This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2014.2373711, IEEE Transactions on Automatic Control

A multi-time-scale generalization of recursive identification algorithm for ARMAX systems⋆

Lirong Huang and Håkan Hjalmarsson

***Abstract*—Recently, [5] presented a new approach to recursive identification for ARMAX systems, which is a three-stage recursive scheme and assumes independent and identically distributed input signals. Here, we observe that, unless the time scale of the algorithm at one stage is reasonably faster than those at the previous stages, convergence to true value may not take place. To remedy this issue, this note proposes a multi-time-scale modification of the algorithm in [5] such that convergence is achieved. In addition, the new scheme handles a wider class of input signals so that the input can be designed for some purpose.The advantage of the multi-time scale algorithm is verified with numerical examples.**

***Index Terms*—ARMAX models, L-mixing processes, multi-time-scale method, quasi-stationary signals, recursive estimation, stochastic approximation.**

I. Introduction

The ARMAX is a widely used stochastic model in economics [1], engineering [12], medicine [23], and science [20] literature and has been intensively studied over several decades, see, e.g., [5], [6], [11], [15]-[19]. One of the challenging problems is to identify the MA-part. The estimation of MA parameters has been a research problem in signal processing and system identification for the past decades (see [5], [11], [8], [9], [16], [21], [24]). The parameter estimation of MA signals from second-order statistics was deemed for a long time to be a difficult nonlinear problem for which no computationally convenient and reliable solution was possible [24]. For the convergence of estimates of MA-part, the strictly positive realness (SPR) condition is normally imposed in the existing results (see [6] ,[15]-[17], [19], [21]). But, as shown in [11] (see also [6]), it is possible to weaken the SPR condition by an over-parametrization technique or an increasing lag method, which, however, may complicate the algorithm and require additional information [5].

It is well known that stochastic recursive algorithms, also known as stochastic approximation [2], [4], [9], [13], [14], [22], have various applications and are a powerful tool for on-line identification that is a key instrument in adaptive control, adaptive filtering, adaptive prediction and adaptive signal processing problems (see [6], [17]-[19]). Recently, [5] presented a new approach of stochastic approximation to recursive identification for ARMAX systems, where the input signal is assumed to be an i.i.d. (independent and identically distributed) sequence and the i.i.d. property plays an important role in the analysis. It is interesting to generalize the result to a case where the input signal is not an i.i.d. sequence but is designed for some purpose, which is also suggested in [5]. To do so, we relax the condition on the concerned signals to a class of quasi-stationary signals (see Lemmas 2.1-2.3), apply L-mixing techniques (see Definition 2.2 and Lemma 2.4), and propose a new scheme for X-part (see (11) below). More importantly, it is noticed that the three-stage recursive scheme in [5] is composed of three coupled recursive algorithms that should be of different time scales, e.g., algorithm (26)-(27) in [5] calculates estimate X_k in the environment where $\widetilde{\zeta}_k$ is recursively updated. When the algorithm operates in a (slowly) varying environment, the time scale of the algorithm should remain reasonably faster than that of the changing environment for otherwise it would never adapt (see [14], [3], [13]). This note proposes a multi-time-scale variant of the recursive algorithm in [5] such that it works with quasi-stationary input signals and hence has wider applicability to the problems of recursive identification for ARMAX systems. Our proposed algorithm may be regarded as a multi-time-scale generalization of that in [5] and is applicable to a wider range of estimation problems, which is explained (Remarks 3.2-3.3) and verified with numerical examples where the scheme in [5] does not work (Section V). To highlight and focus on the new techniques, we consider single-input single-output (SISO) systems but our results can be extended to multi-input multi-output (MIMO) systems of the form in [5].

II. Preliminaries: system and input signal

Our problem will be embedded in an underlying complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\geq 0}, \mathbb{P})$ with a natural filtration $\mathcal{F}_{n_1} \subset \mathcal{F}_{n_2}$ for $n_2 > n_1$, where Ω is the sample space, $\mathcal{F}$ is the σ-algebra that defines events E in Ω which are measurable. Let $\mathbb{E}[\cdot]$ be the expectation operator with respect to the probability measure. Let $(\mathcal{F}_n, \mathcal{F}_n^+)$, $n \geq 0$, be a pair of families of σ-algebras such that (i) $\mathcal{F}_n \subset \mathcal{F}$ is monotone increasing, (ii) $\mathcal{F}_n^+ \subset \mathcal{F}$ is monotone decreasing, and (iii) $\mathcal{F}_n$ and $\mathcal{F}_n^+$ are independent for all $n \geq 0$.

Let us consider the ARMAX-system

$$A^*(q)y_n = B^*(q)u_{n-1} + C^*(q)e_n \tag{1}$$

where $A^*(q)$, $B^*(q)$ and $C^*(q)$ are polynomials in the backward shift operator q^{-1} of degrees p_a, p_b, and p_c, respec-

⋆ This work was partially supported by the European Research Council under the advanced grant LEARN, contract 26738, and by the Swedish Research Council under contract 621-2009-4017.

L. Huang was with ACCESS Linnaeus Center and Automatic Control Lab, KTH, Stockholm, Sweden and is now with Institute of Molecular Systems Biology, ETH Zurich, Switzerland; H. Hjalmarsson is with ACCESS Linnaeus Center and Automatic Control Lab, KTH, Stockholm, Sweden. Email: lirong.huang@imsb.biol.ethz.ch (L. Huang).

Preprint submitted to IEEE Transactions on Automatic Control. Received: November 18, 2014 10:01:06 PST

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2014.2373711, IEEE Transactions on Automatic Control

tively: $A^*(q) = 1 + \sum_{j=1}^{p_a} a_j^* q^{-j}$, $q^{-1}B^*(q) = \sum_{j=1}^{p_b} b_j^* q^{-j}$ and $C^*(q) = 1 + \sum_{j=1}^{p_c} c_j^* q^{-j}$; noise process $\{e_n\}$ is an i.i.d. sequence. Without loss of generality, let $p_c \geq 1$ (see [5]). Assume that the system is at rest prior to time $n = 0$, i.e., $y_n = u_n = e_n = 0$ for $n < 0$. Write $\theta^* = \begin{bmatrix} \theta_A^{*T} & \theta_B^{*T} & \sigma_e^{*2} & \theta_C^{*T} \end{bmatrix}^T \in \mathbb{R}^{p_a+p_b+p_c+1}$ with $\theta_A^* = \begin{pmatrix} a_1^* & a_2^* & \cdots & a_{p_a}^* \end{pmatrix}^T \in \mathbb{R}^{p_a}$, $\theta_B^* = \begin{pmatrix} b_1^* & b_2^* & \cdots & b_{p_b}^* \end{pmatrix}^T \in \mathbb{R}^{p_b}$, $\theta_C^* = \begin{pmatrix} c_1^* & c_2^* & \cdots & c_{p_c}^* \end{pmatrix}^T \in \mathbb{R}^{p_c}$ for the true parameter vector and denote by θ an arbitrary vector of the same structure, $\theta = \begin{bmatrix} \theta_A^T & \theta_B^T & \sigma_e^2 & \theta_C^T \end{bmatrix}^T$ with $\theta_A = (a_1 \ a_2 \ \cdots \ a_{p_a})^T$, $\theta_B = (b_1 \ b_2 \ \cdots \ b_{p_b})^T$ and $\theta_C = (c_1 \ c_2 \ \cdots \ c_{p_c})^T$.

Obviously, in the SISO case (i.e., $m = l = 1$ in [5]), $[a_{p_a}^* \ \ b_s^* b_1^* + c_s^* \sigma_e^{*2}]$ is of full-row-rank because (p_a, p_b, p_c) are the true orders of the system (1), where $s = \max\{p_b, p_c\}$, $b_s^* = 0$ if $s > p_b$ and $c_s^* = 0$ if $s > p_c$(cf. A2 in [5]). Assumptions used in this work are listed as follows, which are the counterpart of [5, Condition A0-A3].

A1. $\mathbb{E}[e_n] = 0$, $\mathbb{E}[e_n^2] = \sigma_e^{*2} > 0$ and $\mathbb{E}[|e_n|^q] < \infty$ for all $q \geq 1$, where σ_e^{*2} is unknown.

A2. $A^*(z)$ is stable (i.e., $A^*(z) \neq 0$ for $|z| \geq 1$) and $C^*(z) \neq 0$ for all $|z| \geq 1$.

A3. $A^*(z)$ and $B^*(z)\Psi_u B^*(z^{-1}) + C^*(z)\sigma_e^{*2} C^*(z^{-1})$ have no common left factor for all $\Psi_u \geq 0$.

A4. The input signal is generated by

$$u_n = f_{n,0}\, w_n \tag{2}$$

where $\{w_n\}$ is an i.i.d. sequence independent of $\{e_n\}$ such that $\mathbb{E}[w_n] = 0$, $\mathbb{E}[w_n^2] = 1$ and $\mathbb{E}[|w_n|^k] < \infty$ for all $k \geq 1$ and $f_{n,0}$ is $\mathcal{F}_{n-1}$ measurable (and hence is independent of w_n) for all $n \geq 1$ with respect to $\mathcal{F}_n = \sigma\{e_t, w_t : 0 \leq t \leq n\}$ and $\mathcal{F}_n^+ = \sigma\{e_t, w_t : t \geq n+1\}$. And $f_{n,0}$ is designed to satisfy conditions (3)-(5) below.

Define $r_{n,j} = \mathbb{E}[u_n u_{n-j} | \mathcal{F}_{n-1}]$ and assume

$$0 < r_{min} \leq r_{n,0} = f_{n,0}^2 \leq r_{max} < +\infty \tag{3}$$

for all $n \geq 1$. So $r_{n,j} = 0$ if $j > 0$ for all $n \geq 1$. Let sequence $\{\lambda_n\}$ be defined by $\lambda_0 = 0$ and

$$\lambda_n = \frac{1}{n}\sum_{k=1}^{n} r_{k,0} = \frac{1}{n}\big[(n-1)\lambda_{n-1} + r_{n,0}\big], \quad n \geq 1. \tag{4}$$

It is also assumed that, at step $n \geq 1$, the input signal is generated by (2) such that

$$\big|r_{n,0} - \lambda_{n-1}\big| \leq \frac{L_\lambda}{n^\alpha}, \tag{5}$$

where $L_\lambda \geq 0$ is usually set large and $\alpha > 0$ small in practice so that, for a control purpose with respect to any experiment length (see, e.g., [10, (34)-(36)]), this constraint is soft enough and it never becomes active throughout the experiment.

Remark 2.1: By the tower property, (3) implies that the auto-correlations $\widehat{r}_{n,j} = \mathbb{E}[u_n u_{n-j}]$ satisfies $0 < r_{min} \leq \widehat{r}_{n,0} = \mathbb{E}[r_{n,0}] = \mathbb{E}[f_{n,0}^2] \leq r_{max} < +\infty$ and $\widehat{r}_{n,j} = 0$ if $j > 0$ for all $n \geq 1$ while (5) implies that $-\frac{L_\lambda}{n^\alpha} \leq \widehat{r}_{n,0} - \widehat{\lambda}_{n-1} \leq \frac{L_\lambda}{n^\alpha}$ for all $n \geq 1$, where $\widehat{\lambda}_0 = 0$ and

$$\widehat{\lambda}_n = \frac{1}{n}\sum_{k=1}^{n} \widehat{r}_{k,0} = \frac{1}{n}\big[(n-1)\widehat{\lambda}_{n-1} + \widehat{r}_{n,0}\big]. \tag{6}$$

It is observed that $\{u_n\}$ is an i.i.d. sequence if $L_\lambda = 0$ [5].

Let us introduce the definition and present the following results on quasi-stationarity (see [18, Definition 2.1, p34]) of our concerned signals.

Definition 2.1: A signal $\{s_n\}$ is said to be quasi-stationary if it is subject to: (i) $\mathbb{E}[s_n] = m_n^s$, $|m_n^s| \leq C$, $\forall n$, (ii) $\mathbb{E}[s_n s_k] = R_{n,k}^s$, $|R_{n,k}^s| \leq C$ and $\overline{\mathbb{E}}[s_k s_{k-\tau}] = \lim_{n\to\infty}(1/n)\sum_{k=1}^{n} R_{k,k-\tau}^s = R_\tau^s$, $\forall \tau$, where $\overline{\mathbb{E}}[\cdot]$ is defined by [18, (2.60), p34], i.e., $\overline{\mathbb{E}}[f(k)] = \lim_{n\to\infty}(1/n)\sum_{k=1}^{n}\mathbb{E}[f(k)]$.

Lemma 2.1: The process $\{u_n\}$ generated by (2) is a quasi-stationary signal if conditions (3)-(5) hold.

Proof. It is obvious that $\big|\mathbb{E}[u_n]\big| \leq \sqrt{\mathbb{E}[u_n^2]} = \sqrt{r_{max}}$ and $\big|\mathbb{E}[u_n u_k]\big| \leq \frac{1}{2}(\mathbb{E}[u_n^2] + \mathbb{E}[u_k^2]) \leq r_{max}$ for all n, k. According to Definition 2.1, now we only need to show that, for any τ, there exists $\bar{r}_\tau$ such that $\bar{r}_\tau = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[u_n u_{n-\tau}] = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\widehat{r}_{n,n-\tau}$. Clearly, this holds for $\tau > 0$ since $\widehat{r}_{n,\tau} = \mathbb{E}[u_n u_{n-\tau}] = 0$ for all n. If $\bar{\tau} = -\tau > 0$, then $\widehat{r}_{n,\tau} = \mathbb{E}[u_n u_{n-\tau}] = \mathbb{E}[u_{n-\tau} u_n] = \widehat{r}_{n+\bar{\tau},\bar{\tau}}$ for all $n \geq 1$. It is easy to see that $\bar{r}_\tau = \lim_{N\to\infty}(1/N)\sum_{n=1}^{N}\widehat{r}_{n,n-\tau} = \lim_{N\to\infty}(1/N)\sum_{n=1}^{N}\widehat{r}_{n+\bar{\tau},\bar{\tau}} = \bar{r}_{\bar{\tau}} = 0$. Notice that $\big\{\sum_{k=1}^{n}\big|\widehat{\lambda}_k - \widehat{\lambda}_{k-1}\big|\big\}$ is a Cauchy sequence since $\big|\widehat{\lambda}_k - \widehat{\lambda}_{k-1}\big| = \frac{1}{k}\big|\widehat{r}_{k,0} - \widehat{\lambda}_{k-1}\big| \leq \frac{L_\lambda}{k^{1+\alpha}}$ for all $k \geq 1$. This implies that $\{\widehat{\lambda}_n\}$ with $\widehat{\lambda}_n = \sum_{k=1}^{n}\widehat{\lambda}_k - \widehat{\lambda}_{k-1}$ is a convergent sequence and hence, by (6),

$$\bar{r}_0 = \lim_{n\to\infty}\widehat{\lambda}_n. \tag{7}$$

So the process $\{u_n\}$ is a quasi-stationary signal. □

Lemma 2.2: The quasi-stationary signals $\{u_n\}$ and $\{e_n\}$ are uncorrelated.

Proof. The jointly quasi-stationary signals $\{u_n\}$ and $\{e_n\}$ are uncorrelated if their cross-covariance function $\overline{\mathbb{E}}[u_n e_{n-\tau}] = 0$ for all τ. Note that $\{u_n\}$ is generated by (2), where $f_{n,0}$ is independent of w_n and $\{w_n\}$ is independent of $\{e_n\}$. So

$$\mathbb{E}[u_n e_{n-\tau}] = \mathbb{E}[f_{n,0} w_n e_{n-\tau}] = \mathbb{E}[f_{n,0} e_{n-\tau}]\mathbb{E}[w_n] = 0 \tag{8}$$

for any τ, which completes the proof. □

Lemma 2.3: $\{y_n\}$ is also a quasi-stationary signal.

Proof. Recall that $A^*(q)$ is stable and quasi-stationary signals $\{u_n\}$ and $\{e_n\}$ are uncorrelated. By [18, Theorem 2.2], both $\{y_{u,n}\}$ and $\{y_{u,n}\}$ are quasi-stationary signals, where $y_n = y_{u,n} + y_{e,n}$, $y_{u,n} = \frac{B^*(q)}{A^*(q)}u_n = G^*(q)u_n$ with $G^*(q) = \sum_{k=0}^{\infty} g_k q^{-k}$, $y_{e,n} = \frac{C^*(q)}{A^*(q)}e_n = L^*(q)e_n$ with $L^*(q) = \sum_{j=0}^{\infty} l_j q^{-j}$. So there is a constant $C_y \geq 2(C_u + C_e)$ such that $\big|\mathbb{E}[y_n]\big| = \big|\mathbb{E}[y_{u,n} + y_{e,n}]\big| \leq \big|\mathbb{E}[y_{u,n}]\big| + \big|\mathbb{E}[y_{e,n}]\big| \leq C_y$ and $\big|\mathbb{E}[y_n y_k]\big| = \big|\mathbb{E}[(y_{u,n} + y_{e,n})(y_{u,k} + y_{e,k})]\big| \leq \big|\mathbb{E}[y_{u,n} y_{u,k}]\big| + \big|\mathbb{E}[y_{e,n} y_{e,k}]\big| + \frac{1}{2}\mathbb{E}[y_{u,n}^2 + y_{u,k}^2 + y_{e,n}^2 + y_{e,k}^2] \leq C_y$, where $\max\{\big|\mathbb{E}[y_{u,n}]\big|, \big|\mathbb{E}[y_{u,n} y_{u,k}]\big|\} \leq C_u$ and $\max\{\big|\mathbb{E}[y_{e,n}]\big|, \big|\mathbb{E}[y_{e,n} y_{e,k}]\big|\} \leq C_e$ for all n, k. Given any τ, $\overline{\mathbb{E}}[y_n y_{n-\tau}] = \overline{\mathbb{E}}[y_{u,n} y_{u,n-\tau}] + \overline{\mathbb{E}}[y_{e,n} y_{e,n-\tau}] + \overline{\mathbb{E}}[y_{u,n} y_{e,n-\tau}] + \overline{\mathbb{E}}[y_{e,n} y_{u,n-\tau}]$ exists if the last two terms $\overline{\mathbb{E}}[y_{u,n} y_{e,n-\tau}]$ and $\overline{\mathbb{E}}[y_{e,n} y_{u,n-\tau}]$ exist. Let $n_\tau = \max\{n, n-\tau\}$. Recall that $y_n = u_n = e_n$ for $n < 0$. Then, by (8), we obtain

$\left|\overline{\mathbb{E}}[y_{u,n}y_{e,n-\tau}]\right| = \left|\mathbb{E}\Big[\sum_{k=0}^{n} g_k u_{n-k}\sum_{j=0}^{n_\tau} l_j e_{n-\tau-j}\Big]\right| = \left|\sum_{k=0}^{n}\sum_{j=0}^{n_\tau} g_k l_j \mathbb{E}[f_{n-k,0}e_{n-\tau-j}]\mathbb{E}[w_{n-k}]\right| = 0$ and hence $\overline{\mathbb{E}}[y_{u,n}y_{e,n-\tau}] = 0$ since both $G^*(q)$ and $L^*(q)$ are stable. Similarly, we have $\overline{\mathbb{E}}[y_{e,n}y_{u,n-\tau}] = 0$. Note that $\overline{\mathbb{E}}[y_n y_{n-\tau}] = \overline{\mathbb{E}}[y_{u,n}y_{u,n-\tau}] + \overline{\mathbb{E}}[y_{e,n}y_{e,n-\tau}]$ and $\{y_{u,n}\}$, $\{y_{e,n}\}$ are quasi-stationary. It follows the desired result. □

We also cite the following definition and result on L-mixing processes (see [7] and [9]) that are useful for the development of our work.

Definition 2.2: A stochastic process $\{s_n\}$ is L-mixing with respect to the σ-algebras $(\mathcal{F}_n, \mathcal{F}_n^+)$ if the following conditions are satisfied: (i) s_n is $\mathcal{F}_n$ measurable, (ii) $\sup_{n\geq 0}\mathbb{E}^{1/k}\left[|s_n|^k\right] < \infty$ for all $1 \leq k < \infty$, (iii) $\sum_{\tau=0}^{\infty}\gamma_k(\tau) < \infty$ for all $1 \leq k < \infty$, where, for $\tau \geq 0$, $\gamma_k(\tau) = \sup_{n\geq\tau}\mathbb{E}^{1/k}\left[|s_n - \mathbb{E}[s_n|\mathcal{F}_n^+]|^k\right]$.

Lemma 2.4: Process $\{z_n\}$ with $z_n = s_n t_n$ is L-mixing if both $\{s_n\}$ and $\{t_n\}$ are L-mixing.

III. Multi-time-scale recursive algorithm

The proposed recursive estimation algorithm is a modification of the one in [5]. As in [5], let $\theta_{A,n}$, $\theta_{B,n}$ and X_n be the estimates for θ_A^*, θ_B^* and $X^* = [\sigma_e^{*2}\ \theta_C^{*T}]^T$ respectively. Let $\{M_n\}$ be a sequence of positive real numbers increasingly diverging to infinity. So there is a finite number n^* such that $|\theta^*| < M_{n^*}$, where $|\cdot|$ is the Euclidean norm of a vector or its induced norm of a matrix. Moreover, denote by $\mathcal{S}_+$ the class of sequences $\{\alpha(n)\}$ satisfying $\alpha(n) > 0$, $\alpha(n) \to 0$ as $n \to \infty$, and $\sum_{n=1}^{\infty}\alpha(n) = \infty$.

A. Recursive algorithm for AR-part

Let $\varphi_n^T = \begin{bmatrix} y_n & y_{n-1} & \cdots & y_{n-p_a+1}\end{bmatrix} \in \mathbb{R}^{p_a}$, $W_n = \mathbb{E}[y_n\varphi_{n-1-s}^T] = \begin{bmatrix} R_{n,s+1} & \cdots & R_{n,s+p_a}\end{bmatrix} \in \mathbb{R}^{p_a}$ and

$$\Gamma_n = \mathbb{E}[\varphi_n\varphi_{n-s}^T] = \begin{bmatrix} R_{n,s} & \cdots & R_{n,s+p_a-1} \\ R_{n,s-1} & \cdots & R_{n,s+p_a-2} \\ \vdots & \ddots & \vdots \\ R_{n,s-p_a+1} & \cdots & R_{n,s} \end{bmatrix} \in \mathbb{R}^{p_a\times p_a}$$

for all $n \geq 1$. Similar to the analysis in [5], we have $\mathbb{E}[B^*(q)u_{n-1}\varphi_{n-1-s}^T] = 0$, $\mathbb{E}[C^*(q)e_n\varphi_{n-1-s}^T] = 0$ and hence $W_n^T = -\Gamma_n^T\theta_A^*$. It is noticed that, unlike the case in [5], $\{y_n\}$ is not stationary. But, since $\{y_n\}$ is a quasi-stationary signal (see [18]), there exist R_τ, W and Γ such that $R_\tau = R_{-\tau} = \overline{\mathbb{E}}[R_{n,\tau}]$ for $\tau \geq 0$, $W = \overline{\mathbb{E}}[W_n]$, $\Gamma = \overline{\mathbb{E}}[\Gamma_n]$. This yields the Yule-Walker equation: $W^T = -\Gamma^T\theta_A^*$ or $\Gamma^T\theta_A^* + W^T = 0$. The recursive algorithm for $\theta_{A,n}$ ($n \geq 1$) is given as follows:

$$\begin{aligned}
&\widetilde{\Gamma}_n = \widetilde{\Gamma}_{n-1} - \frac{1}{n}\big(\widetilde{\Gamma}_{n-1} - \varphi_n\varphi_{n-s}^T\big)\\
&\widetilde{W}_n = \widetilde{W}_{n-1} - \frac{1}{n}\big(\widetilde{W}_{n-1} - y_n\varphi_{n-s-1}^T\big)\\
&\theta_{A,n-} = \theta_{A,n-1} - \alpha_a(n)\cdot\big(\widetilde{\Gamma}_n^T\theta_{A,n-1} + \widetilde{W}_n^T\big)\\
&\theta_{A,n} = \theta_{A,n-}\mathbb{I}_{E_{A,n}} + \theta_{A,0}\mathbb{I}_{E_{A,n}^C}\\
&E_{A,n} = \big\{|\theta_{A,n-}| \leq M_{\lambda_{A,n-1}}\big\}, \quad E_{A,n}^C = \Omega\backslash E_{A,n}\\
&\lambda_{A,n} = \lambda_{A,n-1} + \mathbb{I}_{E_{A,n}^C}, \quad \lambda_{A,0} = 0
\end{aligned} \tag{9}$$

with arbitrary initial values $\Gamma_0 \in \mathbb{R}^{p_a\times p_a}$, $W_0 \in \mathbb{R}^{p_a}$ and $\theta_{A,0} \in \mathbb{R}^{p_a}$, where $\{\alpha_a(n)\} \in \mathcal{S}_+$ and $\mathbb{I}_E$ is the indicator of set E. It is observed that $\widetilde{\Gamma}_n$ and $\widetilde{W}_n$ given above are the recursive expressions of the time averages $(1/n)\sum_{j=1}^{n}\varphi_j\varphi_{j-s}^T$ and $(1/n)\sum_{j=1}^{n} y_j\varphi_{j-s-1}^T$, respectively.

B. Recursive algorithm for X-part

Since the algorithm (16) in [5] cannot be applied to a case where the input is not an i.i.d. signal, we need to propose a new scheme for the quasi-stationary input signals. Let us explain the idea of the algorithm for the X-part as follows. Define $\chi_n = y_n + \varphi_{n-1}^T\theta_A^*$ with $\eta_{n-1} = \begin{bmatrix} u_{n-1} & u_{n-2} & \cdots & u_{n-p_b}\end{bmatrix}^T$. Clearly, $\chi_n = A^*(q)y_n = B^*(q)u_{n-1} + C^*(q)e_n$. Since the input sequence $\{u_n\}$ is generated by (2) where $\{w_n\}$ is an i.i.d. sequence of random variables independent of $\{e_n\}$, we have $\mathbb{E}[\eta_{n-1}\chi_n] = \widehat{r}_{n,0}I_{p_b}\theta_B^*$, where I_{p_b} is the identity matrix of order p_b. By Lemma 2.2, this yields the Yule-Walker equation for the X-part

$$\bar{r}_0 I_{p_b}\theta_B^* - \Xi = 0, \tag{10}$$

where $\bar{r}_0$ is given by (7) and $\Xi = \overline{\mathbb{E}}[\eta_{n-1}\chi_n]$. Therefore, estimating χ_n with $\theta_{A,n}$ by $\widetilde{\chi}_n = y_n + \varphi_{n-1}^T\theta_{A,n} = y_n + \sum_{j=1}^{p_a} a_{j,n}y_{n-j}$, we obtain the recursive estimation for θ_B^*:

$$\begin{aligned}
&\widetilde{\lambda}_n = \widetilde{\lambda}_{n-1} - \frac{1}{n}\big(\widetilde{\lambda}_{n-1} - u_n^2\big), \quad \widetilde{\lambda}_0 = 0\\
&\widetilde{\Xi}_n = \widetilde{\Xi}_{n-1} - \frac{1}{n}\big(\widetilde{\Xi}_{n-1} - \eta_{n-1}\widetilde{\chi}_n\big), \quad \Xi_0 = 0\\
&\theta_{B,n-} = \theta_{B,n-1} - \alpha_b(n)\cdot\Big(\widetilde{\lambda}_n I_{p_b}\theta_{B,n-1} - \widetilde{\Xi}_n\Big)\\
&\theta_{B,n} = \theta_{B,n-}\mathbb{I}_{E_{B,n}} + \theta_{B,0}\mathbb{I}_{E_{B,n}^C}\\
&E_{B,n} = \big\{|\theta_{B,n-}| \leq M_{\lambda_{B,n-1}}\big\}, \quad E_{B,n}^C = \Omega\backslash E_{B,n}\\
&\lambda_{B,n} = \lambda_{B,n-1} + \mathbb{I}_{E_{B,n}^C}, \quad \lambda_{B,0} = 0
\end{aligned} \tag{11}$$

with arbitrary initial value $\theta_{B,0} \in \mathbb{R}^{p_b}$, where $\{\alpha_b(n)\} \in \mathcal{S}_+$ such that $\lim_{n\to\infty}\frac{\alpha_a(n)}{\alpha_b(n)} = 0$.

C. Recursive algorithm for MA-part

In the sequel, we employ a variant of the recursive algorithm proposed in [5] for the MA-part. Define

$$\zeta_n = A^*(q)y_n - B^*(q)u_{n-1} = C^*(q)e_n, \tag{12}$$

which is a stationary ergodic process with correlation function (see [5]) $S = \begin{bmatrix} S(0) & S(1) & \cdots & S(p_c)\end{bmatrix}^T \in \mathbb{R}^{p_c+1}$, where $S(j) = \mathbb{E}[\zeta_n\zeta_{n-j}]$ for $j = 0, 1, \cdots, p_c$. The parameter to be estimated is $X^* = [\sigma_e^{*2}\ \theta_C^{*T}]^T = (\sigma_e^{*2}\ c_1^*\ \cdots\ c_{p_c}^*)^T$, which satisfies the algebraic equation $\Phi(X)X = U(X)S$, where $X = [X(0)\ X(1)\ \cdots\ X(p_c)]^T \in \mathbb{R}^{p_c+1}$, $\Phi(X) \in \mathbb{R}^{(p_c+1)\times(p_c+1)}$ and $U(X) \in \mathbb{R}^{(p_c+1)\times(p_c+1)}$ given as $\Phi(X) = \mathrm{Diag}\{1, X(0), \cdots, X(0)\}$ and

$$U(X) = \begin{bmatrix} U_0(X) & U_1(X) & \cdots & U_{p_c}(X) \\ 0 & U_0(X) & \cdots & U_{p_c-1}(X) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & U_0(X) \end{bmatrix}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2014.2373711, IEEE Transactions on Automatic Control

with entries $U_0(X) = 1$, $U_1(X) = -X(1)$, and $U_j(X) = -X(1)U_{j-1}(X) - X(2)U_{j-2}(X) - \cdots - X(j)$ for $j = 2, \cdots, p_c$. Therefore, X^* should satisfy the equation $\Phi(X)X = \overline{U(X)S}$ (see [5]), where it is not necessary to symmetrize the first element of $\overline{U(X)S} = \left[\sum_{j=0}^{p_c} U_j(X)S(j) \quad \sum_{j=1}^{p_c} U_{j-1}(X)S(j) \quad \cdots \quad S(p_c)\right]^T$.

$$\text{Let} \begin{cases} \widetilde{\zeta}_n = y_n + \varphi_{n-1}^T\theta_{A,n} - \eta_{n-1}^T\theta_{B,n} \\ S_n = \left[S_n(0) \quad S_n(1) \quad \cdots \quad S_n(p_c)\right]^T \end{cases} \tag{13}$$

with $S_n(j) = S_{n-1}(j) - \frac{1}{n}\left(S_{n-1}(j) - \widetilde{\zeta}_n\widetilde{\zeta}_{n-j}\right)$ and $S_0(j) = 0$ for $j = 0, 1, \cdots, p_c$. The recursive algorithm for the estimate X_n $(n \geq 1)$ is given as follows:

$$\begin{aligned} &X_{n-} = X_{n-1} - \alpha_x(n) \cdot \left(\Phi(X_{n-1})X_{n-1} - \overline{U(X_{n-1})S_{n-1}}\right) \\ &X_n = X_{n-}\mathbb{I}_{E_{X,n}} + X_0\mathbb{I}_{E_{X,n}^C} \qquad (14) \\ &E_{X,n} = \left\{|X_{n-}| \leq M_{\lambda_{X,n-1}}\right\}, \quad E_{X,n}^C = \Omega \backslash E_{X,n} \\ &\lambda_{X,n} = \lambda_{X,n-1} + \mathbb{I}_{E_{X,n}^C}, \quad \lambda_{X,0} = 0 \end{aligned}$$

where $X_0 = [\nu \ 0 \ \cdots \ 0]^T \in \mathbb{R}^{p_c+1}$ with $\nu \geq 1$ and $\{\alpha_x(n)\} \in \mathcal{S}_+$ with $\lim_{n\to\infty} \frac{\alpha_b(n)}{\alpha_x(n)} = 0$.

Remark 3.1: A frequently used choice for $\{\alpha_a(n)\}$, $\{\alpha_b(n)\}$ and $\{\alpha_x(n)\}$ is $\alpha_a(n) = n^{-\beta_a}$, $\alpha_b(n) = n^{-\beta_b}$ and $\alpha_x(n) = n^{-\beta_x}$ for all $n \geq 1$, where $0 < \beta_x < \beta_b < \beta_a \leq 1$.

Remark 3.2: The recursive scheme [5, (25)-(27)] is a special case of (14) when $\lim_{n\to\infty} \frac{\alpha_b(n)}{\alpha_x(n)} = 0$ is not imposed. Our variant is more general. To see this, let us consider the difference between $\widetilde{\zeta}_n$ and ζ_n denoted by $\tilde{d}(\widetilde{\zeta}_n, \zeta_n)$, which is large at the beginning and tends to small as n increases to large. Clearly, $\tilde{d}(\widetilde{\zeta}_n, \zeta_n)$ is reflected in $\tilde{d}(X_n, X^*)$. However, the step size δ_x/n, which, in some sense, is the gain of the influence of $\tilde{d}(\widetilde{\zeta}_n, \zeta_n)$ on $\tilde{d}(X_n, X^*)$, diminishes when $\tilde{d}(\widetilde{\zeta}_n, \zeta_n)$ tends to small as n increases to large. This implies that X_n could fail to converge to X^* even when $\widetilde{\zeta}_n \to \zeta_n$ $a.s.$ as $n \to \infty$. A reasonably larger gain δ_x/n^{β_x} with $\beta_x < 1$ used for those data with small $\tilde{d}(\widetilde{\zeta}_n, \zeta_n)$ could lead to an effective algorithm, see the numerical examples below.

Remark 3.3: Our recursive algorithm is a multi-time-scale variant of that in [5], where the components of the iterate are divided into three group and each of them has its own step-size sequence. The multi-time-scale method is usually used (in algorithms) to cope with dynamical systems composed of both fast and slow variables (see, e.g., [14], [13], [3]). In fact, when the algorithm is expected to operate in a slowly varying environment, it is important that the time scale of the algorithm remains reasonably faster than that of the changing environment. Otherwise, it would never adapt. For instance, [5, (26)-(27)] is an algorithm for searching the solutions of [5, Eq.(23)]. However, since ζ_n is unknown, this algorithm employs its estimate $\widetilde{\zeta}_n$ and operates in a varying environment where $\widetilde{\zeta}_n$ is recursively updated at the previous stages. Therefore, the time scale of algorithm [5, (26)-(27)] should remain reasonably faster than those of the previous stages, which lead to the changing environment $\widetilde{\zeta}_n$. Otherwise, convergence to true value could not take place, see Figs. 1-2 in Section V below. It should be also pointed out that good behavior of the coupled schemes depends on reasonable separation of the time scales.

IV. Convergence of estimators

In this section, we consider convergence of the recursive estimators presented above.

Theorem 4.1: Suppose that the assumptions in Section II hold. Then $\{\theta_{A,n}\}$ given by (9) converges to θ_A^* $a.s.$, $\{\theta_{B,n}\}$ given by (11) converges to θ_B^* $a.s.$, and $\{X_n\}$ given by (14) converges to $X^* = [\sigma_e^{*2} \ \theta_C^{*T}]^T$ $a.s.$.

Since the recursive scheme presented in Section III is modified from the one presented in [5], we just outline our proof for the convergence based on the proofs in [5] as follows. Major differences between our scheme and the one in [5] are: (i) multi time scales in class $\mathcal{S}_+$, instead of single scale $1/n$ [5] , in the coupled recursive algorithms, and (ii) quasi-stationarity, instead of stationarity [5], of the input signals. Since condition S1 in [5, Appendix] is satisfied for class $\mathcal{S}_+$, GCT and the related techniques are still applicable to our multi-time-scale scheme. Therefore, the main task of our proof is to show that we still have the convergence in the case where the stationary signals are replaced by the quasi-stationary ones.

Proof. According to the input signal designed in Section II, process $\{u_n\}$ is a quasi-stationary signal with spectrum $\Psi_u(\omega) = \sum_{\bar{\tau}=-\infty}^{\infty} \bar{r}_{\bar{\tau}} e^{-i\bar{\tau}\omega} = \bar{r}_0 \geq r_{min} > 0$, $\forall\omega$, where $\bar{r}_\tau = \bar{r}_{-\tau} = \lim_{N\to\infty}(1/N)\sum_{n=1}^N \widehat{r}_{n,\tau}$ for $\tau \geq 0$ (see, e.g., [18, 2G.2, p47]). By Lemmas 2.2 and 2.3, the spectral function of quasi-stationary process $\{y_n\}$ is $\Psi_y(\omega) = \sum_{\tau=-\infty}^{\infty} \bar{r}_y(\tau)e^{-i\tau\omega} = \left|\frac{B^*(e^{i\omega})}{A^*(e^{i\omega})}\right|^2 \Psi_u(\omega) + \sigma_e^{*2}\left|\frac{C^*(e^{i\omega})}{A^*(e^{i\omega})}\right|^2 > 0$, $\forall\omega$, where $\bar{r}_y(\tau) = \lim_{N\to\infty}(1/N)\sum_{n=1}^N \mathbb{E}[y_n y_{n-\tau}]$ for all τ. By the well-known spectral factorization (see, e.g., [18, p41]), there exists a stationary signal $\bar{y}_n$ such that

$$\Psi_{\bar{y}}(\omega) = \Psi_y(\omega), \ \forall\,\omega \ \Rightarrow \ \bar{r}_y(\tau) = r_{\bar{y}}(\tau) = \mathbb{E}[\bar{y}_n\bar{y}_{n-\tau}], \ \forall\,\tau. \tag{15}$$

Let $\{\bar{\chi}_n\}$ be $\bar{\chi}_n = A^*(q)y_n$, then $\{\bar{\chi}_n\}$ is a stationary signal with spectrum $\Psi_{\bar{\chi}}(\omega) = |A^*(e^{i\omega})|^2\Psi_{\bar{y}}(\omega) > 0$, $\forall\omega$. According to the analysis in the proof of [5, Lemma 1] and (15), the rank of Γ is p_a. But, by [18, Theorem 2.3], $(1/N)\sum_{n=1}^N y_n y_{n-\tau} \to R_\tau$ $a.s.$ as $N \to \infty$, which implies $\mathbb{P}(\Omega_0) = 1$, where $\Omega_0 = \left\{\widetilde{\Gamma}_n \to \Gamma, \widetilde{W}_n \to W \text{ as } n \to \infty\right\}$.

Following the way of Lemma 2 and Theorem 1 in [5], we can apply the GCT with Lyapunov function $v(\theta_A) = |\Gamma\theta_A - W|^2$ and show that $\{\theta_{A,n}\}$ given by (9) converges to θ_A^* $a.s.$

Let us proceed to consider the convergence of the estimates for X-part on Ω_0. Recall that we have the Yule-Walker equation (10) for the X-part. Therefore, we can recursively estimate θ_B^* with

$$\theta_{B,n-} = \theta_{B,n-1} - \alpha_b(n) \cdot (\lambda_n I_{p_b}\theta_{B,n-1} - \Xi_n) \tag{16}$$

where $\Xi_n = (1/n)\sum_{k=1}^n \eta_{n-1}\chi_n$. By Theorem 2.3 in [18], it follows that $(1/n)\sum_{j=1}^n u_j^2 \to \bar{r}_0$ $a.s.$ on Ω_0 as $n \to \infty$. It is observed that $\{\eta_{n-1}\}$ and $\{\chi_n\}$ are jointly quasi-stationary since both $\{\eta_{n-1}\}$ and $\{\chi_n\}$ are quasi-stationary and $\overline{\mathbb{E}}[\eta_{n-1-k}\chi_n] = \overline{\mathbb{E}}[\eta_{n-1-k}B^*(q)u_{n-1}] +$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2014.2373711, IEEE Transactions on Automatic Control

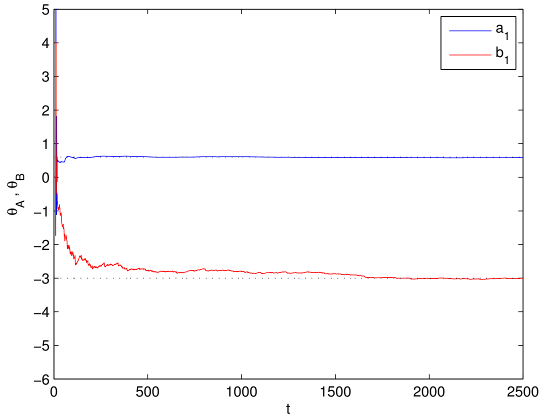

Fig. 1. Solid lines: estimates of θ_A and θ_B by the scheme in [5]. Dotted lines: true values.

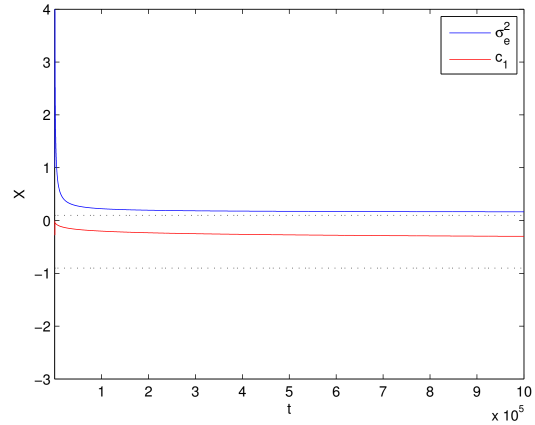

Fig. 2. Solid lines: estimate of X by the scheme in [5]. Dotted lines: true values.

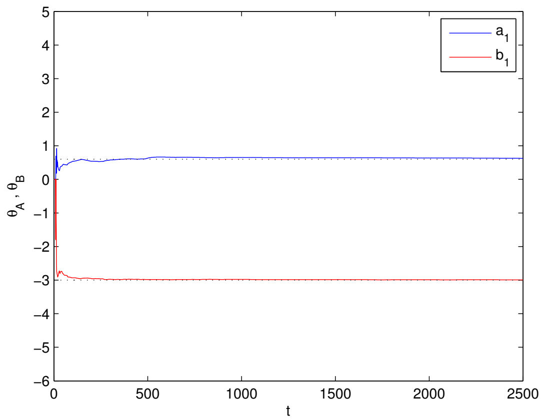

Fig. 3. Solid lines: estimates of θ_A and θ_B by the proposed scheme with i.i.d. input. Dotted lines: true values.

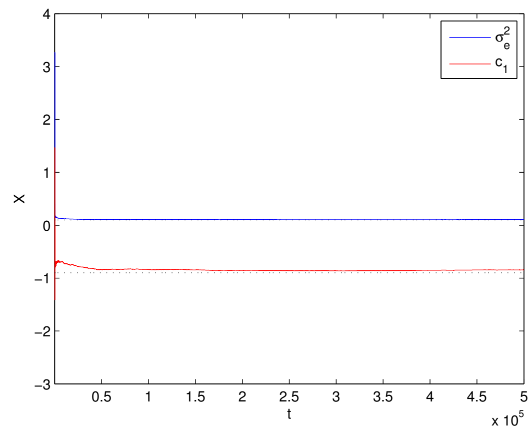

Fig. 4. Solid lines: estimate of X by the proposed scheme with i.i.d. input. Dotted lines: true values.

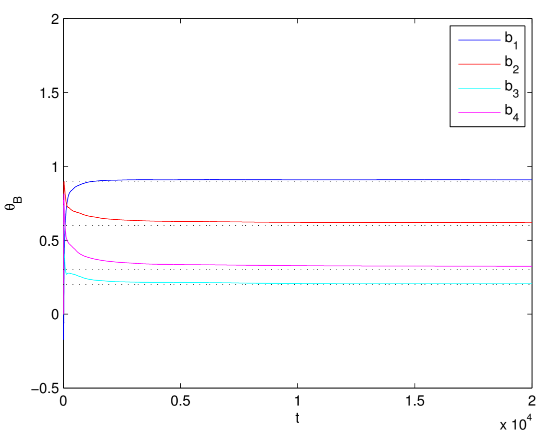

Fig. 5. Solid lines: estimate of θ_B by the proposed scheme with quasi-stationary input. Dotted lines: true values.

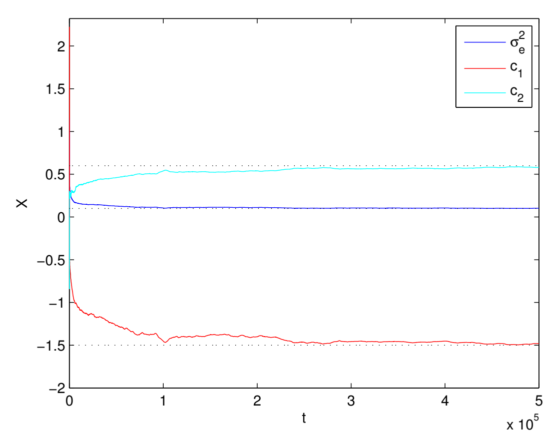

Fig. 6. Solid lines: estimate of X by the proposed scheme with quasi-stationary input. Dotted lines: true values.

$\overline{\mathbb{E}}[\eta_{n-1-k}C^*(q)e_n] = \overline{\mathbb{E}}[\eta_{n-1-k}B^*(q)u_{n-1}]$ exists for all k because, by Lemma 2.2, quasi-stationary signals $\{\eta_{n-1}\}$ and $\{e_n\}$ are uncorrelated. Moreover, it is not difficult to verify that both $\{\eta_{n-1}\}$ and $\{\chi_n\}$ are L-mixing processes with respect to $(\mathcal{F}_n, \mathcal{F}_n^+)$ (see Definition 2.2). But, by Lemma 2.4, $\{\eta_{n-1}\chi_n\}$ is also L-mixing. This implies $\{\xi_n\}$ is an L-mixing process with $\mathbb{E}[\xi_n] = 0$, where $\xi_n = \eta_{n-1}\chi_n - \mathbb{E}[\eta_{n-1}\chi_n]$. According to the strong law of large numbers [7, Corollary 3.1], it follows $\frac{1}{n}\sum_{j=1}^n \xi_j \to 0$ and hence $\frac{1}{n}\sum_{j=1}^n \eta_{j-1}\chi_j \to \Xi$ $a.s.$ on Ω_0 as $n \to \infty$. Note that (3) implies $\bar{r}_0 = \overline{\mathbb{E}}[r_{n,0}] \geq r_{min} > 0$. Similar to the proof of [5, Theorem 1], we can apply the GCT with Lyapunov function $v(\theta_B) = |\bar{r}_0 I_{p_b}\theta_B - \Xi|^2$ to show that $\{\theta_{B,n}\}$ given by (16) converges to θ_B^* $a.s.$. Note that $\{y_n\}$ and hence $\{\varphi_n\}$ are quasi-stationary signals, which, as (39) in [5], yields $\frac{1}{n}\sum_{j=1}^n |\widetilde{\chi}_j - \chi_j|^2 = \frac{1}{n}\sum_{j=1}^n |\varphi_{j-1}^T(\theta_{A,j} - \theta_A^*)|^2 \leq \frac{1}{n}\sum_{j=1}^n |\varphi_{j-1}|^2\, |\theta_{A,j} - \theta_A^*|^2 \to 0$ $a.s.$ as $n \to \infty$ since $\theta_{A,n} \to \theta_A^*$ $a.s.$ as $n \to \infty$. But, by Cauchy-Schwarz

Preprint submitted to IEEE Transactions on Automatic Control. Received: November 18, 2014 10:01:06 PST

inequality, this gives $\left|\frac{1}{n}\sum_{j=1}^{n}\eta_{j-1}\widetilde{\chi}_j - \frac{1}{n}\sum_{j=1}^{n}\eta_{j-1}\chi_j\right| \leq \frac{1}{n}\sum_{j=1}^{n}|\eta_{j-1}(\widetilde{\chi}_j - \chi_j)| \leq \frac{1}{n}\sum_{j=1}^{n}|\eta_{j-1}||\widetilde{\chi}_j - \chi_j| \leq \sqrt{\frac{1}{n}\sum_{j=1}^{n}|\eta_{j-1}|^2}\sqrt{\frac{1}{n}\sum_{j=1}^{n}|\widetilde{\chi}_j - \chi_j|^2} \to 0\ a.s.$ as $n \to \infty$ since $\{u_n\}$ and hence $\{\eta_n\}$ are quasi-stationary signals. This implies $\frac{1}{n}\sum_{j=1}^{n}\eta_{j-1}\widetilde{\chi}_j \to \Xi\ a.s.$ as $n \to \infty$. It immediately follows that $\{\theta_{B,n}\}$ given by (11) converges to $\theta_B^*\ a.s.$.

Since $\theta_{A,n} \to \theta_A^*$ and $\theta_{B,n} \to \theta_B^*\ a.s.$ as $n \to \infty$, by (12) and (13), we have $\mathbb{P}(\Omega_1) = 1$ (see [5, Corollary 1]), where $\Omega_1 = \{S_n \to S \text{ as } n \to \infty\}$. Then we always consider the sample paths on Ω_1.

Now, it can be shown in the way presented in [5, Section IV] that $\{X_n\}$ given by (14) converges to $X^*\ a.s.$, where, particularly, it can be shown as [5, Proposition] that, as $n \to \infty$, $Y(z)$ is stable $a.s.$, i.e., $Y(z) \neq 0$ for all $|z| \geq 1$, with $Y(z)$ defined by $Y(z) = 1 + \sum_{j=1}^{p_c} X(j)z^{-j}$ and $X \in G = \{X \in \mathbb{R}^{p_c+1} : \Phi(X)X = \overline{U(X)S}\}$. □

V. Numerical examples and simulations

Example 5.1: Let us consider ARMAX system (1) with i.i.d. input, where $p_a = p_b = p_c = 1$, $\theta_A^* = 0.6$, $\theta_B^* = -3$, $\theta_C^* = -0.9$, ${\sigma_e^*}^2 = 0.1$ and $X^* = (0.1\ \ -0.9)^T$. All simulations in this note employ initial values $\theta_A = 0$, $\theta_B = 0$, $\theta_C = 0$, $\sigma_e^2 = 1$ and $\{M_n\}_{n\geq 0}$ with $M_n = 1 + n$ (cf. [5]).

The typical realizations of the scheme in [5] show that the estimators for AR-part and X-part converge, see Fig. 1, while that for MA-part does not work, see Fig. 2. The reason for this phenomenon is that the time scale of (25)-(27) is not reasonably faster than that of (16) at the previous stage in [5] and therefore convergence to true value is not guaranteed for [5, (25)-(27)]. Figs. 3-4 are a typical realization of the algorithm presented in Section III with $\beta_a = 1$, $\beta_b = 1/4$ and $\beta_x = 1/7$, which verifies the effectiveness of our result.

Example 5.2: The true parameters of system (1) with orders $p_a = 0$, $p_b = 4$, $p_c = 2$ are $\theta_B^* = (0.9\ 0.6\ 0.2\ 0.3)^T$, $\theta_C^* = (-1.5\ 0.6)^T$, ${\sigma_e^*}^2 = 0.1$, $X^* = (0.1\ -1.5\ 0.6)^T$, which is derived from the FIR example in [10]. Note that $C^*(z)$ does not satisfy the SPR condition while $C^*(z) \neq 0$ for all $|z| \geq 1$.

It can be verified with typical realizations that, again, the algorithm in [5] does not work for the MA-part. Let us consider the modified and generalized algorithm presented above. To use less energy during the process, we employ quasi-stationary input generated by (2) with $r_{1,0} = f_{1,0}^2 = 1$ and $r_{n,0} = f_{n,0}^2 = \max\{r_{min}, \lambda_{n-1} - 1/(4n)\}$ for all $n \geq 1$. Let $r_{min} = \alpha = 10^{-6}$ and $r_{max} = L_\lambda = 10^{10}$. The input $\{u_n\}$ satisfies conditions (3)-(5) and therefore Theorem 4.1 applies, which is verified with the typical realizations of the scheme in Section III with $\beta_b = 1$ and $\beta_x = 1/4$, see Figs. 5-6.

VI. Conclusion

This note proposes a multi-time-scale modification of the recursive algorithm presented in [5]. The conditions on the input signal are relaxed viz-a-viz [5] and convergence to the true parameter values has been established. The advantage of using multiple-time scales has been verified with numerical examples. Our result can be extended to MIMO systems. It would be interesting to study how to find a reasonable (model-complexity-dependent) separation of the time scales for good behavior of the three-stage recursive scheme (9), (11) and (14).

Acknowledgments

The first author would like to thank Dr K.C. Sou for discussions and help with the simulations and acknowledge support from the research project SystemX.Ch-2013/196. The authors gratefully acknowledge the Associate Editor and the Reviewers' helpful comments.

References

[1] R.T. Baillie, Preditions from ARMAX models, *J. Econometrics*, 12 (1980) 365-374.
[2] A. Benveniste, M. Métivier and P. Priouret, *Adaptive algorithms and stochastic approximations*. Berlin, Germany: Spring-Verlag, 1990.
[3] V.S. Borkar, *Stochastic approximation: a dynamical system's viewpoint*. Cambridge, UK: Cambridge University Press, 2008.
[4] H.-F. Chen, , *Stochastic approximation and its applications*, New York, USA: Kluwer Academic Publishers, 2002.
[5] H.-F. Chen, New approach to recursive identification for ARMAX systems, *IEEE Trans. Automatic Control*, 55 (2010) 868-879.
[6] H.-F. Chen and L. Guo, *Stochastic approximations and its applications*. Boston, USA: Birkhäuser, 1991.
[7] L. Gerencsér, On a class of mixing processes, *Stochastics*, 26 (1989) 165-191.
[8] L. Gerencsér, On the martingale approximation of the estimation error of ARMA parameters, *Systems & Control Letters*, 15 (1990) 417-423.
[9] L. Gerencsér, Rate of convergence of recursive estimators, *SIAM J. Control and Optimization*, 30 (1992) 1200-1227.
[10] L. Gerencsér, H. Hjalmarsson and J. Mårtensson, Identification of ARX systems with non-stationary inputs – asymptotic analysis with application to adaptive input design, *Automatica*, 45 (2009) 623-633.
[11] L. Guo and D. Huang, Least-squares identification for ARMAX models without the positive real condition, *IEEE Trans. Automatic Control*, 34 (1989) 1094-1098.
[12] C.-M. Huang, C.-J. Huang and M.-L. Huang, A particle swarm optimization to identifying the ARMAX model for short-term load forecasting, *IEEE Trans. Power Systems*, 20 (2005) 1126-1133.
[13] H. Kushner, Stochastic approximation: a survey, *WIREs Computational Statistics*, 2 (2010) 87-96.
[14] H.J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications (2nd edition)*. New York, USA: Spring, 2003.
[15] T.L. Lai and C.-Z. Wei, Extended least squares and their applications to adaptive control and prediction in linear systems, *IEEE Trans. Automatic Control*, 31 (1986) 898-906.
[16] T.L. Lai and Z. Ying, Recursive solutions of estimating equations and adaptive spectral factorization, *IEEE Trans. Automatic Control*, 37 (1992) 240-243.
[17] I.D. Landau and A. Karimi, A recursive algorithm for ARMAX model identification in closed loop, *IEEE Trans. Automatic Control*, 44 (1999) 840-843.
[18] L. Ljung, *System identification: theory for the user (2nd edition)*. New Jersey, USA: Prentice Hall, 1999.
[19] L. Ljung and T. Söderström, *Theory and practice of recursive identification*. Massachusetts, USA: MIT Press, 1983.
[20] J.-B. Michaud, C.M. Pepin, R. Lecomte and R. Fontaine, ARMAX-RLS parameter-estimation crystal indentification in phoswich PET detectors, *IEEE Trans. Nuclear Science*, 57 (2010) 982-989.
[21] V. Solo, Adaptive spectral factorization, *IEEE Trans. Automatic Control*, 34 (1989) 1047-1051.
[22] J.C. Spall, Adaptive stochastic approximation by the simultaneous perturbation method, *IEEE Trans. Automatic Control*, 45 (2000) 1839-1853.
[23] W. Srikusalanukul, F. De Bruyne and P. McCullagh, Modelling of peripheral lymphocyte migration: system identification approach, *Immunology and Cell Biology*, 78 (2000) 288-293.
[24] P. Stoica, T. Mckelvey and J. Mari, MA estimation in polynomial time, *IEEE Trans. Signal Processing*, 48 (2000) 1999-2012.