the LS and WLS algorithms, the optimality of the tracking, the CLT and the LIL for the estimators of the unknown parameter θ .

We open this section with the definition of an optimal tracking, which provides a performance criteria for the adaptive control:

Definition 2.2.1 *Let $(C_n)_{n \geq 0}$ be the average cost matrix sequence defined by*

$$C_n = \frac{1}{n} \sum_{k=1}^n (X_k - x_k)(X_k - x_k)^t. \quad (2.2.1)$$

The tracking is said to be optimal if C_n converges almost sure to Γ given in section 1.2.

Our first result shows the connexion between the block matrix Λ given by (2.1.5) and the models $ARX_d(p, q)$.

Theorem 2.2.1 *Assume that the matrix polynomial $B(z)$ associated with the $ARX_d(p, q)$ model is causal and that $(\varepsilon_n)_{n \geq 0}$ has finite conditional moment of order $\alpha > 2$. Then, for the LS algorithm, we have*

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \Lambda \quad \text{a.s.} \quad (2.2.2)$$

where the limiting matrix Λ is given by (2.1.5).

Proof. We shall make use of the same approach as Bercu [7] or Guo and Chen [31]. First of all, we recall that for all $n \geq 0$,

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1} \quad (2.2.3)$$

and

$$s_n = \sum_{k=0}^n \|\phi_k\|^2.$$

It follows from (2.2.3) and the strong law of large numbers for martingales (see e.g. Corollary 1.3.25 part 2 of [20]) that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \|X_k\|^2 \geq \text{tr}(\Gamma + \nabla) \quad \text{a.s.}$$

As $\Gamma + \nabla$ is invertible and $s_n \geq \sum_{k=1}^n \|X_k\|^2$ we have that $n = \mathcal{O}(s_n)$ a.s., which implies that

$$s_n \rightarrow \infty \quad \text{a.s.} \quad (2.2.4)$$

Moreover, by Theorem 1 of [6] or Lemma 1 of [31], we have

$$\sum_{k=1}^n (1 - f_k) \|\pi_k\|^2 = \mathcal{O}(\log s_n) \quad \text{a.s.} \quad (2.2.5)$$

where $f_n = \phi_n^t S_n^{-1} \phi_n$. Hence, as ε has finite conditional moment of order $\alpha > 2$, we can show by the minimum phase assumption (A_1) on the matrix polynomial B together with (2.2.5) [31] that $\|\phi_n\|^2 = \mathcal{O}(s_n^\beta)$ a.s. for all

$2\alpha^{-1} < \beta < 1$. In addition, let $g_n = \phi_n^t S_{n-1}^{-1} \phi_n$ and $\delta_n = \text{tr}(S_{n-1}^{-1} - S_n^{-1})$. It follows from Proposition 4.2.12 of [20] that

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

Moreover, as

$$\sum_{n=1}^{\infty} \delta_n = \sum_{n=1}^{\infty} \text{tr}(S_{n-1}^{-1} - S_n^{-1}) \leq \text{tr}(I_\delta) = \delta,$$

the sequence (δ_n) goes to zero a.s. Consequently, as

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

we find that

$$\sum_{k=1}^n \|\pi_k\|^2 = \sum_{k=1}^n (1 - f_k)(1 + g_k) \|\pi_k\|^2,$$

which leads to

$$\sum_{k=1}^n \|\pi_k\|^2 \leq 2 \sum_{k=1}^n (1 - f_k) \|\pi_k\|^2 + \sum_{k=1}^n (1 - f_k) \delta_k \|\Phi_k\|^2 \|\pi_k\|^2. \quad (2.2.6)$$

Hence, we deduce from (2.2.5) and (2.2.6) that

$$\sum_{k=1}^n \|\pi_k\|^2 = o(s_n^\beta \log s_n) \quad \text{a.s.} \quad (2.2.7)$$

Therefore, we obtain from (2.0.1), (2.2.3) and (2.2.7) that

$$\sum_{k=1}^n \|X_{k+1}\|^2 = o(s_n^\beta \log s_n) + \mathcal{O}(n) \quad \text{a.s.} \quad (2.2.8)$$

Furthermore, we infer from assumption (A₁) that

$$U_n = B^{-1}(R)A(R)X_{n+1} - B^{-1}(R)\varepsilon_{n+1}, \quad (2.2.9)$$

so we have from (2.2.9) that

$$\sum_{k=1}^n \|U_k\|^2 \leq 2 \sum_{k=1}^n \|B^{-1}(R)A(R)X_{n+1}\|^2 + 2 \sum_{k=1}^n \|B^{-1}(R)\varepsilon_{n+1}\|^2. \quad (2.2.10)$$

Relation (2.2.8) together with Cauchy-Schwarz inequality and (2.2.10) imply

$$\sum_{k=1}^n \|U_k\|^2 = o(s_n^\beta \log s_n) + \mathcal{O}(n) \quad \text{a.s.} \quad (2.2.11)$$

It remains to put together the two contributions (2.2.8) and (2.2.11) and remember the convergence (2.2.4) to deduce that $s_n = o(s_n) + \mathcal{O}(n)$ a.s. leading to $s_n = \mathcal{O}(n)$. Hence, it follows from this fact, (2.2.7) and (2.2.4) that

$$\sum_{k=1}^n \|\pi_k\|^2 = o(n) \quad \text{a.s.} \quad (2.2.12)$$

Since the reference trajectory is predictable and the law of large numbers (1.2.1) and relation (2.2.12) hold, using the Cauchy-Schwarz inequality, we obtain

$$\sum_{k=1}^n \|\pi_k \varepsilon_{k+1}^t\|^2 = o(n) \quad \text{a.s.} \quad (2.2.13)$$

$$\sum_{k=1}^n \|\pi_k x_{k+1}^t\|^2 = o(n) \quad \text{a.s.} \quad (2.2.14)$$

$$\sum_{k=1}^n \|x_k \varepsilon_{k+1}^t\|^2 = o(n) \quad \text{a.s.} \quad (2.2.15)$$

It follows from these expressions, the strong law of large numbers for ε and convergence (2.0.1) that

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

Observe that

$$\frac{1}{n} \sum_{k=1}^n \varepsilon_{k-m} \varepsilon_{k-j}^t \rightarrow \begin{cases} \Gamma & \text{if } m = j \\ 0 & \text{if } m \neq j \end{cases} \quad \text{a.s.} \quad (2.2.16)$$

which follows for $m = j$ from (1.2.1). On the other hand, if $m \neq j$ (2.2.16) results from the strong law of large numbers for martingales [20]. We also

have that, for all $1 \leq i \leq p-1$ that

$$X_k X_{k-i}^t = (x_k + \pi_{k-1} + \varepsilon_k) (x_{k-i} + \pi_{k-i-1} + \varepsilon_{k-i})^t.$$

Consequently, we obtain from the previous calculations that for all $1 \leq i \leq p-1$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k X_{k-i}^t = 0 \quad \text{a.s.} \quad (2.2.17)$$

which implies that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k^p (X_k^p)^t = L \quad \text{a.s.} \quad (2.2.18)$$

where L is given by (2.1.10).

Furthermore, it follows from (2.2.3) and assumption (A₁) that, for all $n \geq 0$,

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

where

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

Thus, we have

$$\begin{aligned} U_k U_{k-i+1}^t &= V_k V_{k-i+1}^t + V_k Y_{k-i+2}^t + V_k W_{k-i+2}^t + W_{k+1} V_{k-i+1}^t \\ &+ W_{k+1} Y_{k-i+2}^t + W_{k+1} W_{k-i+2}^t + Y_{k+1} V_{k-i+1}^t + Y_{k+1} Y_{k-i+2}^t \\ &+ Y_{k+1} W_{k-i+2}^t \end{aligned}$$

for all $1 \leq i \leq q$. Consequently, as $W_n = P(R)\varepsilon_n$, we deduce from the Cauchy-Schwarz inequality together with (1.5.1), (1.5.3), (2.2.12), the strong law of large numbers for the driven noise and from the fact that x_k is $\mathcal{F}_{k-1}$ measurable that

$$\begin{aligned}
\frac{1}{n} \sum_{k=1}^n V_k V_{k-i+1}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n V_k W_{k-i+2}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n V_k Y_{k-i+2}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n W_{k+1} V_{k-i+1}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n W_{k+1} Y_{k-i+2}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n Y_{k+1} V_{k-i+1}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n Y_{k+1} W_{k-i+2}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n Y_{k+1} Y_{k-i+2}^t &\rightarrow 0 \quad a.s.
\end{aligned}$$

If we recall (2.2.16), then we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n W_{k+1} W_{k-i+2}^t = \sum_{k=i}^{\infty} P_k \Gamma P_{k-i+1}^t. \quad (2.2.19)$$

thus we have obtained

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n U_k U_{k-i+1}^t = H_i \quad a.s. \quad (2.2.20)$$

which ensures that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n U_k^q (U_k^q)^t = H \quad \text{a.s.} \quad (2.2.21)$$

where H is given by (2.1.7). Via similar arguments together with (2.2.3), (1.5.1), (1.5.2), (1.5.3) and (2.2.16), we also find for all $0 \leq m \leq p-2$ and $2 \leq l \leq q$ that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n U_k X_{k-m}^t = K_{m+1} \quad \text{a.s.}$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n U_{k-l} X_n^t = J_{l-1} \quad \text{a.s.}$$

Then we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k^p (U_{k-1}^q)^t = K^t \quad \text{a.s.} \quad (2.2.22)$$

where K is given by (2.1.8) or (2.1.9).

Therefore, it follows from (2.2.18), (2.2.21) and (2.2.22) that

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \Lambda \quad \text{a.s.} \quad (2.2.23)$$

where the limiting matrix Λ is given by (2.1.5). ■

The following Theorem concerns to the almost sure properties of the LS algorithm, the efficiency of the Aström and Wittenmark adaptive tracking control and the LS algorithm itself.

Theorem 2.2.2 *Assume that the $ARX_d(p, q)$ model is strongly controllable and that $(\varepsilon_n)_{n \geq 0}$ has finite conditional moment of order $\alpha > 2$. In addition, if the reference trajectory $x = (x_n)_{n \geq 0}$ satisfies (2.0.1), then for the LS algorithm, we have*

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \Lambda \quad \text{a.s.} \quad (2.2.24)$$

where the limiting matrix Λ is given by (2.1.5).

In addition the tracking is optimal

$$\|C_n - \Gamma_n\| = \mathcal{O}\left(\frac{\log n}{n}\right) \quad a.s. \quad (2.2.25)$$

We can be more precise in (2.2.25) by

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n (X_k - x_k - \varepsilon_k)(X_k - x_k - \varepsilon_k)^t = \delta\Gamma \quad a.s. \quad (2.2.26)$$

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

$$\|\hat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad a.s. \quad (2.2.27)$$

Proof. Convergence (2.2.24) follows directly from Theorem 2.2.1. Hereafter, we recall that the $ARX_d(p, q)$ model is strongly controllable. Thanks to Lemma 2.1.2, the matrix Λ is invertible and Λ^{-1} , given by (2.1.12), may be explicitly calculated. This is the key point for the proof. On the one hand, it follows from (2.2.24) and because Λ is positive definite that $n = \mathcal{O}(\lambda_{\min}(S_n))$, $\|\phi_n\|^2 = o(n)$ a.s. which implies that f_n tends to zero a.s. since

$$\begin{aligned} f_n &= \phi_n^t S_n^{-1} \phi_n \leq \|\phi_n\|^2 \lambda_{\max}(S_n^{-1}) \\ &= \frac{\|\phi_n\|^2}{\lambda_{\min}(S_n)}. \end{aligned}$$

Hence, by (2.2.5), we find that

$$\sum_{k=1}^n \|\pi_k\|^2 = \mathcal{O}(\log n) \quad a.s. \quad (2.2.28)$$

On the other hand, we obviously have from (2.2.3)

$$\|C_n - \Gamma_n\| = \mathcal{O}\left(\frac{1}{n} \sum_{k=1}^n \|\pi_{k-1}\|^2\right) \quad a.s. \quad (2.2.29)$$

Consequently, we immediately obtain the tracking optimality (2.2.25) from (2.2.28) and (2.2.29). Furthermore, by a well-known result of Lai and Wei [45] on the LS estimator, we also have

$$\|\hat{\theta}_{n+1} - \theta\|^2 = \mathcal{O}\left(\frac{\log \lambda_{\max} S_n}{\lambda_{\min} S_n}\right) \quad \text{a.s.} \quad (2.2.30)$$

Hence (2.2.27) clearly follows from (2.2.23) and (2.2.30). Moreover, we also infer from Lemma 1 of Wei [61] together with (2.2.23) that

$$(\hat{\theta}_{n+1} - \theta)^t S_n (\hat{\theta}_{n+1} - \theta) = o(\log n) \quad \text{a.s.} \quad (2.2.31)$$

However, it follows from Theorem 4.3.16 part 4 of [20] that

$$\lim_{n \rightarrow \infty} \frac{1}{\log d_n} \left((\hat{\theta}_{n+1} - \theta)^t S_n (\hat{\theta}_{n+1} - \theta) + \sum_{k=0}^n (1 - f_k) \pi_k \pi_k^t \right) = \Gamma \quad \text{a.s.} \quad (2.2.32)$$

where $d_n = \det(S_n)$. In addition, we deduce from (2.2.24) that

$$\lim_{n \rightarrow \infty} \frac{d_n}{n^\delta} = \det \Lambda \quad \text{a.s.} \quad (2.2.33)$$

Finally, (2.2.31) together with (2.2.32) and (2.2.33) imply (2.2.26), which achieves the proof. ■

Our second set of Theorems is an analogue to the first one, and it is related to the almost sure properties of the WLS algorithm of Bercu and Duflo.

Theorem 2.2.3 *Assume that the matrix polynomial $B(z)$ associated with the $ARX_d(p, q)$ model is causal and that either $(\varepsilon_n)_{n \geq 0}$ is a white noise or $(\varepsilon_n)_{n \geq 0}$ has a finite conditional moment of order $\alpha > 2$. Then, for the LS algorithm, we have*

$$\lim_{n \rightarrow \infty} (\log n)^{1+\gamma} \frac{S_n(a)}{n} = \Lambda \quad \text{a.s.} \quad (2.2.34)$$

where the limiting matrix Λ is given by (2.1.5).

Proof. By Theorem 1 of [8], we have

$$\sum_{n=1}^{\infty} a_n(1 - f_n(a)) \|\pi_n\|^2 < +\infty \quad \text{a.s.} \quad (2.2.35)$$

where $f_n(a) = a_n \phi_n^t S_n^{-1}(a) \phi_n$. Then, as $a_n^{-1} = (\log(s_n))^{1+\gamma}$ with $\gamma > 0$, we clearly obtain $a_n^{-1} = \mathcal{O}(s_n)$ a.s. and $a_n^{-1} \rightarrow \infty$ a.s. As in the proof of Theorem 2.2.2, we get

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

with $g_n(a) = a_n \phi_n^t S_{n-1}^{-1}(a) \phi_n$. However, as $g_n(a) \leq \frac{a_n \|\phi_n\|^2}{\lambda_{\min} S_{n-1}(a)}$, we find that then find that

$$\left(a_n^{-1} + \frac{\|\phi_n\|^2}{\lambda_{\min} S_{n-1}(a)} \right)^{-1} \leq a_n(1 - f_n(a)),$$

which clearly implies

$$\sum_{n=1}^{\infty} \left(a_n^{-1} + \frac{\|\phi_n\|^2}{\lambda_{\min} S_{n-1}(a)} \right)^{-1} \|\pi_n\|^2 < +\infty \quad \text{a.s.}$$

Furthermore, as

$$a_n^{-1} + \frac{\|\phi_n\|^2}{\lambda_{\min} S_{n-1}(a)} = \mathcal{O}(s_n) \quad \text{a.s.}$$

we get that

$$\sum_{n=1}^{\infty} s_n^{-1} \|\pi_n\|^2 < +\infty \quad \text{a.s.} \quad (2.2.36)$$

Hence, it follows from (2.2.36) together with Kronecker's lemma given e.g. by Lemma 1.3.14 of [20] that

$$\sum_{k=1}^n \|\pi_k\|^2 = o(s_n) \quad \text{a.s.} \quad (2.2.37)$$

Therefore, we obtain from (1.5.1), (2.2.3) and (2.2.37) that

$$\sum_{k=1}^n \|X_{k+1}\|^2 = o(s_n) + \mathcal{O}(n) \quad \text{a.s.} \quad (2.2.38)$$

In addition, we deduce from assumption (A₁) as it was done in the proof of Theorem 2.2.2 that

$$\sum_{k=1}^n \|U_k\|^2 = o(s_n) + \mathcal{O}(n) \quad \text{a.s.} \quad (2.2.39)$$

Consequently, we immediately infer from (2.2.38) and (2.2.39) that $s_n = o(s_n) + \mathcal{O}(n)$, so $s_n = \mathcal{O}(n)$ a.s. Hence, (2.2.37) implies that

$$\sum_{k=1}^n \|\pi_k\|^2 = o(n) \quad \text{a.s.} \quad (2.2.40)$$

Proceeding exactly as in the proof of the previous theorem, we find from (2.2.40) that

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

We can also see that

$$S_n(a) = a_{n+1}S_n + \sum_{k=1}^n b_k \frac{S_k}{k} + R \quad (2.2.41)$$

with $b_k = k(a_k - a_{k+1})$ and $R = S_0(a) - a_1S_0$, since

$$\begin{aligned} a_{n+1}S_n + \sum_{k=1}^n b_k \frac{S_k}{k} + R &= \sum_{k=1}^n [a_k - a_{k+1}] S_k + R \\ &= S_n(a) \end{aligned}$$

In addition

$$\sum_{k=1}^n b_k = \sum_{k=1}^n a_k - na_{n+1}$$

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

$$\sum_{k=1}^n b_k \sim (1 + \gamma) \frac{na_n}{\log n} \quad \text{a.s.}$$

which implies

$$\sum_{k=1}^n b_k \rightarrow \infty \quad \text{and} \quad \sum_{k=1}^n b_k = o(na_n) \quad \text{a.s.}$$

By the Toeplitz's lemma, we obtain that

$$\left(\sum_{k=1}^n b_k \right)^{-1} \sum_{k=1}^n b_k \frac{S_k}{k} \rightarrow \Lambda \quad \text{a.s.},$$

thus

$$(na_n)^{-1} \sum_{k=1}^n b_k \frac{S_k}{k} \rightarrow 0 \quad \text{a.s.} \quad (2.2.42)$$

Then (2.2.41) and (2.2.42) imply that

$$\lim_{n \rightarrow \infty} (\log n)^{1+\gamma} \frac{S_n(a)}{n} = \Lambda \quad \text{a.s.} \quad (2.2.43)$$

■

Theorem 2.2.4 *Assume that the $ARX_d(p, q)$ model is strongly controllable. In addition, suppose that either $(\varepsilon_n)_{n \geq 0}$ is a white noise or $(\varepsilon_n)_{n \geq 0}$ has finite conditional moment of order $\alpha > 2$. In addition, we suppose that the reference trajectory $x = (x_n)_{n \geq 0}$ satisfies (2.0.1). Then, for the WLS algorithm, we have that*

$$\lim_{n \rightarrow \infty} (\log n)^{1+\gamma} \frac{S_n(a)}{n} = \Lambda \quad \text{a.s.} \quad (2.2.44)$$

where the limiting matrix Λ is given by (2.1.5). In addition, the tracking is optimal

$$\| C_n - \Gamma_n \| = o \left(\frac{(\log n)^{1+\gamma}}{n} \right) \quad \text{a.s.} \quad (2.2.45)$$

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

$$\| \hat{\theta}_n - \theta \|^2 = \mathcal{O} \left(\frac{(\log n)^{1+\gamma}}{n} \right) \quad \text{a.s.} \quad (2.2.46)$$

Proof. Convergence (2.2.44) follows directly from Theorem 2.2.3 with Λ given by (2.1.5). We have from (2.2.44) that $a_n S_n \leq S_n(a)$ so $na_n = \mathcal{O}(\lambda_{\min}(S_n(a)))$, thus we can obtain that $f_n(a)$ tends to zero a.s. Consequently, we find from (2.2.35) and Kronecker's Lemma that

$$\sum_{k=1}^n \| \pi_k \|^2 = o((\log s_n)^{1+\gamma}) \quad \text{a.s.} \quad (2.2.47)$$

Then, (2.2.45) clearly follows from (2.2.29) and (2.2.47). Finally, by Theorem 1 of [8]

$$\|\hat{\theta}_{n+1} - \theta\|^2 = \mathcal{O}\left(\frac{1}{\lambda_{\min} S_n(a)}\right) \quad \text{a.s.} \quad (2.2.48)$$

Hence, we obtain (2.2.46) from (2.2.44) and (2.2.48), which completes the proof. ■

Now, we are concerned about proving a CLT and a LIL for the LS and the WLS algorithms in the adaptive tracking $ARX_d(p, q)$ framework. These results are a generalisation of those established in [7] and [32].

In order to prove these results, we need the following lemmas on regressive sequences given in [7]. Let $\varepsilon = (\varepsilon_n)_{n \geq 0}$ be a d -dimensional noise, adapted to $\mathcal{F}$, which satisfies $\mathbb{E}[\varepsilon_{n+1} \varepsilon_{n+1}^t | \mathcal{F}_n] = \Gamma$, where Γ is a deterministic covariance matrix. Let $\varphi = (\varphi_n)_{n \geq 0}$ be a δ -dimensional sequence of random vectors, adapted to $\mathcal{F}$. Set, for $n \geq 0$,

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

Lemma 2.2.2 : *Let $(c_n)_{n \geq 0}$ be a deterministic real sequence increasing to infinity. Assume that, for all $\epsilon > 0$,*

$$c_n^{-1} S_{n-1} \xrightarrow{\mathcal{P}} L$$

$$c_n^{-1} \sum_{k=1}^n E \left[\|\Delta M_k\|^2 1_{\{\|\Delta M_k\| \geq \epsilon \sqrt{c_n}\}} | \mathcal{F}_{k-1} \right] \xrightarrow{\mathcal{P}} 0 \quad (2.2.49)$$

where $\Delta M_k = M_k - M_{k-1}$. Then $c_n^{-1} M_n$ tends a.s. toward 0 and

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

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

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

Remark 3 : Assume that ε has a finite conditional moment of order greater 2. Then Lindeberg's condition (2.2.49) is satisfied if $\|\varphi_n\|^2 = o(c_n)$ a.s.

Lemma 2.2.3 : Let $(c_n)_{n \geq 0}$ be a deterministic real sequence increasing to infinity. Assume that the noise ε has finite conditional moment $\alpha > 2$. Also assume that

$$c_n^{-1} S_{n-1} \rightarrow L \quad a.s$$

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

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

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left(\frac{1}{2c_n \log \log c_n} \right)^{1/2} v^t M_n u &= - \liminf_{n \rightarrow \infty} \left(\frac{1}{2c_n \log \log c_n} \right)^{1/2} v^t M_n u \\ &= (v^t L v)^{1/2} (u^t \Gamma u)^{1/2} \quad a.s. \end{aligned} \quad (2.2.50)$$

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

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left(\frac{c_n}{2 \log \log c_n} \right)^{1/2} v^t S_{n-1}^{-1} M_n u &= - \liminf_{n \rightarrow \infty} \left(\frac{c_n}{2 \log \log c_n} \right)^{1/2} v^t S_{n-1}^{-1} M_n u \\ &= (v^t L^{-1} v)^{1/2} (u^t \Gamma u)^{1/2} \quad a.s. \end{aligned} \quad (2.2.51)$$

Now we are in position to prove that the LS and the WLS algorithms share the same CLT and the same LIL.

Theorem 2.2.5 Assume that the $ARX_d(p, q)$ model is strongly controllable and that $(\varepsilon_n)_{n \geq 0}$ has finite conditional moment of order $\alpha > 2$. In addition,

suppose that $(x_n)_{n \geq 0}$ satisfies (2.0.1) and has the same regularity in norm as $(\varepsilon_n)_{n \geq 0}$ which means that for all $2 < \beta < \alpha$,

$$\sum_{k=1}^n \|x_k\|^\beta = \mathcal{O}(n) \quad a.s. \quad (2.2.52)$$

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

$$\sqrt{n}(\hat{\theta}_n - \theta) \stackrel{\mathcal{L}}{\mathcal{N}}(0, \Lambda^{-1} \otimes \Gamma) \quad (2.2.53)$$

where the inverse matrix Λ^{-1} is given by (2.1.12) and the symbol $\otimes$ stands for the matrix Kronecker product. In addition, for any vectors $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^\delta$, they also share the same law of iterated logarithm

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

In particular,

$$\left(\frac{\lambda_{\min} \Gamma}{\lambda_{\max} \Gamma} \right) \leq \limsup_{n \rightarrow \infty} \left(\frac{n}{2 \log \log n} \right) \|\hat{\theta}_n - \theta\|^2 \leq \left(\frac{\lambda_{\max} \Gamma}{\lambda_{\min} \Gamma} \right) \quad a.s. \quad (2.2.54)$$

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

Proof. First of all, it follows from (1.1.22) and (1.3.4) that, for all $n \geq 1$,

$$\hat{\theta}_n - \theta = S_{n-1}^{-1}(a) M_n(a) \quad (2.2.55)$$

where

$$M_n(a) = \hat{\theta}_0 - \theta + \sum_{k=1}^n a_{k-1} \phi_{k-1} \varepsilon_k^t. \quad (2.2.56)$$

On the one hand, for the LS algorithm, we deduce (2.2.53) from convergence (2.2.24) and decomposition (2.2.55) by choosing $\varphi_n = \phi_n$ and $c_n = n$ in Lemma 2.2.2. Next, we make use of the LIL given by Lemma 2.2.3. Since

ε has a finite conditional moment of order $\alpha > 2$, we obtain from Chow's Lemma, given e.g. by Corollary 2.8.5 of [58], that, for all $2 < \beta < \alpha$,

$$\sum_{k=1}^n \|\varepsilon_k\|^\beta = \mathcal{O}(n) \quad \text{a.s.} \quad (2.2.57)$$

Consequently, as the reference trajectory (x_n) satisfies (2.2.52), we deduce from (2.2.3) together with (2.2.24) and (2.2.57) that

$$\sum_{k=1}^n \|X_k\|^\beta = \mathcal{O}(n) \quad \text{a.s.} \quad (2.2.58)$$

Furthermore, it follows from (2.2.9) and (2.2.58) that

$$\sum_{k=1}^n \|U_k\|^\beta = \mathcal{O}(n) \quad \text{a.s.} \quad (2.2.59)$$

Hence, we clearly obtain from (2.2.58) and (2.2.59) that

$$\sum_{k=1}^n \|\phi_k\|^\beta = \mathcal{O}(n) \quad \text{a.s.} \quad (2.2.60)$$

Therefore, as $\beta > 2$, (2.2.60) immediately implies that

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

Finally, we find (2.2.54) via Lemma 2.2.3 and (2.2.24) and equation (2.2.55). On the other hand, for the WLS algorithm, we take $\varphi_n = a_n \phi_n$ and $c_n = n(\log n)^{-2(1+\gamma)}$. Setting

$$Q_n(a) = \sum_{k=0}^n \phi_k \phi_k^t + S = \sum_{k=0}^n a_k^2 \phi_k \phi_k^t + S$$

we can see that

$$Q_n(a) = a_{n+1} S_n(a) + \sum_{k=1}^n b_k \frac{S_k(a)}{k} (\log k)^{\gamma+1} + R$$

where $b_k = \frac{k(a_k - a_{k+1})}{(\log k)^{1+\gamma}}$ and $R = S_0(a) - a_1 S_0$. In addition,

$$\sum_{k=1}^n b_k \sim \sum_{k=1}^n \frac{1}{(\log n)^{2+2\gamma}} - \frac{na_{n+1}}{(\log n)^{1+\gamma}} \quad \text{a.s.}$$

and thus

$$\sum_{k=1}^n b_k \sim \frac{na_n^2}{\log n} (1 + \gamma),$$

which implies

$$\sum_{k=1}^n b_k \rightarrow \infty \quad \text{and} \quad \sum_{k=1}^n b_k = o(na_n^2).$$

By Toeplitz's lemma, we have

$$\left(\sum_{k=1}^n b_k \right)^{-1} \sum_{k=1}^n b_k \frac{S_k(a)}{k} (\log k)^{\gamma+1} \rightarrow \Lambda \quad \text{a.s.},$$

so

$$(na_n^2)^{-1} \sum_{k=1}^n b_k \frac{S_k(a)}{k} (\log k)^{\gamma+1} \rightarrow 0 \quad \text{a.s.} \quad (2.2.61)$$

Then, from (2.2.41) and (2.2.42) we get that

$$\lim_{n \rightarrow \infty} (\log n)^{2+2\gamma} \frac{Q_n(a)}{n} = \Lambda \quad \text{a.s.} \quad (2.2.62)$$

which lead us to obtain (2.2.53), using (2.2.44) and Lemma 2.2.2. If ε has

finite conditional moment of order $\alpha > 2$, we find from Chow's Lemma [58] that, for all $2 < \beta < \alpha$,

$$\sum_{k=1}^n (a_k \parallel \varepsilon_k \parallel)^\beta = \mathcal{O}(n) \quad \text{a.s.} \quad (2.2.63)$$

From (2.2.52) we have that

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

and (2.2.47) implies

$$\sum_{k=1}^n (a_k \parallel \pi_k \parallel)^\beta = \mathcal{O}(n) \quad \text{a.s.}$$

Then we get

$$\sum_{k=1}^n (a_k \parallel \phi_k \parallel)^\beta = \mathcal{O}(n) \quad \text{a.s.}$$

and, finally,

$$\sum_{n=1}^{\infty} \left(\frac{a_n \parallel \phi_n \parallel}{\sqrt{c_n}} \right)^\beta = \mathcal{O}(n) \quad \text{a.s.}$$

Now, we make use of Lemma 2.2.3 to obtain (2.2.54) ■

2.3 Numerical Experiments

The goal of this section is to realize some numerical experiments for illustrating via some pictures the asymptotical results of this Chapter. For the sake of simplicity, in the first and second example, the reference trajectory $(x_n)_{n \geq 0}$ is chosen to be identically zero; i.e, (2.0.1) is satisfied. In all the experiments the driven noise $(\varepsilon_n)_{n \geq 0}$ is white Gaussian $\mathcal{N}(0, I)$. First, we will consider one Strongly Controllable model in dimension $d = 2$. Next, we will consider one non Strongly Controllable $ARX_2(1, 1)$ model in order to show that the Strongly Controllability assumption can not be avoided if we use the control defined in (1.4.4). All the simulations will be based on 400 realizations of a sample size 10000.

Example 1. First, consider the $ARX_2(1, 1)$ model

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

with

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

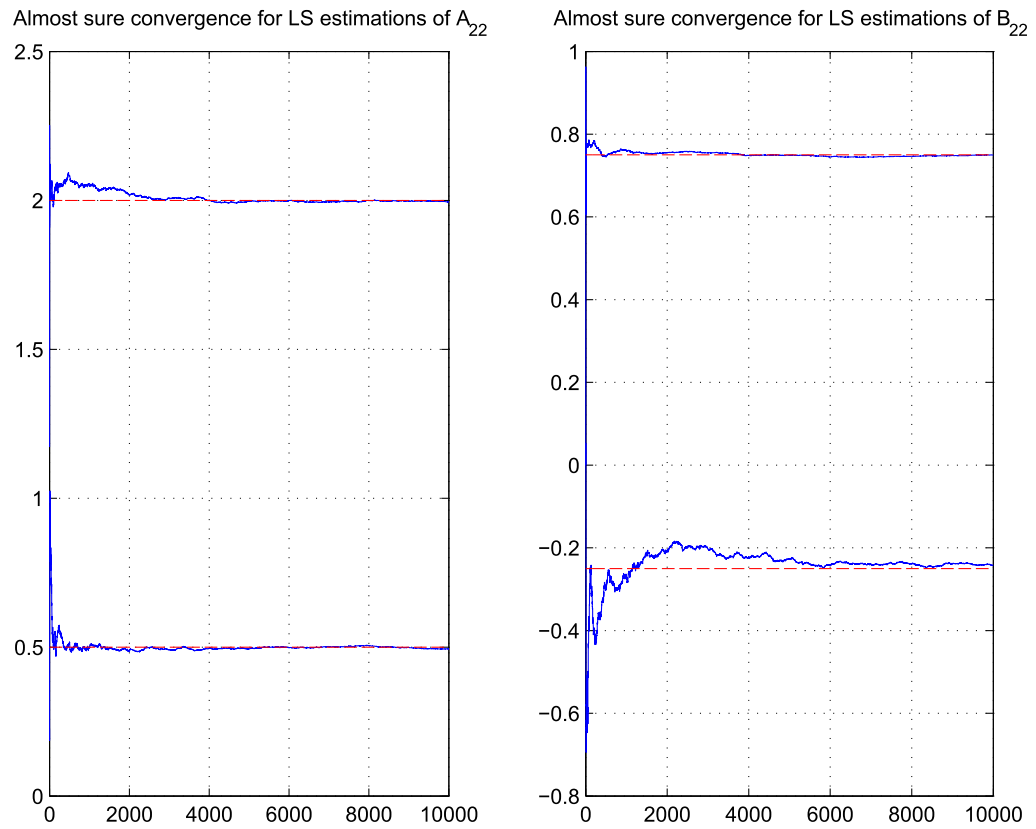


Figure 2.1: Almost sure convergence for LS algorithm, Example 1

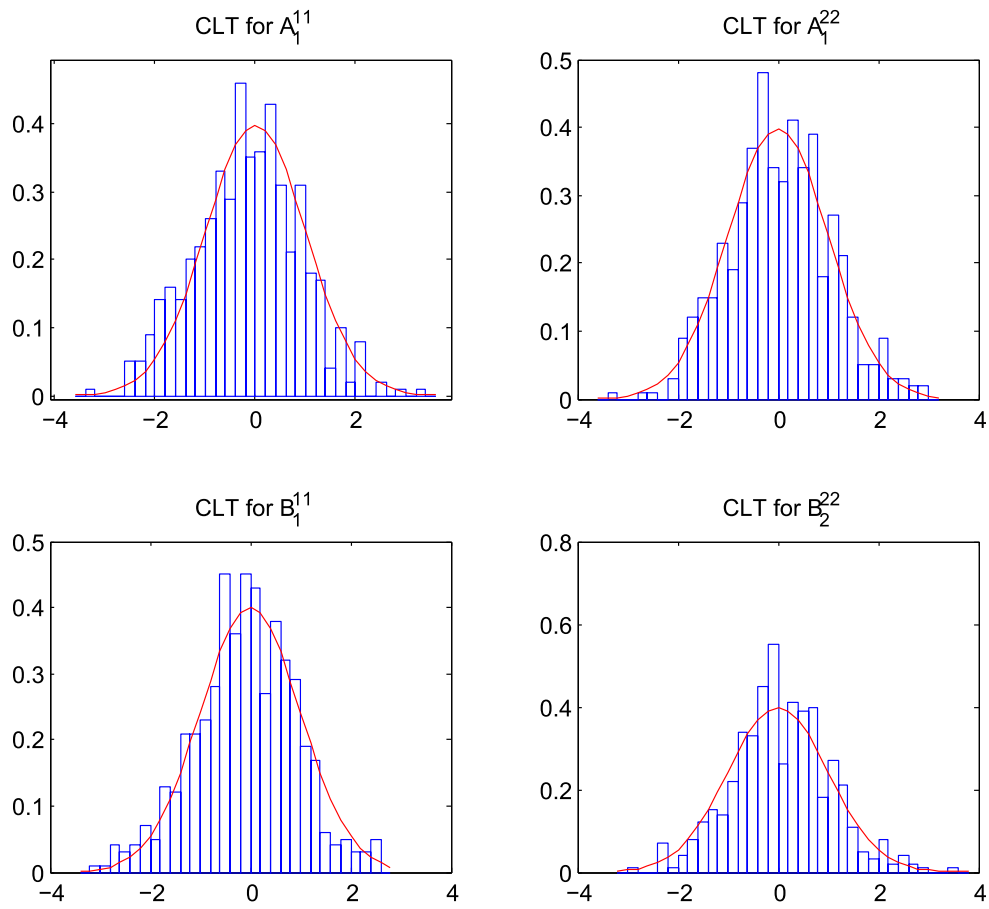


Figure 2.2: CLT for LS algorithm, Example 1

We can see that this $ARX_2(1,1)$ model is Strongly Controllable since $\det(A) = 1 > 0$ and the polynomial $B(z) = I_2 + Bz$ is causal. We have for all $k \geq 1$ that $P_k = -(B)^{k-1}A$. Moreover,

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

consequently

$$H = \begin{pmatrix} 64/7 & 0 \\ 0 & 4/15 \end{pmatrix}.$$

Therefore the limiting matrix Λ is

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 64/7 & 0 \\ 0 & 0 & 0 & 4/15 \end{pmatrix}.$$

One can observe that $\det(\Lambda) \neq 0$ and then Λ^{-1} exists, and we can talk about the CLT for each coordinate of $\hat{\theta}_n$. Figure 2.1 shows the almost sure convergence for LS estimator, for every coordinate which is not zero in matrices A and B . While in Figure 2.2 we can observe that the non zero coordinates of

$$Z_N = \sqrt{N}\Lambda^{1/2}(\hat{\theta}_N - \theta) \tag{2.3.1}$$

have a $\mathcal{N}(0,1)$ distribution, as we expected.

Example 2. In order to show that the strong controllability assumption can not be avoided if we use the adaptive tracking control (1.4.4), let us consider almost the same $ARX_2(1,1)$ model as in example 1 with

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

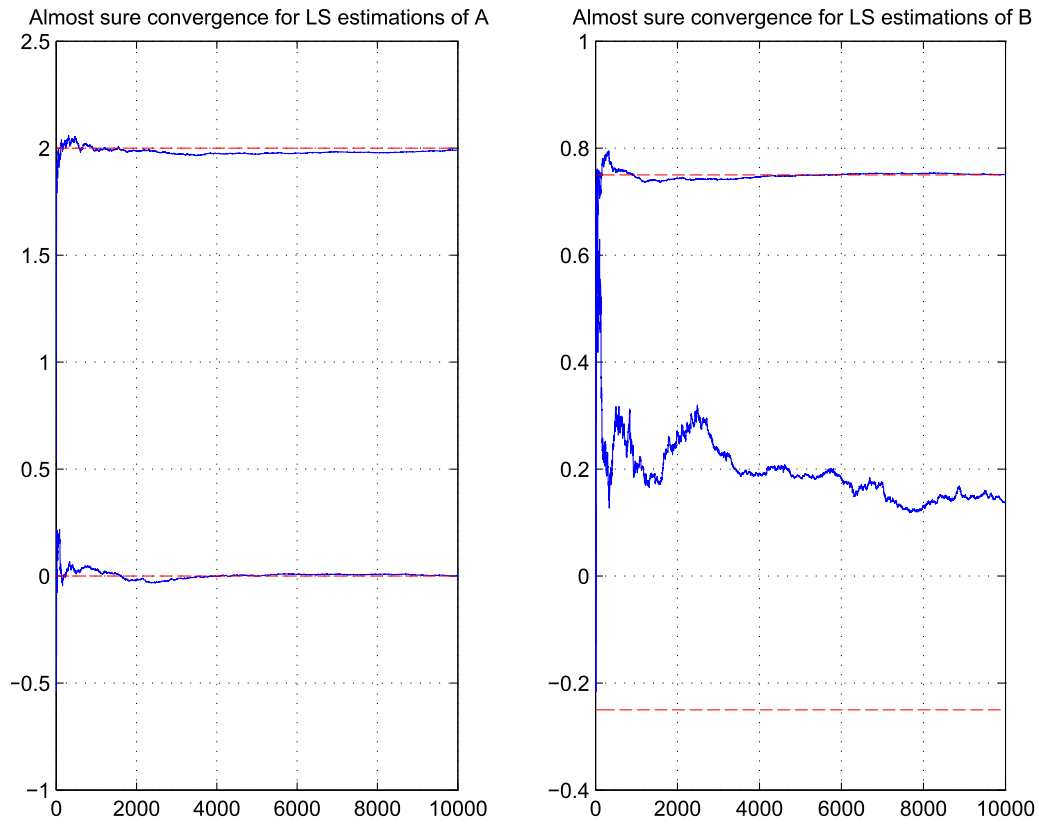


Figure 2.3: Almost sure convergence for LS algorithm of a non Strongly Controllable models, Example 2

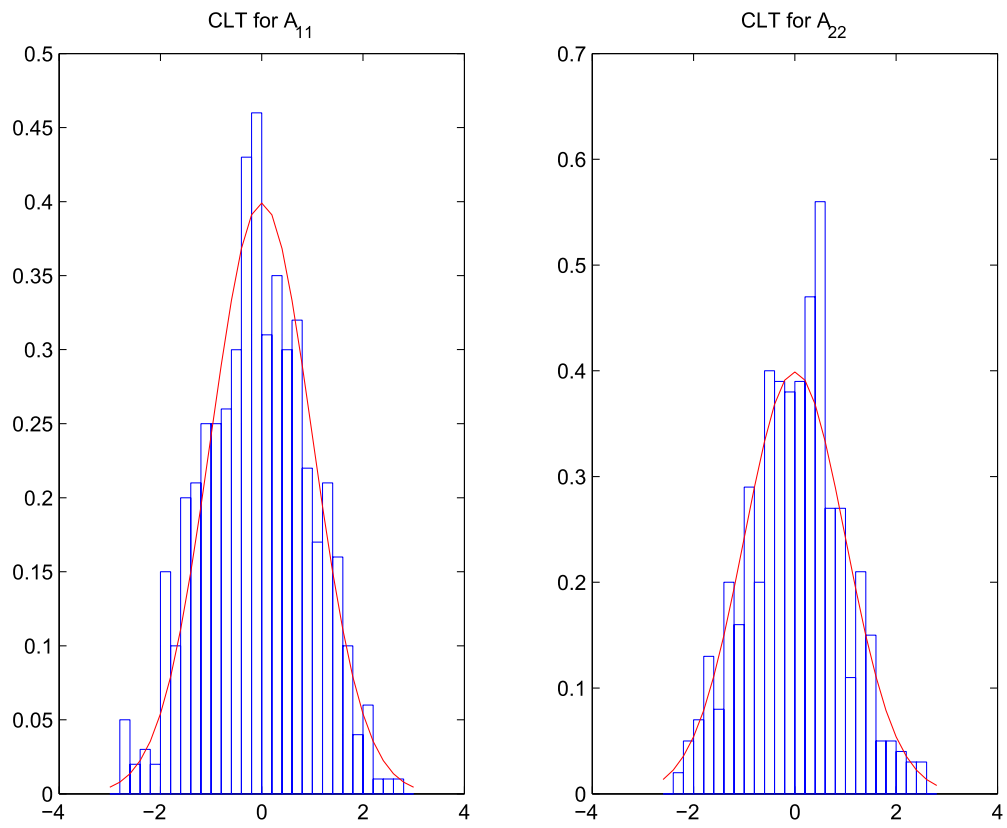


Figure 2.4: CLT for LS algorithm, Example 2

We have changed the second diagonal term of the matrix A . We have that $\det(A) = 0$, thus this $ARX_2(1, 1)$ model is not strongly controllable. In addition we find that

$$H = \begin{pmatrix} 64/7 & 0 \\ 0 & 0 \end{pmatrix}.$$

and

$$\Lambda = \frac{1}{7} \begin{pmatrix} 7 & 0 & 0 & 0 \\ 0 & 7 & 0 & 0 \\ 0 & 0 & 64 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We can observe that this matrix is not invertible, so we can not expect that the strong law of large numbers is satisfied for every entry of $\hat{\theta}_n$. In fact, Figure 2.3 shows that only the matrix A and the first diagonal term of matrix B are properly estimated and we can see in Figure 2.4 that CLT only holds for non zero diagonal coordinates of matrix A .

Chapter 3

Non Strong Controllable Models

We are facing now with the problem in which an $ARX_d(p, q)$ model is not necessarily a strongly controllable one. The complication in this case if we consider a reference trajectory which satisfies (2.0.1) and the control given by (1.4.4) is that even we can have convergences (2.2.24) (LS) or (2.2.44) (WLS), we cannot say that the limit matrix Λ is positive definite. In this Chapter, we use the Aström and Wittenmark adaptive tracking control with a persistent excitation in the multidimensional $ARX_d(p, q)$ framework to be able to avoid the Strongly Controllable condition and obtain similar results for the LS and WLS algorithms of those achieved in the Chapter 2.

3.1 Persistently Excited Tracking

The concept of persistent excitation is well-known in the control community. Since the pioneers works of Anderson [3] and Moore [50], this concept was successfully used in a large variety of fields of application going from economics [2], [18], to adaptive or learning control [17], [46], [47], [26], or mechanical engineering and robotics [25] [37] and [1].

To this end, we use the persistently excited adaptive tracking control given, for all $n \geq 0$, by

$$U_n = x_{n+1} - \hat{\theta}_n^t \phi_n + \xi_{n+1} \quad (3.1.1)$$

where the reference trajectory x still satisfies (2.0.1) and $\xi = (\xi_n)_{n \geq 0}$ is an exogenous noise of dimension d , adapted to $\mathcal{F}$, with mean 0 and positive definite covariance matrix Δ . In addition, we assume that ξ is independent of ε , of x , and of the initial state of the system. We observe that this control is nothing more than the adaptive tracking control of Aström and Wittenmark plus the exogenous noise ξ . Substituting (3.1.1) in (1.1.22) we obtain

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1} + \xi_{n+1} \quad (3.1.2)$$

In this new case, the tracking is said to be residually optimal if C_n given by (2.2.1) converges a.s. to $\Gamma + \Delta$. If the noise ξ satisfies the strong law of large numbers and

$$\Delta_n = \frac{1}{n} \sum_{k=1}^n (\varepsilon_k + \xi_k)(\varepsilon_k + \xi_k)^t, \quad (3.1.3)$$

then using Theorem 4.3.16 of [20], we find that

$$\lim_{n \rightarrow \infty} \Delta_n = \Gamma + \Delta \quad \text{a.s.} \quad (3.1.4)$$

Exactly as in Chapter 2, in order to develop an analogue asymptotical analysis we need to define a suitable matrix Λ^{exc} , and we need to prove the corresponding Lemma about its invertibility. In an informal way, we need to incorporate to the block matrix Λ given by (2.1.5) the effect offered by the exogenous noise. Thus, for all $1 \leq i \leq q$, denote by H_i^{exc} the square matrix of order d

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

where for all $k \geq 0$, $Q_k = D_k + P_k$ are given in the recursions (1.1.26), (1.1.27), (1.1.29) and (1.1.30). In addition, let H^{exc} be the symmetric square

matrix of order dq

$$H^{exc} = \begin{pmatrix} H_1^{exc} & H_2^{exc} & \cdots & H_{q-1}^{exc} & H_q^{exc} \\ (H_2^{exc})^t & H_1^{exc} & H_2^{exc} & \cdots & H_{q-1}^{exc} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (H_{q-1}^{exc})^t & \cdots & (H_2^{exc})^t & H_1^{exc} & H_2^{exc} \\ (H_q^{exc})^t & (H_{q-1}^{exc})^t & \cdots & (H_2^{exc})^t & H_1^{exc} \end{pmatrix}. \quad (3.1.5)$$

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

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

and, if $p \leq q$, by

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

Finally, let L^{exc} be the block diagonal matrix of order dp

$$L^{exc} = \begin{pmatrix} \Gamma + \Delta & 0 & \cdots & 0 & 0 \\ 0 & \Gamma + \Delta & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \Gamma + \Delta & 0 \\ 0 & 0 & \cdots & 0 & \Gamma + \Delta \end{pmatrix} \quad (3.1.6)$$

and denote as in Chapter 2, by Λ^{exc} the symmetric square matrix of order $\delta = d(p + q)$

$$\Lambda^{exc} = \begin{pmatrix} L^{exc} & (K^{exc})^t \\ K^{exc} & H^{exc} \end{pmatrix}. \quad (3.1.7)$$

The corresponding lemma about the invertibility of Λ^{exc} avoids the Strongly Controllable condition, as we can see next.

Lemma 3.1.1 *Let S be the Schur complement of L^{exc} in Λ^{exc}*

$$S = H^{exc} - K^{exc}(L^{exc})^{-1}(K^{exc})^t. \quad (3.1.8)$$

If (A_1) holds, then S and Λ^{exc} are invertible and

$$(\Lambda^{exc})^{-1} = \begin{pmatrix} (L^{exc})^{-1} + (L^{exc})^{-1}(K^{exc})^t S^{-1} K^{exc} (L^{exc})^{-1} & -(L^{exc})^{-1}(K^{exc})^t S^{-1} \\ -S^{-1} K^{exc} (L^{exc})^{-1} & S^{-1} \end{pmatrix}. \quad (3.1.9)$$

The proofs of all the results of this Chapter conserve the spirit and techniques of the ones exposed in the Chapter 2.

Proof. Let M and M' be the infinite-dimensional diagonal square matrices

$$M = \begin{pmatrix} \Gamma & 0 & \dots & \dots & \dots \\ 0 & \Gamma & 0 & \dots & \dots \\ \dots & 0 & \Gamma & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix},$$

$$M' = \begin{pmatrix} \Delta & 0 & \dots & \dots & \dots \\ 0 & \Delta & 0 & \dots & \dots \\ \dots & 0 & \Delta & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

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

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

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

while, if $p \leq q$, by

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

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

In addition, let V be, if $p \geq q$, the matrix with dq rows and dp columns

given, by

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

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

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

Finally, let N be the block diagonal matrix of order dp given by

$$N = \text{diag}(\Sigma, \cdots, \Sigma).$$

where $\Sigma = \Delta - \Delta(\Gamma + \Delta)^{-1}\Delta$ is a positive definite matrix.

First, if $p \geq q$, we can deduce from (3.1.8) after some straightforward, although rather lengthy, linear algebra calculations that

$$S = TMT^t + UM'U^t + VNV^t. \quad (3.1.10)$$

We now focus our attention into the last term in (3.1.10). We can clearly see that VNV^t is positive definite since N is positive definite. Thus S is invertible.

Now, if $p \leq q$, we can see from (3.1.8) that

$$S = TMT^t + UM'U^t + R \quad (3.1.11)$$

where

$$R = \begin{pmatrix} VNV^t & Z \\ Z^t & W \end{pmatrix}$$

with Z equal to the zeros matrix of order $dp \times d(q-p)$, and W is the block diagonal matrix of order $d(q-p)$

$$W = \text{diag}(\Delta, \dots, \Delta). \quad (3.1.12)$$

Then, if we look at the fact that VNV^t and W are positive definite, we can see that R is also positive definite and thus S is invertible.

Finally, as

$$\det(\Lambda) = \det(L) \det(S) = \det(\Gamma + \Delta)^p \det(S), \quad (3.1.13)$$

we obtain from (3.1.13) that Λ is invertible and formula (3.1.9) follows from [36] page 18, which completes the proof of Lemma 3.1.1 ■

3.2 Main Results

The first Theorem of this Chapter, which is an analogue to Theorem 2.2.2 concerns the almost sure properties of the LS algorithm when the persistently excited Aström and Wittenmark adaptive tracking control is used. However, as it was mentioned before, the spirit of the proof of Theorem 2.2.2 is present, this is why we will not display all the details.

Theorem 3.2.1 *Assume that the $ARX_d(p, q)$ model satisfies (A_1) and that (ε_n) has finite conditional moment of order $\alpha > 2$. Then, for the LS algorithm, we have*

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \Lambda^{exc} \quad a.s. \quad (3.2.1)$$

where the limiting matrix Λ^{exc} is given by (3.1.7). In addition, the tracking is residually optimal

$$\|C_n - \Delta_n\| = \mathcal{O}\left(\frac{\log n}{n}\right) \quad a.s. \quad (3.2.2)$$

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

$$\|\widehat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad a.s. \quad (3.2.3)$$

Proof. Once again, we shall make use of the same approach as Bercu [14] or Guo and Chen [31]. First of all, we recall that for all $n \geq 0$,

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1} + \xi_{n+1}. \quad (3.2.4)$$

It follows from (3.2.4) and the strong law of large numbers for martingales given in Corollary 1.3.25 of [20] that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \|X_k\|^2 \geq \text{tr}(\Gamma + \Delta) \quad a.s.$$

As Γ and Δ are positive definite, it clearly implies that $n = \mathcal{O}(s_n)$ a.s.

Moreover, by Theorem 1 of [7] or Lemma 1 of [31], we have

$$\sum_{k=1}^n (1 - f_k) \|\pi_k\|^2 = \mathcal{O}(\log s_n) \quad a.s. \quad (3.2.5)$$

where $f_n = \Phi_n^t S_n^{-1} \Phi_n$. Hence, if (ε_n) has finite conditional moment of order $\alpha > 2$, we can show via the minimum phase assumption on the matrix polynomial B together with (3.2.5) that $\|\Phi_n\|^2 = \mathcal{O}(s_n^\beta)$ a.s. for all $2\alpha^{-1} < \beta < 1$. In addition, let $g_n = \Phi_n^t S_{n-1}^{-1} \Phi_n$ and $\delta_n = \text{tr}(S_{n-1}^{-1} - S_n^{-1})$. It follows from Proposition 4.2.12 of [20] that

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

Moreover, as

$$\sum_{n=1}^{\infty} \delta_n = \sum_{n=1}^{\infty} \text{tr}(S_{n-1}^{-1} - S_n^{-1}) \leq \text{tr}(I_\delta) = \delta,$$

the sequence (δ_n) goes to zero a.s. Consequently, as

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

we find that

$$\sum_{k=1}^n \|\pi_k\|^2 = \sum_{k=1}^n (1 - f_k)(1 + g_k) \|\pi_k\|^2,$$

which leads to

$$\sum_{k=1}^n \|\pi_k\|^2 \leq 2 \sum_{k=1}^n (1 - f_k) \|\pi_k\|^2 + \sum_{k=1}^n (1 - f_k) \delta_k \|\Phi_k\|^2 \|\pi_k\|^2. \quad (3.2.6)$$

Hence, we deduce from (3.2.5) together with (3.2.6) that

$$\sum_{k=1}^n \|\pi_k\|^2 = o(s_n^\beta \log s_n) \quad \text{a.s.} \quad (3.2.7)$$

Therefore, we obtain from (2.0.1), (3.2.4) and (3.2.7) that

$$\sum_{k=1}^n \|X_{k+1}\|^2 = o(s_n^\beta \log s_n) + \mathcal{O}(n) \quad \text{a.s.} \quad (3.2.8)$$

Furthermore, as B is minimum phase, we find from relation (1.1.18) that

$$U_n = B^{-1}(R)A(R)X_{n+1} - B^{-1}(R)\varepsilon_{n+1} \quad (3.2.9)$$

which implies by (3.2.8) that

$$\sum_{k=1}^n \|U_k\|^2 = o(s_n^\beta \log s_n) + \mathcal{O}(n) \quad \text{a.s.} \quad (3.2.10)$$

It remains to put together the two contributions (3.2.8) and (3.2.10) to deduce that $s_n = o(s_n) + \mathcal{O}(n)$ a.s. so that $s_n = \mathcal{O}(n)$ a.s. Hence, it follows from (3.2.7) with $\beta < 1$ that

$$\sum_{k=1}^n \|\pi_k\|^2 = o(n) \quad \text{a.s.} \quad (3.2.11)$$

Consequently, we obtain from (2.0.1), (3.2.4), (3.2.11) and the strong law of large numbers for martingales [20] that

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

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

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

which implies that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k^p (X_k^p)^t = L^{exc} \quad \text{a.s.} \quad (3.2.12)$$

where L^{exc} is given by (3.1.6). Furthermore, it follows from (1.1.18), (3.2.4) and (3.2.9) that for all $n \geq 0$

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

where

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

Consequently, we deduce from the Cauchy-Schwarz inequality together with (2.0.1), (3.2.11), and the strong law of large numbers for martingales [20] that for all $1 \leq i \leq q$

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

which ensures that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n U_k^q (U_k^q)^t = H^{exc} \quad \text{a.s.} \quad (3.2.13)$$

where H is given by (3.1.5). Via the same lines, we also find that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k^p (U_{k-1}^q)^t = (K^{exc})^t \quad \text{a.s.} \quad (3.2.14)$$

Therefore, it follows from the conjunction of (3.2.12), (3.2.13) and (3.2.14) that

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \Lambda^{exc} \quad \text{a.s.} \quad (3.2.15)$$

where the limiting matrix Λ^{exc} is given by (3.1.7). Thanks to Lemma 3.1.1, the matrix Λ^{exc} is invertible. On the one hand, it follows from (3.2.15) that $n = \mathcal{O}(\lambda_{\min}(S_n))$, $\|\Phi_n\|^2 = o(n)$ a.s. which implies that f_n tends to zero a.s. Hence, by (3.2.5), we find that

$$\sum_{k=1}^n \|\pi_k\|^2 = \mathcal{O}(\log n) \quad \text{a.s.} \quad (3.2.16)$$

On the other hand, we obviously have from (3.2.4)

$$\|C_n - \Sigma_n\| = \mathcal{O}\left(\frac{1}{n} \sum_{k=1}^n \|\pi_{k-1}\|^2\right) \quad \text{a.s.} \quad (3.2.17)$$

Consequently, we immediately obtain the tracking residual optimality (3.2.2) from (3.2.16) and (3.2.17). Furthermore, by a well-known result of Lai and Wei [45] on the LS estimator, we also have

$$\|\hat{\theta}_{n+1} - \theta\|^2 = \mathcal{O}\left(\frac{\log \lambda_{\max} S_n}{\lambda_{\min} S_n}\right) \quad \text{a.s.} \quad (3.2.18)$$

Hence (3.2.3) clearly follows from (3.2.15) and (3.2.18), which completes the proof of Theorem 3.2.1. ■

The second Theorem of this Chapter offers similar results of those shown in Theorem 2.2.4 and it is related to the almost sure properties of the WLS algorithm of Bercu and Duflo, when the persistently excited adaptive tracking control expressed in (3.1.1) is employed.

Theorem 3.2.2 *Assume that the $ARX_d(p, q)$ model satisfies (A_1) . In addition, suppose that either (ε_n) is a white noise or (ε_n) has a finite conditional moment of order $\alpha > 2$. Then, for the WLS algorithm, we have*

$$\lim_{n \rightarrow \infty} (\log n)^{1+\gamma} \frac{S_n(a)}{n} = \Lambda^{exc} \quad \text{a.s.} \quad (3.2.19)$$

where the limiting matrix Λ^{exc} is given by (3.1.7). In addition, the tracking is residually optimal

$$\|C_n - \Delta_n\| = o\left(\frac{(\log n)^{1+\gamma}}{n}\right) \quad a.s. \quad (3.2.20)$$

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

$$\|\widehat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{(\log n)^{1+\gamma}}{n}\right) \quad a.s. \quad (3.2.21)$$

Proof. By Theorem 1 of [7], we have

$$\sum_{n=1}^{\infty} a_n(1 - f_n(a)) \|\pi_n\|^2 < +\infty \quad a.s. \quad (3.2.22)$$

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

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

with $\gamma > 0$, we clearly have $a_n^{-1} = \mathcal{O}(s_n)$ a.s. Hence, it follows from (3.2.22) together with Kronecker's lemma given e.g. by Lemma 1.3.14 of [20] that

$$\sum_{k=1}^n \|\pi_k\|^2 = o(s_n) \quad a.s. \quad (3.2.23)$$

Therefore, we obtain from (2.0.1), (3.2.4), (3.2.23) and the strong law of large numbers for martingales given in Theorem 4.3.16 of [20] that

$$\sum_{k=1}^n \|X_{k+1}\|^2 = o(s_n) + \mathcal{O}(n) \quad a.s. \quad (3.2.24)$$

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

$$\sum_{k=1}^n \|U_k\|^2 = o(s_n) + \mathcal{O}(n) \quad a.s. \quad (3.2.25)$$

Consequently, we immediately infer from (3.2.24) and (3.2.25) that $s_n = o(s_n) + \mathcal{O}(n)$ so $s_n = \mathcal{O}(n)$ a.s. Hence, (3.2.23) implies that

$$\sum_{k=1}^n \|\pi_k\|^2 = o(n) \quad \text{a.s.} \quad (3.2.26)$$

Via analogous arguments of those presented in the proof of Theorem 2.2.4 we have that

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

Via an Abel transform, it ensures that

$$\lim_{n \rightarrow \infty} (\log n)^{1+\gamma} \frac{S_n(a)}{n} = \Lambda^{exc} \quad \text{a.s.} \quad (3.2.27)$$

We obviously have from (3.2.27) that $f_n(a)$ tends to zero a.s. Consequently, we obtain from (3.2.22) and Kronecker's Lemma that

$$\sum_{k=1}^n \|\pi_k\|^2 = o((\log s_n)^{1+\gamma}) \quad \text{a.s.} \quad (3.2.28)$$

Then, (3.2.20) clearly follows from (3.2.17) and (3.2.28). Finally, by Theorem 1 of [7]

$$\|\hat{\theta}_{n+1} - \theta\|^2 = \mathcal{O}\left(\frac{1}{\lambda_{\min} S_n(a)}\right) \quad \text{a.s.} \quad (3.2.29)$$

Hence, we obtain (3.2.21) from (3.2.27) and (3.2.29), which completes the proof of Theorem 3.2.2 ■

Next we prove a CLT and a LIL similar to the ones exposed in Theorem 2.2.5.

Theorem 3.2.3 *Assume that the $ARX_d(p, q)$ model satisfies (A_1) and that ε and ξ have both a finite conditional moment of order $\alpha > 2$. In addition, suppose that (x_n) satisfies (2.0.1) and has the same regularity in norm as (ε_n) which means that for all $2 < \beta < \alpha$*

$$\sum_{k=1}^n \|x_k\|^\beta = \mathcal{O}(n) \quad \text{a.s.} \quad (3.2.30)$$

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

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (\Lambda^{exc})^{-1} \otimes \Gamma) \quad (3.2.31)$$

where the inverse matrix $(\Lambda^{exc})^{-1}$ is given by (3.1.9). In addition, for any vectors $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^\delta$, they also share the same law of iterated logarithm

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left(\frac{n}{2 \log \log n} \right)^{1/2} v^t (\hat{\theta}_n - \theta) u &= - \liminf_{n \rightarrow \infty} \left(\frac{n}{2 \log \log n} \right)^{1/2} v^t (\hat{\theta}_n - \theta) u \\ &= \left(v^t (\Lambda^{exc})^{-1} v \right)^{1/2} \left(u^t \Gamma u \right)^{1/2} \text{ a.s.} \end{aligned} \quad (3.2.32)$$

In particular,

$$\left(\frac{\lambda_{\min} \Gamma}{\lambda_{\max} \Lambda^{exc}} \right) \leq \limsup_{n \rightarrow \infty} \left(\frac{n}{2 \log \log n} \right) \|\hat{\theta}_n - \theta\|^2 \leq \left(\frac{\lambda_{\max} \Gamma}{\lambda_{\min} \Lambda^{exc}} \right) \text{ a.s.} \quad (3.2.33)$$

Proof. First of all, it follows from (1.1.22) and (1.3.1) that for all $n \geq 1$

$$\hat{\theta}_n - \theta = S_{n-1}^{-1}(a) M_n(a) \quad (3.2.34)$$

where

$$M_n(a) = \hat{\theta}_0 - \theta + \sum_{k=1}^n a_{k-1} \Phi_{k-1} \varepsilon_k^t. \quad (3.2.35)$$

We now make use of the CLT for multivariate martingales given by Lemma 2.2.2. On the one hand, for the LS algorithm, we clearly deduce (3.2.31) from convergence (3.2.1) and decomposition (3.2.34). On the other hand, for the WLS algorithm, we also infer (3.2.31) from convergence (3.2.19) and (3.2.34). Next, we make use of the LIL for multivariate martingales given by Lemma 2.2.3. For the LS algorithm, since ε has a finite conditional moment of order $\alpha > 2$, we obtain from Chow's Lemma given e.g. by Corollary 2.8.5 of [58] that for all $2 < \beta < \alpha$

$$\sum_{k=1}^n \|\varepsilon_k\|^\beta = \mathcal{O}(n) \quad \text{a.s.} \quad (3.2.36)$$

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

$$\sum_{k=1}^n \|\xi_k\|^\beta = \mathcal{O}(n) \quad \text{a.s.} \quad (3.2.37)$$

Consequently, as the reference trajectory (x_n) satisfies (3.2.30), we deduce from (3.2.4) together with (3.2.16), (3.2.36) and (3.2.37) that for some $2 < \beta < \alpha$

$$\sum_{k=1}^n \|X_k\|^\beta = \mathcal{O}(n) \quad \text{a.s.} \quad (3.2.38)$$

Furthermore, it follows from (2.2.10) and (3.2.38) that

$$\sum_{k=1}^n \|U_k\|^\beta = \mathcal{O}(n) \quad \text{a.s.} \quad (3.2.39)$$

Hence, we clearly obtain from (3.2.38) and (3.2.39) that

$$\sum_{k=1}^n \|\Phi_k\|^\beta = \mathcal{O}(n) \quad \text{a.s.} \quad (3.2.40)$$

Therefore, as $\beta > 2$, (3.2.40) immediately implies that

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

Finally, Lemma 2.2.3 together with convergence (3.2.1) and (3.2.34) lead to (3.2.32). The proof for the WLS algorithm of Bercu and Duflo follows essentially the same arguments than the proof for the LS algorithm and the proof for the WLS of Theorem 2.2.5 ■

3.3 Note on the Reference Trajectory

In the last section of this Chapter, we obtained asymptotical results for non necessarily Strongly Controllable $ARX_d(p, q)$ models tracking a reference

trajectory which satisfies condition (2.0.1) using a persistent excitation in the adaptive tracking control of Aström and Wittenmark. However it is possible to avoid both the Strongly Controllable assumption and the excitation in the control, when the reference trajectory satisfies the properties expressed by (1.5.1), (1.5.2) and (1.5.3).

To be able to develop the asymptotical analysis for the LS and WLS algorithms, we need to investigate the suitable matrix of order δ :

$$\Lambda = \begin{pmatrix} L & K^t \\ K & H \end{pmatrix}, \quad (3.3.1)$$

where for all $1 \leq i \leq q$, H_i is the square matrix of order d :

$$\begin{aligned} H_i &= \sum_{k=i}^{\infty} P_k \Gamma P_{k-i+1}^t + \sum_{k=0}^{i-1} \sum_{l=0}^{\infty} Q_k \nabla_{i+l-k-1} Q_l^t \\ &+ \sum_{k=i}^{\infty} \sum_{l=0}^{k-i+1} Q_k \nabla_{k+1-i-l}^t Q_l^t + \sum_{k=i}^{\infty} \sum_{l=k-i+2}^{\infty} Q_k \nabla_{i+l-k-1} Q_l^t. \end{aligned}$$

In addition, let H be the symmetric square matrix of order dq :

$$H = \begin{pmatrix} H_1 & H_2 & \cdots & H_{q-1} & H_q \\ H_2^t & H_1 & H_2 & \cdots & H_{q-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ H_{q-1}^t & \cdots & H_2^t & H_1 & H_2 \\ H_q^t & H_{q-1}^t & \cdots & H_2^t & H_1 \end{pmatrix}. \quad (3.3.2)$$

We also have $K_0 = \nabla + \sum_{k=1}^{\infty} Q_k \nabla_k^t$ and for all $1 \leq i \leq p$, let

$$K_i = P_i \Gamma + \sum_{k=0}^i Q_k \nabla_{i-k} + \sum_{k=i+1}^{\infty} Q_k \nabla_{k-i}^t.$$

Moreover, we have for all $1 \leq i \leq q-1$,

$$\tilde{J}_i = + \sum_{k=0}^{\infty} Q_k \nabla_{i+k}.$$

meanwhile K denotes the rectangular matrix of dimension $dq \times dp$ given, if $p \geq q$, by

$$K = \begin{pmatrix} K_0 & K_1 & K_2 & \cdots & \cdots & K_{p-2} & K_{p-1} \\ J_1 & K_0 & K_1 & \cdots & \cdots & K_{p-3} & K_{p-2} \\ J_2 & J_1 & K_0 & \cdots & \cdots & K_{p-4} & K_{p-3} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ J_{q-2} & \cdots & K_0 & K_1 & K_2 & \cdots & K_{p-q+1} \\ J_{q-1} & \cdots & J_1 & K_0 & K_1 & \cdots & K_{p-q} \end{pmatrix} \quad (3.3.3)$$

and, if $p \leq q$, by

$$\tilde{K} = \begin{pmatrix} K_0 & K_1 & K_2 & \cdots & \cdots & K_{p-2} & K_{p-1} \\ J_1 & K_0 & K_1 & \cdots & \cdots & K_{p-3} & K_{p-2} \\ J_2 & J_1 & K_0 & \cdots & \cdots & K_{p-4} & K_{p-3} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \cdots & J_1 & K_0 \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & J_1 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ J_{q-2} & \cdots & \cdots & \cdots & \cdots & \cdots & J_{q-p-1} \\ J_{q-1} & \cdots & \cdots & \cdots & \cdots & \cdots & J_{q-p} \end{pmatrix}. \quad (3.3.4)$$

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

$$L = \begin{pmatrix} \Gamma + \nabla & \nabla_1 & \cdots & \cdots & \nabla_{p-1} \\ \nabla_1^t & \Gamma + \nabla & \nabla_1 & \cdots & \nabla_{p-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \nabla_{p-2}^t & \cdots & \nabla_1^t & \Gamma + \nabla & \nabla_1 \\ \nabla_{p-1}^t & \cdots & \cdots & \nabla_1^t & \Gamma + \nabla \end{pmatrix}. \quad (3.3.5)$$

We present the following convergence

Theorem 3.3.1 *Assume that the matrix polynomial $B(z)$ associated with the $ARX_d(p, q)$ model is causal and that $(\varepsilon_n)_{n \geq 0}$ has finite conditional moment of order $\alpha > 2$. Then, for the LS algorithm, we have*

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \Lambda \quad \text{a.s.} \quad (3.3.6)$$

where the limiting matrix Λ is given by (3.3.1).

Proof. We shall make use of the same approach as Bercu [7] or Guo and Chen [31]. First of all, we recall that for all $n \geq 0$,

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1} \quad (3.3.7)$$

and

$$s_n = \sum_{k=0}^n \|\phi_k\|^2.$$

It follows from (3.3.7) and from with the strong law of large numbers for martingales (see e.g. Corollary 1.3.25 part 2 of [20]) that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \|X_k\|^2 \geq \text{tr}(\Gamma + \nabla) \quad \text{a.s.}$$

As $\Gamma + \nabla$ is invertible and $s_n \geq \sum_{k=1}^n \|X_k\|^2$ we have that $n = \mathcal{O}(s_n)$ a.s., which implies that

$$s_n \rightarrow \infty \quad \text{a.s.} \quad (3.3.8)$$

Moreover, by Theorem 1 of [6] or Lemma 1 of [31], we have

$$\sum_{k=1}^n (1 - f_k) \|\pi_k\|^2 = \mathcal{O}(\log s_n) \quad \text{a.s.} \quad (3.3.9)$$

where $f_n = \phi_n^t S_n^{-1} \phi_n$. Hence, as ε has finite conditional moment of order $\alpha > 2$, we can show by the minimum phase assumption (A_1) on the matrix polynomial B together with (3.3.9) [31] that $\|\phi_n\|^2 = \mathcal{O}(s_n^\beta)$ a.s. for all

$2\alpha^{-1} < \beta < 1$. In addition, let $g_n = \phi_n^t S_{n-1}^{-1} \phi_n$ and $\delta_n = \text{tr}(S_{n-1}^{-1} - S_n^{-1})$. It follows from Proposition 4.2.12 of [20] that

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

Moreover, as

$$\sum_{n=1}^{\infty} \delta_n = \sum_{n=1}^{\infty} \text{tr}(S_{n-1}^{-1} - S_n^{-1}) \leq \text{tr}(I_\delta) = \delta,$$

the sequence (δ_n) goes to zero a.s. Consequently, as

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

we find that

$$\sum_{k=1}^n \|\pi_k\|^2 = \sum_{k=1}^n (1 - f_k)(1 + g_k) \|\pi_k\|^2,$$

which leads to

$$\sum_{k=1}^n \|\pi_k\|^2 \leq 2 \sum_{k=1}^n (1 - f_k) \|\pi_k\|^2 + \sum_{k=1}^n (1 - f_k) \delta_k \|\Phi_k\|^2 \|\pi_k\|^2. \quad (3.3.10)$$

Hence, we deduce from (3.3.9) and (3.3.10) that

$$\sum_{k=1}^n \|\pi_k\|^2 = o(s_n^\beta \log s_n) \quad \text{a.s.} \quad (3.3.11)$$

Therefore, we obtain from (1.5.1), (3.3.7) and (3.3.11) that

$$\sum_{k=1}^n \|X_{k+1}\|^2 = o(s_n^\beta \log s_n) + \mathcal{O}(n) \quad \text{a.s.} \quad (3.3.12)$$

Furthermore, we infer from assumption (A₁) that

$$U_n = B^{-1}(R)A(R)X_{n+1} - B^{-1}(R)\varepsilon_{n+1}, \quad (3.3.13)$$

so we have from (3.3.13) that

$$\sum_{k=1}^n \|U_k\|^2 \leq 2 \sum_{k=1}^n \|B^{-1}(R)A(R)X_{n+1}\|^2 + 2 \sum_{k=1}^n \|B^{-1}(R)\varepsilon_{n+1}\|^2. \quad (3.3.14)$$

Relation (3.3.12) together with Cauchy-Schwarz inequality and (3.3.14) imply

$$\sum_{k=1}^n \|U_k\|^2 = o(s_n^\beta \log s_n) + \mathcal{O}(n) \quad \text{a.s.} \quad (3.3.15)$$

It remains to put together the two contributions (3.3.12) and (3.3.15) and remember the convergence (3.3.8) to deduce that $s_n = o(s_n) + \mathcal{O}(n)$ a.s. leading to $s_n = \mathcal{O}(n)$. Hence, it follows from this fact, (3.3.11) and (3.3.8) that

$$\sum_{k=1}^n \|\pi_k\|^2 = o(n) \quad \text{a.s.} \quad (3.3.16)$$

Since the reference trajectory is predictable and the law of large numbers (1.2.1) and relation (3.3.16) hold, using the Cauchy-Schwarz inequality, we obtain

$$\sum_{k=1}^n \|\pi_k \varepsilon_{k+1}^t\|^2 = o(n) \quad \text{a.s.} \quad (3.3.17)$$

$$\sum_{k=1}^n \|\pi_k x_{k+1}^t\|^2 = o(n) \quad \text{a.s.} \quad (3.3.18)$$

$$\sum_{k=1}^n \|x_k \varepsilon_{k+1}^t\|^2 = o(n) \quad \text{a.s.} \quad (3.3.19)$$

It follows from these expressions, the strong law of large numbers for ε and convergence (2.0.1) that

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

Observe that

$$\frac{1}{n} \sum_{k=1}^n \varepsilon_{k-m} \varepsilon_{k-j}^t \rightarrow \begin{cases} \Gamma & \text{if } m = j \\ 0 & \text{if } m \neq j \end{cases} \quad \text{a.s.} \quad (3.3.20)$$

which follows for $m = j$ from (1.2.1). On the other hand, if $m \neq j$ (3.3.20) results from the strong law of large numbers for martingales [20]. We also

have that, for all $1 \leq i \leq p-1$ that

$$X_k X_{k-i}^t = (x_k + \pi_{k-1} + \varepsilon_k) (x_{k-i} + \pi_{k-i-1} + \varepsilon_{k-i})^t.$$

Consequently, we obtain from the previous calculations that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k X_{k-i}^t = \nabla_i \quad \text{a.s.} \quad (3.3.21)$$

which implies that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k^p (X_k^p)^t = L \quad \text{a.s.} \quad (3.3.22)$$

where L is given by (3.3.5).

Furthermore, it follows from (1.1.18), (3.3.7) and assumption (A₁) that, for all $n \geq 0$,

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

where

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

Thus, we have

$$\begin{aligned} U_k U_{k-i+1}^t &= V_k V_{k-i+1}^t + V_k Y_{k-i+2}^t + V_k W_{k-i+2}^t + W_{k+1} V_{k-i+1}^t \\ &+ W_{k+1} Y_{k-i+2}^t + W_{k+1} W_{k-i+2}^t + Y_{k+1} V_{k-i+1}^t + Y_{k+1} Y_{k-i+2}^t \\ &+ Y_{k+1} W_{k-i+2}^t \end{aligned}$$

for all $1 \leq i \leq q$. Consequently, as $W_n = P(R)\varepsilon_n$, we deduce from the Cauchy-Schwarz inequality together with (1.5.1), (1.5.3), (2.2.12), the strong

law of large numbers for the driven noise and from the fact that x_k is $\mathcal{F}_{k-1}$ measurable that

$$\begin{aligned}
\frac{1}{n} \sum_{k=1}^n V_k V_{k-i+1}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n V_k W_{k-i+2}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n V_k Y_{k-i+2}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n W_{k+1} V_{k-i+1}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n W_{k+1} Y_{k-i+2}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n Y_{k+1} V_{k-i+1}^t &\rightarrow 0 \quad a.s. \\
\frac{1}{n} \sum_{k=1}^n Y_{k+1} W_{k-i+2}^t &\rightarrow 0 \quad a.s.
\end{aligned}$$

If we recall (3.3.20), then we have already proved that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n W_{k+1} W_{k-i+2}^t = \sum_{k=i}^{\infty} P_k \Gamma P_{k-i+1}^t. \quad (3.3.23)$$

In addition, one can observe that

$$Y_{k+1} Y_{k-i+1}^t = \left[\sum_{j=0}^k Q_j x_{k-j+1} \right] \left[\sum_{j=0}^{k-i+1} Q_j x_{k-i-j+2} \right]^t. \quad (3.3.24)$$

After some calculation, we rewrite

$$\begin{aligned}
Y_{k+1}Y_{k-i+1}^t &= \sum_{r=0}^{k-i+1} \sum_{j=i+r-1}^k Q_j x_{k-j+1} x_{k-j+r+1}^t Q_{j-r-i+1}^t \\
&+ \sum_{r=1}^k \sum_{j=r}^k Q_{j-r} x_{k-j+r+1} x_{k-j+1}^t Q_{j-i+1}^t \\
&= \sum_{r=0}^{i-1} \sum_{j=0}^{k-i+1} Q_r x_{k-r+1} x_{k-i-j+2}^t Q_j^t \\
&+ \sum_{r=i}^{k-1} \sum_{j=r-i+2}^{k-i+1} Q_r x_{k-r+1} x_{k-i-j+2}^t Q_j^t \\
&+ \sum_{r=i}^k \sum_{j=0}^{r-i+1} Q_r x_{k-r+1} x_{k-i-j+2}^t Q_j^t
\end{aligned}$$

which leads us to conclude that

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Y_{k+1} Y_{k-i+1}^t &= \sum_{r=0}^{i-1} \sum_{j=0}^{\infty} Q_r \nabla_{i+j-r-1} Q_j^t \\
&+ \sum_{r=i}^{\infty} \sum_{j=0}^{r-i+1} Q_r \nabla_{r+1-i-j}^t Q_j^t + \sum_{r=i}^{\infty} \sum_{j=r-i+2}^{\infty} Q_r \nabla_{i+j-r-1} Q_j^t
\end{aligned}$$

with $\nabla_0 = \nabla$. We have obtained that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n U_k U_{k-i+1}^t = H_i \quad \text{a.s.} \quad (3.3.25)$$

which ensures that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n U_k^q (U_k^q)^t = H \quad \text{a.s.} \quad (3.3.26)$$

where H is given by (3.3.2). Via similar arguments together with (3.3.7), (1.5.1), (1.5.2), (1.5.3) and (3.3.20), we also find for all $-1 \leq m \leq p-2$ and $2 \leq l \leq q$ that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n U_k X_{k-m}^t = K_{m+1} \quad \text{a.s.}$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n U_{k-l} X_n^t = J_{l-1} \quad \text{a.s.}$$

Then we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k^p (U_{k-1}^q)^t = K^t \quad \text{a.s.} \quad (3.3.27)$$

where K is given by (3.3.3) or (3.3.4).

Therefore, it follows from (3.3.22), (3.3.26) and (3.3.27) that

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \Lambda \quad \text{a.s.} \quad (3.3.28)$$

where the limiting matrix Λ is given by (3.3.1). ■

If the reference trajectory is such that ∇ is positive definite and for $i \geq 1$ $\nabla_i = 0$, then we have the following simplifications for the matrix Λ given by (3.3.1) which in this case is denoted by Λ_x

$$\begin{aligned} H_i &= \sum_{k=i}^{\infty} P_k \Gamma P_{k-i+1}^t + \sum_{k=i-1}^{\infty} Q_k \nabla Q_{k-i+1}^t \\ K_0 &= \nabla \\ K_r &= P_r \Gamma + Q_r \nabla \\ J_j &= 0 \end{aligned}$$

where for all $k \geq 0$, $Q_k = D_k + P_k$, in addition $1 \leq i \leq q$, $1 \leq j \leq q-1$ and $1 \leq r \leq p$. Finally, the block matrix L is the block diagonal matrix of order pd :

$$L = \begin{pmatrix} \Gamma + \nabla & 0 & \cdots & \cdots & 0 \\ 0 & \Gamma + \nabla & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \Gamma + \nabla & 0 \\ 0 & \cdots & \cdots & 0 & \Gamma + \nabla \end{pmatrix}.$$

We observe that such matrix Λ_x is similar to the matrix Λ^{exc} , we just need to replace Δ by ∇ , then we can prove similar results for both LS and WLS algorithms than the ones exposed in Section 3.2.

Lemma 3.3.1 *Let S be the Schur complement of L in Λ_x*

$$S = H - K(L)^{-1}(K)^t. \quad (3.3.29)$$

If (A_1) holds, then S and Λ_x are invertible and

$$\Lambda_x^{-1} = \begin{pmatrix} (L)^{-1} + (L)^{-1}(K)^t S^{-1} K(L)^{-1} & -(L)^{-1}(K)^t S^{-1} \\ -S^{-1} K(L)^{-1} & S^{-1} \end{pmatrix}. \quad (3.3.30)$$

Theorem 3.3.2 *Assume that the matrix polynomial associated with the $ARX_d(p, q)$ model satisfies the causality condition (A.1) and that $(\varepsilon_n)_{n \geq 0}$ has finite conditional moment of order $\alpha > 2$. In addition, if the reference trajectory $x = (x_n)_{n \geq 0}$ satisfies (1.5.1), (1.5.2) and (1.5.3) with for all $i \geq 1$ $\nabla_i = 0$ and ∇ a positive definite matrix, then for the LS algorithm, we have*

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \Lambda_x \quad a.s. \quad (3.3.31)$$

where the limiting matrix Λ_x is given in Remark 4. In addition, the tracking is optimal

$$\|C_n - \Gamma_n\| = \mathcal{O}\left(\frac{\log n}{n}\right) \quad a.s. \quad (3.3.32)$$

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

$$\|\hat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad a.s. \quad (3.3.33)$$

Theorem 3.3.3 *Assume that the matrix polynomial $B(z)$ associated with the $ARX_d(p, q)$ model is causal. In addition, suppose that either $(\varepsilon_n)_{n \geq 0}$ is a white noise or $(\varepsilon_n)_{n \geq 0}$ has a finite conditional moment of order $\alpha > 2$. Then, for the, we have*

$$\lim_{n \rightarrow \infty} (\log n)^{1+\gamma} \frac{S_n(a)}{n} = \Lambda_x \quad a.s. \quad (3.3.34)$$

In addition, the tracking is optimal

$$\|C_n - \Gamma_n\| = o\left(\frac{(\log n)^{1+\gamma}}{n}\right) \quad a.s. \quad (3.3.35)$$

Finally, $\hat{\theta}_n$ converges almost sure to θ

$$\|\hat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{(\log n)^{1+\gamma}}{n}\right) \quad a.s. \quad (3.3.36)$$

Theorem 3.3.4 *Assume that the matrix polynomial $B(z)$ associated with the $ARX_d(p, q)$ model is causal and that $(\varepsilon_n)_{n \geq 0}$ has a finite conditional moment of order $\alpha > 2$. In addition, suppose that $(x_n)_{n \geq 0}$ has the same regularity in norm as $(\varepsilon_n)_{n \geq 0}$ which means that for all $2 < \beta < \alpha$,*

$$\sum_{k=1}^n \|x_k\|^\beta = \mathcal{O}(n) \quad a.s. \quad (3.3.37)$$

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

$$\sqrt{n}(\hat{\theta}_n - \theta) \stackrel{\mathcal{L}}{\rightarrow} \mathcal{N}(0, \Lambda_x^{-1} \otimes \Gamma) \quad (3.3.38)$$

In addition, for any vectors $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^\delta$, they also share the same law of iterated logarithm

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

In particular,

$$\left(\frac{\lambda_{\min} \Gamma}{\lambda_{\max} \Lambda_x}\right) \leq \limsup_{n \rightarrow \infty} \left(\frac{n}{2 \log \log n}\right) \|\hat{\theta}_n - \theta\|^2 \leq \left(\frac{\lambda_{\max} \Gamma}{\lambda_{\min} \Lambda_x}\right) \quad a.s. \quad (3.3.39)$$

3.4 Numerical Experiments

In this section we present some numerical experiments in order to illustrate the results obtained in this Chapter corresponding to the LS algorithms using a persistently excited adaptive tracking control. First we consider a Strongly Controllable model for showing that we still have the asymptotic properties obtained in the previous Chapter but with a different associated limit matrix. Next, we consider a non Strongly Controllable model and we will obtain the almost sure convergence of the LS algorithm as well as a CLT, in the third experiment, we chose the reference trajectory to be a random *i.i.d.* sequence, while in the last one it is chosen constant. All the simulations will be based on 400 realizations of a sample size 10000. For sake of simplicity we consider in the first two experiments the reference trajectory to be identically zero and the driven and exogenous noises as white Gaussians $\mathcal{N}(0, I_2)$.

Example 1. We consider now the $ARX_2(1, 1)$ model,

$$X_{n+1} = AX_n + U_n + BU_{n-1} + \varepsilon_{n+1} \quad (3.4.1)$$

with

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \text{and} \quad B = \frac{1}{4} \begin{pmatrix} 1/5 & 0 \\ 0 & -1/3 \end{pmatrix}$$

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

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

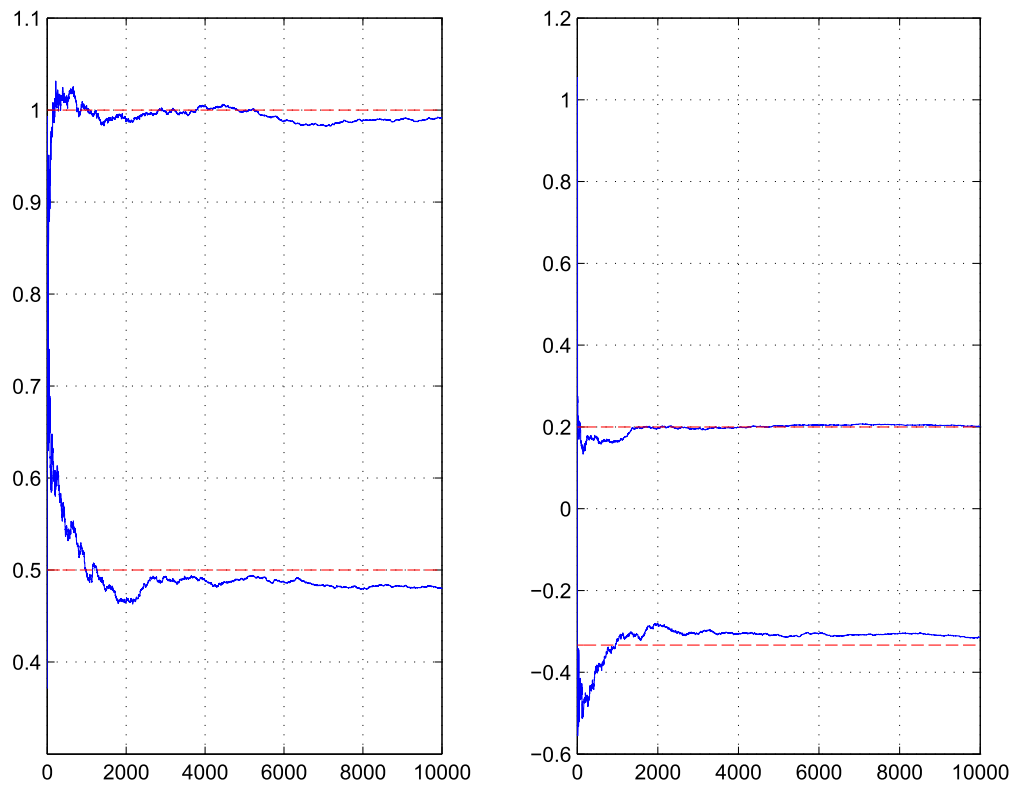


Figure 3.1: Almost sure convergence of LS algorithm for a Strongly Controllable model

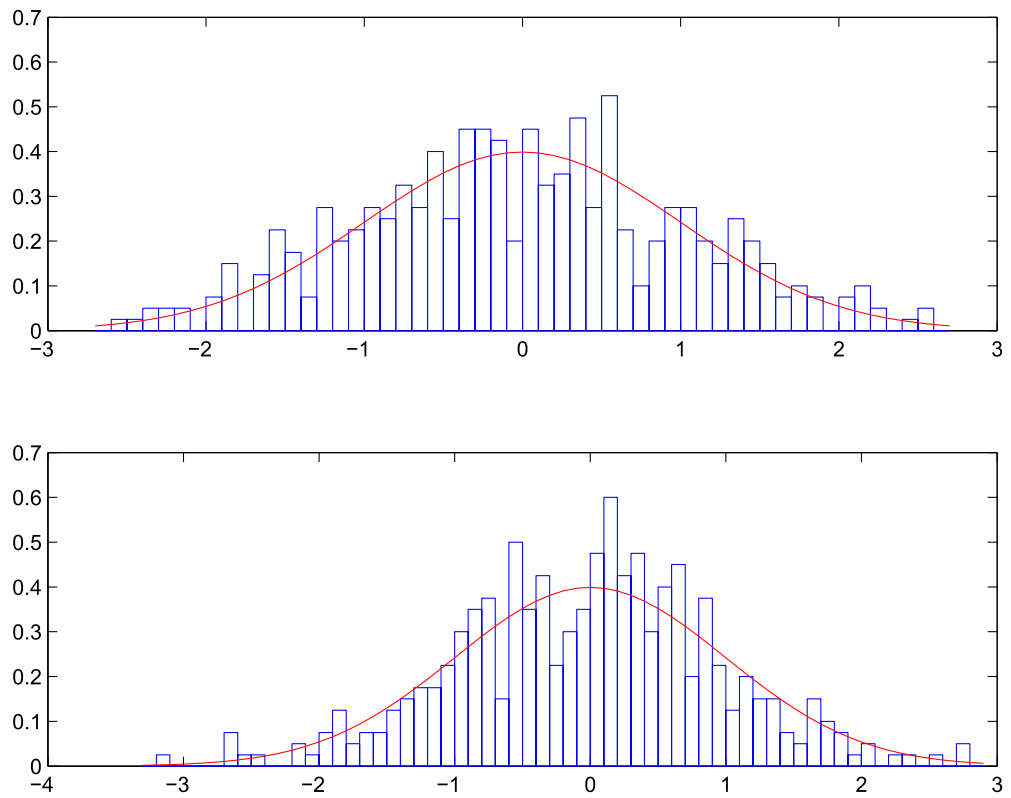


Figure 3.2: CLT of LS algorithm for a Strongly Controllable model

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

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

Then, we can obtain that the matrices H^{exc} , L^{exc} , K^{exc} and Λ^{exc} are respectively given by

$$H^{exc} = \frac{1}{48} \begin{pmatrix} 170 & 0 \\ 0 & 63 \end{pmatrix}.$$

$$L^{exc} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

$$K^{exc} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

$$\Lambda^{exc} = \frac{1}{48} \begin{pmatrix} 96 & 0 & 48 & 0 \\ 0 & 96 & 0 & 48 \\ 48 & 0 & 170 & 0 \\ 0 & 48 & 0 & 63 \end{pmatrix}.$$

We observe that $\det(\Lambda^{exc}) \neq 0$. We show in Figures 3.1 and 3.2 respectively the almost sure convergence and the CLT of the LS algorithm for the diagonal coordinates of matrices A and B .

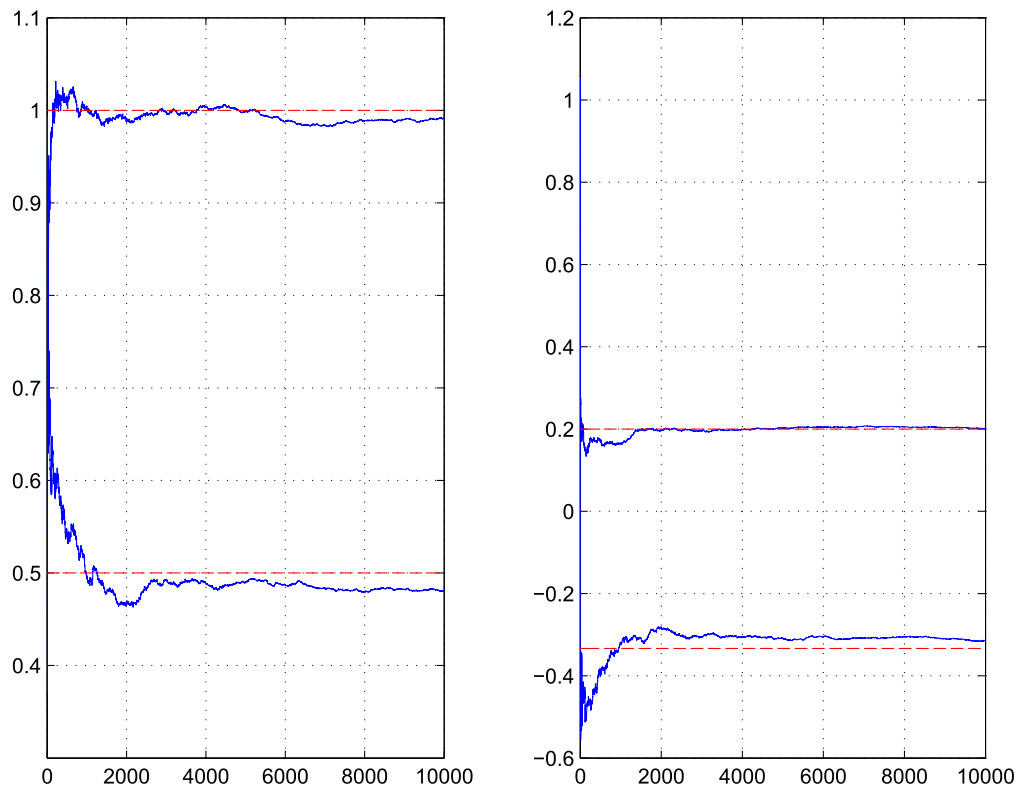


Figure 3.3: Almost sure convergence of the LS algorithm

Example 2. We consider now the same $ARX_2(1, 1)$ model than in example 2 of Chapter 2

$$X_{n+1} = AX_n + U_n + BU_{n-1} + \varepsilon_{n+1} \quad (3.4.2)$$

with

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

since this $ARX_2(1, 1)$ process is not strongly controllable, we use the persistently excited adaptive tracking control given in (3.1.1). In this example, we have that

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

Therefore, the limiting matrix Λ^{exc} given by (3.1.7) is

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

We can see that $\det(\Lambda^{exc}) \neq 0$. It is interesting to remember that for the same model with the Aström and Wittenmark adaptive tracking control without the persistent excitation the matrix Λ is not invertible. We can see in Figure 3.3 that the parameter θ is correctly estimated, while in Figures 3.4 and 3.5 we can see that the CLT is satisfied for the estimation of the four diagonal elements of matrices A and B .

Example 3. In order to illustrate the almost sure convergence (3.3.33) and the CLT shown in Theorem 3.3.4, consider now, a non Strongly Controllable $ARX_2(1, 1)$ model with

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

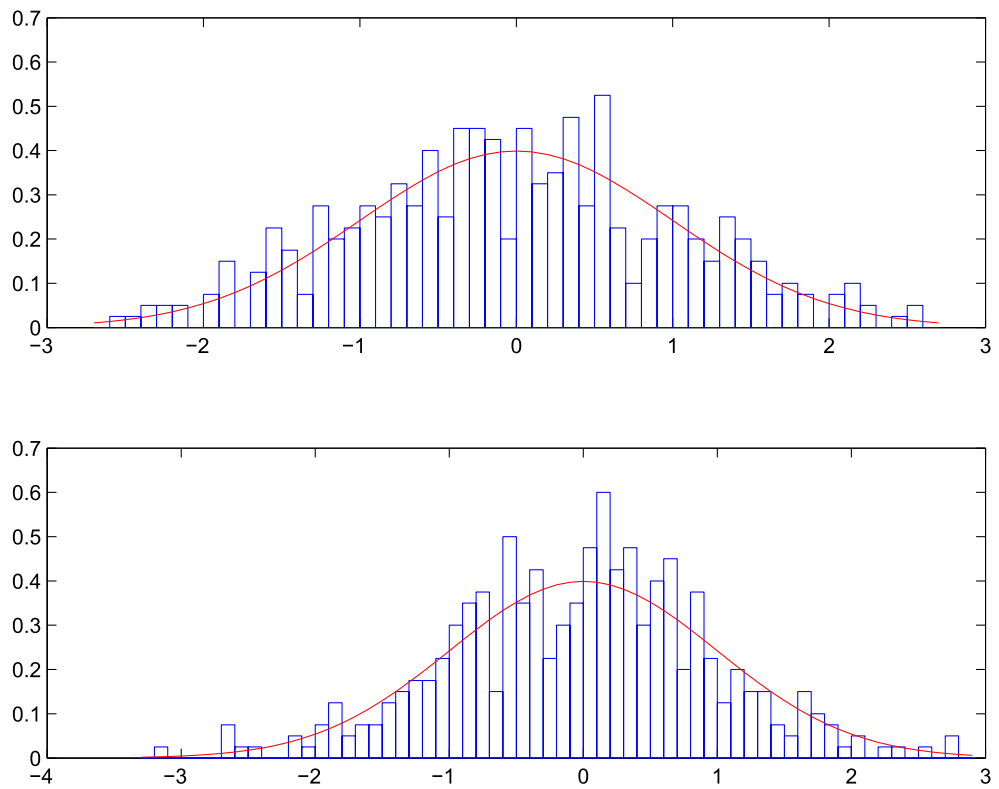


Figure 3.4: CLT of the LS algorithm for matrix A

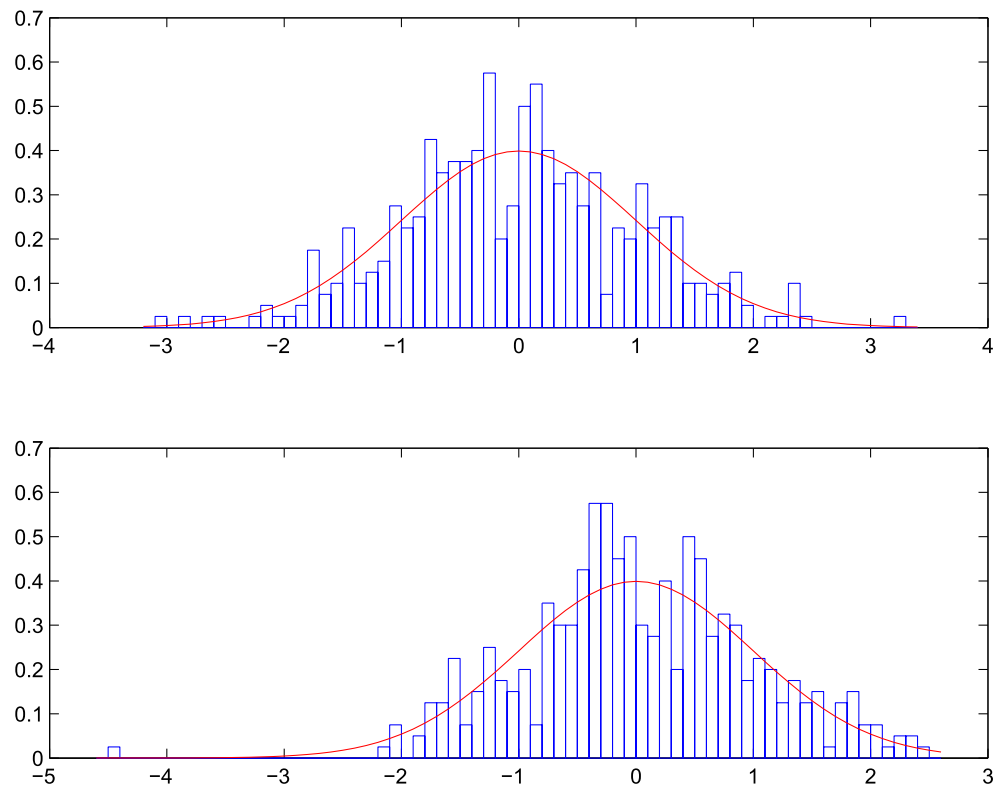


Figure 3.5: CLT of the LS algorithm for matrix B

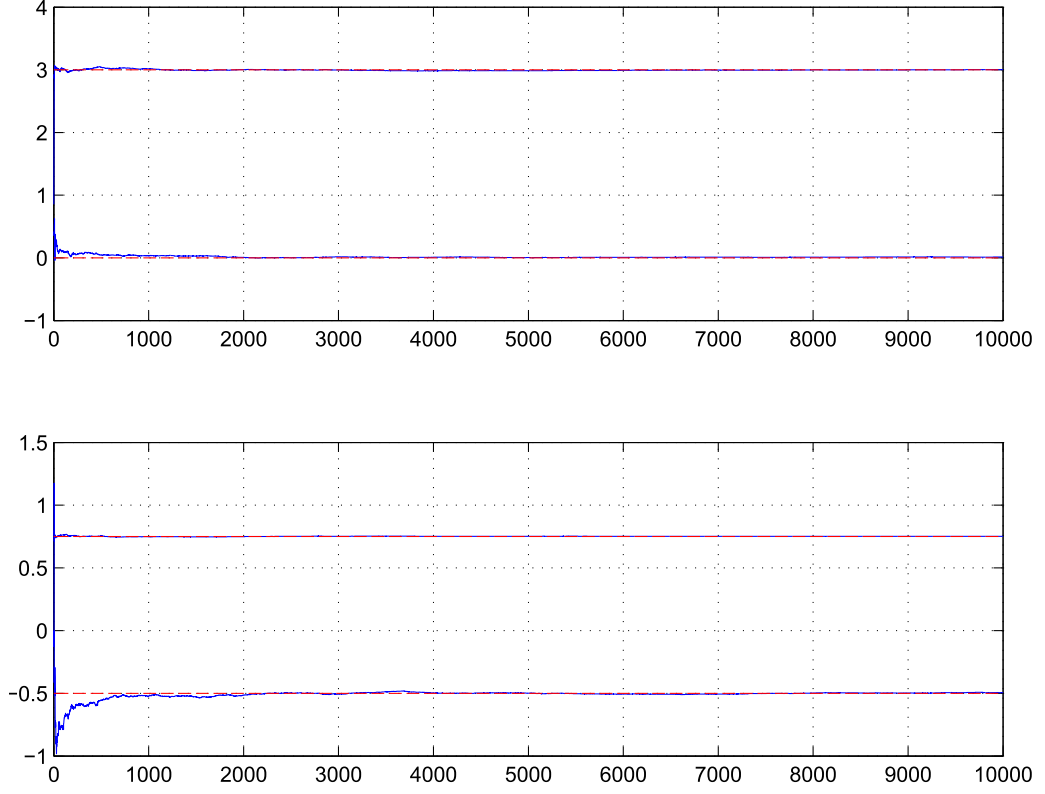


Figure 3.6: Almost sure convergence for LS algorithm

the reference trajectory $x = (x_n)_{n \geq 0}$ is an *i.i.d.* sequence with zero mean such that $\mathbb{E}[x_{n+1}x_{n+1}^t | \mathcal{F}_n] = I_2$ a.s. We can compute the matrix Λ_x noticing that $L = 2I_2$, $K = K^t = I_2$ and

$$H = \frac{1}{21} \begin{pmatrix} 1128 & 0 \\ 0 & 28 \end{pmatrix}. \quad (3.4.3)$$

We find that

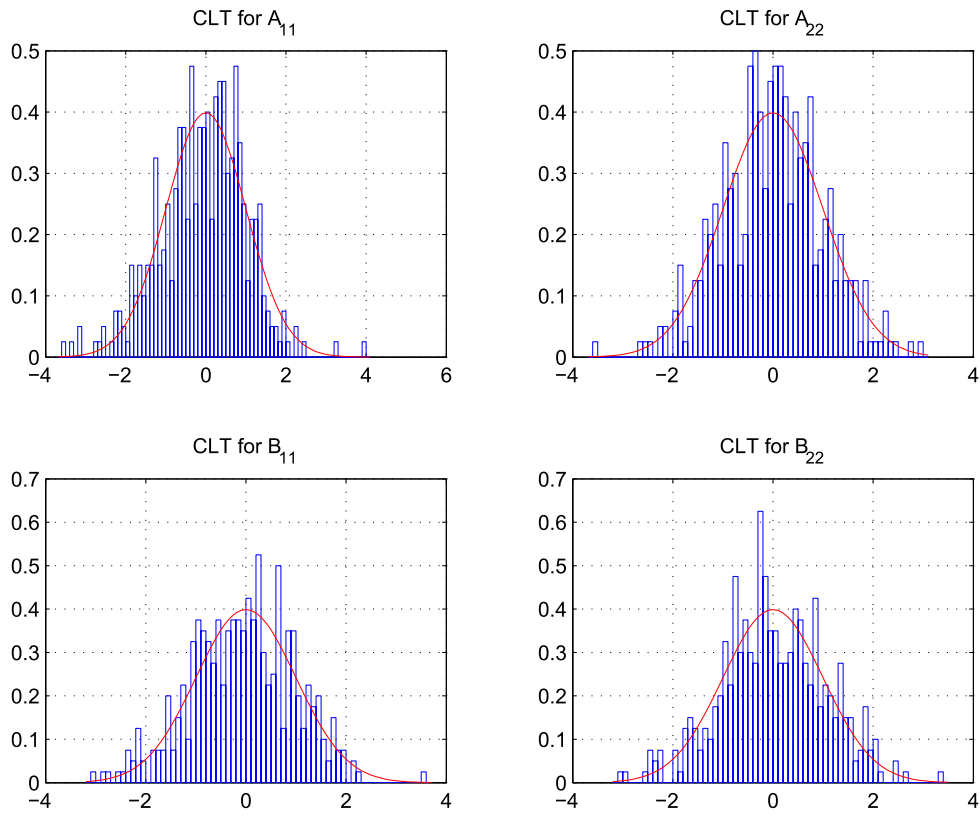


Figure 3.7: CLT for LS algorithm

$$\Lambda_x = \frac{1}{21} \begin{pmatrix} 42 & 0 & 21 & 0 \\ 0 & 42 & 0 & 21 \\ 21 & 0 & 1128 & 0 \\ 0 & 21 & 0 & 28 \end{pmatrix}.$$

We have also that $\det(H) = 71.619$, $\det(H - KL^{-1}K^t) = 44.3452$ and $\det(\Lambda_x) = 177.3810$. Figure 3.6 shows the almost sure convergence of the LS algorithm, while in Figure 3.7, we observe the CLT for every coordinate of the LS estimator.

Example 4. Even though we did not prove any asymptotical result when the reference trajectory is such that $\nabla \neq 0$ and $\nabla_i \neq 0$ but not invertible, we present the following experiment. Considerer the same $ARX_2(1,1)$ model as in example 3; same matrices A and B , with a constant reference trajectory

$$x_n = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (3.4.4)$$

We clearly have

$$\nabla = \nabla_i = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.4.5)$$

We find that $L = I_2 + \nabla$, $K = K_0 = \nabla - (I_2 + B)^{-1}(A + B)\nabla$. Consequently

$$L = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad (3.4.6)$$

$$K = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}. \quad (3.4.7)$$

In addition

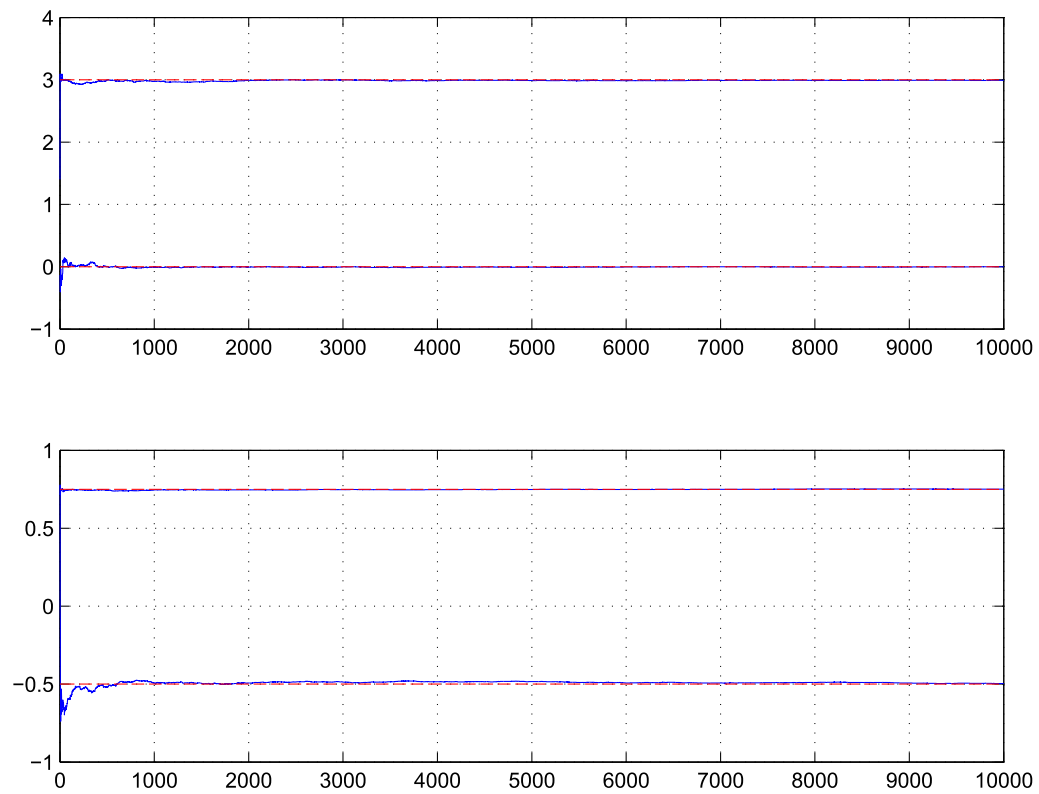


Figure 3.8: Almost sure convergence for LS algorithm

$$H = \frac{1}{7} \begin{pmatrix} 144 & 0 \\ 0 & 28 \end{pmatrix} \quad (3.4.8)$$

and

$$\Lambda = \frac{1}{7} \begin{pmatrix} 7 & 0 & 0 & 0 \\ 0 & 14 & 0 & 14 \\ 0 & 0 & 144 & 0 \\ 0 & 14 & 0 & 28 \end{pmatrix}$$

We have $\det(H) = 82.2857$, $\det(H - KL^{-1}K^t) = 41.1429$ and $\det(\Lambda) = 82.2857$. We have found that a non Strongly Controllable $ARX_d(p, q)$ model tracking a non zero constant reference trajectory holds similar asymptotical results of those proved in this Chapter. As a matter of fact, we observe in Figures 3.8 and 3.9 respectively the almost sure convergence and the CLT of the LS algorithm. Those simulation results are of course not surprising since $\det(\Lambda) > 0$.

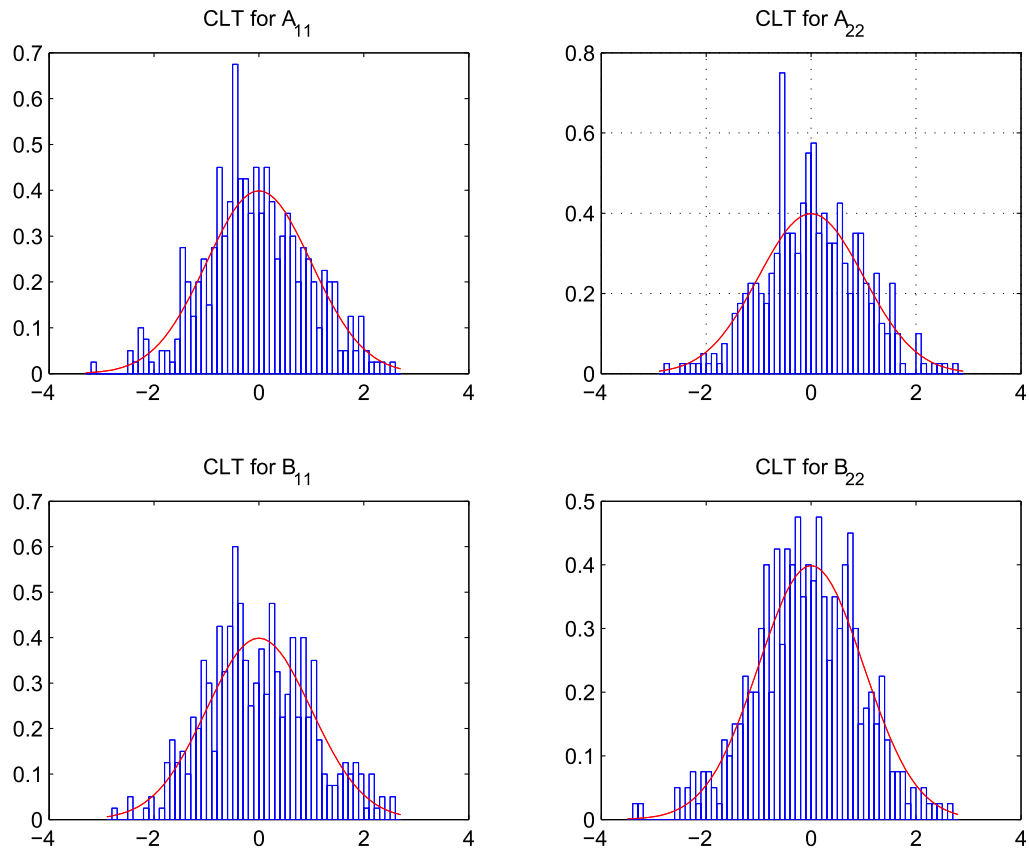


Figure 3.9: CLT for LS algorithm

Chapter 4

Work in Progress on the Durbin Watson Statistic

The Durbin-Watson statistic has been used to detect the presence of a first order autocorrelated noise in linear regression models with the independent variables regarded as "fixed", and it was introduced by J. Durbin and G. S. Watson in their works [22], [23]. Since their pioneer work the procedure around this statistic has been the main topic of a large variety of theoretical papers [21], [51], [52], [54], [57], [62]. In addition the Durbin Watson procedure has been employed in application areas such as Electrochemical Science [24], Crystallography [34], Pharmaceutics [60] and Chemometrics [63].

In this Chapter, we focus our attention on the Durbin-Watson statistic, the LS algorithm and their asymptotic behavior in the $AR_1(p)$ adaptive tracking framework with a noise which may have a first order serial autocorrelation.

4.1 Preliminaries

We will work on the following one dimensional regression models:

$$X_{n+1} = \theta^t \phi_n + U_n + \varepsilon_{n+1}, \quad (4.1.1)$$

with $\theta^t = (a_1, \dots, a_p)$ and $\phi_n^t = (X_n, \dots, X_{n-p+1})$.

In addition, we suppose that the driven noise $(\varepsilon_n)_{n \geq 0}$ satisfies the autoregressive $AR_1(1)$ model given, for all $n \geq 0$, by

$$\varepsilon_{n+1} = \rho \varepsilon_n + V_{n+1}, \quad (4.1.2)$$

where $|\rho| < 1$ is the first-order autocorrelation parameter and $(V_n)_{n \geq 0}$ is a martingale difference sequence adapted to a filtration $\mathbb{F}$ such that for all $n \geq 0$, $\mathbb{E}[V_{n+1}^2 | \mathcal{F}_n] = \sigma^2 > 0$ a.s. Moreover, if

$$\gamma_n = \frac{1}{n} \sum_{k=1}^n V_k^2, \quad (4.1.3)$$

we suppose that $(\gamma_n)_{n \geq 0}$ converges almost surely to σ^2 .

We are now in position to list some straightforward properties of the driven noise $(\varepsilon_n)_{n \geq 0}$:

1. The initial value $\varepsilon_0 = V_0$ and for all $n \geq 0$

$$\varepsilon_n = \sum_{k=0}^n \rho^{n-k} V_k \quad (4.1.4)$$

2. For all $n \geq 0$ and $h \geq 0$,

$$\begin{aligned} \mathbb{E}[\varepsilon_n] &= 0, \\ \text{Var}[\varepsilon_n] &= \mathbb{E}[\varepsilon_n^2] = \frac{\sigma^2 (1 - \rho^{2n+1})}{1 - \rho^2}, \\ \text{Cov}[\varepsilon_n, \varepsilon_{n+h}] &= \mathbb{E}[\varepsilon_n \varepsilon_{n+h}] = \rho^h \sigma^2 \frac{1 - \rho^{2n+1}}{1 - \rho^2} \end{aligned}$$

which leads us to

$$\begin{aligned} \text{Cov}(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n) = \\ \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix} 1 - \rho^2 & \rho(1 - \rho^2) & \rho^2(1 - \rho^2) & \dots & \rho^n(1 - \rho^2) \\ \rho(1 - \rho^2) & 1 - \rho^4 & \rho(1 - \rho^4) & \dots & \rho^{n-1}\rho(1 - \rho^4) \\ \rho^2(1 - \rho^2) & \rho(1 - \rho^4) & 1 - \rho^6 & \dots & \rho^{n-2}(1 - \rho^6) \\ \dots & \dots & \dots & \dots & \dots \\ \rho^n(1 - \rho^2) & \rho^{n-1}(1 - \rho^4) & \rho^{n-2}(1 - \rho^6) & \dots & 1 - \rho^{2(n+1)} \end{pmatrix}. \end{aligned}$$

3. We can easily prove the following convergences:

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n \varepsilon_k &\rightarrow 0 \quad \text{a.s.} \\ \frac{1}{n} \sum_{k=1}^n \varepsilon_k^2 &\rightarrow \sigma^2(1 - \rho^2)^{-1} \quad \text{a.s.} \end{aligned}$$

4. For all $n \geq 0$ we also have

$$\begin{aligned} \mathbb{E}[\varepsilon_{n+1} | \mathcal{F}_n] &= \rho \varepsilon_n \quad \text{a.s.} \\ \mathbb{E}[\varepsilon_{n+1}^2 | \mathcal{F}_n] &= \rho^2 \varepsilon_n^2 + \sigma^2 \quad \text{a.s.} \end{aligned}$$

Remark 4 *One can observe in the $AR_1(p)$ case that, if $\rho = 0$, then we obtain exactly the same model studied in [14] in dimension $d = 1$. Moreover in the $ARX_1(p, q)$ framework, if $\rho = 0$, then we are in the cases considered in the previous Chapters.*

The Durbin-Watson procedure is based on the LS or WLS residuals which are defined next:

Definition 4.1.1 *Suppose that we estimate the parameter of a linear model of the form given in equation (4.1.1) by the LS or the WLS algorithm, then the residuals of such algorithm are defined for all $n \geq 1$ by*

$$\hat{\varepsilon}_n = X_n - U_{n-1} - \hat{\theta}_n^t \phi_{n-1}. \quad (4.1.5)$$

We can define now the Durbin-Watson statistic in order to detect the first order serial correlated noise:

Definition 4.1.2 *Assume that we estimate the parameter of a linear model of the form given in equation (4.1.1) by the LS or the WLS algorithm, then the Durbin-Watson statistic is defined by*

$$D_n = \frac{\sum_{k=1}^n (\hat{\varepsilon}_k - \hat{\varepsilon}_{k-1})^2}{\sum_{k=1}^n \hat{\varepsilon}_k^2} \quad (4.1.6)$$

where $\hat{\varepsilon}_0 = 0$ and $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ are given in Definition 4.1.1.

In addition, in the classical regression framework, the correlation parameter ρ is often estimated by [62]

$$\hat{\rho}_n = \frac{\sum_{k=1}^n \hat{\varepsilon}_k \hat{\varepsilon}_{k-1}}{\sum_{k=1}^n \hat{\varepsilon}_k^2}. \quad (4.1.7)$$

We can obtain the following relation between the Durbin-Watson statistic and the estimator of the correlation parameter at step $n \geq 1$:

$$D_n = 2(1 - \hat{\rho}_n) + \delta_n \quad (4.1.8)$$

where

$$\delta_n = \frac{\hat{\varepsilon}_n^2}{\sum_{k=1}^n \hat{\varepsilon}_k^2}. \quad (4.1.9)$$

This expression shows a duality between D_n and $\hat{\rho}_n$; as a matter of fact, small values of $\hat{\rho}_n$ lead us to obtain values for D_n close to two, while an estimation of ρ not so close to zero indicates that D_n is far from two.

If in fact we have $\rho = 0$ and we work with an $AR_1(p)$ or an $ARX_1(p, q)$ models then it is not hard to see that:

$$\hat{\rho}_n \rightarrow 0 \quad \text{a.s.} \quad (4.1.10)$$

$$D_n \rightarrow 2 \quad \text{a.s.} \quad (4.1.11)$$

The Durbin-Watson statistic has been successfully used for contrasting the hypothesis

$$H_0 : \rho = 0 \quad \text{vs} \quad H_1 : \rho \neq 0; \quad (4.1.12)$$

when the regression variables can be treated as constants [21], [22], [23] . It has been shown that exact critical values for this test can not be obtained since its distribution depends on the regression variables [55]. However, Durbin and Watson [23] found bounds D_U and D_L to those critical values, such bounds have been tabulated independently of the regression variables [23]. In this case the test says that if the observed statistic D_n is less than D_L or bigger than $4 - D_L$ we reject H_0 , if D_n lies between D_U and $4 - D_U$ we conclude H_1 and the test is inconclusive otherwise. In addition, notice that $D_n = 0$ when $\hat{\varepsilon}_k = \hat{\varepsilon}_{k-1}$ for all $k \geq 0$ and $D_n = 4$ when $\hat{\varepsilon}_k = -\hat{\varepsilon}_{k-1}$ for all $k \geq 0$, while a value of D_n close to 2 indicates a low or zero valued ρ .

It has also been argued that the Durbin-Watson statistic is biased toward 2 when lagged dependent variables are included in the model and a least squares estimation is employed [21], [51], [54], making the Durbin-Watson test inappropriate for the $AR_1(p)$ scenario. However, it is interesting for us to investigate such convergences and the asymptotic properties of the LS algorithm keeping in mind the words of Johnston [40] "...if we combine the two complications of lagged variables and autocorrelated residuals, things get really bad" .

4.2 The DW Statistic in the $AR_1(p)$ Framework

In this section, we present a series of numerical results which will lead us to formulate some conjectures about the asymptotical properties of the Durbin-Watson statistic and the LS algorithm in the $AR_1(p)$ adaptive framework.

a	σ^2	ρ	$\hat{a}_n$	$\hat{\rho}_n$	D_n	S_n/n
2.5	0.5	-0.7	1.802	-0.006	2.01	0.3393
2.5	1	-0.7	1.8002	-0.008	2.015	1.29
2.5	3	-0.7	1.804	-0.015	2.03	12.8653
0.5	1	-0.7	-0.202	-0.0035	2.006	1.323
1	1	-0.7	0.3018	-0.0036	2.007	1.342
1	1	0.3	1.2989	-0.0004	2.0071	1.0082
1	1	0.8	1.799	-0.0025	1.9998	1.6938
1	1	-0.1	0.9019	0.0021	1.996	1.0001
-1.5	1	-0.1	-1.5999	0.0017	1.9965	1.0001
0.3	1	-0.4	-0.0982	-0.0018	2.0015	1.0263

Chart 1: Average of 20 repetitions of sample size 10000, $AR_1(1)$.

We consider first some numerical experiments for the $AR_1(1)$ case:

$$\begin{aligned} X_{n+1} &= aX_n + U_n + \varepsilon_{n+1}, \quad \text{and} \\ \varepsilon_{n+1} &= \rho\varepsilon_n + V_{n+1}, \end{aligned}$$

where $(V_n)_{n \geq 0}$ is an i.i.d. $N(0, \sigma^2)$ sequence and for the sake of simplicity the reference trajectory is chosen identically zero. In Chart 1 we present twelve sets of parameters for the model presented above, for every one of them we present the average of 20 repetitions of a sample of size 10 000 for the LS estimation for the parameter a , the statistics $\hat{\rho}_n$, D_n , and the value of the matrix S_n/n .

We observe from Chart 1 that the value of $\hat{a}_n$ is not affected by σ^2 , in particular, we observe this from the first three rows. In addition, we may conjecture from the fourth column that

$$\lim_{n \rightarrow \infty} \hat{a}_n = a + \rho \quad \text{a.s.} \quad (4.2.1)$$

From the first three rows we conjecture that the LS estimation, $\hat{\rho}_n$ and D_n do not depend on the value of σ^2 , only the matrix S_n/n depends on its value. We already expected this situation since Bercu [7] and ourselves in Chapter

a_1	a_2	a_3	σ^2	ρ	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{\rho}_n$	D_n	S_n/n
1	1.6	-2.5	0.25	-0.8	0.1963	2.2195	-2.9419	-0.0765	2.1125	0.4185
1	1.6	-2.5	1	-0.8	0.1984	2.2305	-2.9793	-0.0453	2.0905	1.1816
1	1.6	-2.5	3	-0.8	0.1989	2.3337	-2.9839	-0.0348	2.0695	3.6001
1	1.6	-2.5	0.5	-0.8	0.1977	2.2288	-2.9764	-0.0414	2.0825	0.6055
2	1.6	-2.5	1	-0.8	1.1951	2.2320	-2.9818	-0.0483	2.0964	1.1954
1.5	1.6	-2.5	1	-0.8	0.7	2.2236	-2.9700	-0.0515	2.1028	1.2001
1	-1.1	-2.5	1	-0.8	0.205	-0.4888	-2.9566	-0.0708	2.1414	1.2111
1	-0.5	-2.5	1	-0.8	0.1979	0.1236	-2.9707	-0.0510	2.1019	1.1885
1	1.6	-1.5	1	-0.8	0.2003	2.2294	-1.9800	-0.0455	2.0908	1.1988
1	1.6	1	1	-0.8	0.1946	2.2401	0.5115	-0.0254	2.0505	1.2000
1	1.6	-2.5	1	0.8	1.8034	2.2239	-2.0366	0.0623	1.8851	1.2117
1	1.6	-2.5	1	0.5	1.4918	1.8315	-2.4013	0.0234	1.9529	1.0124
1	1.6	-2.5	1	0.2	1.1992	1.6365	-2.4960	0.0090	1.9819	1.0024

Chart 2: Average of 20 repetitions of sample size 10000, $AR_1(3)$.

2 obtained, when $\rho = 0$, that in this case S_n/n converges almost surely to σ^2 .

In addition, we can glimpse that the Durbin-Watson statistic converges almost surely to 2 and $\hat{\rho}_n$ converges almost surely to zero.

We put now our attention on the $AR_1(3)$ model:

$$\begin{aligned} X_{n+1} &= a_1 X_n + a_2 X_{n-1} + a_3 X_{n-2} + U_n + \varepsilon_{n+1}, \quad \text{and} \\ \varepsilon_{n+1} &= \rho \varepsilon_n + V_{n+1}, \end{aligned}$$

where once again $(V_n)_{n \geq 0}$ is an i.i.d. $N(0, \sigma^2)$ sequence and the reference trajectory is chosen identically zero. In Chart 2 we present fifteen combinations of the parameters of the previous model, for every one of them we provide the average of 20 repetitions of a sample of size 10 000 for the LS estimation for the parameter θ , the statistics $\hat{\rho}_n$, D_n and the value of the matrix S_n/n .

As it was done with Chart 1 we can make some conjectures about Chart 2. It can be seen from rows 1 – 4 that only the value of S_n/n is affected

significantly by changes in the value of σ^2 . We also observe that the LS estimations for a_1 have a similar behavior than the ones for a in Chart 1, more precisely the LS estimations for a_1 seems to be very close to $a_1 + \rho$, less obvious is the behavior of the LS estimations for a_2 and a_3 , but it can be seen that they are very close to $a_2 + \rho^2$ and $a_3 + \rho^3$ respectively, as a matter of fact if we consider the parameters of the sixth rows of Chart 2; $a_1 = 2$, $a_2 = 1.6$, $a_3 = -2.5$, $\sigma^2 = 1$ and $\rho = -0.8$, then we have from the same row of the same Chart that

$$\begin{aligned}\hat{a}_1 - (a_1 + \rho) &= -0.0049, \\ \hat{a}_2 - (a_2 + \rho^2) &= -0.008, \\ \hat{a}_3 - (a_3 + \rho^3) &= 0.0302.\end{aligned}$$

Finally, we conjecture that the Durbin-Watson statistic may be asymptotically biased toward two and thus $\hat{\rho}_n$ is biased toward zero.

In order to attack the general $AR_1(p)$ tracking problem, we need to recall some notation from [7] and Chapter 1. Consider the $AR_1(p)$ model obtained from relations (1.1.2), (1.1.3), (1.1.4) and (1.1.5), where the noise $(\varepsilon_n)_{n \geq 0}$ satisfies (4.1.2) and U_n is the adaptive tracking control given for all $n \geq 0$ in relation (1.4.4), and the reference trajectory satisfies the convergence (2.0.1). In addition, let us define the following parameter

$$\theta^* = \theta + \mu \tag{4.2.2}$$

where $\theta^t = (a_1, \dots, a_p)$ and $\mu^t = (\rho, \rho^2, \dots, \rho^p)^t$. Next let us define the sequence $(\pi_n)_{n \geq 0}$ in the following way

$$\pi_n = \left(\theta^* - \hat{\theta}_n \right)^t \phi_n. \tag{4.2.3}$$

Remark 5 From relations (1.1.2), (1.1.3), (1.1.4), (1.1.5), (1.4.4), (4.1.2) and (4.2.3) we obtain

$$\begin{aligned}
X_{n+1} &= \theta^t \phi_n + U_n + \varepsilon_{n+1} \\
X_{n+1} &= \theta^t \phi_n + x_{n+1} - \widehat{\theta}_n^t \phi_n + \varepsilon_{n+1} \\
X_{n+1} &= \theta^t \phi_n + x_{n+1} - \widehat{\theta}_n^t \phi_n + \rho \varepsilon_n + V_{n+1} \\
X_{n+1} &= \theta^t \phi_n + x_{n+1} - \widehat{\theta}_n^t \phi_n + \rho [X_n - \theta^t \phi_{n-1} - U_{n-1}] + V_{n+1} \\
X_{n+1} &= \theta^t \phi_n + x_{n+1} - \widehat{\theta}_n^t \phi_n + \rho [X_n - \theta^t \phi_{n-1} - x_n + \widehat{\theta}_{n-1}^t \phi_{n-1}] + V_{n+1} \\
X_{n+1} &= (\theta - \widehat{\theta}_n)^t \phi_n + \rho \left[X_n - (\theta - \widehat{\theta}_{n-1})^t \phi_{n-1} \right] + [x_{n+1} - \rho x_n] + V_{n+1} \\
X_{n+1} &= [\pi_n - \rho \pi_{n-1}] + \rho X_n + [x_{n+1} - \rho x_n] + \rho \mu^t \phi_{n-1} - \mu^t \phi_n + V_{n+1} \\
X_{n+1} &= [\pi_n - \rho \pi_{n-1}] + [x_{n+1} - \rho x_n] + \rho^{p+1} X_{n-p} + V_{n+1} \\
X_{n+1} &= y_n + z_n + \rho^{p+1} X_{n-p} + V_{n+1},
\end{aligned}$$

where $y_n = \pi_n - \rho \pi_{n-1}$ and $z_n = x_{n+1} - \rho x_n$.

In order to obtain results about the almost sure convergence of S_n/n , $\widehat{\theta}_n$, D_n and $\widehat{\rho}_n$ and following the same approach developed in Chapters 2 and 3 we need to prove beforehand that:

$$\sum_{k=1}^n \pi_k^2 = \mathcal{O}(\log s_n) \quad \text{a.s.} \quad (\text{E.1}) \quad (4.2.4)$$

where

$$s_n = \sum_{k=1}^n \|\phi_k\|^2. \quad (4.2.5)$$

However the techniques followed by Dufflo [20], Guo and Chen [31] and Bercu [6] respectively for proving (E.1) are supported by the fact that $(\varepsilon_n)_{n \geq 0}$ satisfies the properties establish in Section 1.2. It is particularly important the one that assures that if $\rho = 0$ and $(\alpha_n)_{n \geq 0}$ is a sequence adapted to the filtration $\mathbb{F}$ then

$$M_n = \sum_{k=0}^n \alpha_k \varepsilon_{k+1}, \quad (4.2.6)$$

is a martingale. When $\rho \neq 0$ we can not say that M_n is a martingale since in this case

$$\begin{aligned} M_n &= \sum_{k=0}^n \alpha_k \varepsilon_{k+1} \\ &= \sum_{k=0}^n \alpha_k V_{k+1} + \sum_{j=1}^{n+1} \rho^j \sum_{k=j-1}^n \alpha_k V_{k-j+1}, \end{aligned}$$

and we observe that the first sum of this new expression of M_n is in fact a martingale but we can not say too much about the remaining sums.

In the following result we present the almost surely convergence of the matrix sequence (S_n/n) and the LS algorithm.

Theorem 4.2.1 *Consider the $AR_1(p)$ adaptive tracking framework. Assume that $(\varepsilon_n)_{n \geq 0}$ satisfies the relation given by (4.1.2) and that $(V_n)_{n \geq 0}$ has a finite conditional moment of order $\alpha > 2$. In addition, suppose that the reference trajectory $(x_n)_{n \geq 0}$ satisfies the condition given by (2.0.1). Then, for the LS algorithm, we have*

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \frac{\sigma^2}{1 - \rho^{2(p+1)}} I_p \quad a.s. \quad (4.2.7)$$

*In addition the LS algorithm converges almost surely to θ^**

$$\hat{\theta}_n \rightarrow \theta^* \quad a.s. \quad (4.2.8)$$

Proof. Recalling that for all $n \geq 0$

$$X_{n+1} - x_{n+1} = \pi_n - \mu^t \phi_n + \varepsilon_{n+1}. \quad (4.2.9)$$

It follows from conjecture (E.1) that

$$\sum_{k=1}^n \pi_k^2 = \mathcal{O}(\log s_n) \quad a.s. \quad (4.2.10)$$

Hence, we deduce from (4.2.9) that

$$\sum_{k=1}^n X_{k+1}^2 = \mathcal{O}(\log s_n) + \mathcal{O}(n) \quad \text{a.s.} \quad (4.2.11)$$

Finally, we obtain that $s_n = o(s_n) + \mathcal{O}(n)$, so $s_n = \mathcal{O}(n)$. Then from (4.2.10) we have that

$$\sum_{k=1}^n \pi_k^2 = o(n) \quad \text{a.s.} \quad (4.2.12)$$

Assuming that

$$\frac{1}{n} \sum_{k=1}^n X_k^2 \quad (4.2.13)$$

converges almost surely, from Remark 5 we have that

$$\begin{aligned} X_n^2 &= y_{n-1}^2 + z_{n-1}^2 + \rho^{2p+2} X_{n-p-1}^2 + V_n^2 + \\ &\quad 2[y_{n-1} + z_{n-1}] [\rho^{p+1} X_{n-p-1} + V_n] + \\ &\quad 2\rho^{2p+2} X_{n-p-1} V_n \end{aligned}$$

Then since $(V_n)_{k \geq 0}$ holds the strong law of large numbers, from (2.0.1) and 4.2.12) it follows that

$$(1 - \rho^{2p+2}) \sum_{k=1}^n X_k^2 = \sum_{k=1}^n V_k^2 + o(n) + 2\rho^p \sum_{k=p+1}^n X_{k-p-1} V_k. \quad (4.2.14)$$

We observe that the last sum in (4.2.14) is a martingale such that

$$2\rho^p \sum_{k=p+1}^n X_{k-p-1} V_k = o(n) \quad \text{a.s.} \quad (4.2.15)$$

Then we have obtained that

$$\lim_{n \rightarrow \infty} \frac{1 - \rho^{2p+2}}{n} \sum_{k=1}^n X_k^2 = \sigma^2 \quad \text{a.s.}, \quad (4.2.16)$$

which leads us to

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k^2 = \frac{\sigma^2}{1 - \rho^{2p+2}} \quad \text{a.s.} \quad (4.2.17)$$

In addition, from Remark 5 we have that

$$\begin{aligned}
\sum_{k=1}^n X_k X_{k-1} &= \sum_{k=1}^n [y_{k-1} + z_{k-1}] X_{k-1} + \rho^{p+1} \sum_{k=1}^n X_{k-p-1} X_{k-1} + \sum_{k=1}^n X_{k-1} V_k \\
\sum_{k=1}^n X_k X_{k-2} &= \sum_{k=1}^n [y_{k-1} + z_{k-1}] X_{k-2} + \rho^{p+1} \sum_{k=1}^n X_{k-p-1} X_{k-2} + \sum_{k=1}^n X_{k-2} V_k \\
&\vdots \\
\sum_{k=1}^n X_k X_{k-p+1} &= \sum_{k=1}^n [y_{k-1} + z_{k-1}] X_{k-p+1} + \rho^{p+1} \sum_{k=1}^n X_{k-p-1} X_{k-p+1} + \sum_{k=1}^n X_{k-p+1} V_k,
\end{aligned}$$

observing that for all $1 \leq i-1 \leq p$

$$\sum_{k=1}^n X_{k-i} V_k = o(n) \quad \text{a.s.} \quad (4.2.18)$$

we find from (2.0.1) and (4.2.12) that

$$\begin{aligned}
\sum_{k=1}^n X_k X_{k-1} &= \rho^{p+1} \sum_{k=1}^n X_{k-p-1} X_{k-1} + o(n) \\
\sum_{k=1}^n X_k X_{k-2} &= \rho^{p+1} \sum_{k=1}^n X_{k-p-1} X_{k-2} + o(n) \\
&\vdots \\
\sum_{k=1}^n X_k X_{k-p+1} &= \rho^{p+1} \sum_{k=1}^n X_{k-p-1} X_{k-p+1} + o(n).
\end{aligned}$$

If we set for all $1 \leq i \leq p-1$

$$l_i = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k X_{k-i} \quad (4.2.19)$$

then we obtain the homogeneous system

$$l_1 - \rho^{p+1} l_p = 0 \quad (4.2.20)$$

$$l_2 - \rho^{p+1} l_{p-1} = 0 \quad (4.2.21)$$

$$\vdots \quad (4.2.22)$$

$$l_{p-1} - \rho^{p+1} l_1 = 0, \quad (4.2.23)$$

which is solved only if for all $1 \leq i \leq p-1$

$$l_i = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k X_{k-i} = 0 \quad \text{a.s.} \quad (4.2.24)$$

Hence, we have found from (4.2.17) and (4.2.24) the convergence (4.2.7).

In order to prove the almost sure convergence of the LS algorithm, we recall that for all $n \geq 0$

$$\hat{\theta}_n = S_{n-1}^{-1} \sum_{k=1}^n \phi_{k-1} (X_k - U_{k-1}), \quad (4.2.25)$$

then

$$S_{n-1} \hat{\theta}_n = \sum_{k=1}^n \phi_{k-1} (X_k - U_{k-1}). \quad (4.2.26)$$

We can observe that we need to investigate the asymptotic behavior of

$$\frac{1}{n} \sum_{k=1}^n X_{k-i} U_{k-1}. \quad (4.2.27)$$

From (1.4.4) we have for all $1 \leq i \leq p$ that

$$X_{k-i} U_{k-1} = X_{k-i} \left[x_k - \hat{\theta}_{k-1}^t \phi_{k-1} \right] \quad (4.2.28)$$

$$X_{k-i} U_{k-1} = X_{k-i} x_k + X_{k-i} \pi_{k-1} - X_{k-i} (\theta^*)^t \phi_{k-1}. \quad (4.2.29)$$

from (2.0.1), (4.2.12), (4.2.17) and (4.2.24), we find for all $1 \leq i \leq p$ that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_{k-i} U_{k-1} = \frac{-(a_i + \rho^i) \sigma^2}{1 - \rho^{2p+2}} \quad \text{a.s.} \quad (4.2.30)$$

Finally, from (4.2.7) and (4.2.30) we obtain the convergence (4.2.8) ■

Next, we show the bias of both $\hat{\rho}_n$ and D_n .

Theorem 4.2.2 *Consider the $AR_1(p)$ adaptive tracking framework. Assume that $(\varepsilon_n)_{n \geq 0}$ satisfies the relation given by (4.1.2) and that $(V_n)_{n \geq 0}$ has a finite conditional moment of order $\alpha > 2$. In addition, suppose that the reference*

trajectory $(x_n)_{n \geq 0}$ satisfies the condition given by (2.0.1). Then, we have the following almost sure convergences

$$\widehat{\rho}_n \rightarrow 0 \quad \text{a.s., and} \quad D_n \rightarrow 2 \quad \text{a.s.} \quad (4.2.31)$$

i.e., ρ_n and D_n are biased toward 0 and 2 respectively.

Proof. From (4.1.5) we have that:

$$\widehat{\varepsilon}_n = X_n - U_{n-1} - \widehat{\theta}_n^t \phi_{n-1} \quad (4.2.32)$$

$$= X_n - x_n - \check{\theta}_n^t \phi_{n-1} - \pi_{n-1}, \quad (4.2.33)$$

where $\check{\theta}_n = (\widehat{\theta}_n - \theta^*)$. We can see then that

$$\begin{aligned} \widehat{\varepsilon}_n^2 &= X_n^2 + x_n^2 + \pi_{n-1}^2 + \left(\check{\theta}_n^t \phi_{n-1} \right)^2 - 2X_n \left(\check{\theta}_n^t \phi_{n-1} \right) + 2x_n \pi_{n-1} \\ &\quad - 2X_n x_n - 2X_n \pi_{n-1} + 2 \left(\check{\theta}_n^t \phi_{n-1} \right) \pi_{n-1} + 2x_n \left(\check{\theta}_n^t \phi_{n-1} \right), \end{aligned}$$

and from (2.0.1), (4.2.7) and (4.2.12) we find that

$$\frac{1}{n} \sum_{k=1}^n \widehat{\varepsilon}_k^2 \rightarrow \frac{\sigma^2}{1 - \rho^{2p+2}} \quad \text{a.s.} \quad (4.2.34)$$

In a similar way we find from (2.0.1), (4.2.7), (4.2.12) and (4.2.24) that

$$\frac{1}{n} \sum_{k=1}^n \widehat{\varepsilon}_k \widehat{\varepsilon}_{k-1} \rightarrow 0 \quad \text{a.s.} \quad (4.2.35)$$

Hence, convergences (4.2.34) and (4.2.35) imply that

$$\widehat{\rho}_n \rightarrow 0 \quad \text{a.s.} \quad (4.2.36)$$

In addition, from relation (4.1.8) we have already proved that

$$D_n \rightarrow 2 \quad \text{a.s.} \quad (4.2.37)$$

■

In order to deal with a CLT for the LS algorithm, we notice as in Theorem 2.2.5 that

$$\begin{aligned}\widehat{\theta}_n - \theta^* &= S_{n-1}^{-1} \sum_{k=1}^n \phi_{k-1} \varepsilon_k - \mu \\ &= S_{n-1}^{-1} \sum_{k=1}^n \rho \phi_{k-1} \varepsilon_{k-1} + S_{n-1}^{-1} M_n - \mu\end{aligned}$$

where

$$M_n = \sum_{k=1}^n \phi_{k-1} V_k. \quad (4.2.38)$$

We observe that $(M_n)_{n \geq 0}$ is a vector martingale, from Lemma 2.2.2 and convergence (4.2.7) we have that

$$\sqrt{n} S_{n-1}^{-1} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, (1 - \rho^{2p+2}) I_p). \quad (4.2.39)$$

However, we can not make a similar conclusion for

$$\begin{aligned}\sqrt{n} S_{n-1}^{-1} \sum_{k=1}^n \rho \phi_{k-1} \varepsilon_{k-1} &= \sqrt{n} S_{n-1}^{-1} \sum_{k=1}^n \rho \phi_{k-1} X_{k-1} \\ &\quad - \sqrt{n} S_{n-1}^{-1} \sum_{k=1}^n \rho \phi_{k-1} (\pi_{k-2} + x_{k-1})\end{aligned}$$

since it is necessary to prove if the sequences

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n \phi_{k-1} X_{k-1} \quad (4.2.40)$$

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n \phi_{k-1} (\pi_{k-2} + x_{k-1}) \quad (4.2.41)$$

converge almost surely or not.

We know that if actually $\rho = 0$ then [7]

ρ	$\bar{Z}_1$	$\bar{Z}_2$	$\bar{Z}_3$	$V[Z_1]$	$V[Z_2]$	$V[Z_3]$
0	-0.0322	0.06	-0.0124	0.9655	0.9575	1.0414
-0.05	-0.0164	0.016	-0.082	0.9046	0.9919	0.9991
-0.1	0.049	-0.07	0.013	1.0825	0.9979	1.0362
0.2	-0.1906	-0.1684	-0.1679	0.9963	1.0951	1.1759
0.5	-0.2336	-0.6997	-0.9860	1.2191	2.2467	2.3680
0.8	0.5723	-0.3325	-2.6568	1.2672	3.3323	8.6544

Chart 3: Summary of the LS-CLT experiment, $AR_1(3)$.

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 I_p) \quad (4.2.42)$$

i.e., we only conserve the martingale given by (4.2.38).

In order to glimpse the behavior of a CLT for the LS algorithm via numerical experiments we consider the $AR_1(3)$ model, given for all $n \geq 0$ by

$$\begin{aligned} X_{n+1} &= a_1 X_n + a_2 X_{n-1} + a_3 X_{n-2} + U_n + \varepsilon_{n+1}, \quad \text{and} \\ \varepsilon_{n+1} &= \rho \varepsilon_n + V_{n+1}. \end{aligned}$$

where $a_1 = 1$, $a_2 = 1.6$, $a_3 = -0.5$ and $(V_n)_{n \geq 0}$ is an i.i.d. $N(0, 3.24)$ sequence, in addition for the sake of simplicity we consider the reference trajectory to be identically zero. This experiment is based in 400 samples of size $N=10000$. In order to simplify the presentation of the results of the experiment, let us define Z_1 , Z_2 and Z_3 as the first, second and third coordinate of

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \quad (4.2.43)$$

respectively, and denote by $\bar{Z}_i$ and $V[Z_i]$ for $i = 1, 2, 3$ their means and variances. In Chart 3 we present five no zero choices for the autocorrelation parameter and we can compare the mean and variance of Z_1 , Z_2 and Z_3 for each case with the reference case offered by $\rho = 0$. We observe that the results of the experiment for small values of ρ are closer to the ones for $\rho = 0$

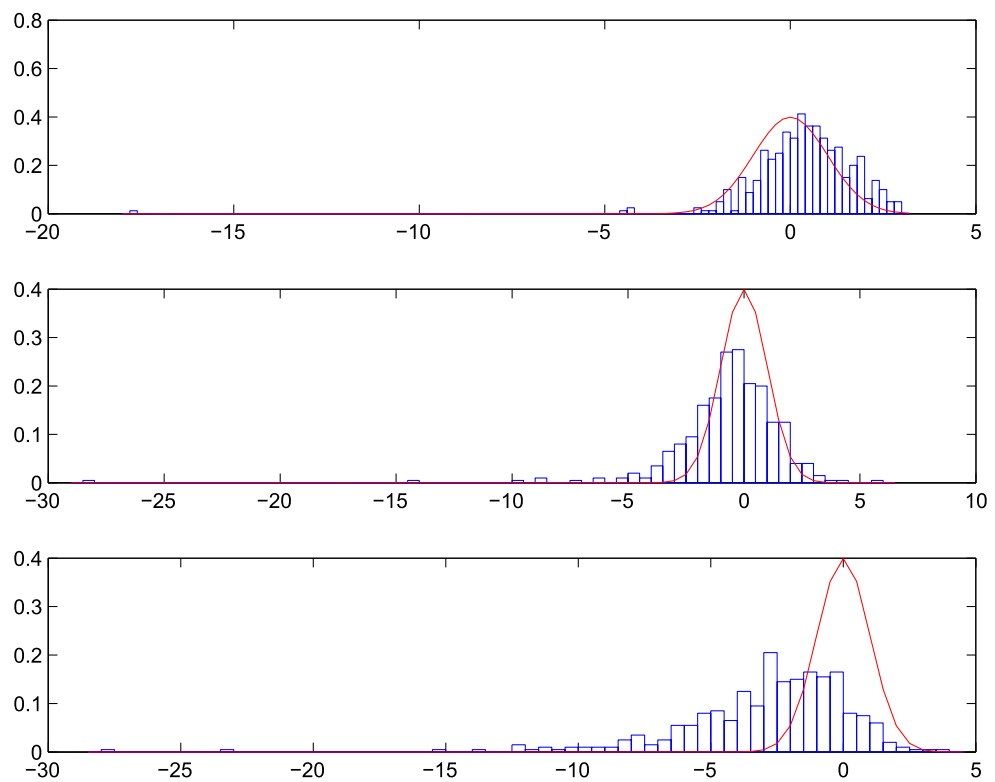
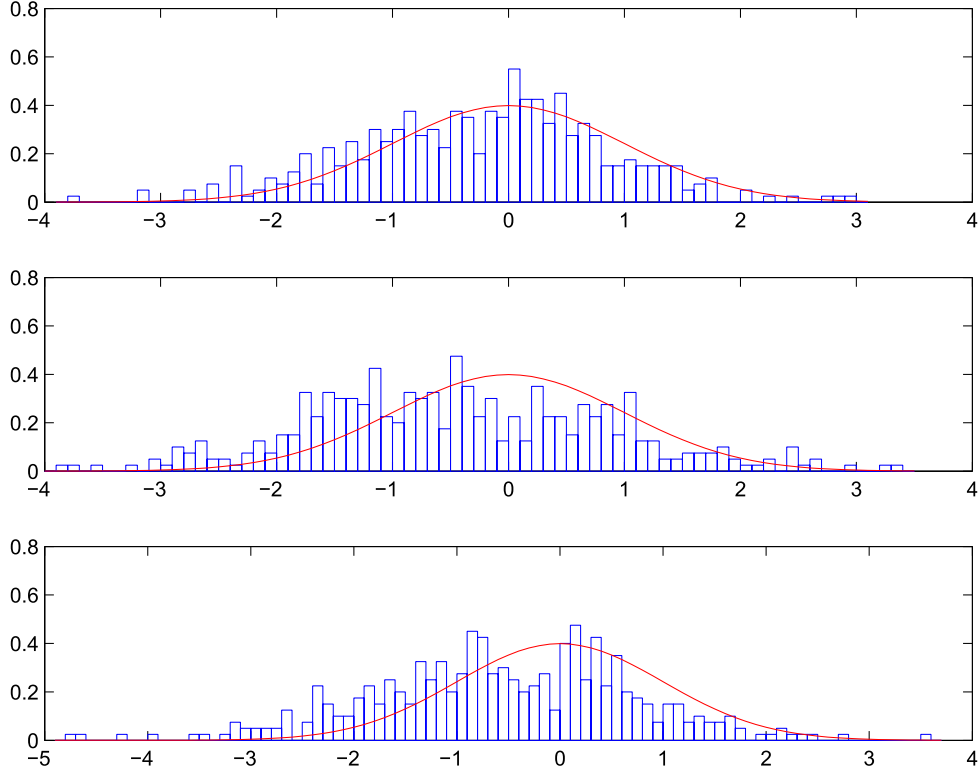


Figure 4.1: CLT for $\rho = 0.8$

Figure 4.2: CLT for $\rho = 0.4$

while bigger values for the autocorrelation parameter have significantly more distant results from the ones of the reference case. Furthermore in Figures 4.1, 4.2 and 4.3 we display histograms for illustrating the CLT for the choices $\rho = 0.8$, $\rho = 0.4$ and $\rho = -0.1$. From the discussion presented above and Figures 4.1, 4.2 and 4.3 we may conjecture in the one hand that smaller values in modulus for the autocorrelation parameter leads to the distribution of (4.2.43) to be closer to a $\mathcal{N}(0, I_3)$, while values for ρ closer to one present a normal distribution which does not seems to be $\mathcal{N}(0, I_3)$.

At this point, we may ask to ourselves about the possibility of having a CLT for the Durbin-Watson statistic and the estimator of the autocorrelation parameter given for all $n \geq 0$ by (4.1.7), due to the relation between this

statistics we only work on the limit distribution of $\widehat{\rho}_n$. For this, we notice that

$$\begin{aligned}
\sum_{k=1}^n \widehat{\varepsilon}_k \widehat{\varepsilon}_{k-1} &= \sum_{k=1}^n X_k \left(\pi_{k-2} - \check{\theta}_{k-1}^t \phi_{k-2} - x_{k-1} \right) \\
&+ \sum_{k=1}^n \left(\pi_{k-1} - x_k - \check{\theta}_k^t \phi_{k-1} \right) \left(X_{k-1} - x_{k-1} - \check{\theta}_{k-1}^t \phi_{k-2} + \pi_{k-2} \right) \\
&+ \sum_{l=0}^{[n/2]} (\rho^{p+1})^{2l} M_{l,n} + \sum_{l=0}^{[n/2]} (\rho^{p+1})^{2l+1} N_{l,n} + \sum_{l=0}^{[n/2]} (\rho^{p+1})^{2l} \eta_{l,n} \\
&+ \sum_{l=0}^{[n/2]} (\rho^{p+1})^{2l+1} \mu_{l,n},
\end{aligned}$$

where for all $l = 1, 2, \dots, [n/2]$, and

$$M_{l,n} = \sum_{k=1}^n X_{k-l(p+1)-1} V_{k-l(p+1)} \quad (4.2.44)$$

$$N_{l,n} = \sum_{k=1}^n X_{k-p(l+1)-1-l} V_{k-l(p+1)-1} \quad (4.2.45)$$

$$\eta_{l,n} = \sum_{k=1}^n X_{k-l(p+1)-1} \omega_{k-l(p+1)} \quad (4.2.46)$$

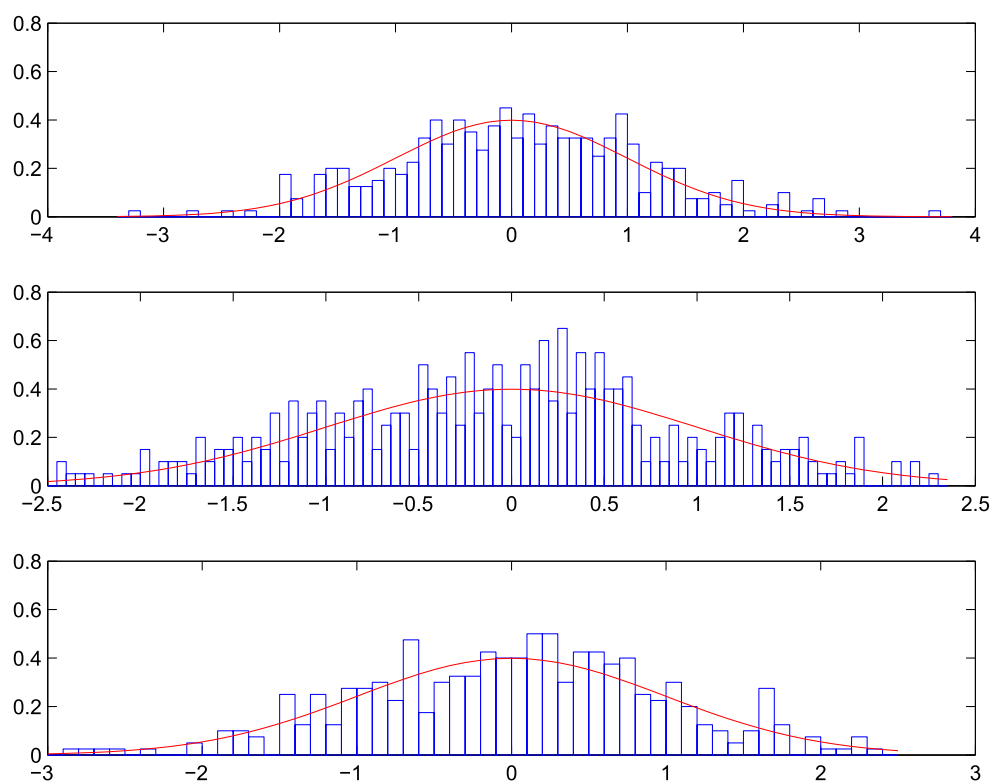
$$\mu_{l,n} = \sum_{k=1}^n X_{k-p(l+1)-1-l} \omega_{k-l(p+1)-1} \quad (4.2.47)$$

and for all $k \geq 0$ $\omega_k = y_k + z_k$, with y_k and z_k given in Remark 4. We find that for all $l = 1, 2, \dots, [n/2]$ both $M_{l,n}$ and $N_{l,n}$ are martingales, then from Lemma 2.2.2 and convergence (4.2.7) we have that

$$\frac{1}{\sqrt{n}} (\rho^{p+1})^{2l} M_{l,n} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \sigma^4 (\rho^{2p+2})^{2l} (1 - \rho^{2p+2})^{-1} \right) \quad (4.2.48)$$

and

$$\frac{1}{\sqrt{n}} (\rho^{p+1})^{2l+1} N_{l,n} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \sigma^4 (\rho^{2p+2})^{2l+1} (1 - \rho^{2p+2})^{-1} \right). \quad (4.2.49)$$

Figure 4.3: CLT for $\rho = -0.1$

which imply

$$\frac{(\rho^{p+1})^{2l} M_{l,n}}{\sqrt{n} \sum_{k=1}^n \widehat{\varepsilon}_k^2} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, (\rho^{2p+2})^{2l} (1 - \rho^{2p+2}) \right) \quad (4.2.50)$$

and

$$\frac{(\rho^{p+1})^{2l+1} N_{l,n}}{\sqrt{n} \sum_{k=1}^n \widehat{\varepsilon}_k^2} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, (\rho^{2p+2})^{2l+1} (1 - \rho^{2p+2}) \right) \quad (4.2.51)$$

all this lead us to think that the limit distribution of $\sqrt{n}\widehat{\rho}_n$ is normal, however we have similar problems to the ones had for the LS algorithm as we can see in (4.2.44). For trying to see the behavior of the Durbin-Watson statistic via a numerical experiment, we set

$$Z_4 = \sqrt{n}\widehat{\rho}_n \quad (4.2.52)$$

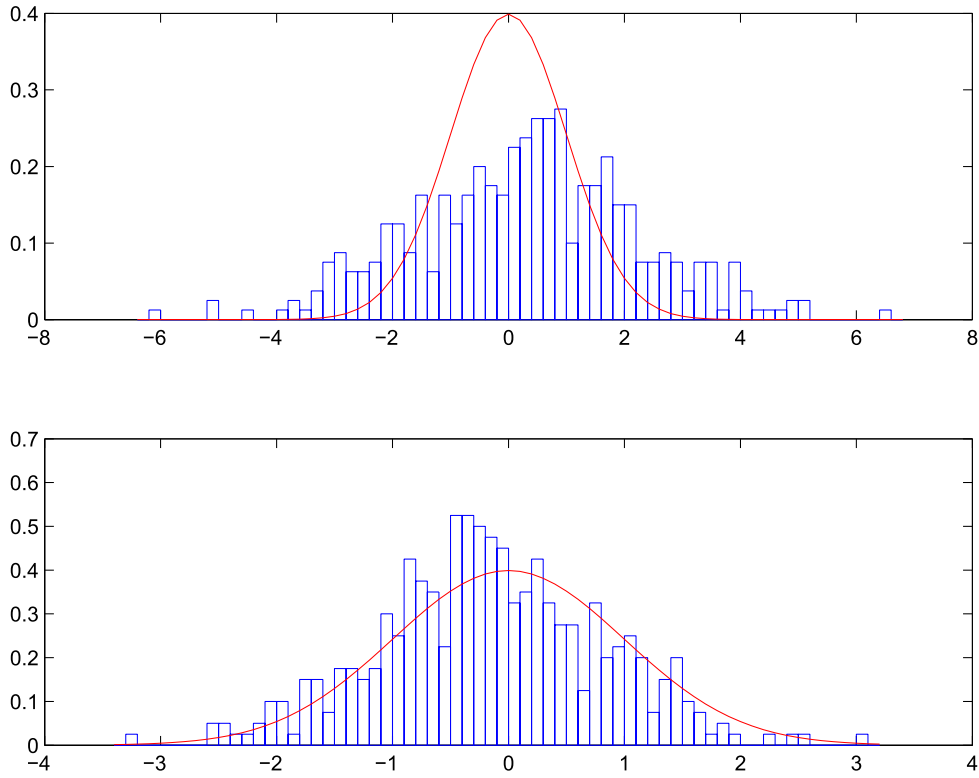
$$Z_5 = \sqrt{n} (D_n - 2) \quad (4.2.53)$$

and consider the $AR_1(3)$ process given for all $n \geq 0$ by

$$\begin{aligned} X_{n+1} &= a_1 X_n + a_2 X_{n-1} + a_3 X_{n-2} + U_n + \varepsilon_{n+1}, \\ \varepsilon_{n+1} &= \rho \varepsilon_n + V_{n+1} \end{aligned}$$

where $a_1 = 1$, $a_2 = 1.6$, $a_3 = -0.5$ and $(V_n)_{n \geq 0}$ is an i.i.d. $N(0, 3.24)$ sequence and the reference trajectory is identically zero. In Chart 4, we observe that small values of ρ have closer results to the ones obtained when $\rho = 0$ than the corresponding models for $\rho = 0.8$ and $\rho = 0.5$. In addition we observe in Figures 4.4, 4.5 and 4.6 that the distribution of Z_4 and Z_5 is normal, but both mean and variance depends on ρ in a direct way.

ρ	$\bar{Z}_4$	$\bar{Z}_5$	$V[Z_4]$	$V[Z_5]$
0	0.1071	-0.2262	1.0294	4.1214
-0.05	-0.1069	0.2005	0.9416	3.7635
-0.1	-0.199	0.3843	1.0597	4.2379
0.2	0.4303	-0.8744	1.0646	4.2557
0.5	1.4540	-2.9194	2.403	9.6127
0.8	-6.4644	3.2264	9.0894	36.3521

Chart 4: Summary of the DW-CLT experiment, $AR_1(3)$.Figure 4.4: CLT for $\rho = -0.1$

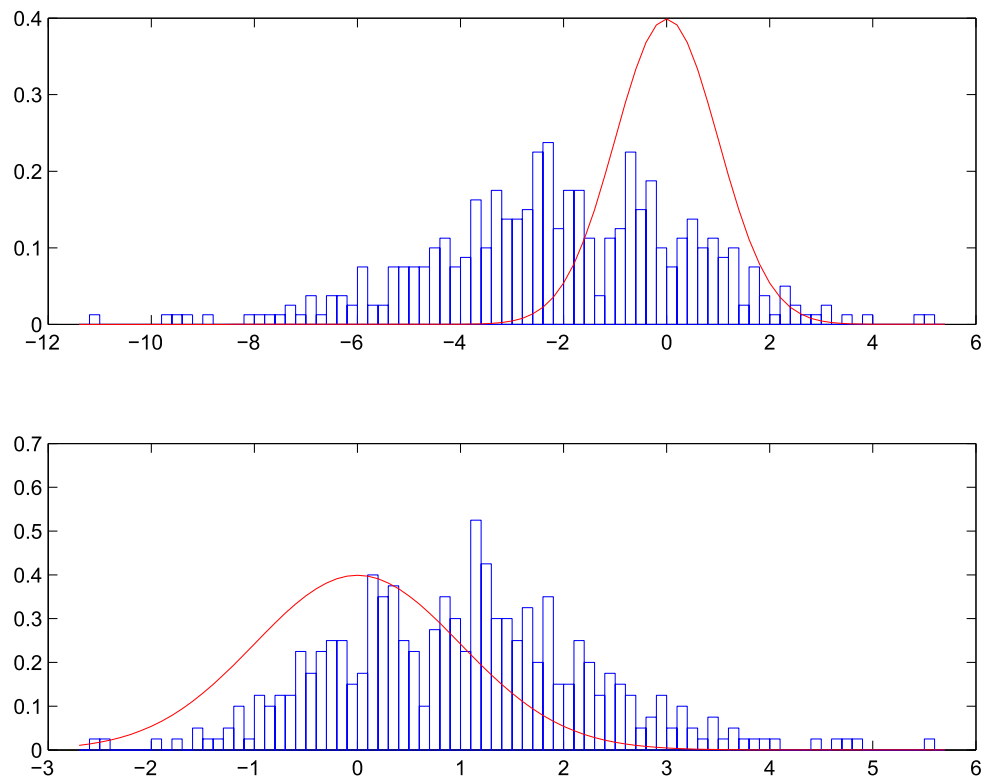


Figure 4.5: CLT for $\rho = 0.4$

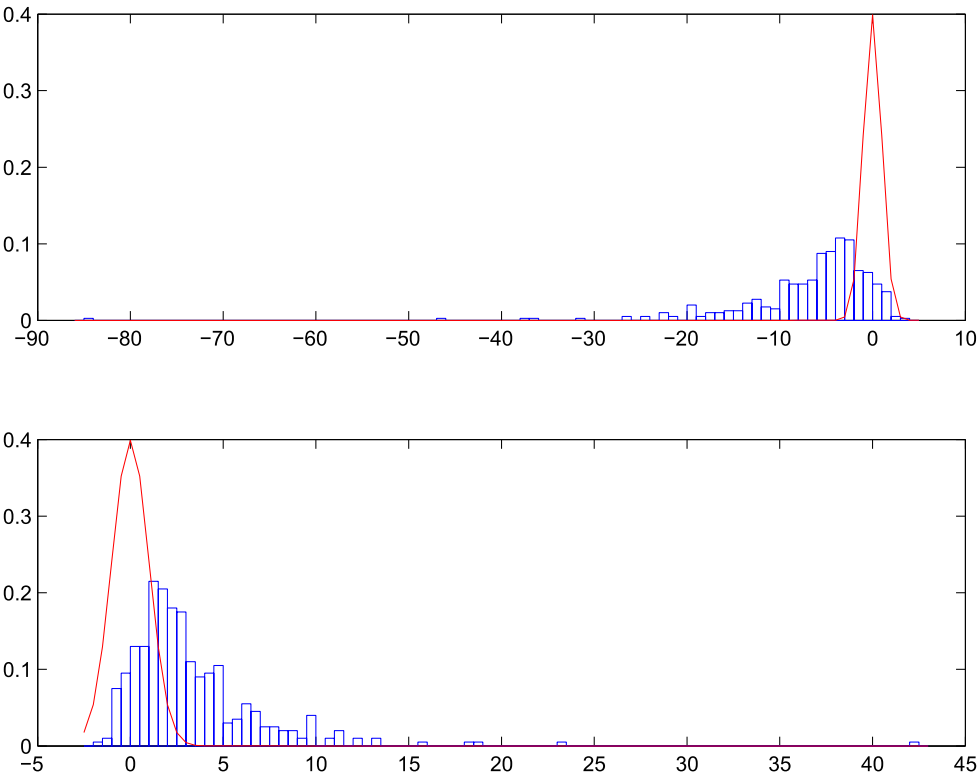


Figure 4.6: CLT for $\rho = 0.8$

Conclusions

In Chapter 2 via our new concept of strong controllability, we have extended the analysis of the almost sure convergence for both LS and WLS algorithms in the multidimensional ARX framework. It enables us to provide a positive answer to a conjecture in [5] by establishing a CLT and a LIL for these two stochastic algorithms.

In Chapter 3 via the use of a persistently excited adaptive tracking control, we have shown that it was possible to get ride of the strong controllability assumption proposed in Chapter 2. We have established the almost sure convergence for the LS and WLS estimators in the multidimensional ARX framework. In addition, we have shown the residual optimality of the adaptive tracking. Moreover, both LS and the WLS estimators share the same CLT and LIL.

In Chapter 4 via numerical experiments we conjecture about the asymptotic properties of the LS algorithm and the Durbin-Watson statistic in the $AR_1(p)$ adaptive tracking framework, in addition under additional conditions some of such conjectures were proved. The problematic introduced by the autocorrelated noise in order to obtain a CLT for the LS algorithm and the Durbin-Watson statistic is also explained.

Further Work

It would be a great challenge for the control community to carry out similar analysis to the exposed in Chapters 2 and 3 for $ARX_d(p, q)$ models with unknown high frequency gain and to extend it to $ARMAX_d(p, q, r)$ models.

It also would be interesting to extend the Section 3.3 in order to obtain the invertibility of the matrix Λ_x and obtain analogue results of the exposed in Chapters 2 and 3 for reference trajectories which no necessarily satisfy (2.0.1).

Finally it is necessary to prove condition (E.1) for proving in a full way Theorems 4.2.1 and 4.2.2 and produce the background needed for solving the complications carried by the autocorrelated noise.

Matlab Codes

Program 1. Example 1 of Chapter 2.

```

*****
* ARX
* d=2, p=1, q=1
*  $X_{n+1} = AX_n + U_n + BU_{n-1} + eps_{n+1}$ 
*****

clear all
NR=400; * The number of repetitions
N=10000; * The sample size
d=2; * The dimension we are working on
*****

* The parameters of the model
A=[2 0;0 1/2];
A1=A(1,1);A2=A(2,2);
B=[3/4 0;0 -1/4];
B1=B(1,1);B2=B(2,2);
THETA = [A B]';
SIGMA=1;
delta=d+d;
L=[eye(d) zeros(d);zeros(d) A*A*inv(eye(d)-B*B)];
K=inv(L);
*****

* The reference trajectory
x=[0;0];
*****

* We reserve some space for the CLT of the LS algorithm
Z1=[ ]; Z2=[ ]; Z3=[ ]; Z4=[ ];
*****

* Start of the program
*****

Implementation of the Least Squares algorithm and the Aström and Wittenmark tracking
Construction of THETAN and SN
*****

for j=1:NR,
* The initial values of XN, UN, PHIN and SN in every repetition

```

```

XN = ones(d,1);
UN = zeros(d,1); UNM = zeros(d,1);
PHIN = [XN' UNM']';
SPHIN = PHIN*PHIN'+eye(delta);
RPHIN = inv(SPHIN);
SN = SPHIN;
* The initial values of THETAN
THETAN = [zeros(d) zeros(d)]';
AN = zeros(d);
BN = zeros(d);
* We reserve some space for the LLN for LS algorithm
ANT=[]; BNT=[];
for i=1:N
*****
* Computation of XN at step i
EPSN =SIGMA*randn(d,1) ;
XN = THETA'*PHIN + UN + EPSN;
*****
* The LS algorithm at step i
THETAN = THETAN + RPHIN*PHIN*(XN - UN - THETAN'*PHIN)';
AN = THETAN(1:d,:)' ;
BN = THETAN(d+1:2*d,:)' ;
ANT = [ANT AN];
BNT = [BNT BN];
*****
* The adaptive tracking control and regression vector
PHIN = [XN' UN']';
UN=x-THETAN'*PHIN;
*****
* Computation of the matrix SN at step i
SN = SN + PHIN*PHIN';
RPHIN = RPHIN - (1/(1+PHIN'*RPHIN*PHIN))*RPHIN*PHIN*PHIN'*RPHIN;
end;
* Central Limit Theorem for LS algorithm
Z1=[Z1 sqrt(N)*(AN(1,1)-A1)/sqrt(K(1,1))];
Z2=[Z2 sqrt(N)*(AN(2,2)-A2)/sqrt(K(2,2))];
Z3=[Z3 sqrt(N)*(BN(1,1)-B1)/sqrt(K(3,3))];
Z4=[Z4 sqrt(N)*(BN(2,2)-B2)/sqrt(K(4,4))];

```

```

end
* Graphic illustrations
*****

ANT1=ANT(1:1,1:d:d*N);
ANT2=ANT(2:2,2:d:d*N);
BNT1=BNT(1:1,1:d:d*N);
BNT2=BNT(2:2,2:d:d*N);
* LLN for LS algorithm
clf; tp=[1:N];
figure(1)
subplot(2,1,1);
plot(tp,ANT1(tp),'b',tp,A1*ones(N,1),'r-');
title('Almost Sure convergence for A');
hold on
plot(tp,ANT2(tp),'b',tp,A2*ones(N,1),'r-');
subplot(2,1,2);
plot(tp,BNT1(tp),'b',tp,B1*ones(N,1),'r-');
hold on
plot(tp,BNT2(tp),'b',tp,B2*ones(N,1),'r-');
title('Almost Sure convergence for B');
* CLT for LS algorithm
NC=sqrt(N);
figure(2)
subplot(2,2,1);
[EZ1,CZ1]=histo(Z1,NC,0,1); hold on;
plot(CZ1,dnorm(CZ1),'r-'); hold off;
title('CLT for  $A'_{11}$ ');
subplot(2,2,2); [EZ2,CZ2]=histo(Z2,NC,0,1); hold on;
plot(CZ2,dnorm(CZ2),'r-'); hold off;
title('CLT for  $A'_{22}$ ');
subplot(2,2,3);
[EZ3,CZ3]=histo(Z3,NC,0,1); hold on;
plot(CZ3,dnorm(CZ3),'r-'); hold off;
title('CLT for  $B'_{11}$ ');
subplot(2,2,4);
[EZ4,CZ4]=histo(Z4,NC,0,1); hold on;
plot(CZ4,dnorm(CZ4),'r-'); hold off;
title('CLT for  $B'_{22}$ ');

```

Program 2. Example 2 of Chapter 2.

```

*****
* ARX
* d=2, p=1, q=1
*  $X_{n+1} = AX_n + U_n + BU_{n-1} + eps_{n+1}$ 
*****

clear all
NR=400; * The number of repetitions
N=10000; * The sample size
d=2; * The dimension we are working on
*****

* The parameters of the model
A=[2 0;0 0];
A1=A(1,1);A2=A(2,2);
B=[3/4 0;0 -1/4];
B1=B(1,1);B2=B(2,2);
THETA = [A B]';
SIGMA=1;
delta=d+d;
L=[eye(d) zeros(d);zeros(d) A*A*inv(eye(d)-B*B)];
K=inv(L);
*****

* The reference trajectory
x=[0;0];
*****

* We reserve some space for the CLT of the LS algorithm
Z1=[]; Z2=[]; Z3=[]; Z4=[];
*****

* Start of the program
*****

Implementation of the Least Squares algorithm and the Aström and Wittenmark tracking
Construction of THETAN and SN
*****

for j=1:NR,
* The initial values of XN, UN, PHIN and SN in every repetition
XN = ones(d,1);
UN = zeros(d,1); UNM = zeros(d,1);

```

```

PHIN = [XN' UNM']';
SPHIN = PHIN*PHIN'+eye(delta);
RPHIN = inv(SPHIN);
SN = SPHIN;
* The initial values of THETAN
THETAN = [zeros(d) zeros(d)]';
AN = zeros(d);
BN = zeros(d);
* We reserve some space for the LLN for LS algorithm
ANT=[]; BNT=[];
for i=1:N
*****

* Computation of XN at step i
EPSN =SIGMA*randn(d,1) ;
XN = THETA'*PHIN + UN + EPSN;
*****

* The LS algorithm at step i
THETAN = THETAN + RPHIN*PHIN*(XN - UN - THETAN'*PHIN)';
AN = THETAN(1:d,:)';
BN = THETAN(d+1:2*d,:)';
ANT = [ANT AN];
BNT = [BNT BN];
*****

* The adaptive tracking control and regression vector
PHIN = [XN' UN']';
UN=x-THETAN'*PHIN;
*****

* Computation of the matrix SN at step i
SN = SN + PHIN*PHIN';
RPHIN = RPHIN - (1/(1+PHIN'*RPHIN*PHIN))*RPHIN*PHIN*PHIN'*RPHIN;
end;

* Central Limit Theorem for LS algorithm
Z1=[Z1 sqrt(N)*(AN(1,1)-A1)/sqrt(K(1,1))];
Z2=[Z2 sqrt(N)*(AN(2,2)-A2)/sqrt(K(2,2))];
Z3=[Z3 sqrt(N)*(BN(1,1)-B1)/sqrt(K(3,3))];
Z4=[Z4 sqrt(N)*(BN(2,2)-B2)/sqrt(K(4,4))];
end

* Graphic illustrations

```

```

*****

ANT1=ANT(1:1,1:d:d*N);
ANT2=ANT(2:2,2:d:d*N);
BNT1=BNT(1:1,1:d:d*N);
BNT2=BNT(2:2,2:d:d*N);
* LLN for LS algorithm
clf; tp=[1:N];
figure(1)
subplot(2,1,1);
plot(tp,ANT1(tp),'b',tp,A1*ones(N,1),'r-');
title('Almost Sure convergence for A');
hold on
plot(tp,ANT2(tp),'b',tp,A2*ones(N,1),'r-');
subplot(2,1,2);
plot(tp,BNT1(tp),'b',tp,B1*ones(N,1),'r-');
hold on
plot(tp,BNT2(tp),'b',tp,B2*ones(N,1),'r-');
title('Almost Sure convergence for B');
* CLT for LS algorithm
NC=sqrt(N);
figure(2)
subplot(2,2,1);
[EZ1,CZ1]=histo(Z1,NC,0,1); hold on;
plot(CZ1,dnorm(CZ1),'r-'); hold off;
title('CLT for A11');
subplot(2,2,2); [EZ2,CZ2]=histo(Z2,NC,0,1); hold on;
plot(CZ2,dnorm(CZ2),'r-'); hold off;
title('CLT for A22');
subplot(2,2,3);
[EZ3,CZ3]=histo(Z3,NC,0,1); hold on;
plot(CZ3,dnorm(CZ3),'r-'); hold off;
title('CLT for B11');
subplot(2,2,4);
[EZ4,CZ4]=histo(Z4,NC,0,1); hold on;
plot(CZ4,dnorm(CZ4),'r-'); hold off;
title('CLT for B22');

```

Program 3. Example 1 of Chapter 3.

```

*****
*
* ARX
* d=2 p=1 q=1
*  $X_{n+1} = AX_n + U_n + BU_{n-1} + eps_{n+1}$ 
*
*
*****

NR=400; * The number of repetitions
N=10000; * The sample size
d=2; * The dimension we are working on
*****

* The parameters of the model
A=[1 0;0 1/2];
A1=A(1,1);A2=A(2,2);
B=[1/20 0;0 -1/12]; B1=B(1,1);B2=B(2,2);
THETA = [A B]';
SIGMA=1;
SIGMA1=1;
delta=d+d;
L=[2*eye(d) eye(d);eye(d) A*A*inv(eye(d)-B*B)];
K=inv(L);
*****

* The reference trajectory
x=[0;0];
*****

* We reserve some space for the CLT of the LS algorithm
Z1=[ ]; Z2=[ ]; Z3=[ ]; Z4=[ ];
*****

* Start of the program
*****

Implementation of the Least
Squares algorithm and
the Aström and Wittenmark tracking

Construction of THETAN and SN
*****

for j=1:NR,

```

```

* The initial values of XN, UN, PHIN and SN in every repetition
XN = ones(d,1);
UN = zeros(d,1); UNM = zeros(d,1);
PHIN = [XN' UNM']';
SPHIN = PHIN*PHIN'+eye(delta);
RPHIN = inv(SPHIN);
SN = SPHIN;

* The initial values of THETAN
THETAN = [zeros(d) zeros(d)]';
AN = zeros(d);
BN = zeros(d);

* We reserve some space for the LLN for LS algorithm
ANT=[]; BNT=[];
for i=1:N
*****

* Computation of XN at step i
EPSN =SIGMA*randn(d,1) ;
XN = THETA'*PHIN + UN + EPSN;
***** * The LS algorithm at step i

THETAN = THETAN + RPHIN*PHIN*(XN - UN - THETAN'*PHIN)';
AN = THETAN(1:d,:)' ;
BN = THETAN(d+1:2*d,:)' ;
ANT = [ANT AN];
BNT = [BNT BN];
*****

* The adaptive tracking control and regression vector
PHINM=PHIN;
PHIN = [XN' UN']';
UN=x-THETAN'*PHIN+SIGMA1*randn(d,1);
*****

* Computation of the matrix SN at step i
SN = SN + PHIN*PHIN';
RPHIN = RPHIN - (1/(1+PHIN'*RPHIN*PHIN))*RPHIN*PHIN*PHIN'*RPHIN;
end; * Central Limit Theorem for LS algorithm

Z1=[Z1 sqrt(N)*(AN(1,1)-A1)/sqrt(K(1,1))];

```

```

Z2=[Z2 sqrt(N)*(AN(2,2)-A2)/sqrt(K(2,2))];
Z3=[Z3 sqrt(N)*(BN(1,1)-B1)/sqrt(K(3,3))];
Z4=[Z4 sqrt(N)*(BN(2,2)-B2)/sqrt(K(4,4))];
end
* Graphic illustrations
*****

ANT1=ANT(1:1,1:d:d*N);
ANT2=ANT(2:2,2:d:d*N);
BNT1=BNT(1:1,1:d:d*N);
BNT2=BNT(2:2,2:d:d*N);
* LLN for LS algorithm
clf; tp=[1:N];
figure(1)
subplot(2,1,1);
plot(tp,ANT1(tp),'b',tp,A1*ones(N,1),'r-');
title('Almost Sure convergence for A');
hold on
plot(tp,ANT2(tp),'b',tp,A2*ones(N,1),'r-');
subplot(2,1,2);
plot(tp,BNT1(tp),'b',tp,B1*ones(N,1),'r-');
hold on
plot(tp,BNT2(tp),'b',tp,B2*ones(N,1),'r-');
title('Almost Sure convergence for B');
* CLT for LS algorithm
NC=sqrt(N);
figure(3)
subplot(2,2,1);
[EZ1,CZ1]=histo(Z1,NC,0,1); hold on;
plot(CZ1,dnorm(CZ1),'r-'); hold off;
title('CLT for A11');
subplot(2,2,2);
[EZ2,CZ2]=histo(Z2,NC,0,1); hold on;
plot(CZ2,dnorm(CZ2),'r-'); hold off;
title('CLT for A22');
subplot(2,2,3);
[EZ3,CZ3]=histo(Z3,NC,0,1); hold on;
plot(CZ3,dnorm(CZ3),'r-'); hold off;
title('CLT for B11');

```

```
subplot(2,2,4);
[EZ4,CZ4]=histo(Z4,NC,0,1); hold on;
plot(CZ4,dnorm(CZ4),'r-'); hold off;
title('CLT for B'_{22});
```

Program 4. Example 2 of Chapter 3.

```
*****
*
* ARX
* d=2 p=1 q=1
*  $X_{n+1} = AX_n + U_n + BU_{n-1} + eps_{n+1}$ 
*
*
*****

NR=400; * The number of repetitions
N=10000; * The sample size
d=2; * The dimension we are working on
*****

The parameters of the model
A=[2 0;0 0];
A1=A(1,1);A2=A(2,2);
B=[3/4 0;0 -1/2];
B1=B(1,1);B2=B(2,2);
THETA = [A B]';
SIGMA=1;
SIGMA1=1;
delta=d+d;
L=[2*eye(d) eye(d);eye(d) eye(d)+(A*A+(A+B)*(A+B))*inv(eye(d)-B*B)];
K=inv(L);
*****

The reference trajectory x=[0;0];
*****

* We reserve some space for the CLT of the LS algorithm Z1=[ ]; Z2=[ ]; Z3=[ ]; Z4=[ ];
*****

* Start of the program
*****

Implementation of the Least
Squares algorithm and
```

the Aström and Wittenmark tracking

Construction of THETAN and SN

for j=1:NR,

* The initial values of XN, UN, PHIN and SN in every repetition

XN = ones(d,1);

UN = zeros(d,1); UNM = zeros(d,1);

PHIN = [XN' UNM]';

SPHIN = PHIN*PHIN'+eye(delta);

RPHIN = inv(SPHIN);

SN = SPHIN;

* The initial values of THETAN

THETAN = [zeros(d) zeros(d)]';

AN = zeros(d);

BN = zeros(d);

* We reserve some space for the LLN for LS algorithm

ANT=[]; BNT=[];

for i=1:N

* Computation of XN at step i

EPSN =SIGMA*randn(d,1) ;

XN = THETA'*PHIN + UN + EPSN;

* The LS algorithm at step i

THETAN = THETAN + RPHIN*PHIN*(XN - UN - THETAN'*PHIN)';

AN = THETAN(1:d,:);

BN = THETAN(d+1:2*d,:);

ANT = [ANT AN];

BNT = [BNT BN];

* The adaptive tracking control and regression vector

PHINM=PHIN;

PHIN = [XN' UN]';

UN=x-THETAN'*PHIN+SIGMA1*randn(d,1);

* Computation of the matrix SN at step i

SN = SN + PHIN*PHIN'; RPHIN = RPHIN - (1/(1+PHIN'*RPHIN*PHIN))*RPHIN*PHIN*PHIN'*RPHIN

```

end;
* Central Limit Theorem for LS algorithm
Z1=[Z1 sqrt(N)*(AN(1,1)-A1)/sqrt(K(1,1))];
Z2=[Z2 sqrt(N)*(AN(2,2)-A2)/sqrt(K(2,2))];
Z3=[Z3 sqrt(N)*(BN(1,1)-B1)/sqrt(K(3,3))];
Z4=[Z4 sqrt(N)*(BN(2,2)-B2)/sqrt(K(4,4))];
end
* Graphic illustrations
*****
ANT1=ANT(1:1,1:d:d*N); ANT2=ANT(2:2,2:d:d*N);
BNT1=BNT(1:1,1:d:d*N); BNT2=BNT(2:2,2:d:d*N);
* LLN for LS algorithm
clf; tp=[1:N];
figure(1)
subplot(2,1,1);
plot(tp,ANT1(tp),'b',tp,A1*ones(N,1),'r-');
title('Almost Sure convergence for A');
hold on
plot(tp,ANT2(tp),'b',tp,A2*ones(N,1),'r-');
subplot(2,1,2);
plot(tp,BNT1(tp),'b',tp,B1*ones(N,1),'r-');
hold on
plot(tp,BNT2(tp),'b',tp,B2*ones(N,1),'r-');
title('Almost Sure convergence for B');
* CLT for LS algorithm
NC=sqrt(N);
figure(3)
subplot(2,2,1);
[EZ1,CZ1]=histo(Z1,NC,0,1); hold on;
plot(CZ1,dnorm(CZ1),'r-'); hold off;
title('CLT for A11');
subplot(2,2,2);
[EZ2,CZ2]=histo(Z2,NC,0,1); hold on;
plot(CZ2,dnorm(CZ2),'r-'); hold off;
title('CLT for A22');
subplot(2,2,3);
[EZ3,CZ3]=histo(Z3,NC,0,1); hold on;
plot(CZ3,dnorm(CZ3),'r-'); hold off;

```

```

title('CLT for B'_{11});
subplot(2,2,4);
[EZ4,CZ4]=histo(Z4,NC,0,1); hold on;
plot(CZ4,dnorm(CZ4),'r-'); hold off;
title('CLT for B'_{22});

```

Program 5. Example 3 of Chapter 3.

```

*****

*

* ARX
* d=2 p=1 q=1
*  $X_{n+1} = AX_n + U_n + BU_{n-1} + eps_{n+1}$ 
*
*

*****

NR=400; * The number of repetitions
N=10000; * The sample size
d=2; * The dimension we are working on
*****

The parameters of the model
A=[3 0;0 0];
A1=A(1,1);A2=A(2,2);
B=[3/4 0;0 -1/4];
B1=B(1,1);B2=B(2,2);
THETA = [A B]';
SIGMA=1;
SIGMA1=1;
delta=d+d;
L=[2*eye(d) eye(d);eye(d) eye(d)+(A*A+(A+B)*(A+B))*inv(eye(d)-B*B)];
K=inv(L);
*****

* We reserve some space for the CLT of the LS algorithm
Z1=[ ]; Z2=[ ]; Z3=[ ]; Z4=[ ];
*****

* Start of the program
*****

Implementation of the Least
Squares algorithm and

```

the Aström and Wittenmark tracking

Construction of THETAN and SN

```
*****
for j=1:NR,
* The initial values of XN, UN, PHIN and SN in every repetition
XN = ones(d,1);
UN = zeros(d,1); UNM = zeros(d,1);
PHIN = [XN' UNM']';
SPHIN = PHIN*PHIN'+eye(delta);
RPHIN = inv(SPHIN);
SN = SPHIN;
* The initial values of THETAN
THETAN = [zeros(d) zeros(d)]';
AN = zeros(d);
BN = zeros(d);
* We reserve some space for the LLN for LS algorithm
ANT=[]; BNT=[];
for i=1:N
*****
* Computation of XN at step i
EPSN =SIGMA*randn(d,1) ;
XN = THETA'*PHIN + UN + EPSN;
*****
* The LS algorithm at step i
THETAN = THETAN + RPHIN*PHIN*(XN - UN - THETAN'*PHIN)';
AN = THETAN(1:d,:)';
BN = THETAN(d+1:2*d,:)';
ANT = [ANT AN];
BNT = [BNT BN];
*****
* The adaptive tracking control and regression vector
PHINM=PHIN;
PHIN = [XN' UN']';
UN=SIGMA1*randn(d,1)-THETAN'*PHIN;
*****
* Computation of the matrix SN at step i
SN = SN + PHIN*PHIN';
```

```

RPHIN = RPHIN - (1/(1+PHIN'*RPHIN*PHIN))*RPHIN*PHIN*PHIN'*RPHIN;
end;
* Central Limit Theorem for LS algorithm
Z1=[Z1 sqrt(N)*(AN(1,1)-A1)/sqrt(K(1,1))];
Z2=[Z2 sqrt(N)*(AN(2,2)-A2)/sqrt(K(2,2))];
Z3=[Z3 sqrt(N)*(BN(1,1)-B1)/sqrt(K(3,3))];
Z4=[Z4 sqrt(N)*(BN(2,2)-B2)/sqrt(K(4,4))];
end
* Graphic illustrations
*****

ANT1=ANT(1:1,1:d:d*N);
ANT2=ANT(2:2,2:d:d*N);
BNT1=BNT(1:1,1:d:d*N);
BNT2=BNT(2:2,2:d:d*N);
* LLN for LS algorithm
clf; tp=[1:N];
figure(1)
subplot(2,1,1);
plot(tp,ANT1(tp),'b',tp,A1*ones(N,1),'r-');
title('Almost Sure convergence for A');
hold on
plot(tp,ANT2(tp),'b',tp,A2*ones(N,1),'r-');
subplot(2,1,2);
plot(tp,BNT1(tp),'b',tp,B1*ones(N,1),'r-');
hold on
plot(tp,BNT2(tp),'b',tp,B2*ones(N,1),'r-');
title('Almost Sure convergence for B');
* CLT for LS algorithm
NC=sqrt(N);
figure(3)
subplot(2,2,1);
[EZ1,CZ1]=histo(Z1,NC,0,1); hold on;
plot(CZ1,dnorm(CZ1),'r-'); hold off;
title('CLT for A11');
subplot(2,2,2);
[EZ2,CZ2]=histo(Z2,NC,0,1); hold on;
plot(CZ2,dnorm(CZ2),'r-'); hold off;
title('CLT for A22');

```

```

subplot(2,2,3);
[EZ3,CZ3]=histo(Z3,NC,0,1); hold on;
plot(CZ3,dnorm(CZ3),'r-'); hold off;
title('CLT for B'_{11});
subplot(2,2,4);
[EZ4,CZ4]=histo(Z4,NC,0,1); hold on;
plot(CZ4,dnorm(CZ4),'r-'); hold off;
title('CLT for B'_{22});

```

Program 6. Example 4 of Chapter 3.

```

*****
*
* ARX
* d=2 p=1 q=1
*  $X_{n+1} = AX_n + U_n + BU_{n-1} + \epsilon_{n+1}$ 
*
*
*****
NR=400; * The number of repetitions
N=10000; * The sample size
d=2; * The dimension we are working on
*****

The parameters of the model
A=[3 0;0 0];
A1=A(1,1);A2=A(2,2);
B=[3/4 0;0 -1/4];
B1=B(1,1);B2=B(2,2);
THETA = [A B]';
SIGMA=1;
SIGMA1=1;
delta=d+d;
* The reference trajectory x=[0;1]; nablax=x*x'
L1=A*A*inv(eye(2)-B*B);
L2=eye(2)-(A+B)*inv(eye(2)+B);
L3=(A+B)*(A+B)*inv(eye(2)+B)*(inv(eye(2)+B)-inv(eye(2)-B*B));
L4=(A+B)*(A+B)*inv(eye(2)+B)*inv(eye(2)-B*B);
L5=-(A+B)*inv(eye(2)+B);

```

```

L=[1 0;0 1]+nabla nabla-inv(eye(2)+B)*(A+B)*nabla;nabla-inv(eye(2)+B)*(A+B)*nabla'
L1+nabla*(L2+L3+L4+L5)];
K=inv(L);
*****

* We reserve some space for the CLT of the LS algorithm
Z1=[]; Z2=[]; Z3=[]; Z4=[];
*****

* Start of the program
*****

Implementation of the Least
Squares algorithm and
the Aström and Wittenmark tracking

Construction of THETAN and SN
*****

for j=1:NR,
* The initial values of XN, UN, PHIN and SN in every repetition
XN = ones(d,1);
UN = zeros(d,1); UNM = zeros(d,1);
PHIN = [XN' UNM]';
SPHIN = PHIN*PHIN'+eye(delta);
RPHIN = inv(SPHIN);
SN = SPHIN;
* The initial values of THETAN
THETAN = [zeros(d) zeros(d)]';
AN = zeros(d);
BN = zeros(d);
* We reserve some space for the LLN for LS algorithm
ANT=[]; BNT=[];
for i=1:N
*****

* Computation of XN at step i
EPSN =SIGMA*randn(d,1) ;
XN = THETA'*PHIN + UN + EPSN;
*****

* The LS algorithm at step i
THETAN = THETAN + RPHIN*PHIN*(XN - UN - THETAN'*PHIN)';
AN = THETAN(1:d,:);

```

```

BN = THETAN(d+1:2*d,:);
ANT = [ANT AN];
BNT = [BNT BN];
*****
* The adaptive tracking control and regression vector
PHINM=PHIN;
PHIN = [XN' UN']';
UN=x-THETAN'*PHIN;
*****
* Computation of the matrix SN at step i
SN = SN + PHIN*PHIN';
RPHIN = RPHIN - (1/(1+PHIN'*RPHIN*PHIN))*RPHIN*PHIN*PHIN'*RPHIN;
end;
* Central Limit Theorem for LS algorithm
Z1=[Z1 sqrt(N)*(AN(1,1)-A1)/sqrt(K(1,1))];
Z2=[Z2 sqrt(N)*(AN(2,2)-A2)/sqrt(K(2,2))];
Z3=[Z3 sqrt(N)*(BN(1,1)-B1)/sqrt(K(3,3))];
Z4=[Z4 sqrt(N)*(BN(2,2)-B2)/sqrt(K(4,4))];
end
* Graphic illustrations
*****
ANT1=ANT(1:1,1:d*d*N);
ANT2=ANT(2:2,2:d*d*N);
BNT1=BNT(1:1,1:d*d*N);
BNT2=BNT(2:2,2:d*d*N);
* LLN for LS algorithm
clf; tp=[1:N];
figure(1)
subplot(2,1,1);
plot(tp,ANT1(tp),'b',tp,A1*ones(N,1),'r-');
title('Almost Sure convergence for A');
hold on
plot(tp,ANT2(tp),'b',tp,A2*ones(N,1),'r-');
subplot(2,1,2);
plot(tp,BNT1(tp),'b',tp,B1*ones(N,1),'r-');
hold on
plot(tp,BNT2(tp),'b',tp,B2*ones(N,1),'r-');
title('Almost Sure convergence for B');

```

```

* CLT for LS algorithm
NC=sqrt(N);
figure(3)
subplot(2,2,1);
[EZ1,CZ1]=histo(Z1,NC,0,1); hold on;
plot(CZ1,dnorm(CZ1),'r-'); hold off;
title('CLT for A'_{11});
subplot(2,2,2);
[EZ2,CZ2]=histo(Z2,NC,0,1); hold on;
plot(CZ2,dnorm(CZ2),'r-'); hold off;
title('CLT for A'_{22});
subplot(2,2,3);
[EZ3,CZ3]=histo(Z3,NC,0,1); hold on;
plot(CZ3,dnorm(CZ3),'r-'); hold off;
title('CLT for B'_{11});
subplot(2,2,4);
[EZ4,CZ4]=histo(Z4,NC,0,1); hold on;
plot(CZ4,dnorm(CZ4),'r-'); hold off;
title('CLT for B'_{22});

```

Program 7. First example. Chapter 4.

```

*****

```

DURBIN WATSON

```

p=1 q=1

```

```

*****

```

```

N = 10000; * The sample size

```

```

NR =400; * The number of repetitions

```

```

*****

```

```

* The parameters of the model

```

```

THETA = 1; RHO =0.8;

```

```

THETASTAR = (THETA + RHO);

```

```

RHOSTAR = 0; DSTAR = 2;

```

```

SIGMA=1;

```

```

* We reserve some space for the CLT of the LS algorithm

```

```

Z1=[ ]; DD=[ ]; RHOCLT=[ ];

```

```

*****

```

```

* Start of the program

```

```

*****

```

Implementation of the Least
Squares algorithm and
the Aström and Wittenmark tracking

```

Construction of THETAN, SN , RHON and DN
for j=1:NR
* The initial values of XN, UN, PHIN and SN in every repetition
XN = 1; SN = XN*XN;
UN = 0;
EPSN =SIGMA*randn;
THETAN = 0; RHON = 0; DN = 0;
EPSHATN = randn; QN = EPSHATN*EPSHATN;
THETANT = [ ]; RHONT = [ ]; DNT = [ ];
for i=1:N
EPSN = RHO*EPSN +SIGMA*randn; XNM = XN; XN = THETA*XNM + UN +
EPSN; X=[X XN];
*****

* The LS algorithm at step i
THETANM = THETAN;
THETAN = THETANM + XNM*(XN - UN - THETANM*XNM)/SN;
RHONM = RHON;
EPSHATNM = EPSHATN;
EPSHATN = XN - UN - THETAN*XNM;
*****

* The adaptive tracking control *
UN = - THETAN*XN;
*****

* Computation of the matrix SN, QN and the DW statistic at step i
SN = SN + XN*XN;
QN = QN + EPSHATN*EPSHATN;
RHON = RHONM + EPSHATN*(EPSHATNM - RHONM*EPSHATN)/QN;
DN = DN + ( (EPSHATN - EPSHATNM)2 - (EPSHATN2) * DN)/QN;
*****

* Stockage
THETANT = [THETANT THETAN];
RHONT = [RHONT RHON];
DNT = [DNT DN];
end

```

```

* Central Limit Theorem
Z1=[Z1 sqrt(N)*(THETAN-THETASTAR)];
DD=[DD sqrt(N)*(DN-2)/2];
RHOCLT=[RHOCLT sqrt(N)*RHON];
end
* Graphic illustrations
*****

clf; tp=[1:N];
figure(1)
subplot(3,1,1)
plot(tp,THETANT(tp),'b',tp,THETASTAR*ones(N,1),'r-');
subplot(3,1,2)
plot(tp,RHONT(tp),'b',tp,RHOSTAR*ones(N,1),'r-');
subplot(3,1,3)
plot(tp,DNT(tp),'b',tp,DSTAR*ones(N,1),'r-');
NC=sqrt(N);
figure(3)
[EZ1,CZ1]=histo(Z1,NC,0,1);hold on
plot(CZ1,dnorm(CZ1),'r-');
figure(4)
subplot(2,1,1);
[EZ4,CZ4]=histo(DD,NC,0,1);hold on
plot(CZ4,dnorm(CZ4),'r-');
subplot(2,1,2);
[EZ5,CZ5]=histo(RHOCLT,NC,0,1);hold on
plot(CZ5,dnorm(CZ5),'r-');

```

Program 8. Examples 2, 3 and of Chapter 4.

```

*****
*
* DURBIN WATSON
* p=3 q=1
*****

N = 10000; * The sample size
NR =400; * The number of repetitions
*****

* The parameters of the model
A=1.0;

```

```

B=1.6;
C=-0.5;
THETA = [A;B;C]; RHO =0;
THETASTAR = THETA + [RHO;RHO^2;RHO^3];
SIGMA=sqrt(3.24);
RHOSTAR = 0; DSTAR = 2;
p=3;
* We reserve some space for the CLT of the LS algorithm
Z1=[ ]; Z2=[ ]; Z3=[ ];
DD=[ ]; RHOCCLT=[ ];
*****

* Start of the program
*****

Implementation of the Least
Squares algorithm and
the Aström and Wittenmark tracking
Construction of THETAN, SN , RHON and DN
for j=1:NR
* The initial values
XN = 1; XNM=1; XNM1=1;
UN = 0;
EPSN =SIGMA*randn;
THETAN = [0 0 0]'; RHON = 0; DN = 0;
EPSHATN = randn;
QN = EPSHATN*EPSHATN;
PHIN=[XN XNM XNM1]';
SN=PHIN*PHIN'+eye(3);
SPHIN = PHIN*PHIN'+eye(3);
RPHIN = inv(SPHIN);
THETANT = [ ]; RHONT = [ ]; DNT = [ ];
for i=1:N
EPSN = RHO*EPSN +SIGMA*randn;
XNM1=XNM;
XNM = XN;
XN = THETA'*PHIN + UN + EPSN;
PHINM=PHIN;
PHINM1=PHINM;
PHIN=[XN;XNM;XNM1];

```

```

*****

* The LS algorithm at step i
THETANM = THETAN;
THETAN = THETANM + RPHIN*PHINM*XN;
*****

* Computation of the matrix SN, QN and the DW statistic at step i
RHONM = RHON;
EPSHATNM = EPSHATN;
EPSHATN=XN*(1-PHINM'*inv(SN)*PHINM);
QNM=QN;
QN = QN + EPSHATN*EPSHATN;
SN = SN + PHIN*PHIN'; RHON = RHONM + EPSHATN*(EPSHATNM - RHONM*EPSHATN)/QN;
DN = DN + ( (EPSHATN - EPSHATNM)^2 - (EPSHATN^2) * DN)/QN;
RPHIN = RPHIN - (1/(1+PHIN'*RPHIN*PHIN))*RPHIN*PHIN*PHIN'*RPHIN;
*****

* The adaptive tracking control
UN = - THETAN'*PHIN;
*****

* Stockage
THETANT = [THETANT THETAN];
RHONT = [RHONT RHON];
DNT = [DNT DN];
end

* Central Limit Theorem
Z1=[Z1 sqrt(N)*(THETAN(1,1)-THETASTAR(1,1))];
Z2=[Z2 sqrt(N)*(THETAN(2,1)-THETASTAR(2,1))];
Z3=[Z3 sqrt(N)*(THETAN(3,1)-THETASTAR(3,1))];
DD=[DD sqrt(N)*(DN-2)];
RHOCLT=[RHOCLT sqrt(N)*RHON];
end

* Graphic illustrations
*****

clf; tp=[1:N];
figure(1)
subplot(3,1,1)
plot(tp,THETANT(1,tp),'b',tp,THETASTAR(1)*ones(N,1),'r-');
subplot(3,1,2)
plot(tp,THETANT(2,tp),'b',tp,THETASTAR(2)*ones(N,1),'r-');

```

```

subplot(3,1,3)
plot(tp,THETANT(3,tp),'b',tp,THETASTAR(3)*ones(N,1),'r-');
figure(2)
subplot(2,1,1)
plot(tp,RHONT(tp),'b',tp,RHOSTAR*ones(N,1),'r-');
subplot(2,1,2)
plot(tp,DNT(tp),'b',tp,DSTAR*ones(N,1),'r-');
NC=sqrt(N);
figure(3)
subplot(3,1,1);
[EZ1,CZ1]=histo(Z1,NC,0,1);hold on
plot(CZ1,dnorm(CZ1),'r-');
subplot(3,1,2);
[EZ2,CZ2]=histo(Z2,NC,0,1);hold on
plot(CZ2,dnorm(CZ2),'r-');
subplot(3,1,3);
[EZ3,CZ3]=histo(Z3,NC,0,1);hold on
plot(CZ3,dnorm(CZ3),'r-');
figure(4)
subplot(2,1,1);
[EZ4,CZ4]=histo(DD,NC,0,1);hold on
plot(CZ4,dnorm(CZ4),'r-');
subplot(2,1,2);
[EZ5,CZ5]=histo(RHOCLT,NC,0,1);hold on
plot(CZ5,dnorm(CZ5),'r-');

```

Bibliography

- [1] A. A. Abouelsoud, M. F. Hassan and M. A. Sultan, An adaptive observer for robots with persistent excitation, *International Journal of Systems Science*, Vol. 27, pp. 809-820, 1996.
- [2] E. Aggelogiannaki, P. Doganis and H. Sarimveis, An adaptive model predictive control configuration for production-inventory systems, *International Journal of Production Economics*, Vol. 114, pp. 165-178, 2008.
- [3] B. D. O. Anderson, Exponential convergence and persistent excitation, 21th IEEE Conference on Decision and Control, 1982.
- [4] K. J. Aström and B. Wittenmark, *Adaptive Control*, 2nd edition, Addison-Wesley, New York, 1995.
- [5] P. Baraldi, A. A. Manginelli, M. Maieron, D. Liberati, and C. A. Porro, An ARX model-based approach to trial by trial identification of fMRI-BOLD responses, *NeuroImage*, 37, pp 189-201, 2007.
- [6] B. Bercu, Sur l'estimateur des moindres carrés généralisés d'un modèle ARMAX. Application à l'identification des modèles ARMA. *Annales de l'Institut Henri Poincaré*, 27, pp. 425-443, 1991.
- [7] B. Bercu, Central limit theorem and law of iterated logarithm for least squares algorithms in adaptive tracking, *SIAM J. Control Optim.*, 36, pp. 910-928, 1998.
- [8] B. Bercu, Weighted estimation and tracking for ARMAX models, *SIAM J. Control Optim.*, 33, pp. 89-106, 1995.
- [9] B. Bercu and V. Vázquez, Further results for ARX models in adaptive tracking, 47th IEEE Conference on Decision and Control, 2008.
- [10] B. Bercu and V. Vázquez, On the usefulness of persistent excitation in ARX adaptive tracking, 48th IEEE Conference on Decision and Control, 2009.

- [11] B. Bercu and V. Vazquez, On the usefulness of persistent excitation in ARX adaptive tracking, *International Journal of Control*, 83, 2010.
- [12] B. Bercu and V. Vázquez, A new concept of strong controllability via the Schur complement in adaptive tracking, accepted for publication in *Automatica*, 2010.
- [13] B. Bercu and M. Duflo, Moindres carrés pondérés et poursuite, *Annals de l'Institut Henri Poincaré*, 28, pp. 403-430, 1992.
- [14] B. Bercu and B. Portier, Adaptive control of parametric nonlinear autoregressive models via a new martingale approach, *IEEE Trans. Automat. Control*, 47, pp. 1524-1528, 2002.
- [15] H. Chang and P. K. Tzenog, Analysis of the dynamic characteristics of pressure sensors using ARX system identification, *Sensors and Actuators A: Physical*, 141, pp. 367-375, 2008.
- [16] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*, Birkhäuser, Boston, 1991.
- [17] C. Chengyu, N. Hovakimyan and W. Jiang, Intelligent excitation for adaptive control with unknown parameters in reference input, *IEEE Trans. Automat. Control*, Vol. 52, pp. 1525-1532, 2007.
- [18] L. Corrado and S. Holly, The linearisation and optimal control of large non-linear rational expectations models by persistent excitation source, *Computational Economics*, Vol. 28, pp. 139-153, 2006.
- [19] M. Dehghani, M. Karrari, W. Rosehart and O.P. Malik. Synchronous machine model parameters estimation by a time-domain identification method. *International Journal of Electrical Power and Energy Systems*, In Press, 2009.
- [20] M. Duflo, *Random Iterative Models*, Springer Verlag, Berlin, 1997.
- [21] J. Durbin, Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables. *Econometrica*. Vol. 38 no. 3, pp. 410-421, 1970.
- [22] J. Durbin and G.S. Watson. Testing for serial correlation in Least Squares regression I. *Biometrika* , Vol. 37, pp. 409-428, 1950.
- [23] J. Durbin and G.S. Watson. Testing for serial correlation in Least Squares regression II. *Biometrika* , Vol. 38, pp. 159-178, 1951.
- [24] Thomas Z. Fahidy. An application of Durbin Watson statistic to electrochemical science. *Electrochimica Acta* 51, pp. 3516-3520, 2006.

- [25] H. Fang, R. X. Fan, B. Thuilot and P. Marinet, Lateral control for autonomous vehicles with sliding angle reconstruction, Chinese Control and Decision Conference, 2008.
- [26] A. Farrell, Persistent excitation conditions in passive learning control, *Automatica*, Vol. 33, pp. 699-703, 1997.
- [27] T. C. Fonseca, J. R. P. Mendes, Ad. B. S. Serapiao, I. R. Guilherme, A Genetic Neuro-Model Reference Adaptive Controller for Petroleum Wells Drilling Operations, cimca, p. 3, International Conference on Computational Intelligence for Modelling Control and Automation and International Conference on Intelligent Agents Web Technologies and International Commerce (CIMCA'06), 2006
- [28] J. Fresnel, Anneaux, Hermann, Paris, 2001.
- [29] K. Fukata, T. Washio, and H. Motoda, A method to search ARX model orders and its application to sales dynamics analysis, *ICDM Workshops*, pp. 590-595, 2006.
- [30] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, Resultants, and Multidimensional Determinants*, Birkhäuser, Boston, 1994.
- [31] L. Guo and H. F. Chen, The Aström Wittenmark self-tuning regulator revisited and ELS-based adaptive trackers, *IEEE Trans. Automat. Control*, 36, pp. 802-812, 1991.
- [32] L. Guo, Further results on least squares based adaptive minimum variance control, *SIAM J. Control Optim.*, 32, pp. 187-212, 1994.
- [33] P. Hall and C. C. Heyde, *Martingale limit theory and its application*, Academic Press, New York, 1980.
- [34] R.J. Hill and H.D. Flack. The use of the Durbin Watson d statistic in Rietveld analysis. *Journal of Applied Crystallography*. Vol. 20 part 5, pp. 356-361, 1987.
- [35] Ho Pham Huy Anh and Kyoung Kwan Ahn. Identification of pneumatic artificial muscle manipulators by a MGA-based nonlinear NARX fuzzy model, *Machronics*, Vol. 19 Issue 1, pp. 106-133, 2009.
- [36] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 1990.
- [37] J. T. Huang, Persistent excitation in a shunt DC motor under adaptive control, *Asian Journal of Control*, Vol. 9, pp. 37-44, 2008.
- [38] K. Y. Huang and C.-J. Jane, A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories, To appear in *Expert Systems with Applications*, 2008.

- [39] Y.J. Jeon, Y.s. Eom, J.O. Hwang, H.K. Yang, J.J. Im, S.H. Park and N.S. Jung. Estimation of an Aortic Pressure Waveform from Radial Pulse using ARX Model. Proceeding of the World Congress on Medical Physics and Biomedical Engineering, Vol. 14, Part 7, pp. 183-186, 2006.
- [40] Johnston J. Econometric Methods. McGraw-Hill, 1963.
- [41] M.L. King, The alternative Durbin Watson test. An assessment of Durbin and Watson's choice of test statistic. Journal of Econometrics. Vol. 17, pp. 5-66, 1981.
- [42] A. Kireçci, M. Topalbekiroglu, İ. Eker, Experimental evaluation of a model reference adaptive control for a hydraulic robot: a case study, Robotica, Volume 21 , Issue 1, pp. 71-78, 2003.
- [43] G. Klančar and I. Skrjanc. Predictive Trajectory Tracking Control for Mobile Robots. 12th International Power Electronics and Motion Control Conference, pp. 373-378, 2006.
- [44] Y.H. Kuang and J. Chuen-Jiuan. A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories. Expert Systems with Applications, Vol. 36, Issue 3, Part 1, pp. 5387-5392, 2009.
- [45] T. L. Lai and C. Z. Wei, Extended least squares and their applications to adaptive control and prediction in linear systems, IEEE Trans. Automat. Control, 31, pp. 898-906, 1986.
- [46] D. Levanoy, On persistent excitation conditions for the filtering of convergent semi martingales, 43th IEEE Conference on Decision and Control, 2004.
- [47] D. Levanoy and P. E. Caines, On persistent excitation for linear systems with stochastic coefficients, SIAM J. Control Optim., Vol. 40, pp. 882-897, 2001.
- [48] T. Li and J.-F. Zhang, Decentralized tracking-type games for multi-agent systems with coupled ARX models: Asymptotic Nash equilibria, Automatica, 44, pp. 713-725, 2008.
- [49] Y. Liu, A. A. Birch, and R. Allen, Dynamic cerebral auto regulation assessment using an ARX model: comparative study using step response and phase shift analysis, Medical Engineering and Physics, 25, pp. 647-653, 2003.
- [50] J. B. Moore, Persistency of excitation in extended least squares, IEEE Trans. Automat. Control, Vol. 28, pp. 60-68, 1983.
- [51] M. Nerlove and K.F. Wallis. Use of the Durbin Watson statistic in inappropriate situations. Econometrica. Vol. 34 no. 1, pp. 235-238. 1966.

- [52] P.C.B. Phillips and M. Loretan. The Durbin Watson ratio under infinite variance errors. *Journal of Econometrics*. Vol. 47, pp. 85-114, 1991.
- [53] Poloni T., Rohal-Ilkiv B., Johansen T.A. Multiple ARX model-based air-fuel ratio predictive control for SI engines. *Advanced Fuzzy and Neural Control*, Vol. 3, Part 3, 2007.
- [54] A. Sapir. Use of the Durbin Watson statistic with lagged dependent variables. *Metroeconomica*. Vol. 29 issue 1-2-3, pp. 169-172, 2006.
- [55] A.K. Sen, M.S. Srivastava, M. Srivastava. *Regression analysis: theory, methods and applications*. Springer Verlag. 1990.
- [56] S.B. Shah, K. M. Moudgalya and K. Ramamritham. Feedback Control of Internet Applications involving the Tracking of Dynamic Data. *Proceedings of the 17th World Congress of the International Federation of Automatic Control*, Seoul Korea, 2008.
- [57] M.S. Srivastava. Asymptotic distribution of Durbin Watson statistic. *Economics Letters*. Vol. 24, pp. 157-160, 1987.
- [58] W. F. Stout, *Almost sure convergence*, Academic Press, New York, 1974.
- [59] R. Vidal, Recursive identification of switched ARX systems, To appear in *Automatica*, 2008.
- [60] H. van der Voeta, P. de Haanb and D.A. Doornbosa. The use of the Durbin-Watson statistic for testing the validity of kinetic models for dissolution. *International Journal of Pharmaceutics*. Vol. 14, issues 2-3, pp. 291-298, 1983.
- [61] R. Wei, J. Joong Im, S. Ha Park, N. Sik Jung, Estimation of Central Blood Pressure Using Radial Pulse Waveform, *isite*, pp. 250-253, 2007 *International Symposium on Information Technology Convergence (ISITC 2007)*, 2007.
- [62] K.J. White. The Durbin Watson test for autocorrelation in nonlinear models. *The Review of Economics and Statistics*. Vol. 47 no. 2, pp. 370-373, 1992.
- [63] W. Windig. The use of the Durbin-Watson criterion for noise and background reduction of complex liquid chromatography/mass spectrometry data and a new algorithm to determine sample differences. *Chemometrics and Intelligent Laboratory Systems*. Vol. 77, Issues 1-2, pp. 206-214, 2005.
- [64] S. Zheng, W. Zheng and X. Ji. Estimation of ARX parametric model in regional economic systems. *Systems Engineering* Vol. 10, Issue 4, pp. 290-296, 2007.
- [65] R. Zhu, D. Sun and Z. Zhou, Integrated design of trajectory planning and control for micro air vehicles. *Mechatronics*, Vol. 17, pp. 245-253, 2007.