Chin. Ann. Math. Ser. B
44(5), 2023, 687–702
DOI: 10.1007/s11401-023-0038-5

Chinese Annals of Mathematics, Series B

Estimation of IIR Systems with Binary-Valued Observations*

Ruifen DAI[1] Lei GUO[2]

Abstract Estimation and control problems with binary-valued observations exist widely in practical systems. However, most of the related works are devoted to finite impulse response (FIR for short) systems, and the theoretical problem of infinite impulse response (IIR for short) systems has been less explored. To study the estimation problems of IIR systems with binary-valued observations, the authors introduce a projected recursive estimation algorithm and analyse its global convergence properties, by using the stochastic Lyapunov function methods and the limit theory on double array martingales. It is shown that the estimation algorithm has similar convergence results as those for FIR systems under a weakest possible non-persistent excitation condition. Moreover, the upper bound for the accumulated regret of adaptive prediction is also established without resorting to any excitation condition.

Keywords Binary-valued observations, Infinite impulse response, Adaptive estimation, Double array martingales, Adaptive prediction
2000 MR Subject Classification 93E12, 93C10, 93E24, 62L12

1 Introduction

In many practical situations such as classification, quantization, stimulus-response problems, the outputs of systems cannot be measured accurately, and what can be known is only whether or the outputs belong to some known sets (see [15]). A typical situation is the binary-valued observation (BVO for short) systems, which exist widely in application fields including neural network (see [3]), sensor network (see [15]), gas content sensors in gas and oil industry (see [12]), traffic condition indicators in the asynchronous transmission mode network (see [10]), switching sensors for shift-by-wire in automotive applications (see [14]) and so on.

Compared to the traditional case where the true values of the system outputs can be observed or estimated directly or indirectly, the estimation problems for BVO systems are more complicated since one only has limited information available. Nevertheless, a great deal of research efforts has been devoted to the investigation of BVO systems, which can be classified

Manuscript received January 20, 2023. Revised April 1, 2023.
[1]Data Science Institute, Shandong University, Jinan 250100, China.
E-mail: rfdai@mail.sdu.edu.cn
[2]Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. E-mail: lguo@amss.ac.cn
*This work was supported by the National Natural Science Foundation of China (No. 12288201).

into two typical classes: Finite impulse response (FIR for short) systems and infinite impulse response (IIR for short) systems.

First, for FIR systems, Wang et al. [14] considered BVO systems, which has inspired much subsequent investigations. However, almost all the related works require either periodicity (see e.g., [14, 19]), or independent and identically distributed (i.i.d. for short) property (see e.g., [9, 11]) or some strong persistence of excitation (PE for short) conditions (see e.g., [4, 15]) on the inputs signals to ensure the strong consistency of parameter estimation algorithms. As pointed out in [2], though these conditions may be satisfied for some open-loop or off-line identification, they are difficult to be satisfied or verified for closed-loop system identification, since the inputs are usually generated from non-stationary and strongly correlated signals of stochastic feedback control systems. Recently, Zhang et al. [17] constructed a recursive projected Quasi-Newton estimation algorithm, and obtained the strong consistency of this algorithm under the following non-persistent excitation condition:

$$\log n = o\Big(\lambda_{\min}\Big\{\sum_{i=1}^{n}\phi_i\phi_i^{\mathrm{T}}\Big\}\Big), \quad \text{a.s.}, \tag{1.1}$$

which is equivalent to the weakest possible excitation condition for the least square (LS for short) algorithm of linear stochastic regression systems (see [8]). In [17], the convergence of adaptive predictors and the corresponding applications in adaptive control were also derived without any excitation condition. Besides, the use of other methods, such as quadratic programming-based method, can be found in [13].

Second, for IIR systems, there are extensive investigations and abundant results in the traditional case where the true outputs can be observed directly or indirectly (see, e.g., [2]). In particular, by establishing some limit theory on double array martingales, Guo et al. [6] investigated the estimation of IIR systems with regular output observations and established the strong consistency of a LS type estimation algorithm under a general non-PE condition. Besides, other methods have also been used to estimate the parameters in IIR systems, for examples, a stochastic approximation algorithm has been used for stochastic Wiener systems (the noise term is outside the saturation function) (see [18]), and a kernel-based method has been used with quantized observations in [1], however, Gaussian assumptions are needed for either input signals or impulse responses.

It can be seen from the above overview that most of the existing works on estimation of systems with BVO require quite strong input conditions to guarantee the convergence of parameter estimates, and there are only a few results on adaptive prediction with BVO. To overcome these limitations, we consider in this paper the adaptive estimation problems for IIR systems with BVO under a more general input signal condition which does not exclude applications to feedback signals, by using some basic methods and results developed in [6–7, 17].

The main contributions of this paper are as follows. We will introduce a projected recursive estimation algorithm for IIR systems with BVO, will prove the strong consistency of the algorithm under a general non-PE condition on the input signals, and will show that the asymptotic order of the accumulated regret of adaptive prediction can be bounded by $O(\log^{2a+1} n)$ without any excitation condition. These results can be regarded as extensions of the related results on FIR systems with BVO in [17], and are established by using both the stochastic Lyapunov function methods and the limit theory on double array martingales developed in [6–7].

The paper is organized as follows. Section 2 gives the problem formulation for the estimation of IIR systems with BVO; Section 3 introduces a projected recursive estimation algorithm; Section 4 presents the main results of this paper, followed by Section 5 where the proof of the main results is given; Section 6 concludes this paper with some remarks.

2 Problem Formulation

Consider the following IIR systems with BVO:

$$y_k = \sum_{i=1}^{\infty} b_i u_{k-i} + v_k, \quad k \geq 0, \tag{2.1}$$

$$s_{k+1} = I(y_{k+1} \geq c_k) = \begin{cases} 1, & y_{k+1} \geq c_k; \\ 0, & y_{k+1} < c_k, \end{cases} \tag{2.2}$$

where $y_k \in \mathbb{R}^1$, $u_k \in \mathbb{R}^1$, $v_k \in \mathbb{R}^1$ are the systems output, input and random noise, respectively, s_k is the BVO and c_k is a time-varying threshold sequence. It is assumed that $v_k = u_k = 0$ for all $k < 0$. The coefficients $b_i \in \mathbb{R}^1, i \geq 0$ are unknown, and satisfy the following summability condition:

$$\sum_{i=1}^{\infty} |b_i| < \infty. \tag{2.3}$$

Remark 2.1 Let us consider the following stochastic systems with rational transfer functions:

$$y_k = A^{-1}(z)C(z)u_{k-1} + v_k, \tag{2.4}$$

where $A(z) = 1 + a_1 z + \cdots + a_p z^p$ $(a_p \neq 0)$ is stable, i.e., $A(z) \neq 0$ for any $|z| \leq 1$, $C(z) = c_0 + c_1 z + \cdots + c_q z^q$, and $a_i, i = 1, \cdots, p$, and $c_j, j = 1, \cdots, q$, are unknown coefficients, z is the backward-shift operator. Then (2.4) can be written as (2.1) and satisfies (2.3), because there exists a $\lambda \in (0, 1)$ such that $|b_i| = O(\lambda^i), i \geq 0$.

We introduce the unknown parameter vector

$$\theta = [b_1, b_2, b_3, \cdots]^{\mathrm{T}} \in \mathbb{R}^{\infty} \tag{2.5}$$

and the corresponding regression vector

$$\phi_{k-1} = [u_{k-1}, u_{k-2}, u_{k-3}, \cdots]^{\mathrm{T}} \in \mathbb{R}^{\infty}, \tag{2.6}$$

so that (2.1) can be succinctly rewritten as

$$y_k = \theta^{\mathrm{T}} \phi_{k-1} + v_k, \quad k \geq 0. \tag{2.7}$$

To analyse adaptive estimation problems for the above system, we introduce the following notations and assumptions.

Throughout the paper, let $\|\cdot\|$ denote the Euclidean norm of vectors or matrices. The H_∞-norm is denoted by

$$\|g(z)\|_\infty = \operatorname*{ess\,sup}_{\theta \in [0,2\pi]} |g(\mathrm{e}^{\mathrm{i}\theta})|,$$

where $g(z)$ is a complex function which is analytic in $|z| < 1$ and bounded almost everywhere on $|z| = 1$. The L_2-norm of $g(z)$ is defined as

$$\|g(z)\|_2 = \Big\{ \frac{1}{2\pi} \int_0^{2\pi} |g(\mathrm{e}^{\mathrm{i}\theta})|^2 \mathrm{d}\theta \Big\}^{\frac{1}{2}}.$$

The maximum and minimum eigenvalues of a square matrix X are denoted by $\lambda_{\max}\{X\}$ and $\lambda_{\min}\{X\}$, respectively.

Assumption 2.1 *The system input $u_k \in \mathscr{F}_k$ for $k \geq 0$, and satisfies*

$$\sup_{k \geq 0} |u_k| \leq M < \infty, \quad a.s., \tag{2.8}$$

where $\{\mathscr{F}_k\}$ is a sequence of non-decreasing σ-algebras, and M is a constant.

Assumption 2.2 *The unknown parameter vector θ belongs to a known bounded convex set $D \subseteq \mathbb{R}^\infty$ and we assume that*

$$\sup_{x \in D} \|x\| \leq L < \infty, \tag{2.9}$$

where L is a constant.

Assumption 2.3 *The threshold $c_k \in \mathscr{F}_k$ for $k \geq 0$, and satisfies*

$$\sup_{k \geq 0} |c_k| \leq C < \infty, \quad a.s., \tag{2.10}$$

where C is a constant.

Assumption 2.4 *The random noise v_k is integrable and $\mathscr{F}_k$-measurable. For any $k \geq 1$, the conditional probability density function $f_k(\cdot)$ of v_k given $\mathscr{F}_{k-1}$ is known and satisfies*

$$\inf_{|x| \leq LM + C} \{f_k(x)\} > 0, \quad k = 1, 2, \cdots, \quad a.s., \tag{2.11}$$

where M, L and C are defined by (2.8)–(2.10), respectively.

3 Estimation Algorithm

Let $\{p_n\}$ be any non-decreasing sequence of positive integers satisfying $p_n \leq n$ for $n > 0$. Set

$$\theta(n) = [b_1, \cdots, b_{p_n}]^{\mathrm{T}} \tag{3.1}$$

and

$$\phi_k(n) = [u_k, \cdots, u_{k-p_n+1}]^{\mathrm{T}}, \quad k = 0, 1, \cdots, n, \tag{3.2}$$

the estimate $\widehat{\theta}_k(n)$ for $\theta(n)$ can be written as

$$\widehat{\theta}_k(n) = [\, b_1(k), \cdots, b_{p_n}(k)\,]^{\mathrm{T}}, \quad k = 0, 1, \cdots, n. \tag{3.3}$$

We will follow the ideas in the construction of the projected recursive estimation algorithm for FIR systems in [17], for which, we need to introduce a suitable projection operator.

First, we need to introduce a set $D_n \subseteq \mathbb{R}^{p_n}$ as follows:

$$D_n = \{(x_1, \cdots, x_{p_n})^{\mathrm{T}} \in \mathbb{R}^{p_n} \mid (x_1, \cdots, x_{p_n}, 0, 0, \cdots)^{\mathrm{T}} \in D\}, \tag{3.4}$$

where the set D is defined in Assumption 2.2. Obviously, D_n is also a bounded convex set and satisfies $\sup\limits_{x \in D_n} \|x\| \leq L < \infty, \forall n > 0$.

For the above defined convex set $D_n \in \mathbb{R}^{p_n}$ and for any given positive definite matrix Q, we define the corresponding projection operator as

$$\Pi_{n,Q}\{x\} = \arg\min_{\omega \in D_n} \|x - \omega\|_Q, \quad \forall x \in \mathbb{R}^{p_n}, \tag{3.5}$$

where the weighted norm $\|\cdot\|_Q$ is defined as

$$\|y\|_Q = (y^{\mathrm{T}} Q y)^{\frac{1}{2}}, \quad \forall y \in \mathbb{R}^{p_n}. \tag{3.6}$$

Now, for any given $n > 0$, the projected estimation algorithm is recursively defined for $0 \leq k \leq n$ as follows:

$$\widehat{\theta}_{k+1}(n) = \Pi_{n, P_{k+1}^{-1}(n)}\{\widehat{\theta}_k(n) + \beta_k b_k(n) P_k(n) \phi_k(n) e_{k+1}(n)\}, \tag{3.7}$$

$$P_{k+1}(n) = P_k(n) - \beta_k^2 b_k(n) P_k(n) \phi_k(n) \phi_k^{\mathrm{T}}(n) P_k(n), \tag{3.8}$$

$$e_{k+1}(n) = s_{k+1} - 1 + F_{k+1}(c_k - \phi_k^{\tau}(n) \widehat{\theta}_k(n)), \tag{3.9}$$

$$b_k(n) = \frac{1}{1 + \beta_k^2 \phi_k^{\tau}(n) P_k(n) \phi_k(n)}, \tag{3.10}$$

$$\beta_{k+1} = \min\Big\{\beta_k, \inf_{|x| \leq LM + C} f_{k+2}(x)\Big\}, \tag{3.11}$$

where $\widehat{\theta}_k(n)$ in (3.7) is the estimate for $\theta(n)$ at instant k, the initial value $\widehat{\theta}_0(n)$ can be chosen arbitrarily in D_n; F_{k+1} in (3.9) is the conditional distribution function of v_{k+1} given $\mathscr{F}_k$; β_0

can be chosen in $\big(0, \min\big\{1, \inf_{|x|\le LM+C} f_1(x)\big\}\big)$ arbitrarily; $P_0(n)=\beta I$ with real number $\beta>0$ chosen arbitrarily; $\Pi_{n,P_{k+1}^{-1}(n)}$ in (3.7) is the projection operator defined by (3.5), the well-posedness of $\Pi_{n,P_{k+1}^{-1}(n)}$ is ensured by the positive definite property of $P_{k+1}^{-1}(n)$, because by the matrix inversion formula, $P_{k+1}^{-1}(n)$ can be explicitly written as

$$P_{k+1}^{-1}(n)=P_k^{-1}(n)+\beta_k^2\phi_k(n)\phi_k^{\mathrm{T}}(n). \tag{3.12}$$

For convenience of analysis, we need to introduce the following notations to be used throughout the sequel:

$$\widetilde{\theta}_k(n)=\theta(n)-\widehat{\theta}_k(n), \tag{3.13}$$

$$\gamma_k={\beta_k}^{-1}, \tag{3.14}$$

$$w_{k+1}=s_{k+1}-1+F_{k+1}(c_k-\phi_k^{\mathrm{T}}\theta), \tag{3.15}$$

$$\psi_k(n)=F_{k+1}(c_k-\phi_k^{\mathrm{T}}(n)\widehat{\theta}_k(n))-F_{k+1}(c_k-\phi_k^{\mathrm{T}}\theta), \tag{3.16}$$

$$\delta_n=\Big(\sum_{i=p_n+1}^{\infty}|b_i|\Big)^2, \tag{3.17}$$

$$\varepsilon_k(n)=\sum_{j=p_n+1}^{\infty}b_j u_{k-j+1}. \tag{3.18}$$

We also need the following notations for polynomials in the backward-shift operator z:

$$B(z)=\sum_{i=1}^{\infty}b_iz^i,\quad B_n(z)=\sum_{i=1}^{p_n}b_iz^i,\quad \widehat{B}_n(z)=\sum_{i=1}^{p_n}b_i(n)z^i. \tag{3.19}$$

4 Main Results

In this section, we will establish the strong consistency of the algorithm introduced in Section 3 under a weak non-persistent excitation condition on the input signals, and will analyse the accuracy of the adaptive predictors without resorting to any excitation condition. The proof of the main results will be given in the next section.

Theorem 4.1 *Consider the system* (2.1)–(2.2) *under Assumptions* 2.1–2.4. *Then as* $n\to\infty$, *the upper bound of parameter estimation error produced by the algorithm* (3.7)–(3.11) *is as follows*:

$$\|\widehat{B}_n(z)-B(z)\|_\infty^2=O\Big(\frac{p_n}{\lambda_{\min}\{P_n^{-1}(n)\}}\{p_n\log n+\delta_n^{\frac{1}{2}}n\}\Big),\quad a.s. \tag{4.1}$$

Remark 4.1 The related result for FIR systems with BVO in [17] can be included if we take p_n as an upper bound for the order of the FIR system, since $\delta_n=0$ and $\|\widehat{B}_n(z)-B(z)\|_2$ is bounded by $\|\widehat{B}_n(z)-B(z)\|_\infty$.

Corollary 4.1 *Let the conditions of Theorem* 4.1 *hold,* $\{f_k(x)\}$ *given in Assumption* 2.4 *satisfy*

$$\inf_{|x|\le LM+C, k\ge 1}\{f_k(x)\} > 0, \quad a.s., \tag{4.2}$$

and let $|b_i| = O(\lambda^i)$ *hold for all* $i \ge 0$ *and for some* $0 < \lambda < 1$. *If we take* $p_n = [\log^a n], a > 1$, *where* $[x]$ *is the integer part of* x. *Then as* $n \to \infty$,

$$\|\widehat{B}_n(z) - B(z)\|_\infty^2 = O\Big(\frac{p_n^2 \log n}{\lambda_{\min}\Big\{\beta I + \sum\limits_{i=0}^{n-1} \phi_i(n)\phi_i^{\mathrm{T}}(n)\Big\}}\Big), \quad a.s. \tag{4.3}$$

Remark 4.2 From (4.3) we know that the estimation error will converge to zero if we have

$$p_n^2 \log n = o\Big(\lambda_{\min}\Big\{\sum_{i=0}^{n-1} \phi_i(n)\phi_i^{\mathrm{T}}(n)\Big\}\Big), \quad a.s. \tag{4.4}$$

This condition is much weaker than the usual PE condition, and is a natural generalization of the non-PE condition (1.1) for FIR systems with BVO in [17].

Next, we analyse the accuracy of the adaptive prediction. Note that for any $n > 0$ and $0 \le k \le n$, the system output y_k in (2.1) can be rewritten as follows:

$$y_{k+1} = \phi_k^{\mathrm{T}}(n)\theta(n) + \varepsilon_k(n) + v_{k+1}, \quad k = 0, \cdots, n, \tag{4.5}$$

which can be regarded as a linear stochastic regression model of order p_n plus a residual term $\varepsilon_k(n)$ defined in (3.18). Taking conditional expectations on both sides of (4.5), we can obtain the best prediction for y_{k+1} in the mean square sense:

$$E(y_{k+1}|\mathscr{F}_k) = \phi_k^{\mathrm{T}}(n)\theta(n) + \varepsilon_k(n) + E(v_{k+1}|\mathscr{F}_k), \quad k = 0, \cdots, n. \tag{4.6}$$

Replacing the unknown $\theta(n)$ in (4.6) by its estimate $\widehat{\theta}_k(n)$ and omitting the residual term $\varepsilon_k(n)$, we can define the following adaptive predictor for the output:

$$\widehat{y}_{k+1}(n) = \phi_k^{\mathrm{T}}(n)\widehat{\theta}_k(n) + E(v_{k+1}|\mathscr{F}_k), \quad k = 0, \cdots, n. \tag{4.7}$$

Usually, the difference between the best prediction and the adaptive prediction is referred to as regret. From (4.6) and (4.7), the regret can be defined as follows:

$$\begin{aligned} R_k(n) &= [E(y_{k+1}|\mathscr{F}_k) - \widehat{y}_{k+1}(n)]^2 \\ &= \{\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n) + \varepsilon_k(n)\}^2, \quad k = 0, \cdots, n. \end{aligned} \tag{4.8}$$

Naturally, we hope that the regret is small in a certain sense. Fortunately, this can be realized by analysing the asymptotic upper bound for the accumulated regret as will be shown in the following theorem.

Theorem 4.2 *Consider the system* (2.1)–(2.2) *under Assumptions* 2.1–2.4. *Then the accumulated regret has the following upper bound as* $n \to \infty$:

$$\sum_{k=1}^{n} R_k(n) = O(\gamma_n^2 \{p_n \log n + \delta_n^{\frac{1}{2}} n\}), \quad a.s. \tag{4.9}$$

According to Theorem 4.2, we can immediately deduce the following corollary.

Corollary 4.2 *Let the conditions of Theorem* 4.2 *hold and* $\{f_k(x)\}$ *given in Assumption* 2.4 *satisfy* (4.2). *Then the following result holds as* $n \to \infty$:

$$\sum_{k=1}^{n} R_k(n) = O(p_n \log n + \delta_n^{\frac{1}{2}} n), \quad a.s. \tag{4.10}$$

Remark 4.3 Consider the case in Remark 2.1. If we take $p_n = [\log^a n], a > 1$, then it follows from Corollary 4.2 that

$$\sum_{k=1}^{n} R_k(n) = O(p_n \log n), \quad \text{a.s.} \tag{4.11}$$

The result (4.11) coincides with the related result for ARMAX models with regular outputs in [7], where it has shown that the asymptotic order $O(p_n \log n)$ of the accumulated regret is the best possible.

To make the above results on the accumulated regret conveniently applied to adaptive control, we need to get an upper bound for the following accumulated "synchronized regret" $\sum\limits_{k=1}^{n} R_k(k)$.

Theorem 4.3 *Under the conditions of Corollary* 4.2, *let* $|b_i| = O(\lambda^i)$ *hold for all* $i \geq 0$ *and for some* $0 < \lambda < 1$. *If we take* $p_n = [\log^a n], a > 1$, *then the accumulated "synchronized regret" of adaptive prediction has the following upper bound as* $n \to \infty$:

$$\sum_{k=0}^{n-1} R_k(k) = O(p_n^2 \log n), \quad a.s. \tag{4.12}$$

5 Proof of the Main Results

We need the following lemmas to prove the main results in Section 4.

Lemma 5.1 (see [7]) *Assume that* $\{w_n, \mathscr{F}_n\}$ *is a martingale difference sequence satisfying*

$$\sup_j E[\|w_{j+1}\|^2 \mid \mathscr{F}_j] < \infty \quad and \quad \|w_n\| = o(\varphi(n)), \quad a.s., \tag{5.1}$$

where $\varphi(\cdot)$ *is a deterministic, positive, nondecreasing function and satisfies*

$$\sup_k \varphi(\mathrm{e}^{k+1})/\varphi(\mathrm{e}^k) < \infty. \tag{5.2}$$

Let $\{f_j(k)\}, j,k = 1,2,\cdots$ *be a* $\mathscr{F}_j$*-measurable random sequence. Then for* $p_n = O([\log^a n]), a > 1$*, the following property holds as* $n \to \infty$:

$$\max_{1\le k\le p_n}\max_{1\le i\le n}\Big\|\sum_{j=1}^{i} f_j(k)w_{j+1}\Big\| = O(a_n\log a_n) + o(a_n\varphi(n)\log\log n), \quad a.s., \tag{5.3}$$

where

$$a_i = \max_{1\le k\le p_n} g_i(k), \quad g_i(k) = \Big[\sum_{j=1}^{i}\|f_j(k)\|^2 + 1\Big]^{\frac{1}{2}}, \quad g_0(k) = 1. \tag{5.4}$$

Lemma 5.2 (see [6]) *Let* $\{f_k(n)\}, 1 \le k \le n$ *be a* $\mathscr{F}_k$*-measurable random vector sequence in* $\mathbb{R}^{p_n}, p_n \ge 1$*, and* $M_k(n) = \gamma I + \sum_{j=1}^{k} f_j(n)f_j^{\mathrm{T}}(n), \gamma > 0, 1 \le k \le n$*. Then as* $n \to \infty$*,*

$$\sum_{k=1}^{n} f_k^{\mathrm{T}}(n)M_k^{-1}(n)f_k(n) = O(\log^+\{\det(M_n(n))\} + 1), \tag{5.5}$$

where det $\big(M_n(n)\big)$ *denotes the determinant of* $M_n(n)$*, and* $\log^+\{\cdot\}$ *denotes the positive part of* $\log\{\cdot\}$.

Lemma 5.3 (see [7]) *Assume that* $w_n, \varphi(\cdot), f_k(n), M_k(n)$ *and* p_n *are the same as those in Lemmas* 5.1–5.2*. Then as* $n \to \infty$*,*

$$\begin{aligned}&\sum_{i=1}^{n} f_k^{\mathrm{T}}(n)M_k^{-1}(n)f_k(n)\|w_{k+1}\|^2\\&= o(\varphi^2(n)\log\log n) + O(p_n\log^+\lambda_{\max}\{M_n(n)\}), \quad a.s.\end{aligned} \tag{5.6}$$

Next, we present a fundamental lemma, which can be regarded as an extension of those in Guo et al. [6], Guo [5] and Zhang et al. [17].

Lemma 5.4 *Let Assumptions* 2.1–2.4 *be satisfied for the system* (2.1)–(2.2)*. Then the estimation error produced by the algorithm* (3.7)–(3.11) *satisfies the following property*:

$$\widetilde{\theta}_n^{\mathrm{T}}(n)P_n^{-1}(n)\widetilde{\theta}_n(n) + \Big(1 - \frac{2}{b}\Big)\sum_{k=0}^{n-1}\beta_k^2[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2 = O(p_n\log n + \delta_n^{\frac{1}{2}}n), \quad a.s., \tag{5.7}$$

where $b > 2$ *is a constant.*

Proof For any given $n > 0$, we consider the stochastic Lyapunov function $V_k(n) = \widetilde{\theta}_k^{\mathrm{T}}(n)P_k^{-1}(n)\widetilde{\theta}_k(n)$. By the properties of the projection operator, and by (3.5)–(3.8), (3.12) and (3.15)–(3.16), we have

$$\begin{aligned}V_{k+1}(n) &= \|\widetilde{\theta}_{k+1}(n)\|^2_{P_{k+1}^{-1}(n)}\\&= \|\theta(n) - \Pi_{n,P_{k+1}^{-1}(n)}\{\widehat{\theta}_k(n) + b_k(n)\beta_kP_k(n)\phi_k(n)e_{k+1}(n)\}\|^2_{P_{k+1}^{-1}(n)}\\&\le \|\theta(n) - \widehat{\theta}_k(n) - b_k(n)\beta_kP_k(n)\phi_k(n)e_{k+1}(n)\|^2_{P_{k+1}^{-1}(n)}\end{aligned}$$

$$
\begin{aligned}
&= \{\widetilde{\theta}_k(n) - b_k(n)\beta_k P_k(n)\phi_k(n)[\psi_k(n) + w_{k+1}]\}^{\mathrm{T}} P_{k+1}^{-1}(n) \\
&\quad \{\widetilde{\theta}_k(n) - b_k(n)\beta_k P_k(n)\phi_k(n)[\psi_k(n) + w_{k+1}]\} \\
&= \widetilde{\theta}_k^{\mathrm{T}}(n) P_{k+1}^{-1}(n)\widetilde{\theta}_k(n) - 2b_k(n)\beta_k\widetilde{\theta}_k^{\mathrm{T}}(n)P_{k+1}^{-1}(n)P_k(n)\phi_k(n)\psi_k(n) \\
&\quad - 2b_k(n)\beta_k\widetilde{\theta}_k^{\mathrm{T}}(n)P_{k+1}^{-1}(n)P_k(n)\phi_k(n)w_{k+1} \\
&\quad + b_k^2(n)\beta_k^2\psi_k^2(n)\phi_k^{\mathrm{T}}(n)P_k(n)P_{k+1}^{-1}(n)P_k(n)\phi_k(n) \\
&\quad + b_k^2(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)P_{k+1}^{-1}(n)P_k(n)\phi_k(n)w_{k+1}^2 \\
&\quad + 2b_k^2(n)\beta_k^2\psi_k(n)\phi_k^{\mathrm{T}}(n)P_k(n)P_{k+1}^{-1}(n)P_k(n)\phi_k(n)w_{k+1}. \qquad (5.8)
\end{aligned}
$$

We first simplify the terms in (5.8). By (3.10) and (3.12), we have

$$\widetilde{\theta}_k^{\mathrm{T}}(n)P_{k+1}^{-1}(n)\widetilde{\theta}_k(n) = \widetilde{\theta}_k^{\mathrm{T}}(n)P_k^{-1}(n)\widetilde{\theta}_k(n) + \beta_k^2[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2, \qquad (5.9)$$

$$b_k(n)P_{k+1}^{-1}(n)P_k(n)\phi_k(n) = \phi_k(n). \qquad (5.10)$$

Substituting the above into (5.8), we know that

$$
\begin{aligned}
V_{k+1}(n) \le V_k(n) &+ \beta_k^2[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2 \\
&- 2\beta_k\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)\psi_k(n) \\
&- 2\beta_k\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)w_{k+1} \\
&+ b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)\psi_k^2(n) \\
&+ b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)w_{k+1}^2 \\
&+ 2b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)\psi_k(n)w_{k+1}. \qquad (5.11)
\end{aligned}
$$

Moreover, by the definition of $\psi_k(n)$, we have

$$|\psi_k(n)|^2 \le 1, \qquad (5.12)$$

which implies that

$$b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)\psi_k^2(n) \le b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n). \qquad (5.13)$$

By the differential mean value theorem, from (3.11) and (3.16), we see that

$$|\psi_k(n)| \ge \beta_k|\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n) + \varepsilon_k(n)|. \qquad (5.14)$$

By (3.16) and the differential mean value theorem again, we know that $\psi_k(n)$ and $[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)+\varepsilon_k(n)]$ have the same sign, and from (5.14) we get

$$-2\beta_k[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n) + \varepsilon_k(n)]\psi_k(n) \le -2\beta_k^2[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n) + \varepsilon_k(n)]^2. \qquad (5.15)$$

Substituting this and (5.13) into (5.11), and summing both sides of (5.11) from $k = 0$ to $n - 1$, we obtain

$$V_n(n) \le V_0(n) + \sum_{k=0}^{n-1}\beta_k^2[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2 - 2\sum_{k=0}^{n-1}\beta_k^2[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n) + \varepsilon_k(n)]^2$$

$$
\begin{aligned}
&+ 2\sum_{k=0}^{n-1}\beta_k\psi_k(n)\varepsilon_k(n) + \sum_{k=0}^{n-1} b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)\\
&- 2\sum_{k=0}^{n-1}\beta_k\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)w_{k+1}\\
&+ \sum_{k=0}^{n-1} b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)w_{k+1}^2\\
&+ 2\sum_{k=0}^{n-1} b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)\psi_k(n)w_{k+1}.
\end{aligned} \tag{5.16}
$$

We now estimate the right-hand side of (5.16) term by term. By (3.17)–(3.18) we have

$$
\sum_{k=0}^{n}\varepsilon_k^2(n) \le \sum_{k=0}^{n}\Big(\sum_{j=p_n+1}^{\infty}|b_j||u_{k-j+1}|\Big)^2 = O(\delta_n n). \tag{5.17}
$$

Moreover, by this, the boundedness of both β_k and $\psi_k(n)$, and the Schwarz inequality, we have

$$
\sum_{k=0}^{n-1}\beta_k\psi_k(n)\varepsilon_k(n) = O\Big(\sum_{k=0}^{n-1}\varepsilon_k(n)\Big) = O\Big(\Big[n\sum_{k=0}^{n-1}\varepsilon_k^2(n)\Big]^{\frac{1}{2}}\Big) = O(\delta_n^{\frac{1}{2}} n). \tag{5.18}
$$

By the elementary inequality $2xy \le \frac{x^2}{b} + by^2$, we know that for any given $b > 2$ we have

$$
-4\sum_{k=0}^{n-1}\beta_k^2\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)\varepsilon_k(n) \le \frac{2}{b}\sum_{k=0}^{n-1}\beta_k^2[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2 + 2b\sum_{k=0}^{n-1}\beta_k^2\varepsilon_k^2(n). \tag{5.19}
$$

Note that $\beta_k^2 \le 1$, from (5.17) and (5.19) we obtain

$$
\begin{aligned}
&- 2\sum_{k=0}^{n-1}\beta_k^2[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n) + \varepsilon_k(n)]^2\\
&\le \Big(\frac{2}{b} - 2\Big)\sum_{k=0}^{n-1}\beta_k^2[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2 + (2b-2)\sum_{k=0}^{n-1}\beta_k^2\varepsilon_k^2(n)\\
&= \Big(\frac{2}{b} - 2\Big)\sum_{k=0}^{n-1}\beta_k^2[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2 + O(\delta_n n).
\end{aligned} \tag{5.20}
$$

We proceed to estimate the fifth term in (5.16). Let $f_k(n) = \beta_k\phi_k(n)$ in Lemma 5.2, since $b_k(n) < 1$, we obtain

$$
\begin{aligned}
&\sum_{k=0}^{n-1} b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)\\
&\le \sum_{k=0}^{n-1} f_k^{\mathrm{T}}(n)P_k(n)f_k(n)\\
&= O(\log^+\{\det(P_n^{-1}(n))\})\\
&= O(p_n\log^+\lambda_{\max}\{P_n^{-1}(n)\}).
\end{aligned} \tag{5.21}
$$

Moreover, by (5.21) and the fact that $b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n) < 1$, we have

$$\sum_{k=0}^{n-1}[b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)]^2 = O(p_n \log^+ \lambda_{\max}\{P_n^{-1}(n)\}). \tag{5.22}$$

Next, we estimate the sixth term in (5.16). By the definition of w_{k+1} we obtain

$$\begin{aligned} &E(w_{k+1}|\mathscr{F}_k)\\ &= E[s_{k+1}|\mathscr{F}_k] - E[P(y_{k+1} \geq c_k|\mathscr{F}_k)|\mathscr{F}_k]\\ &= E[I(y_{k+1} \geq c_k)|\mathscr{F}_k] - E[E(I(y_{k+1} \geq c_k)|\mathscr{F}_k)|\mathscr{F}_k]\\ &= 0, \end{aligned} \tag{5.23}$$

which implies that $\{w_k, \mathscr{F}_k\}$ is a martingale difference sequence. Since $|w_{k+1}| \leq 1$, we know that

$$|w_n| = o([\log\log(n+\mathrm{e})]^{\frac{1}{2}}) \quad \text{and} \quad \sup_j E[|w_{j+1}|^2 \mid \mathscr{F}_j] < \infty, \quad \text{a.s.} \tag{5.24}$$

Consequently, by (5.24) and Lemma 5.1, we get

$$\begin{aligned} &\sum_{k=0}^{n-1}\beta_k\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)w_{k+1}\\ &= O\Big(\Big\{\sum_{k=0}^{n-1}[\beta_k\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2\Big\}^{\frac{1}{2}}\Big) + o\Big(\Big\{\sum_{k=0}^{n-1}[\beta_k\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2\Big\}^{\frac{1}{2}}[\log\log n]^{\frac{3}{2}}\Big)\\ &= o\Big(\sum_{k=0}^{n-1}[\beta_k\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2\Big) + o([\log\log n]^3), \quad \text{a.s.} \end{aligned} \tag{5.25}$$

We now estimate the seventh term in (5.16). Note that $b_k(n) < 1$, by (5.24) and Lemma 5.3, we see that

$$\begin{aligned} &\sum_{k=0}^{n-1} b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)w_{k+1}^2\\ &= O(p_n \log^+ \lambda_{\max}\{P_n^{-1}(n)\}) + o([\log\log n]^2), \quad \text{a.s.} \end{aligned} \tag{5.26}$$

As for the last term in (5.16), by (5.12), (5.22) and Lemma 5.1 again, we know that

$$\begin{aligned} &\sum_{k=0}^{n-1} b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)\psi_k(n)w_{k+1}\\ &= o\Big(\sum_{k=0}^{n-1}[b_k(n)\beta_k^2\phi_k^{\mathrm{T}}(n)P_k(n)\phi_k(n)]^2\Big) + o([\log\log n]^3)\\ &= o(p_n \log^+ \lambda_{\max}\{P_n^{-1}(n)\}) + o([\log\log n]^3), \quad \text{a.s.} \end{aligned} \tag{5.27}$$

Note that by (3.2) and (3.12), we have

$$\log^+ \lambda_{\max}\{P_n^{-1}(n)\}$$

$$
\begin{aligned}
&\leq \log^+ \operatorname{tr}\Big[P_0^{-1}(n) + \sum_{i=0}^{n-1} \beta_i^2 \phi_i(n)\phi_i^{\mathrm{T}}(n)\Big] \\
&= \log^+\Big[\operatorname{tr}(P_0^{-1}(n)) + \sum_{i=0}^{n-1}\beta_i^2 \sum_{j=0}^{p_n-1} u_{i-j}^2\Big] \\
&= O(\log^+(p_n + p_n n)) \\
&= O(\log n), \quad \text{a.s.}
\end{aligned} \tag{5.28}
$$

Finally, substituting (5.18), (5.20)–(5.21) and (5.25)–(5.28) into (5.16), we have

$$
\begin{aligned}
&\widetilde{\theta}_n^{\mathrm{T}}(n)P_n^{-1}(n)\widetilde{\theta}_n(n) + \Big(1 - \frac{2}{b}\Big)\sum_{k=0}^{n-1}\beta_k^2[\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2 \\
&= O(p_n \log n + (\log\log n)^2 + (\log\log n)^3 + \delta_n^{\frac{1}{2}} n) \\
&= O(p_n \log n + \delta_n^{\frac{1}{2}} n), \quad \text{a.s.}
\end{aligned} \tag{5.29}
$$

Hence Lemma 5.4 is true.

Proof of Theorem 4.1 Noticing that

$$
\lambda_{\min}\{P_n^{-1}(n)\}\|\widetilde{\theta}_n(n)\|^2 \leq \widetilde{\theta}_n^{\mathrm{T}}(n)P_n^{-1}(n)\widetilde{\theta}_n(n), \tag{5.30}
$$

it follows from this and Lemma 5.4 that

$$
\|\theta(n) - \widehat{\theta}_n(n)\|^2 = O\Big(\frac{1}{\lambda_{\min}\{P_n^{-1}(n)\}}\{p_n \log n + \delta_n^{\frac{1}{2}} n\}\Big), \quad \text{a.s.} \tag{5.31}
$$

Combining (3.17), (3.19), (3.1) and (3.3) with (5.31), we have

$$
\begin{aligned}
&\|\widehat{B}_n(z) - B(z)\|_\infty^2 \\
&= \|\widehat{B}_n(z) - B_n(z) + B_n(z) - B(z)\|_\infty^2 \\
&\leq 2\|\widehat{B}_n(z) - B_n(z)\|_\infty^2 + 2\|B_n(z) - B(z)\|_\infty^2 \\
&\leq 2\Big[\sum_{i=1}^{p_n}|b_i(n) - b_i|\Big]^2 + 2\Big[\sum_{i=p_n+1}^{\infty}|b_i|\Big]^2 \\
&\leq 2p_n \sum_{i=1}^{p_n}[b_i(n) - b_i]^2 + 2\delta_n \\
&\leq 2p_n\|\theta(n) - \widehat{\theta}_n(n)\|^2 + 2\delta_n \\
&= O\Big(\frac{p_n}{\lambda_{\min}\{P_n^{-1}(n)\}}\{p_n \log n + \delta_n^{\frac{1}{2}} n\}\Big), \quad \text{a.s.,}
\end{aligned} \tag{5.32}
$$

where the last equality holds because $\delta_n = O(1)$ and $\lambda_{\min}\{P_n^{-1}(n)\} = O(n)$.

Proof of Corollary 4.1 By the fact that δ_k is non-increasing, and by the choice of p_n and property of $|b_i|$, from (3.17) we have

$$
\delta_n^{\frac{1}{2}} n \leq \sum_{k=1}^{n}\delta_k^{\frac{1}{2}} = O\Big(\sum_{k=1}^{n}\lambda^{p_k}\Big) = O\Big(\sum_{k=1}^{n} k^{(\log\lambda)(\log^{a-1} k)}\Big) < \infty. \tag{5.33}
$$

Noticing that (4.2) implies $\beta_k \nrightarrow 0$, by Theorem 4.1 and (5.33), we see that the assertion of Corollary 4.1 holds.

Proof of Theorem 4.2 From Lemma 5.4 we know that

$$\sum_{k=0}^{n-1} \beta_k^2 [\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2 = O(p_n \log n + \delta_n^{\frac{1}{2}} n), \quad \text{a.s.} \tag{5.34}$$

Noting that $[\widetilde{\theta}_n^{\mathrm{T}}(n)\phi_n(n)]^2 = O(p_n)$, by the non-increasing property of β_k, and $\beta_n = \gamma_n^{-1}$, we have

$$\sum_{k=1}^{n} [\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2 = O(\gamma_n^2 \{p_n \log n + \delta_n^{\frac{1}{2}} n\}), \quad \text{a.s.} \tag{5.35}$$

Combining this with (5.17), it is not difficult to see that

$$\sum_{k=1}^{n} R_k(n) = O(\gamma_n^2 \{p_n \log n + \delta_n^{\frac{1}{2}} n\}), \quad \text{a.s.} \tag{5.36}$$

Proof of Theorem 4.3 Since $\beta_k \nrightarrow 0$, it is not difficult to see from (5.33) and (5.35) that

$$\sum_{k=1}^{n} [\widetilde{\theta}_k^{\mathrm{T}}(n)\phi_k(n)]^2 = O(\log^{a+1} n), \quad \text{a.s.} \tag{5.37}$$

Note that by the definition of p_k, $p_k = [\log^a k], a > 1$. Let us denote the "inverse" function of p_k as $q_k = [\mathrm{e}^{k^{\frac{1}{a}}}]$. Now for any integer $k \geq 1$, there exists an integer $i \geq 0$ such that

$$q_i \leq k < q_{i+1}. \tag{5.38}$$

It is not difficult to show that $p_k = i$ for all suitably large integer k or i, hence by the definition of $\widetilde{\theta}_k(n)$ and $\phi_k(n)$, we know that

$$\widetilde{\theta}_k^{\mathrm{T}}(k)\phi_k(k) = \widetilde{\theta}_k^{\mathrm{T}}(q_{i+1} - 1)\phi_k(q_{i+1} - 1). \tag{5.39}$$

Now, for any $n > 1$, let m be the positive integer such that $q_m \leq n < q_{m+1}$, by (5.37) and (5.39), we obtain

$$\begin{aligned}
&\sum_{k=0}^{n-1} [\widetilde{\theta}_k^{\mathrm{T}}(k)\phi_k(k)]^2 \\
&= \sum_{i=0}^{m} \sum_{k=q_i}^{q_{i+1}-1} \{\widetilde{\theta}_k^{\mathrm{T}}(q_{i+1} - 1)\phi_k(q_{i+1} - 1)\}^2 + O(1) \\
&= O\Big(\sum_{i=0}^{m-1} \log^{a+1}(q_{i+1})\Big) \\
&= O(m \log^{a+1} q_m) \\
&= O(p_n^2 \log n), \quad \text{a.s.}
\end{aligned} \tag{5.40}$$

Moreover, by (3.17)–(3.18), the choice of p_n and property of $|b_i|$, we have

$$\begin{aligned}
&\sum_{k=0}^{n-1}\varepsilon_k^2(k)\\
&\leq \sum_{k=0}^{n-1}\Big(\sum_{j=p_k+1}^{\infty}|b_j||u_{k-j+1}|\Big)^2\\
&= O\Big(\sum_{k=0}^{n-1}\delta_k\Big)\\
&= O\Big(\sum_{k=0}^{n-1}\lambda^{2p_k}\Big)\\
&= O\Big(\sum_{k=0}^{n-1}k^{2(\log\lambda)(\log^{a-1}k)}\Big)\\
&= O(1).
\end{aligned}\tag{5.41}$$

Therefore, by (5.40)–(5.41) we have the desired result

$$\begin{aligned}
&\sum_{k=0}^{n-1}R_k(k)\\
&\leq 2\Big\{\sum_{k=0}^{n-1}[\widetilde{\theta}_k^{\mathrm{T}}(k)\phi_k(k)]^2+\sum_{k=0}^{n-1}\varepsilon_k^2(k)\Big\}\\
&= O(p_n^2\log n),\quad \text{a.s.}
\end{aligned}\tag{5.42}$$

6 Conclusions

This paper has considered the estimation problems for IIR systems with BVO, by using a projected recursive estimation algorithm. The convergence and the convergence rate of the estimation algorithm have been established under a general and weakest possible excitation condition. Moreover, the accumulated regret of adaptive prediction has been shown to be bounded by $O(\log^{2a+1}n)$, which implies that the averaged regret converges to 0. For further investigation, it will be interesting to consider adaptive control problems, and to extend the related results to IIR systems with more general observations, such as saturated observations recently considered in [16].

Declarations

Conflicts of interest Lei GUO is an editorial board member for Chinese Annals of Mathematics Series B and was not involved in the editorial review or the decision to publish this article. All authors declare that there are no conflicts of interest.

References

[1] Bottegal, G., Hjalmarsson, H. and Pillonetto, G., A new kernel-based approach to system identification with quantized output data, *Automatica*, **85**, 2017, 145–152.

[2] Chen, H. F. and Guo, L., Identification and Stochastic Adaptive Control, Birkhauser, Boston, 1991.

[3] Ghysen, A., The origin and evolution of the nervous system, *Int. J. Dev. Biol.*, **47**(7–8), 2003, 555–562.

[4] Guo, J. and Zhao, Y., Recursive projection algorithm on FIR system identification with binary-valued observations, *Automatica*, **49**(11), 2013, 3396–3401.

[5] Guo, L., Convergence and logarithm laws of self-tuning regulators, *Automatica*, **31**(3), 1995, 435–450.

[6] Guo, L., Huang, D. W. and Hannan, E. J., On ARX(∞) approximation, *J. Multivar. Anal.*, **32**(1), 1990, 17–47.

[7] Huang, D. and Guo, L., Estimation of nonstationary ARMAX models based on the Hannan-Rissanen method, *Ann. Stat.*, **18**(4), 1990, 1729–1756.

[8] Lai, T. L. and Wei, C. Z., Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems, *Ann. Stat.*, **10**(1), 1982, 154–166.

[9] Marelli, D., You, K. and Fu, M., Identification of ARMA models using intermittent and quantized output observations, *Automatica*, **49**(2), 2013, 360–369.

[10] Schwiebert, L. and Wang, L. Y., Robust control and rate coordination for efficiency and fairness in ABR traffic with explicit rate marking, *Comput. Commun.*, **24**(13), 2001, 1329–1340.

[11] Song, Q., Recursive identification of systems with binary-valued outputs and with ARMA noises, *Automatica*, **93**, 2018, 106–113.

[12] Sun, J., Yong, W. K. and Wang, L. Y., Aftertreatment control and adaptation for automotive lean burn engines with HEGO sensors, *Int. J. Adapt. Control Signal Process.*, **18**(2), 2004, 145–166.

[13] Wang, J. and Zhang, Q., Identification of FIR systems based on quantized output measurements: A quadratic programming-based method, *IEEE Trans. Autom. Control*, **60**(5), 2014, 1439–1444.

[14] Wang, L. Y., Zhang, J. F. and Yin, G. G., System identification using binary sensors, *IEEE Trans. Autom. Control*, **48**(11), 2003, 1892–1907.

[15] Zhang, H., Wang, T. and Zhao, Y., Asymptotically efficient recursive identification of FIR systems with binary-valued observations, *IEEE Trans. Syst. Man Cybern. -Syst.*, **51**(5), 2021, 2687–2700.

[16] Zhang, L. and Guo, L., Adaptive identification with guaranteed performance under saturated-observation and non-persistent excitation, 2022, https://arxiv.org/abs/2207.02422.

[17] Zhang, L., Zhao, Y. and Guo, L., Identification and adaptation with binary-valued observations under non-persistent excitation condition, *Automatica*, **138**, 2022, 110158.

[18] Zhao, W. X. and Chen, H. F., Markov chain approach to identifying Wiener systems, *Sci. China Inf. Sci.*, **55**(5), 2012, 1201–1217.

[19] Zhao, Y. L., Wang, L. Y., Yin, G. G. and Zhang, J. F., Identification of Wiener systems with binary-valued output observations, *Automatica*, **43**(10), 2007, 1752–1765.