Received: 27 November 2018 | Revised: 10 May 2019 | Accepted: 10 May 2019
DOI: 10.1002/acs.3013

SPECIAL ISSUE ARTICLE

WILEY

On the almost sure central limit theorem for ARX processes in adaptive tracking

Bernard Bercu[1] | Víctor Vázquez[2]

[1]Institut de Mathématiques de Bordeaux, UMR CNRS 5251, Université de Bordeaux, Talence, France

[2]Facultad de Ciencias Físico Matemáticas, Universidad Autónoma de Puebla, Puebla, Mexico

Correspondence
Bernard Bercu, Institut de Mathématiques de Bordeaux, UMR CNRS 5251, Université de Bordeaux, 33405 Talence, France.
Email: bernard.bercu@u-bordeaux.fr

Summary

The goal of this paper is to highlight the almost sure central limit theorem (ASCLT) for martingales to the control community. We shall establish the ASCLT for the least squares estimator of the unknown parameter of a controllable ARX(p, q) process in adaptive tracking. The usual notion of controllability for ARX(p, q) processes allows us to avoid the use of a persistent excitation in the adaptive tracking control. We shall also provide strongly consistent estimators of the even moments of the driven noise as well as two quadratic strong laws for the average costs and the estimation errors sequences. Our theoretical results are illustrated by numerical experiments.

KEYWORDS

almost sure central limit theorem, ARX process, controllability, least squares estimation

1 | INTRODUCTION

Zadeh[1] introduced the concept of system identification as the determination, on the basis of input and output, of a system within a specified class of systems to which the system under test is equivalent. Hence, any theoretical result that leads us to make such determination in a more precise way will be a step into the spirit of Zadeh's definition. In order to track some of the most relevant results in system identification, we may find exhaustive and very useful reviews summarizing the most important contributions in this area of applied mathematics. To the best of our knowledge, the most relevant of them are the survey of Åström and Eykhoff,[2] the excellent book of Åström and Wittenmark,[3] the work of Ljung[4] devoted to adaptive tracking in system identification, and the beautiful and captivating book-chapter of Gevers,[5] which deals with the recent developments in identification theory. We also refer the reader to the work of Hong et al[6] for an overview of basic research on model selection approaches for linear systems and to the work of Pillonetto et al[7] for Kernel methods in system identification, machine learning, and functional estimation.

Since the pioneer works on system identification and adaptive control, there has been a great deal of activity from the control community on the theoretical aspects as well as on the practical applications. Recently, Cho et al[8] proposed a new parameter estimation method in the framework of composite model reference adaptive control to improve parameter estimation without persistent excitation. Heydari[9] investigated the stability of adaptive optimal control using value iteration, initiated from a stabilizing control policy. Jaramillo-Lopez et al[10] presented an adaptive control framework for compensation of uncertainties and perturbations that satisfy the matching condition on a class of nonlinear dynamic systems. Moreover, Zhu et al[11] proposed an adaptive model predictive control for unconstrained discrete-time linear systems with parametric uncertainties. We also refer the reader to the work of Gao et al,[12] who investigated the problem of adaptive tracking control for a class of stochastic uncertain nonlinear systems in the presence of input saturation; to the work of Zhao et al,[13] who studied the adaptive control for linear systems with set-valued observations to track a given periodic target; and to the work of Tao et al,[14] for the higher-order tracking properties of model reference adaptive control systems.

Bercu and Vázquez[15,16] investigated the asymptotic behavior of the least squares estimator for ARX process in adaptive tracking.[3] More precisely, a new notion of strong controllability for multidimensional ARX processes was proposed in the work of Bercu and Vázquez.[15] In addition, via a persistently excited version of the adaptive control, it has been shown in another work of Bercu and Vázquez[16] how to avoid this strong controllability condition. In the scalar framework, a serial correlation noise was considered in Bercu et al[17] for ARX processes (see also the work of Bercu et al[18]). The asymptotic behavior of the least squares estimator was analyzed together with the almost sure convergence of the Durbin-Watson statistics as well as its asymptotic normality. It led us to propose a bilateral statistical test for testing whether or not the serial correlation parameter is equal to some nonzero fixed value. We also refer the reader to related works,[19-26] where the asymptotic behavior of the least squares estimator has been extensively investigated in engineering science contexts.

The primary goal of this paper is to highlight the almost sure central limit theorem (ASCLT) for martingales to the control community and to show the usefulness of this result for the system identification of a controllable ARX(p, q) process in adaptive tracking. The ASCLT has been widely investigated in stochastic approximation theory[27-29] and in statistics.[30,31] On the one hand, a large literature is available on the ASCLT for sums of independent random variables.[32-35] On the other hand, it is also possible to find many references on the ASCLT for martingales.[36-39]

Surprisingly, the deep impact of the ASCLT has not deeply reached to the control community. To the best of our knowledge, no reference is available in the engineering literature dealing the ASCLT. Hence, the aim of this paper is to show how the ASCLT for martingales could provide interesting results for the system identification of ARX processes in adaptive tracking by finding consistent estimators of the even moments of driven noise and increasing our knowledge on its distribution. New numerical methods could also be developed for controllable ARX(p, q) processes without persistent excitation in the adpative tracking control.

This paper in organized as follows. Section 2 is devoted to the one-dimensional ARX(p, q) processes in adaptive tracking, whereas the ASCLT for the least squares estimator is given in Section 3. Our theoretical results are illustrated by numerical experiments in Section 4. A short conclusion is given in Section 5. The ASCLT for martingales is provided in Appendix A, whereas all technical proofs are postponed to Appendix B.

2 | ARX PROCESSES

In this section, we focus our attention on the one-dimensional ARX(p, q) processes in adaptive tracking, given for all $n \geq 0$ by

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

where R stands for the shift-back operator; X_n, U_n, and ε_n are the system output, input, and driven noise, respectively. The polynomials A and B are given for all $z \in \mathbb{C}$ by

$$A(z) = 1 - a_1 z - \dots - a_p z^p,$$
$$B(z) = 1 + b_1 z + \dots + b_q z^q,$$

where a_i and b_j are typically unknown real numbers. In all the sequel, we shall make use of the well-known causality assumption on B, also known as the minimum phase condition, as well as the usual notion of controllability for one-dimensional ARX processes. To be more precise, we assume that the polynomial $B(z)$ only has zeros with modulus > 1 and that polynomials $A(z) - 1$ and $B(z)$ are coprime. Relation (1) may be rewritten in the compact form

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

where $\theta^T = (a_1, \dots, a_p, b_1, \dots, b_q)$ and $\Phi_n^T = (X_n^p, U_{n-1}^q)$ with $X_n^p = (X_n, \dots, X_{n-p+1})$ and $U_n^q = (U_n, \dots, U_{n-q+1})$. We shall assume that the driven noise (ε_n) is a martingale difference sequence adapted to the filtration $\mathbb{F} = (\mathcal{F}_n)$, where $\mathcal{F}_n$ is for the σ-algebra of the events occurring up to time n, which means that, for all $n \geq 0$, $\mathbb{E}[\varepsilon_{n+1} | \mathcal{F}_n] = 0$ a.s. Moreover, we assume that, for all $n \geq 0$, $\mathbb{E}[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \sigma^2$ a.s., where $\sigma^2 > 0$. Finally, we assume that (ε_n) satisfies, for some integer $m \geq 1$ and some real number $a > 2m$,

$$\sup_{n \geq 0} \mathbb{E}\left[|\varepsilon_{n+1}|^a | \mathcal{F}_n \right] < \infty \qquad \text{a.s.} \tag{3}$$

The goal of adaptive tracking is to regulate the dynamics of the process (X_n) by forcing the output X_n to track, step by step, a predictable reference trajectory (x_n) such that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k^2 = 0 \qquad \text{a.s.}$$

Moreover, at the same time, we shall also estimate the unknown parameter θ.

First, we focus our attention on the estimation of θ. We shall make use of the least squares estimator, which satisfies, for all $n \geq 0$,

$$\hat{\theta}_{n+1} = \hat{\theta}_n + S_n^{-1}\Phi_n\left(X_{n+1} - U_n - \hat{\theta}_n^T\Phi_n\right), \tag{4}$$

$$S_n = \sum_{k=0}^{n}\Phi_k\Phi_k^T + I_d,$$

where the initial value $\hat{\theta}_0$ may be arbitrarily chosen and I_d is the identity matrix of order $d = p + q$.

Next, we are concerned with the choice of the adaptive control sequence (U_n). We shall make use of the adaptive tracking control proposed by Åström and Wittenmark[3] given, for all $n \geq 0$, by

$$U_n = x_{n+1} - \hat{\theta}_n^T\Phi_n. \tag{5}$$

It has already been shown that this choice of the adaptive tracking control is optimal[15] in the sense that

$$\lim_{n\to\infty} \frac{1}{\log n}\sum_{k=1}^{n}(X_k - x_k - \varepsilon_k)^2 = (p+q)\sigma^2 \quad \text{a.s.}$$

By substituting (5) into (2), we obtain the closed-loop system

$$X_{n+1} - x_{n+1} = \pi_n + \varepsilon_{n+1}, \tag{6}$$

where prediction error $\pi_n = (\theta - \hat{\theta}_n)^T\Phi_n$. Finally, for any integer $m \geq 1$, denote by $(C_n(m))$ and $(G_n(m))$ the sequences of average costs and estimation errors respectively given by

$$C_n(m) = \frac{1}{n}\sum_{k=1}^{n}(X_k - x_k)^{2m} \tag{7}$$

and

$$G_n(m) = \sum_{k=1}^{n} k^{m-1}\|\hat{\theta}_k - \theta\|^{2m}. \tag{8}$$

We assume that the polynomial $B(z)$ only has zeros with modulus > 1. If $r > 1$ is strictly less than the smallest modulus of the zeros of $B(z)$, then $B(z)$ is invertible in the ball with center zero and radius r and $B^{-1}(z)$ is a holomorphic function. For all $z \in \mathbb{C}$ such that $|z| \leq r$, denote

$$P(z) = B^{-1}(z)(A(z) - 1) = \sum_{k=1}^{\infty} P_k z^k. \tag{9}$$

All the coefficients P_k may be explicitly calculated as functions of the coefficients $a_1, \dots, a_p$ and $b_1, \dots, b_q$ in the polynomials A and B (see the work of Bercu and Vázquez[15]). For example, we always have $P_1 = -a_1$. In addition, if $p = q = 1$, then, for all $k \geq 2$, $P_k = -a_1(-b_1)^{k-1}$, whereas, if $p = 2$, $q = 1$, $P_k = (a_1b_1 - a_2)(-b_1)^{k-2}$. Moreover, if $p = 1$, $q = 2$, then $P_2 = a_1b_1$ and $P_3 = a_1(b_2 - b_1^2)$, whereas, if $p = 2$, $q = 2$, $P_2 = a_1b_1 - a_2$, and $P_3 = a_1(b_2 - b_1^2) + a_2b_1$. For any $1 \leq i \leq q$, let

$$H_i = \sum_{k=i}^{\infty} P_k P_{k-i+1}. \tag{10}$$

In addition, denote by H the square matrix of order q,

$$H = \begin{pmatrix} H_1 & H_2 & \dots & H_{q-1} & H_q \\ H_2 & H_1 & H_2 & \dots & H_{q-1} \\ \dots & \dots & \dots & \dots & \dots \\ H_{q-1} & \dots & H_2 & H_1 & H_2 \\ H_q & H_{q-1} & \dots & H_2 & H_1 \end{pmatrix}.$$

Let K be the rectangular matrix of dimension $q \times p$ given, if $p \geq q$, by

$$K = \begin{pmatrix} 0 & P_1 & P_2 & \dots & \dots & P_{p-2} & P_{p-1} \\ 0 & 0 & P_1 & \dots & \dots & P_{p-3} & P_{p-2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & P_1 & P_2 & \dots & P_{p-q+1} \\ 0 & \dots & 0 & 0 & P_1 & \dots & P_{p-q} \end{pmatrix},$$

whereas, if $p \leq q$, by

$$K = \begin{pmatrix} 0 & P_1 & \dots & P_{p-2} & P_{p-1} \\ 0 & 0 & P_1 & \dots & P_{p-2} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & P_1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

Finally, denote by L the square matrix of order $d = p + q$, ie,

$$L = \begin{pmatrix} I_p & K^T \\ K & H \end{pmatrix}. \tag{11}$$

3 | MAIN RESULTS

Our first result deals with the ASCLT for the least squares estimator $\widehat{\theta}_n$ of the unknown parameter θ. Roughly speaking, the ASCLT asserts that, observing a single sample path of the ARX(p, q) process (X_n), we can show that a logarithmic mean of $\sqrt{n}(\widehat{\theta}_n - \theta)$ converges to a Gaussian distribution. In practice, this result can be very useful if it is only possible to observe a single trajectory of (X_n).

Theorem 1. *Assume that the ARX(p, q) process is causal and controllable. Moreover, assume that, for some real number $a > 2$,*

$$\sup_{n \geq 0} \mathbb{E}\left[|\varepsilon_{n+1}|^a | \mathcal{F}_n\right] < \infty \qquad a.s. \tag{12}$$

Then, we have the ASCLT

$$\frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \delta_{\sqrt{k}(\hat{\theta}_k - \theta)} \Rightarrow \mathcal{N}_d(0, L^{-1}) \qquad a.s., \tag{13}$$

where δ is the Dirac delta function. In other words, for any bounded continuous function h,

$$\lim_{n \to \infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} h\left(\sqrt{k}(\hat{\theta}_k - \theta)\right) = \int_{\mathbb{R}^d} h(x) dG(x) \qquad a.s.,$$

where G stands for the $\mathcal{N}_d(0, L^{-1})$ Gaussian measure.

Remark 1. One can observe that the scalar variance σ^2 vanishes in the ASCLT.

Corollary 1. *Assume that the ARX(p, q) process is causal and controllable. Moreover, assume that (ε_n) satisfies, for some integer $m \geq 1$, condition (3). Then, we have*

$$\lim_{n \to \infty} \frac{1}{\log n} \sum_{k=1}^{n} k^{m-1} \left((\hat{\theta}_k - \theta)^T L (\hat{\theta}_k - \theta)\right)^m = \ell(m) \qquad a.s., \tag{14}$$

where

$$\ell(m) = (p + q) \prod_{k=1}^{m-1} (p + q + 2k). \tag{15}$$

Remark 2. In the special case $m = 1$, we can deduce from (14) the quadratic strong law

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} (\widehat{\theta}_k - \theta)^T L(\widehat{\theta}_k - \theta) = p + q \qquad \text{a.s.} \tag{16}$$

We now focus our attention on the average costs and estimation errors sequences $(C_n(m))$ and $(G_n(m))$, respectively given by (7) and (8). First of all, it was proven in lemma 3 in the work of Bercu and Vázquez[15] that matrix L is positive definite. Hence, (14) immediately implies that

$$G_n(m) = \mathcal{O}(\log n) \qquad \text{a.s.}$$

Furthermore, denote

$$\Gamma_n(m) = \frac{1}{n} \sum_{k=1}^{n} \varepsilon_k^{2m}.$$

The asymptotic behavior of $(C_n(m))$ is as follows.

Corollary 2. *Assume that the ARX(p, q) process is causal and controllable. Moreover, assume that (ε_n) satisfies, for some integer $m \geq 1$, condition (3). Suppose that it exists some integer $1 \leq s \leq m$ such that $\mathbb{E}[\varepsilon_{n+1}^{2s} | \mathcal{F}_n] = \sigma(2s)$ a.s. Then, $C_n(s)$ is a strongly consistent estimator of $\sigma(2s)$, ie,*

$$\lim_{n\to\infty} C_n(s) = \sigma(2s) \qquad \textit{a.s.} \tag{17}$$

More precisely, for all $0 < b < 1$ such that $2m < ab$, we also have

$$(C_n(s) - \Gamma_n(s))^2 = o\left(n^{b-1}\right) \qquad \textit{a.s.} \tag{18}$$

4 | NUMERICAL EXPERIMENTS

We provide now some numerical experiments to illustrate the most relevant almost sure results of Section 3. More precisely, we shall focus attention on the quadratic strong law given by (16) as well as on the almost sure convergence of even moments given by convergence (17) for different values of m. For the sake of simplicity, we assume that the driven noise (ε_n) is a sequence of independent and identically distributed random variables sharing the same $\mathcal{N}(0, \sigma^2)$ distribution with $\sigma = 0.8$, and the reference trajectory (x_n) is identically zero. Consider the ARX(2, 2) process given by (1), where

$$A(z) = 1 + \frac{6}{5}z - \frac{1}{2}z^2 \qquad \text{and} \qquad B(z) = 1 + \frac{2}{5}z + \frac{1}{4}z^2.$$

One may observe that B is causal since its complex roots have modulus 2. Moreover, one may easily check that the process is controllable and matrix L is given by

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 6/5 & 0 \\ 0 & 6/5 & 244/99 & -628/495 \\ 0 & 0 & -628/495 & 244/99 \end{pmatrix}.$$

We already saw in Section 2 that the adaptive tracking control given in (5) is optimal.[15] Figure 1 shows one single sample path of the system output (X_n) (blue line) and the system input (U_n) (red line) with sample size $n = 100$. One may observe the effect of the adaptive tracking control U_n on the behavior of X_n.

In order to illustrate the quadratic strong law given by (16), the sample size will increase from $n = 100$ to $n = 5000$, and we shall denote by Δ_n the average of $N = 100$ values of

$$\frac{1}{\log n} \sum_{k=1}^{n} (\widehat{\theta}_k - \theta)^T L(\widehat{\theta}_k - \theta).$$

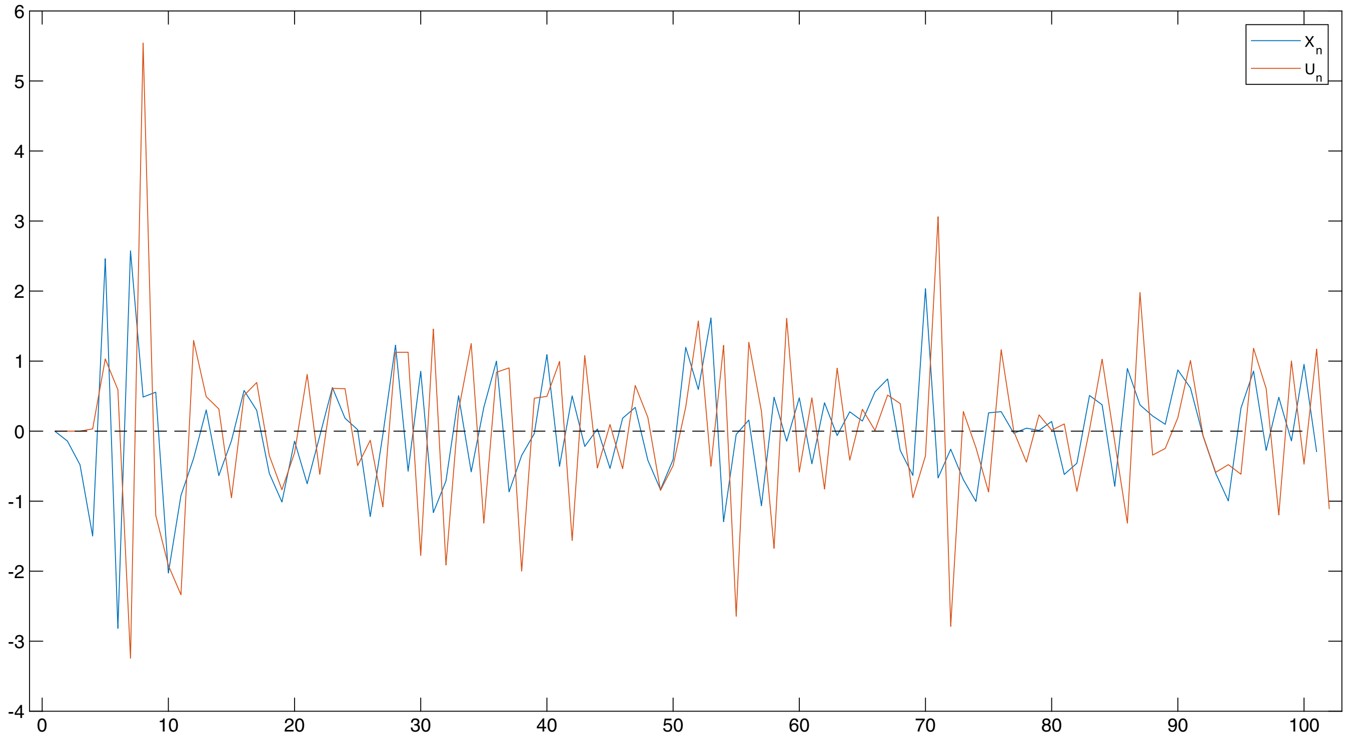

FIGURE 1 Tracking performance [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Quadratic strong law

n	Δ_n	Relative Error
100	4.33	8.14%
500	4.14	3.51%
1 000	3.97	0.69%
2 000	4.03	0.78%
5 000	3.972	0.70%

TABLE 2 Convergence of even moments

m	$C_n(m)$	$\sigma(2m)$	Relative Error
1	0.6426	0.64	0.41%
2	1.245	1.229	1.84%
3	4.07	3.932	3.51%
4	18.81	17.62	6.75%
5	110.61	101.47	9.00%

We may conclude by observing Table 1 that, even with the slow growth of the logarithmic function, relative errors are small, and the quadratic strong law is nicely shown. We recall here that the almost sure limit given by (16) is $p + q = 4$.

Let us deal now with the almost sure convergence of even moments given by convergence (17). For that purpose, we shall consider the average of $N = 100$ values of sample size $n = 10\,000$ of $C_n(m)$. The corresponding results are presented in Table 2, where the values of m increase from 1 to 5.

We observe that as the value of m increases, the relative error also increases. In other words, it is necessary to take large sample sizes to estimate large order even moments. For example, choose the value of $m = 5$ and consider the large sample sizes $n = 20\,000$, $30\,000$, and $50\,000$, as indicated in Table 3. As expected, for large sample sizes values, the almost sure convergence of even moments can be improved substantially.

TABLE 3 Estimation of $\sigma(10)$

n	$C_n(5)$	Relative Error
20 000	108.54	6.90%
30 000	106.15	4.61%
50 000	105.44	3.91%

5 | CONCLUSION

In this paper, we established the ASCLT for the least squares estimator of the unknown parameter of a controllable ARX(p, q) process in adaptive tracking. We have also provided strongly consistent estimators for the even moments of the driven noise. Such convergence results have been obtained without a persistent excitation in the adaptive tracking control and only the usual notion of controllability was required. Even when most of the engineering methods do not consider high-order moments, it is useful to go deeper into the knowledge of the driven noise distribution through the estimation of such moments because it gives us a better notion of the underlying uncertainty. One possible direction for future research concerns the extension of the asymptotical results of this paper to ARMAX(p, q, r) processes in adaptive tracking. Another interesting topic is the application of our results to real-life example arising from the adaptive control area. Finally, if the controllability assumption is not satisfied or cannot be guaranteed, then two strategy are available. The first one is to introduce a persistent excitation in the adaptive tracking control.[16] The second one is to make use of the less restrictive notion of interval excitation, as recently introduced in the work of Pan et al.[21]

ACKNOWLEDGEMENTS

The authors would like to thank the associate editor and the three anonymous reviewers for their constructive comments and suggestions, which helped to improve this paper substantially.

ORCID

Bernard Bercu https://orcid.org/0000-0002-8118-5232
Víctor Vázquez https://orcid.org/0000-0001-6602-8733

REFERENCES

1. Zadeh LA. From circuit theory to system theory. *Proceedings IRE*. 1962;50:856-865.
2. Åström KJ, Eykhoff P. System identification—a survey. *Automatica*. 1971;7:123-162.
3. Åström KJ, Wittenmark B. *Adaptive Control*. New York, NY: Addison-Wesley; 1995.
4. Ljung L, Gunnarsson S. Adaptation and tracking in system identification—a survey. *Automatica*. 1990;26:7-21.
5. Gevers M. System identification without Lennart Ljung: what would have been different?. In: *Forever Ljung in System Identification*. Lund, Sweden: Studentlitteratur; 2006.
6. Hong X, Mitchell RJ, Chen S, Harris CJ, Li K, Irwin GW. Model selection approaches for non-linear system identification: a review. *Int J Syst Sci*. 2008;39:925-946.
7. Pillonetto G, Dinuzzo F, Chen T, De Nicolao G, Ljung L. Kernel methods in system identification, machine learning and function estimation: a survey. *Automatica*. 2014;50:657-682.
8. Cho N, Shin HS, Kim Y, Tsourdos A. Composite model reference adaptive control with parameter convergence under finite excitation. *IEEE Trans Autom Control*. 2018;63(3):811-818.
9. Heydari A. Stability analysis of optimal adaptive control using value iteration with approximation errors. *IEEE Trans Autom Control*. 2018;63(9):3119-3126.
10. Jaramillo-Lopez F, Kenne G, Lamnabhi-Lagarrigue F. Adaptive control for a class of uncertain nonlinear systems: application to photovoltaic control systems. *IEEE Trans Autom Control*. 2017;62:393-398.
11. Zhu B, Xia X. Adaptive model predictive control for unconstrained discrete-time linear systems with parametric uncertainties. *IEEE Trans Autom Control*. 2016;61:3171-3176.
12. Gao YF, Sun XM, Wen C, Wang W. Adaptive tracking control for a class of stochastic nonlinear systems with input saturation. *IEEE Trans Autom Control*. 2017;62:2498-2504.
13. Zhao Y, Guo J, Zhang JF. Adaptive tracking control of linear systems with binary-valued observations and periodic target. *IEEE Trans Autom Control*. 2013;58:1293-1298.
14. Tao G, Song G. Higher order tracking properties of model reference adaptive control systems. *IEEE Trans. Autom Control*. 2018;63(11):3912-3918.

15. Bercu B, Vázquez V. A new concept of strong controllability via the Schur complement in adaptive tracking. *Automatica*. 2010;46:1799-1805.
16. Bercu B, Vázquez V. On the usefulness of persistent excitation in ARX adaptive tracking. *Int J Control*. 2010;83:1145-1154.
17. Bercu B, Portier B, Vázquez V. On the asymptotic behavior of the Durbin-Watson statistic for ARX processes in adaptive tracking. *Int J Adapt Control Signal Process*. 2014;28:1002-1023.
18. Bercu B, Portier B, Vázquez V. A Durbin-Watson serial correlation test for ARX processes via excited adaptive tracking. *Int J Control*. 2015;88(12):2611-2618.
19. Cheng MY, Prayogo D, Wu YW. Prediction of permanent deformation in asphalt pavements using a novel symbiotic organisms search-least squares support vector regression. *Neural Comput Appl*. 2018. https://doi.org/10.1007/s00521-018-3426-0
20. Meng XL, Shi FG, Yao JC. An inequality approach for evaluating decision making units with a fuzzy output. *J Intell Fuzzy Syst*. 2018;34(1):459-465.
21. Pan Y, Sun T, Yu H. On parameter convergence in least squares identification and adaptive control. *Int J Robust Nonlinear Control*. 2019. Early access. https://doi.org/10.1002/rnc.4527
22. de Jesús Rubio J. SOFMLS: online self-organizing fuzzy modified least-squares network. *IEEE Trans Fuzzy Syst*. 2009;17(6):1296-1309.
23. de Jesús Rubio J. Stable Kalman filter and neural network for the chaotic systems identification. *J Frankl Inst*. 2017;354(16):7444-7462.
24. de Jesús Rubio J, Lughofer E, Meda-Campaña JA, Páramo LA, Novoa JF, Pacheco J. Neural network updating via argument Kalman filter for modeling of Takagi-Sugeno fuzzy models. *J Intell Fuzzy Syst*. 2018;35(2):2585-2596.
25. Wang X, Ding F. Convergence of the recursive identification algorithms for multivariate pseudo-linear regressive systems. *Int J Adapt Control Signal Process*. 2016;30:824-842.
26. Zhang XM, Han QL. State estimation for static neural networks with time-varying delays based on an improved reciprocally convex inequality. *IEEE Trans Neural Netw Learn Syst*. 2018;29(4):1376-1381.
27. Bercu B, Cénac P, Fayolle G. On the almost sure central limit theorem for vector martingales: convergence of moments and statistical applications. *J Appl Probab*. 2009;46:151-169.
28. Cénac P. On the convergence of moments in the almost sure central limit theorem for stochastic approximation algorithms. *ESAIM Probab Stat*. 2013;17:179-194.
29. Pelletier M. An almost sure central limit theorem for stochastic approximation algorithms. *J Multivar Anal*. 1999;71:76-93.
30. Bin T, Zuoxiang P, Nadarajah S. An extension of almost sure central limit theorem for order statistics. *Extremes*. 2009;12:201-209.
31. Holzmann H, Koch S, Min A. Almost sure limit theorems for U-statistics. *Stat Probab Lett*. 2004;69:261-269.
32. Berkes I, Csáki E. A universal result in almost sure central limit theory. *Stoch Process Their Appl*. 2001;94:105-134.
33. Brosamler GA. An almost everywhere central limit theorem. *Math Proc Cambridge Philos Soc*. 1988;104:561-574.
34. Lacey MT, Phillip W. A note on the almost sure central limit theorem. *Stat Probab Lett*. 1990;9:201-205.
35. Schatte P. On strong versions of the central limit theorem. *Math Nachr*. 1988;137:249-256.
36. Bercu B, Fort JC. A moment approach for the almost sure central limit theorem for martingales. *Stud Sci Math Hung*. 2008;45:139-159.
37. Chaâbane F, Maâouia F. Théorèmes limites avec poids pour les martingales vectorielles. *ESAIM Probab Stat*. 2000;4:137-189.
38. Chaâbane F. Invariance principles with logarithmic averaging for martingales. *Stud Sci Math Hung*. 2001;37:21-52.
39. Lifshits MA. Almost sure limit theorem for martingales. *Limite Theorems Probab Stat*. 2002;2:367-390.
40. Bercu B. On the convergence of moments in the almost sure central limit theorem for martingales with statistical applications. *Stoch Process Their Appl*. 2004;111:157-173.
41. Chaâbane F, Maâouia F, Touati A. Généralisation du théorème de la limite centrale presque-sûr pour les martingales vectorielles. *Comptes Rendus Acad Sci-Ser I-Math*. 1998;326:229-232.
42. Duflo M. *Random Iterative Models*. Berlin, Germany: Springer-Verlag; 1997.

How to cite this article: Bercu B, Vázquez V. On the almost sure central limit theorem for ARX processes in adaptive tracking. *Int J Adapt Control Signal Process*. 2019;33:1901–1911. https://doi.org/10.1002/acs.3013

APPENDIX A

ON THE ALMOST SURE CENTRAL LIMIT THEOREM FOR MARTINGALES

The goal of this Appendix is to highlight the ASCLT for martingales[40,27,37-39] to the control community. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space endowed with a filtration $\mathbb{F} = (\mathcal{F}_n)$, where $\mathcal{F}_n$ is for the σ-algebra of the events occurring up to time n. Assume that (M_n) be a sequence of integrable random vectors in $\mathbb{R}^d$ such that, for all $n \geq 0, M_n$ is $\mathcal{F}_n$-measurable. We shall say that (M_n) is a martingale with respect to the filtration $\mathbb{F}$ if, for all $n \geq 0$, $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$ almost surely. Throughout this Appendix, (ε_n) is a martingale difference sequence adapted to $\mathbb{F}$. It means that, for all $n \geq 0$, $\mathbb{E}[\varepsilon_{n+1}|\mathcal{F}_n] = 0$ a.s. We also assume that, for all $n \geq 0$, $\mathbb{E}[\varepsilon_{n+1}^2|\mathcal{F}_n] = \sigma^2$ a.s., where $\sigma^2 > 0$. Let (Φ_n) be a sequence of random vectors of $\mathbb{R}^d$, adapted to $\mathbb{F}$. Denote by (M_n) the locally square integrable martingale

$$M_n = M_0 + \sum_{k=1}^{n} \Phi_{k-1}\varepsilon_k,$$

where the initial value M_0 can be taken arbitrarily. Its increasing process $\langle M\rangle_n$ is defined, for all $n \geq 1$, by

$$\langle M\rangle_n = \sum_{k=1}^{n} \mathbb{E}\left[\Delta M_k \Delta M_k^T|\mathcal{F}_{k-1}\right]$$

where $\Delta M_k = M_k - M_{k-1}$. We clearly have $\langle M\rangle_n = \sigma^2 S_{n-1}$, where

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

A simplified version of the ASCLT for multivariate martingales is as follows.[41]

Theorem 2. *Assume that it exists a positive definite symmetric matrix L such that*

$$\lim_{n\to\infty} \frac{1}{n} S_n = L \quad a.s. \tag{A1}$$

Moreover, assume that (M_n) satisfies Lindeberg's condition which means that, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[||\Delta M_k||^2 \mathrm{I}_{\left\{||\Delta M_k|| \geq \varepsilon\sqrt{n}\right\}}|\mathcal{F}_{k-1}\right] = 0 \quad a.s. \tag{A2}$$

Then, we have the ASCLT

$$\frac{1}{\log n}\sum_{k=1}^{n} \frac{1}{k}\delta_{M_k/\sqrt{k}} \Rightarrow \mathcal{N}_d(0, \sigma^2 L) \quad a.s.,$$

where δ is the Dirac delta function. In other words, for any bounded continuous function h,

$$\lim_{n\to\infty} \frac{1}{\log n}\sum_{k=1}^{n} \frac{1}{k} h\left(\frac{M_k}{\sqrt{k}}\right) = \int_{\mathbb{R}^d} h(x)dG(x) \quad a.s.,$$

where G stands for the $\mathcal{N}_d(0, \sigma^2 L)$ Gaussian measure.

The convergence of the even moments in the ASCLT for multivariate martingales was established in the work of Bercu et al.[27]

Theorem 3. *Assume that the almost sure convergence (A1) is satisfied. In addition, suppose that, for some integer $m \geq 1$ and for some real number $a > 2m$,*

$$\sup_{n\geq 0} \mathbb{E}\left[|\varepsilon_{n+1}|^a|\mathcal{F}_n\right] < \infty \qquad a.s. \tag{A3}$$

Then, we have

$$\lim_{n\to\infty} \frac{1}{\log n}\sum_{k=1}^{n} \frac{1}{k}\left(M_k^T S_{k-1}^{-1} M_k\right)^m = \ell(m) \qquad a.s., \tag{A4}$$

where

$$\ell(m) = d\sigma^{2m}\prod_{k=1}^{m-1}(d + 2k). \tag{A5}$$

Remark 3. The limit $\ell(m)$ corresponds exactly to the mean value of $||Z||^{2m}$, where Z has a standard $\mathcal{N}_d(0, \sigma^2 I_d)$ distribution. Consequently, Theorem 3 can be seen as the convergence of moments of order $2m$ in the ASCLT for multivariate martingales.

APPENDIX B

PROOFS OF OUR MAIN RESULTS

Proof of Theorem 1. It follows from (2) and (4) that, for all $n \geq 1$,

$$\widehat{\theta}_n - \theta = S_{n-1}^{-1} M_n, \tag{B1}$$

where

$$M_n = M_0 + \sum_{k=1}^{n} \Phi_{k-1} \varepsilon_k,$$

with $M_0 = \widehat{\theta}_0 - \theta$. It was proven in theorem 5 in the work of Bercu and Vázquez[15] that

$$\lim_{n\to\infty} \frac{S_n}{n} = \Lambda \qquad \text{a.s.,} \tag{B2}$$

where $\Lambda = \sigma^2 L$ and the limiting matrix L is given by (11). Moreover, we can deduce from (12) that Lindeberg's condition (A2) is satisfied. Consequently, we obtain (13) from Theorem 2 together with (B1) and (B2). ☐

Proof of Corollary 1. The almost sure convergence (14) follows from the conjunction of Theorem 1 together with (B1) and (B2), using the same arguments as in the proof of corollary 3.3 in the work of Bercu et al.[27] ☐

Proof of Corollary 2. For any integer $1 \leq s \leq m$, we obtain from (6) that

$$\begin{aligned} n(C_n(s) - \Gamma_n(s)) &= \sum_{k=1}^{n} (X_k - x_k)^{2s} - \sum_{k=1}^{n} \varepsilon_k^{2s} \\ &= \sum_{k=1}^{n} (\pi_{k-1} + \varepsilon_k)^{2s} - \sum_{k=1}^{n} \varepsilon_k^{2s} \\ &= \sum_{k=1}^{n} \pi_{k-1}^{2s} + R_n(s), \end{aligned} \tag{B3}$$

where

$$R_n(s) = \sum_{l=1}^{2s-1} \sum_{k=1}^{n} \binom{2s}{l} \pi_{k-1}^{2s-l} \varepsilon_k^l.$$

On the one hand, in the special case $s = 1$, we deduce from Theorem 3 with $m = 1$ that

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \pi_k^2 = \sigma^2 (p+q) \qquad \text{a.s.} \tag{B4}$$

In addition, we find from the strong law of large numbers for martingales[42] that

$$R_n(1) = o(\log n) \quad \text{a.s.} \tag{B5}$$

Hence, we obtain from (B3) together with (B4) and (B5) that

$$\lim_{n\to\infty} \frac{n}{\log n} (C_n(1) - \Gamma_n(1)) = \sigma^2 (p+q) \qquad \text{a.s.} \tag{B6}$$

Therefore, as

$$\lim_{n\to\infty} \Gamma_n(1) = \sigma^2 \qquad \text{a.s.},$$

convergence (B6) clearly leads to (17) since $\sigma(2) = \sigma^2$. On the other hand, for $2 \leq s \leq m$, it follows from convergence (B2) that $\log d_n \sim (p+q)\log n$, where $d_n = \det(S_n)$. Consequently, corollary 3.1 in the work of Bercu et al[27] lead us to

$$\lim_{n\to\infty} \frac{1}{\log n}\sum_{k=1}^{n}\pi_k^{2s} = 0 \qquad \text{a.s.} \tag{B7}$$

Moreover, it is not hard to see that the remainder term $R_n(s)$ plays a negligible role. Hence, as

$$\lim_{n\to\infty} \Gamma_n(s) = \sigma(2s) \qquad \text{a.s.},$$

we obtain (17) from (B3) and (B7). Finally, we deduce (18) from remark 3.1 of Bercu et al,[27] which completes the proof of Corollary 2. □