# Recursive Identification Based on Local Likelihood Function with Binary-Valued Observations

Xin Li, *Member, IEEE*, Mingjie Shao, *Member, IEEE*, Ji-Feng Zhang, *Fellow, IEEE*, and Yanlong Zhao, *Senior Member, IEEE*

***Abstract*— This paper studies the control-oriented recursive identification of finite impulse response systems with binary-valued observations. Inspired by the Maximum Likelihood method, a novel recursive algorithm is proposed using the statistical property of system noises and observations. Unlike existing research, the gradient of the proposed algorithm is derived from the local likelihood function, which has not been previously considered. The core advantage of the algorithm is the adaptation of the recursive weight term, and especially, it has an accelerating effect when the estimated value deviates far from the true value. Besides, compared with existing algorithm based on time-varying thresholds, the proposed algorithm makes it applicable to fixed threshold scenarios through weighting, thus avoiding the complexity caused by time-varying thresholds. The proposed algorithm is proved to be convergent in both almost sure and mean square sense. Furthermore, the almost sure and mean square convergence rates are also obtained under some mild conditions. Two simulations are presented to demonstrate the effectiveness of the proposed algorithm and advantage of convergence rate over existing algorithm.**

***Index Terms*— Binary-valued observations, Likelihood function, Stochastic approximation, System identification.**

## I. Introduction

With the continuous development of information technology and digital communication, set-valued systems are increasingly integrated into our daily lives, due to their widespread applications [1]–[3]. Different from the traditional accurate output systems, the information provided by set-valued systems is the set to which the output belongs. The emergence of set-valued systems is mainly owing to the inherent features of systems themselves (such as determining disease and health states in complex disease diagnosis [4]–[6], truth and falsehood of radar targets [7], digital signal in computers [8], [9]), as well as the cost constraints of sensors (such as switch sensors for measuring automobile exhaust [10], [11], limited communication resource [12]–[16], etc.). When the output is limited into two sets, the system is named as a binary-valued system. The arising of set-valued systems identification problem, due to limited observational information, makes the existing identification methods that rely on accurate output no longer applicable. Consequently, it presents new challenges and demands for identification theory.

There are many offline algorithm studies for identification of quantized systems, which have demonstrated excellent performance in identification tasks. For example, with periodic inputs, the maximum likelihood (ML) solution can be obtained directly for the identification of finite impulse response (FIR) system with binary-valued observations [1], which is named as the empirical measure method. [17] constructed an optimal quasi-convex combination estimation for the empirical measure algorithm, and achieved its asymptotic effectiveness. [18] used weighted least squares (LS) as the optimization objective function approach and designed an offline parameter estimation algorithm through iterative optimization. [19] provided numerical solutions of the ML method for this quantized identification problem by designing iterative algorithm. However, offline algorithms based on periodic inputs or iterative optimization are difficult to apply to control-oriented problems, due to fully determined input of the former and geometrically increasing complexity over time of the latter [20]–[23]. Therefore, research on control-oriented identification with binary-valued observations mainly focuses on recursive algorithms.

Recursive identification algorithms have attracted extensive research in adaptive control problems [24], due to their significant advantages of requiring less storage and low complexity. For the identification of quantized FIR system, there is no explicit solution of the LS and ML method with normal input conditions. Based on the similar LS objective function of [18], [25] constructed a recursive identification algorithm for noise-

The work is supported by National Natural Science Foundation of China under Grants 62025306, 62433020 and T2293770, CAS Project for Young Scientists in Basic Research under Grant YSBR-008. The material in this paper was not presented at any conference. Corresponding author: Yanlong Zhao.

Xin Li is with the State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (e-mail: lixin2020@amss.ac.cn).

Mingjie Shao and Yanlong Zhao are with the State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Mathematics Sciences, University of Chinese Academy of Sciences, Bejing 100149, China (e-mail: mingjieshao@amss.ac.cn; ylzhao@amss.ac.cn).

Ji-Feng Zhang is with the School of Automation and Electrical Engineering, Zhongyuan University of Technology, ZhengZhou 450007, China, and also with the State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (e-mail: jif@iss.ac.cn).

free FIR systems with the known norm of unknown parameters under binary observations. Furthermore, [26] improved the algorithm by using adaptive regulative coefficient, and proves that the estimation algorithm can converge to the true value. [27] proposed a recursive identification algorithm using a known norm of parameters, which was modified by the recursive weighted LS algorithm, and provided the bound of the estimation error. However, it is difficult to achieve noise-free conditions and a known norm of parameters in practical systems.

Under bounded persistent excitations, a innovative gradient utilizing the distribution function of noise was constructed in [28], and a recursive projection algorithm was designed to identify the FIR system with binary-valued observations. The mean square and almost sure convergence rates of the algorithm were proved to be $O\left(\ln k/k\right)$, where $k$ denotes the number of observations. This type of recursive algorithm has attracted extensive research attention. [29] further proved that the mean square convergence rate of the recursive projection algorithm can reach $O\left(1/k\right)$, and demonstrated that this rate depends on the true parameters. [30] proposed a stochastic gradient projection algorithm, and proved that by designing appropriate weights, the algorithm can achieve asymptotic effectiveness. But, this type of algorithms [28]–[30] has a significant problem, which is that in order to achieve bounded estimation required for algorithm analysis, a projection operator is added to the algorithm. It increases the complexity of the algorithm, and also requires the unknown parameters belong to a known compact set. To remove the projection operator, [31] innovatively introduced an auxiliary stochastic process, and proved that the algorithm can converge without the need for a projection operator. However, it is important to note that the mean square convergence rate can only be guaranteed when the estimated value falls within a bounded region, or in other words, when $k$ is big enough. Conversely, when the estimate deviates far from the true value, the convergence rate tends to decrease.

Besides, some studies were conducted using time-varying thresholds [32]–[35]. Compared to fixed thresholds, designable thresholds can provide richer information, while introducing considerable complexity to the quantizer. Given that thresholds are designed using parameter estimates [32]–[34], the quantizer requires continuous access to these estimates, either by transmitting the estimates to the quantizer or by enabling the quantizer itself to compute the estimates. However, this requirement is difficult to implement in a real system. Consequently, this paper focuses on refining the sign-error type algorithm proposed by [33] and making it suitable for fixed thresholds scenarios.

In this paper, it is noted that the recursive algorithm constructed in many existing works [28]–[31], [36] utilizing the gradient derived from local objective function of LS method, whereas the gradient of local objective function of ML method has not been considered. Based on the gradient derived by local likelihood function, a novel recursive identification algorithm is designed under fixed thresholds scenarios. The main contributions are listed as follows.

i. This paper studies the recursive identification of FIR systems with binary-valued observations. A recursive identification algorithm based on local likelihood function is proposed using the statistical property of the system noises and observations with fixed threshold. The core of the proposed algorithm is a weighting approach for the binary-valued observations. Compared to the existing sign-error type recursive algorithm that employs time-varying thresholds [33], this approach avoids the complexity of computation or transmission for quantizers, and makes the sign-error type algorithm applicable to fixed thresholds scenarios.

ii. The gradient of the proposed algorithm in this paper is derived from the local likelihood function, which has not been previously considered, compared to existing research [28]–[31], [36] deriving gradient from local objective function of LS method. The main advantage of the algorithm is the adaptation of recursive weight, and especially, it has an accelerating effect when the estimated value deviates far from the true value. In addition, the proposed algorithm does not require a priori information on parameter location.

iii. Through analyzing the properties of the adaptive weight, the proposed algorithm is proved to be convergent in both almost sure and mean square sense under bounded persistent excitations. Furthermore, despite being constrained by limited information, the almost sure convergence rate of the proposed algorithm is proved to achieve $O(\sqrt{\ln k/k})$. And the mean square convergence rate reaches $O\left(1/k\right)$, which is the best performance with quantized and even accurate observations in the sense of the Cramér-Rao (CR) lower bound.

The rest of this paper is organized as follows. Section II formulates the identification problem considered in this paper. Section III introduces the design idea of the proposed algorithm. Section IV gives the properties of the identification algorithm, including convergence and convergence rate. In Section V, two simulations are given to demonstrate the theoretical results. At last, Section VI summarizes the conclusions and future works of this paper.

## II. Problem formulation

Consider the FIR system:

$$y_k = \phi_k^\top \theta + d_k, k \geq 1, \tag{1}$$

where $\phi_k = [u_k, \ldots, u_{k-n+1}]^\top \in \mathbb{R}^n$ is the input, $\theta \in \mathbb{R}^n$ is the unknown parameter, and $d_k$ is the system noise, respectively. The output $y_k$ of the system (1) cannot be measured directly, and only can be observed by a binary-valued sensor with a known and fixed threshold $C$, which can be represented by a sign function

$$s_k = \operatorname{sgn}(y_k - C) = \begin{cases} 1, & y_k \geq C; \\ -1, & y_k < C. \end{cases} \tag{2}$$

Our goal is to estimate the unknown parameter $\theta$ utilizing the input $\phi_k$ and the binary observation $s_k$. We have the following assumptions on the noise and input.

*Assumption 1:* The noise sequence $\{d_k, k \geq 1\}$ is independent and identically distributed with $d_1 \sim N(0, \sigma^2)$, where $\sigma$ is known. The cumulative distribution and probability density function of $d_1$ are denoted as $F(\cdot)$ and $f(\cdot)$, respectively.

*Assumption 2:* The input $\phi_k$ is $\mathscr{F}_{k-1}$-measurable with $\mathscr{F}_{k-1}$ being the $\sigma$-algebra generated by $\{d_1, \ldots, d_{k-1}\}$, and follows

$$\sup_{k\geq 1} \|\phi_k\| \leq M < \infty,$$

where $\|\cdot\|$ is the Euclidean norm in this paper, and there exists positive integer $N \geq n$ and positive constant $\delta$, such that

$$\frac{1}{N}\sum_{j=1}^{N} \phi_{k+j}\phi_{k+j}^{\top} \geq \delta I_n, k \geq 0, \tag{3}$$

where $I_n$ is the $n \times n$ identity matrix.

*Remark 1:* Assumption 2 is the same as the assumption about input in [28]. The input is an adaptive sequence, which is more applicable to control-oriented tasks, compared to the deterministic input used by the existing algorithm without projection in [31].

## III. Algorithm

In this section, we will present the main idea of algorithm design. Recursive algorithms require the design of innovation (direction of algorithm) and step size to update the estimate from the previous time step. Let $\hat{\theta}_k$ represents the estimation of $\theta$ at time $k$.

### A. Existing algorithm for the identification with binary-valued observations

For the recursive identification of FIR system with binary-valued observations, most of the existing works [28]–[31], [36] design algorithm by using LS method.

Specifically, when receiving the $k$-th observation $s_k$, according to the local optimization objective of LS, that is, the part of the objective function that only involves the $k$-th observation $s_k$,

$$\min_{\theta\in\mathbb{R}^n} g_k(\theta) := \left(s_k - \mathbb{E}[s_k|\theta]\right)^2, \tag{4}$$

where

$$\begin{aligned}\mathbb{E}[s_k|\theta] &= P(s_k = 1) - P(s_k = -1)\\ &= P(\phi_k^{\top}\theta + d_k \geq C) - P(\phi_k^{\top}\theta + d_k < C)\\ &= F(-C + \phi_k^{\top}\theta) - F(C - \phi_k^{\top}\theta)\\ &= 2F(\phi_k^{\top}\theta - C) - 1,\end{aligned}$$

the negative gradient direction based on the $\hat{\theta}_{k-1}$ is

$$\begin{aligned}&-\left.\frac{\partial g_k(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}_{k-1}}\\ =&4\left(s_k + 1 - 2F(\phi_k^{\top}\hat{\theta}_{k-1} - C)\right) f(\phi_k^{\top}\hat{\theta}_{k-1} - C)\phi_k.\end{aligned}$$

Due to $f(\cdot)$ always being greater than 0 under Assumption 1, many researchers use the above gradient and employ stochastic approximation method, to design recursive algorithm in the following form,

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \frac{\beta\phi_k}{k}\left(s_k + 1 - 2F(\phi_k^{\top}\hat{\theta}_{k-1} - C)\right), \tag{5}$$

with any given initial value $\hat{\theta}_0$ and appropriate step size $\beta$. For the convenience of analysis, many existing works [28]–[30] have added a projection operator to ensure that the estimated values are uniformly bounded, which requires a priori information of the parameter location. [31] proved that without the projection operator, as shown in algorithm (5), the convergence and convergence rate can also be obtained.

*Remark 2:* The stochastic approximation method [37] is suitable for some situations with random interference, and it requires the step size $a_k$ to satisfy

$$\sum_{k=1}^{\infty} a_k = \infty, \ \sum_{k=1}^{\infty} a_k^2 < \infty.$$

A typical and simple step size is designed as $a_k = \dfrac{\beta}{k}$, where $\beta$ is a positive step size that can be freely designed.

### B. A recursive identification algorithm based on local likelihood function

Based on the above discussion, it is natural to consider the log-likelihood function of ML method. Since $P(s_k = 1) = F(\phi_k^{\top}\theta - C)$ and $P(s_k = -1) = F(C - \phi_k^{\top}\theta)$, the local optimization objective of log-likelihood function that only involves $s_k$ is

$$\max_{\theta\in\mathbb{R}^n} L_k(\theta) := \log F\left(s_k(\phi_k^{\top}\theta - C)\right). \tag{6}$$

Further, the gradient direction based on the $\hat{\theta}_{k-1}$ is

$$\left.\frac{\partial L_k(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}_{k-1}} = \frac{f\left(s_k(\phi_k^{\top}\hat{\theta}_{k-1} - C)\right)}{F\left(s_k(\phi_k^{\top}\hat{\theta}_{k-1} - C)\right)} s_k\phi_k.$$

Define

$$p(x) = \frac{f(x)}{F(x)}, \tag{7}$$

which is named as Inverse Mills' Ratio [44]. A recursive identification algorithm based on local likelihood function can be designed as

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \frac{\beta\phi_k}{k} s_k p\left(s_k(\phi_k^{\top}\hat{\theta}_{k-1} - C)\right), \tag{8}$$

with any given initial value $\hat{\theta}_0$ and appropriate positive step size $\beta$. $\beta$ plays an important role in the convergence rate, which will be shown in Theorems 2 and 3.

We have the following lemma and corollary about $p(x)$.

*Lemma 1:* (Remark 4, [19]) Under Assumption 1 and definition (7), for $x \in (-\infty, \infty)$, the function $\dfrac{\mathrm{d}p(x)}{\mathrm{d}x}$ is a strictly increasing function and $\dfrac{\mathrm{d}p(x)}{\mathrm{d}x} \in (-\dfrac{1}{\sigma^2}, 0)$.

*Corollary 1:* Under the conditions of Lemma 1, we have

i. $|p(x) - p(y)| \leq \dfrac{|x-y|}{\sigma^2}$, for any $x, y \in (-\infty, \infty)$;

ii. $p(x) \le p(0)+\frac{|x|}{\sigma^2} = \frac{\sqrt{2}}{\sqrt{\pi}\sigma}+\frac{|x|}{\sigma^2}$, and $p^2(x) \le \frac{4}{\pi\sigma^2}+\frac{2x^2}{\sigma^4}$, for any $x \in (-\infty, \infty)$.

The proof of Corollary 1 is omitted.

### C. Difference from the sign-error type algorithm with time-varying thresholds

For the identification problem with binary-valued observations using time-varying thresholds, the sign-error type algorithm proposed in [33] can be written as follows:

$$\begin{cases} \hat{\theta}_k = \Pi_\Theta \left( \hat{\theta}_{k-1} + \frac{\beta\phi_k}{r_k} s_k \right), \\ s_k = \operatorname{sgn}(y_k - \phi_k^\top \hat{\theta}_{k-1}), \\ r_k = 1 + \sum_{l=1}^{k} \phi_l^\top \phi_l, \end{cases} \tag{9}$$

where $\phi_k^\top \hat{\theta}_{k-1}$ is designed time-varying threshold, and $\Pi_\Theta(\cdot)$ is a projection mapping from $\mathbb{R}_n$ to a compact set $\Theta$, to ensure that $\hat{\theta}_k$ is uniformly bounded. $r_k$ represents the sum of the squared Euclidean norms of $\{\phi_l, l = 1, \ldots, k\}$. Under Assumption 2, we have that $r_k$ is of the same order as $k$, because

$$\begin{aligned} r_k &= 1 + \sum_{l=1}^{k} \operatorname{tr}\left(\phi_l \phi_l^\top\right) \\ &\ge 1 + \left\lfloor \frac{k}{N} \right\rfloor Nn\delta \ge 1 + (k-N)n\delta, \end{aligned}$$

and

$$r_k \le 1 + kM^2.$$

It can be seen that the differences between algorithm (8) and (9) lie in that (8) has a weighted term $p\left(s_k(\phi_k^\top \hat{\theta}_{k-1} - C)\right)$ on the binary-valued observations, does not require priori parameter information (no projection operator) and uses a fixed threshold $C$ rather than time-varying thresholds designed by $\phi_k^\top \hat{\theta}_{k-1}$. Designable thresholds $C_k$ contains estimated values $\hat{\theta}_{k-1}$, which are difficult to implement in a real system, either through transmitting the estimates to the quantizer or by the quantizer itself being able to compute the parameter estimates. To overcome this hurdle, a weighting approach for the binary-valued observations is applied in algorithm (8), making the sign-error type algorithm (9) applicable for fixed threshold scenarios.

## IV. Properties of the Identification Algorithm

In this section, the almost sure and mean square convergence of the identification algorithm will be proved and the convergence rates will also be obtained. First, define

$$q(x) = \frac{f(x)}{F(x)\left[1 - F(x)\right]}, \tag{10}$$

and an important lemma about $q(x)$ is introduced to help us analyze the convergence of the proposed algorithm.

*Lemma 2:* Under Assumption 1, we have the following properties for $q(x)$:

i. $q(x)$ is an even function;
ii. $q(x) > \frac{2}{\sigma}\sqrt{\frac{2}{\pi}}$, when $x > 0$;
iii. $q(x) < \frac{2}{\sigma}\sqrt{\frac{2}{\pi}} + \frac{x}{\sigma^2}$, when $x > 0$;
iv. $q(x) > \frac{x}{\sigma^2}$, when $x > 0$.

The proof of Lemma 2 is given in Appendix I-A.

*Remark 3:* $q(x)$ has a positive consistent lower bound globally, and

$$\lim_{x\to+\infty} \frac{q(x)\sigma^2}{x} = 1.$$

The function curve of $q(x)$ with $\sigma = 1$ is shown in Fig. 1.

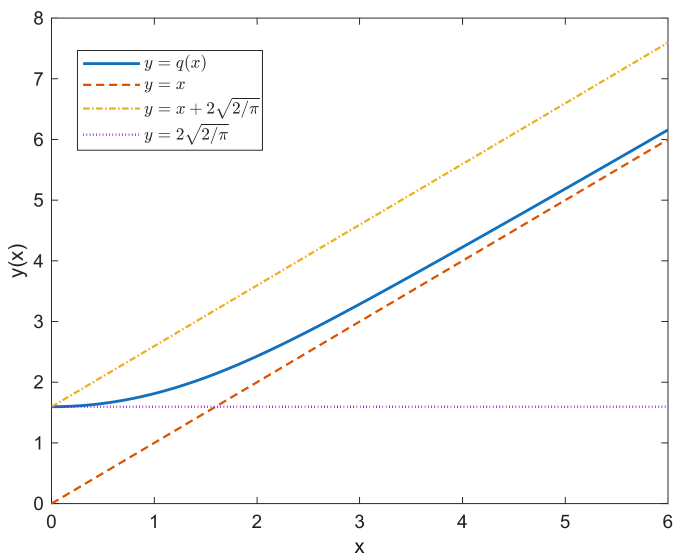


Fig. 1. $q(x)$ on positive real axis with $\sigma = 1$.

### A. Convergence

This subsection will present the conclusions of convergence properties of the proposed algorithm (8).

Let $\tilde{\theta}_k = \hat{\theta}_k - \theta$ denote the estimation error at time $k$. Then, from (1), (2), and the proposed algorithm (8), we have

$$\tilde{\theta}_k = \tilde{\theta}_{k-1} + \frac{\beta s_k p\left(s_k(\phi_k^\top \hat{\theta}_{k-1} - C)\right)\phi_k}{k}, \tag{11}$$

and the following lemma.

*Lemma 3:* Under Assumptions 1 and 2, we have

$$\|\tilde{\theta}_k - \tilde{\theta}_{k-1}\| \le \frac{\beta M |\phi_k^\top \tilde{\theta}_{k-1}|}{k\sigma^2} + O\left(\frac{1}{k}\right),$$

and further

$$\|\tilde{\theta}_k - \tilde{\theta}_{k-1}\|^2 \le \frac{2\beta^2 M^2 |\phi_k^\top \tilde{\theta}_{k-1}|^2}{k^2\sigma^4} + O\left(\frac{1}{k^2}\right).$$

The proof of Lemma 3 is given in Appendix I-B.

*Lemma 4:* Under Assumption 1, for any fixed $b > 0$, when $|x| \le \frac{b}{2}$, there exists $B = F\left(\frac{3b}{2}\right) - F\left(\frac{b}{2}\right)$, such that

$$\begin{cases} F(x) - F(x+\alpha) \le -B, \text{ if } \alpha > b; \\ F(x) - F(x+\alpha) \ge B, \text{ if } \alpha < -b. \end{cases}$$

The proof of Lemma 4 is given in Appendix I-C.

*Theorem 1:* Under Assumptions 1 and 2, the parameter estimation error $\tilde{\theta}_k$ of the proposed algorithm (8) for the systems (1) and (2) converges to 0 in both mean square and almost sure sense, i.e.,

$$\lim_{k\to\infty} \mathbb{E}\tilde{\theta}_k^\top \tilde{\theta}_k = 0,$$

and

$$\lim_{k\to\infty} \tilde{\theta}_k = 0, a.s.$$

*Proof:* It follows from (11) that

$$\begin{aligned}
&\mathbb{E}\left[\|\tilde{\theta}_k\|^2|\mathscr{F}_{k-1}\right]\\
=&\|\tilde{\theta}_{k-1}\|^2+\mathbb{E}\left[\frac{2\beta s_k p\left(s_k(\phi_k^\top\hat{\theta}_{k-1}-C)\right)\phi_k^\top\tilde{\theta}_{k-1}}{k}|\mathscr{F}_{k-1}\right]\\
&+\mathbb{E}\left[\frac{\beta^2p^2\left(s_k(\phi_k^\top\hat{\theta}_{k-1}-C)\right)\phi_k^\top\phi_k}{k^2}|\mathscr{F}_{k-1}\right].
\end{aligned} \tag{12}$$

Let us analyze the second term on the right-hand-side of (12). Under Assumption 2, given a constant $b$ that satisfies $b\geq 2(M\|\theta\|+|C|)\geq 2|\phi_k^\top\theta+C|$, then we get

$$\begin{aligned}
&\mathbb{E}\left[s_kp\left(s_k(\phi_k^\top\hat{\theta}_{k-1}-C)\right)\phi_k^\top\tilde{\theta}_{k-1}|\mathscr{F}_{k-1}\right]\\
=&\phi_k^\top\tilde{\theta}_{k-1}F\left(\phi_k^\top\theta-C\right)\frac{f\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)}{F\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)}\\
&-\phi_k^\top\tilde{\theta}_{k-1}\left(1-F\left(\phi_k^\top\theta-C\right)\right)\frac{f\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)}{1-F\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)}\\
=&\phi_k^\top\tilde{\theta}_{k-1}q\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\left(F\left(\phi_k^\top\theta-C\right)-F\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\right)\\
=&\phi_k^\top\tilde{\theta}_{k-1}q\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\cdot\\
&\left\{\left(F\left(\phi_k^\top\theta-C\right)-F\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\right)I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|>b\}}\right.\\
&\left.+\left(F\left(\phi_k^\top\theta-C\right)-F\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\right)I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|\leq b\}}\right\},
\end{aligned} \tag{13}$$

where $q(\cdot)$ is defined in (10).

When $|\phi_k^\top\tilde{\theta}_{k-1}|>b$, due to $|\phi_k^\top\theta-C|\leq\frac{b}{2}$, by Lemma 4, there exists $B>0$, such that if $\phi_k^\top\tilde{\theta}_{k-1}>b$,

$$\left(F\left(\phi_k^\top\theta-C\right)-F\left(\phi_k^\top\theta-C+\phi_k^\top\tilde{\theta}_{k-1}\right)\right)\leq -B,$$

and if $\phi_k^\top\tilde{\theta}_{k-1}<-b$,

$$\left(F\left(\phi_k^\top\theta-C\right)-F\left(\phi_k^\top\theta-C+\phi_k^\top\tilde{\theta}_{k-1}\right)\right)\geq B.$$

It means that

$$\begin{aligned}
&\phi_k^\top\tilde{\theta}_{k-1}\left(F(\phi_k^\top\theta-C)-F\left(\phi_k^\top\theta-C+\phi_k^\top\tilde{\theta}_{k-1}\right)\right)I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|>b\}}\\
&\leq -B|\phi_k^\top\tilde{\theta}_{k-1}|I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|>b\}}.
\end{aligned} \tag{14}$$

Besides, when $|\phi_k^\top\tilde{\theta}_{k-1}|>b$, since

$$|\phi_k^\top\theta-C|\leq\frac{b}{2}<\frac{|\phi_k^\top\tilde{\theta}_{k-1}|}{2},$$

and by Lemma 2, we have

$$\begin{aligned}
&q\left(\phi_k^\top\theta-C+\phi_k^\top\tilde{\theta}_{k-1}\right)I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|>b\}}\\
\geq&\frac{|\phi_k^\top\theta-C+\phi_k^\top\tilde{\theta}_{k-1}|}{\sigma^2}I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|>b\}}\\
\geq&\frac{|\phi_k^\top\tilde{\theta}_{k-1}|-|\phi_k^\top\theta-C|}{\sigma^2}I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|>b\}}\\
\geq&\frac{|\phi_k^\top\tilde{\theta}_{k-1}|}{2\sigma^2}I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|>b\}}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\phi_k^\top\tilde{\theta}_{k-1}q\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\cdot\\
&\left(F\left(\phi_k^\top\theta-C\right)-F\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\right)I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|>b\}}\\
\leq&-B\frac{|\phi_k^\top\tilde{\theta}_{k-1}|^2}{2\sigma^2}I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|>b\}}.
\end{aligned} \tag{15}$$

In addition, when $|\phi_k^\top\tilde{\theta}_{k-1}|\leq b$, using the differential mean value theorem (Theorem 5.3.1, [40]), Assumption 1 and Lemma 2, there exists $\xi_k$ between $\phi_k^\top\theta-C$ and $\phi_k^\top\theta-C+\phi_k^\top\tilde{\theta}_{k-1}$, such that

$$\begin{aligned}
&\phi_k^\top\tilde{\theta}_{k-1}q\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\cdot\\
&\left(F\left(\phi_k^\top\theta-C\right)-F\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\right)I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|\leq b\}}\\
=&-\left(\phi_k^\top\tilde{\theta}_{k-1}\right)^2f(\xi_k)q\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|\leq b\}}\\
\leq&-\left(\phi_k^\top\tilde{\theta}_{k-1}\right)^2\frac{2}{\sigma}\sqrt{\frac{2}{\pi}}f(|\phi_k^\top\theta-C|+|\phi_k^\top\tilde{\theta}_{k-1}|)I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|\leq b\}}\\
\leq&-\left(\phi_k^\top\tilde{\theta}_{k-1}\right)^2\frac{2}{\sigma}\sqrt{\frac{2}{\pi}}f\left(\frac{3}{2}b\right)I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|\leq b\}}.
\end{aligned} \tag{16}$$

Let

$$\gamma=\min\left\{\frac{B}{2\sigma^2},\frac{2}{\sigma}\sqrt{\frac{2}{\pi}}f\left(\frac{3}{2}b\right)\right\}. \tag{17}$$

Then, combining (15), (16) and (17), (13) becomes

$$\begin{aligned}
&\mathbb{E}\left[s_kp\left(s_k(\phi_k^\top\hat{\theta}_{k-1}-C)\right)\phi_k^\top\tilde{\theta}_{k-1}|\mathscr{F}_{k-1}\right]\\
=&\phi_k^\top\tilde{\theta}_{k-1}q\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\cdot\\
&\left\{\left(F\left(\phi_k^\top\theta-C\right)-F\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\right)I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|>b\}}\right.\\
&\left.+\left(F\left(\phi_k^\top\theta-C\right)-F\left(\phi_k^\top\hat{\theta}_{k-1}-C\right)\right)I_{\{|\phi_k^\top\tilde{\theta}_{k-1}|\leq b\}}\right\}\\
\leq&-\gamma\left(\phi_k^\top\tilde{\theta}_{k-1}\right)^2.
\end{aligned} \tag{18}$$

Let's move to the third term on the right-hand-side of (12).

Using Corollary 1, we can obtain

$$\begin{aligned}
&\mathbb{E}\left[\frac{\beta^2p^2\left(s_k(\phi_k^\top\hat{\theta}_{k-1}-C)\right)\phi_k^\top\phi_k}{k^2}|\mathscr{F}_{k-1}\right]\\
\leq&\frac{\beta^2M^2}{k^2}\mathbb{E}\left[\left(\frac{\sqrt{2}}{\sqrt{\pi}\sigma}+\frac{|\phi_k^\top\hat{\theta}_{k-1}-C|}{\sigma^2}\right)^2|\mathscr{F}_{k-1}\right]\\
\leq&\frac{\beta^2M^2}{k^2}\left(\frac{\sqrt{2}}{\sqrt{\pi}\sigma}+\frac{|\phi_k^\top\tilde{\theta}_{k-1}|+|\phi_k^\top\theta-C|}{\sigma^2}\right)^2\\
\leq&\frac{\beta^2M^2}{k^2}\left(\frac{2\left(\phi_k^\top\tilde{\theta}_{k-1}\right)^2}{\sigma^4}+2\left(\frac{\sqrt{2}}{\sqrt{\pi}\sigma}+\frac{|\phi_k^\top\theta-C|}{\sigma^2}\right)^2\right)\\
=&\frac{2\beta^2M^2\left(\phi_k^\top\tilde{\theta}_{k-1}\right)^2}{k^2\sigma^4}+O\left(\frac{1}{k^2}\right).
\end{aligned}\tag{19}$$

Combining (12) with (18) and (19), we get

$$\begin{aligned}
&\mathbb{E}\left[\|\tilde{\theta}_k\|^2|\mathscr{F}_{k-1}\right]\\
\leq&\|\tilde{\theta}_{k-1}\|^2+\left(-\frac{2\beta\gamma}{k}+\frac{2\beta^2M^2}{k^2\sigma^4}\right)\left(\phi_k^\top\tilde{\theta}_{k-1}\right)^2+O\left(\frac{1}{k^2}\right),
\end{aligned}\tag{20}$$

and

$$\begin{aligned}
&\mathbb{E}\|\tilde{\theta}_k\|^2\leq\mathbb{E}\|\tilde{\theta}_{k-1}\|^2\\
&+\left(-\frac{2\beta\gamma}{k}+\frac{2\beta^2M^2}{k^2\sigma^4}\right)\mathbb{E}\left(\phi_k^\top\tilde{\theta}_{k-1}\right)^2+O\left(\frac{1}{k^2}\right).
\end{aligned}\tag{21}$$

Let $V_k=\sum_{j=1}^N\mathbb{E}\|\tilde{\theta}_{k+j}\|^2$, where $N$ is from Assumption 2. Then, we have

$$\begin{aligned}
V_k\leq&V_{k-1}-\frac{2\beta\gamma\sum_{j=1}^N\mathbb{E}\left(\phi_{k+j}^\top\tilde{\theta}_{k+j-1}\right)^2}{k}\\
&+V_{k-1}O\left(\frac{1}{k^2}\right)+O\left(\frac{1}{k^2}\right).
\end{aligned}\tag{22}$$

For any $1\leq i\neq j\leq N$, without loss of generality, let $j>i$. Using Lemma 3, we can get

$$\begin{aligned}
&\|\tilde{\theta}_{k+j-1}-\tilde{\theta}_{k+i-1}\|^2\\
=&\left\|\sum_{l=i}^{j-1}\left(\tilde{\theta}_{k+l}-\tilde{\theta}_{k+l-1}\right)\right\|^2\\
\leq&(j-i)\sum_{l=i}^{j-1}\|\tilde{\theta}_{k+l}-\tilde{\theta}_{k+l-1}\|^2\\
\leq&\sum_{l=i}^{j-1}\|\tilde{\theta}_{k+l-1}\|^2O\left(\frac{1}{k^2}\right)+O\left(\frac{1}{k^2}\right),
\end{aligned}\tag{23}$$

and

$$\begin{aligned}
&\left|\left(\phi_{k+j}^\top\tilde{\theta}_{k+j-1}\right)^2-\left(\phi_{k+j}^\top\tilde{\theta}_{k+i-1}\right)^2\right|\\
=&\left|\left(\tilde{\theta}_{k+j-1}-\tilde{\theta}_{k+i-1}+\tilde{\theta}_{k+i-1}\right)^\top\phi_{k+j}\cdot\right.\\
&\left.\phi_{k+j}^\top\left(\tilde{\theta}_{k+j-1}-\tilde{\theta}_{k+i-1}+\tilde{\theta}_{k+i-1}\right)-\left(\phi_{k+j}^\top\tilde{\theta}_{k+i-1}\right)^2\right|\\
=&\left|\left(\left(\tilde{\theta}_{k+j-1}-\tilde{\theta}_{k+i-1}\right)^\top\phi_{k+j}\right)^2\right.\\
&\left.+2\left(\tilde{\theta}_{k+j-1}-\tilde{\theta}_{k+i-1}\right)^\top\phi_{k+j}\phi_{k+j}^\top\tilde{\theta}_{k+i-1}\right|\\
\leq&M^2\|\tilde{\theta}_{k+j-1}-\tilde{\theta}_{k+i-1}\|^2+M^2k\|\tilde{\theta}_{k+j-1}-\tilde{\theta}_{k+i-1}\|^2\\
&+\frac{M^2}{k}\|\tilde{\theta}_{k+i-1}\|^2\\
\leq&\sum_{l=i}^{j-1}\|\tilde{\theta}_{k+l-1}\|^2O\left(\frac{1}{k}\right)+O\left(\frac{1}{k}\right).
\end{aligned}\tag{24}$$

Then, we can get

$$\begin{aligned}
&N\sum_{j=1}^N\left(\phi_{k+j}^\top\tilde{\theta}_{k+j-1}\right)^2\\
\geq&\sum_{j=1}^N\sum_{i=1}^N\left(\phi_{k+j}^\top\tilde{\theta}_{k+i-1}\right)^2-\sum_{i=1}^{N-1}\|\tilde{\theta}_{k+i-1}\|^2O\left(\frac{1}{k}\right)-O\left(\frac{1}{k}\right)\\
=&\sum_{i=1}^N\tilde{\theta}_{k+i-1}^\top\sum_{j=1}^N\left(\phi_{k+j}\phi_{k+j}^\top\right)\tilde{\theta}_{k+i-1}\\
&-\sum_{i=1}^{N-1}\|\tilde{\theta}_{k+i-1}\|^2O\left(\frac{1}{k}\right)-O\left(\frac{1}{k}\right),
\end{aligned}\tag{25}$$

and by Assumption 2,

$$\sum_{i=1}^N\tilde{\theta}_{k+i-1}^\top\sum_{j=1}^N\left(\phi_{k+j}\phi_{k+j}^\top\right)\tilde{\theta}_{k+i-1}\geq N\delta\sum_{i=1}^N\|\tilde{\theta}_{k+i-1}\|^2.\tag{26}$$

Inequality (22) together with (25) and (26) indicates

$$\begin{aligned}
V_k\leq&V_{k-1}-\frac{2\beta\gamma\sum_{j=1}^N\mathbb{E}\left(\phi_{k+j}^\top\tilde{\theta}_{k+j-1}\right)^2}{k}+\\
&V_{k-1}O\left(\frac{1}{k^2}\right)+O\left(\frac{1}{k^2}\right)\\
\leq&V_{k-1}-\frac{2\beta\gamma\delta V_{k-1}}{k}+V_{k-1}O\left(\frac{1}{k^2}\right)+O\left(\frac{1}{k^2}\right)\\
=&\left(1-\frac{2\beta\gamma\delta}{k}+O\left(\frac{1}{k^2}\right)\right)V_{k-1}+O\left(\frac{1}{k^2}\right),
\end{aligned}\tag{27}$$

which implies that $V_k$ is monotonically decreasing, when $k$ is big enough. And further we can get

$$V_k\leq V_0+\sum_{l=1}^k\left(-\frac{2\beta\gamma\delta}{l}+O\left(\frac{1}{l^2}\right)\right)V_{l-1}+\sum_{l=1}^kO\left(\frac{1}{l^2}\right).\tag{28}$$

Due to $V_k \geq 0$, we have

$$\sum_{l=1}^{k}\left(\frac{2\beta\gamma\delta}{l}+O\left(\frac{1}{l^2}\right)\right)V_{l-1}<\infty,$$

which together with the monotonicity of $V_k$ when $k$ is big enough, leads to

$$\lim_{k\to\infty} V_k = 0,$$

and

$$\lim_{k\to\infty}\mathbb{E}\|\tilde{\theta}_k\|^2=0.$$

It follows from (20) that

$$\mathbb{E}\left[\|\tilde{\theta}_k\|^2|\mathscr{F}_{k-1}\right]\leq\|\tilde{\theta}_{k-1}\|^2+\frac{2\beta^2M^2\left(\phi_k^\top\tilde{\theta}_{k-1}\right)^2}{k^2\sigma^4}+O\left(\frac{1}{k^2}\right), \tag{29}$$

and

$$\begin{aligned}&\mathbb{E}\sum_{l=1}^{k}\left\{\frac{2\beta^2M^2(\phi_l^\top\tilde{\theta}_{l-1})^2}{l^2\sigma^4}+O\left(\frac{1}{l^2}\right)\right\}\\ \leq&\sum_{l=1}^{k}\left\{\frac{2\beta^2M^4}{l^2\sigma^4}\mathbb{E}\|\tilde{\theta}_{l-1}\|^2+O\left(\frac{1}{l^2}\right)\right\}\\ <&\infty.\end{aligned} \tag{30}$$

In view of Lemma 1.2.2 in [37], we have that $\|\tilde{\theta}_k\|$ converges to a bounded limit, a.s.

Notice that $\lim_{k\to+\infty}\mathbb{E}\|\tilde{\theta}_k\|^2=0$. Consequently, $\tilde{\theta}_k$ converges to 0, a.s. The proof is completed. ■

### *B. Comparison of Transient Convergence*

In this subsection, the transient performance of existing algorithm (5) and the proposed algorithm (8) will be discussed.

First of all, the recursive LS algorithm (Section 11.2, [38]) based on accurate observations is presented below,

$$\begin{cases}\hat{\theta}_k=\hat{\theta}_{k-1}+P_k\phi_k\left(y_k-\phi_k^\top\hat{\theta}_{k-1}\right),\\ P_k=\left(\sum_{l=1}^{k}\phi_k\phi_k^\top\right)^{-1},\end{cases} \tag{31}$$

where $P_k$ is of the same order as $\frac{1}{k}$ under Assumption 2.

Compared to the existing algorithm (5) and the proposed algorithm (8) for the identification with binary-valued observations, the main difference lies in: $y_k-\phi_k^\top\hat{\theta}_{k-1}$ of (31), $s_k+1-2F(\phi_k^\top\hat{\theta}_{k-1}-C)$ of (5) and $s_kp\left(s_k(\phi_k^\top\hat{\theta}_{k-1}-C)\right)$ of (8). They have similar estimation update abilities, but due to distinct types of observations, their transient convergence characteristics vary. Specifically, the innovation (i.e., the difference between the output and its estimate) of the recursive LS algorithm (31) satisfies

$$\mathbb{E}\left[y_k-\phi_k\hat{\theta}_{k-1}|\mathscr{F}_{k-1}\right]=-\phi_k^\top\tilde{\theta}_{k-1}, \tag{32}$$

which is a linear function of $\tilde{\theta}_{k-1}$, and the innovation of the existing algorithm (5) satisfies

$$\left|s_k+1-2F(\phi_k^\top\hat{\theta}_{k-1}-C)\right|\leq 2. \tag{33}$$

We can see that the innovation (32) with the same order as $\phi_k^\top\tilde{\theta}_{k-1}$, has stronger estimation updating ability, compared to the innovation (33) with a consistent upper bound. It means that the identification algorithm (5) has a drawback that it suffers from slow transient convergence rate when the estimation is poor.

Furthermore, according to (18), it is noted that the conditional expectation $\mathbb{E}\left[s_kp\left(s_k(\phi_k^\top\hat{\theta}_{k-1}-C)\right)|\mathscr{F}_{k-1}\right]$ of the corresponding term in algorithm (8) maintains the same order as $-\phi_k^\top\tilde{\theta}_{k-1}$, which is similar to the innovation (32) of the recursive LS algorithm.

Therefore, the proposed algorithm (8) overcomes the drawback of existing algorithm (5), and has an accelerating effect when the estimated value deviates far from the true value. This will be verified in Section V-B. It should also be pointed out that since $|s_k|=1$, the acceleration ability is attributed to the weighted term $p\left(s_k(\phi_k^\top\hat{\theta}_{k-1}-C)\right)$ completely.

### *C. Mean square convergence rate*

This subsection will obtain the mean square convergence rate of the proposed algorithm (8).

*Theorem 2:* Under Assumptions 1 and 2, the parameter estimation error $\tilde{\theta}_k$ of the proposed algorithm (8) for the systems (1) and (2) have the following mean square convergence rate,

$$\mathbb{E}\|\tilde{\theta}_k\|^2=\begin{cases}O\left(\frac{1}{k}\right), & 2\beta\gamma\delta>1;\\ O\left(\frac{\ln k}{k}\right), & 2\beta\gamma\delta=1;\\ O\left(\frac{1}{k^{2\beta\gamma\delta}}\right), & 2\beta\gamma\delta<1,\end{cases} \tag{34}$$

where $\beta$ is the step size of proposed algorithm (8), $\gamma$ is defined by (17), and $\delta$ is given in Assumption 2.

*Proof:* By (27), one can get

$$\begin{aligned}V_k\leq&\left(1-\frac{2\beta\gamma\delta}{k}+O\left(\frac{1}{k^2}\right)\right)V_{k-1}+O\left(\frac{1}{k^2}\right)\\ \leq&V_0\prod_{l=1}^{k}\left(1-\frac{2\beta\gamma\delta}{l}+O\left(\frac{1}{l^2}\right)\right)\\ &+\sum_{l=1}^{k}\prod_{j=l+1}^{k}\left(1-\frac{2\beta\gamma\delta}{j}+O\left(\frac{1}{j^2}\right)\right)O\left(\frac{1}{l^2}\right)\\ =&V_0e^{\sum_{l=1}^{k}\ln\left(1-\frac{2\beta\gamma\delta}{l}+O\left(\frac{1}{l^2}\right)\right)}\\ &+\sum_{l=1}^{k}O\left(\frac{1}{l^2}\right)e^{\sum_{j=l+1}^{k}\ln\left(1-\frac{2\beta\gamma\delta}{j}+O\left(\frac{1}{j^2}\right)\right)}\\ =&O\left(k^{-2\beta\gamma\delta}\right)+O\left(k^{-2\beta\gamma\delta}\sum_{l=1}^{k}l^{2\beta\gamma\delta-2}\right)\\ =&\begin{cases}O\left(\frac{1}{k}\right), & 2\beta\gamma\delta>1;\\ O\left(\frac{\ln k}{k}\right), & 2\beta\gamma\delta=1;\\ O\left(\frac{1}{k^{2\beta\gamma\delta}}\right), & 2\beta\gamma\delta<1.\end{cases}\end{aligned} \tag{35}$$

Therefore, $\mathbb{E}\|\tilde{\theta}_k\|^2 \le V_k = \begin{cases} O\left(\frac{1}{k}\right), & 2\beta\gamma\delta > 1; \\ O\left(\frac{\ln k}{k}\right), & 2\beta\gamma\delta = 1; \\ O\left(\frac{1}{k^{2\beta\gamma\delta}}\right), & 2\beta\gamma\delta < 1. \end{cases}$

The proof is completed. ■

*Remark 4:* It is worth noting that, $\gamma$ defined in (17) is related to the unknown parameter $\theta$, so the convergence rate is affected by $\theta$. In addition, to guarantee a fast algorithm convergence rate, on the one hand, we can set a big enough setp size $\beta$ according to the range of the $\theta$. On the other hand, when there is no priori knowledge of $\theta$, the step size can be adjusted using the square of the estimated value, such that $\beta$ satisfies the condition of Theorem 2 and the following Theorem 3.

*Remark 5:* The mean square convergence rate $O(1/k)$ when $2\beta\gamma\delta > 1$ in Theorem 2 is the best rate under quantized and even accurate observations in the sense of CR lower bound. This is because the CR lower bound of estimating $\theta$ utilizing binary-valued observations $s_1, \ldots, s_k$ [30] is

$$\sigma^2_{CR}(s_1, \ldots, s_k) = \left(\sum_{l=1}^{k} \frac{f^2(\phi_l^\top \theta - C)}{F(\phi_l^\top\theta - C)F(C - \phi_l^\top\theta)} \phi_l \phi_l^\top\right)^{-1} = O\left(\frac{1}{k}\right),$$

and the CR lower bound of estimating $\theta$ utilizing accurate observations $y_1, \ldots, y_k$ is

$$\sigma^2_{CR}(y_1, \ldots, y_k) = \left(\sum_{l=1}^{k} \frac{\phi_l\phi_l^\top}{\sigma^2}\right)^{-1} = O\left(\frac{1}{k}\right).$$

## D. Almost sure convergence rate

In this subsection, we will get the almost sure convergence rate of the proposed algorithm (8).

*Theorem 3:* Under Assumptions 1 and 2, the parameter estimation error $\tilde{\theta}_k$ of the proposed algorithm (8) for the systems (1) and (2) have the following almost sure convergence rate

$$\|\tilde{\theta}_k\| = O\left(\sqrt{\ln k / k}\right), a.s.,$$

when $\beta > \frac{1}{2\gamma\delta}$, where $\beta$ is the step size of proposed algorithm (8), $\gamma$ is defined by (17), and $\delta$ is given in Assumption 2.

*Proof:* By the conclusion of Theorem 1, $\|\tilde{\theta}_k\|$ converges to 0 almost surely. Then for any $\epsilon > 0$, there exists $k_\epsilon > 0$, such that $\|\tilde{\theta}_k\| \le \epsilon$ when $k > k_\epsilon$.

By (11) and Corollary 1, when $k > k_\epsilon$, we get

$$\begin{aligned} &\|\tilde{\theta}_k\|^2 \\ =&\|\tilde{\theta}_{k-1}\|^2 + \frac{2\beta s_k p\left(s_k(\phi_k^\top\hat{\theta}_{k-1} - C)\right)\phi_k^\top\tilde{\theta}_{k-1}}{k} \\ &+ \frac{\beta^2 p^2\left(s_k(\phi_k^\top\hat{\theta}_{k-1} - C)\right)\phi_k^\top\phi_k}{k^2} \\ \le&\|\tilde{\theta}_{k-1}\|^2 + \frac{2\beta s_k p\left(s_k(\phi_k^\top\hat{\theta}_{k-1} - C)\right)\phi_k^\top\tilde{\theta}_{k-1}}{k} + O\left(\frac{1}{k^2}\right). \end{aligned} \tag{36}$$

Then, by (18), we have

$$\begin{aligned} &k\|\tilde{\theta}_k\|^2 - (k-1)\|\tilde{\theta}_{k-1}\|^2 \\ \le&\|\tilde{\theta}_{k-1}\|^2 + 2\beta s_k p\left(s_k(\phi_k^\top\hat{\theta}_{k-1} - C)\right)\phi_k^\top\tilde{\theta}_{k-1} + O\left(\frac{1}{k}\right) \\ =&\|\tilde{\theta}_{k-1}\|^2 + 2\beta\phi_k^\top\tilde{\theta}_{k-1} q\left(\phi_k^\top\hat{\theta}_{k-1} - C\right) \cdot \\ &\left(F\left(\phi_k^\top\theta - C\right) - F\left(\phi_k^\top\hat{\theta}_{k-1} - C\right)\right) \\ &+ 2\beta s_k p\left(s_k(\phi_k^\top\hat{\theta}_{k-1} - C)\right)\phi_k^\top\tilde{\theta}_{k-1} \\ &- 2\beta\phi_k^\top\tilde{\theta}_{k-1} q\left(\phi_k^\top\hat{\theta}_{k-1} - C\right) \cdot \\ &\left(F\left(\phi_k^\top\theta - C\right) - F\left(\phi_k^\top\hat{\theta}_{k-1} - C\right)\right) + O\left(\frac{1}{k}\right) \\ \le&\|\tilde{\theta}_{k-1}\|^2 - 2\beta\gamma\left(\phi_k^\top\tilde{\theta}_{k-1}\right)^2 + O\left(\frac{1}{k}\right) \\ &+ 2\beta\phi_k^\top\tilde{\theta}_{k-1}\Big\{s_k p\left(s_k(\phi_k^\top\hat{\theta}_{k-1} - C)\right) - \\ &q\left(\phi_k^\top\hat{\theta}_{k-1} - C\right)\left(F\left(\phi_k^\top\theta - C\right) - F\left(\phi_k^\top\hat{\theta}_{k-1} - C\right)\right)\Big\}. \end{aligned} \tag{37}$$

Let $k_a = k - N\lfloor \frac{k-k_\epsilon}{N} \rfloor$. Then, $k_a > k_\epsilon$. When $k > k_a$, by virtue of $\|\tilde{\theta}_k\| \le \epsilon$, and using (25) and (26), we obtain

$$\begin{aligned} &- \sum_{l=k_a+1}^{k} \left(\phi_l^\top\tilde{\theta}_{l-1}\right)^2 \\ \le& -\delta\sum_{l=k_a+1}^{k}\|\tilde{\theta}_{l-1}\|^2 + \sum_{i=1}^{\frac{k-k_a}{N}} O\left(\frac{1}{k_a + (i-1)N}\right) \\ =& -\delta\sum_{l=k_a+1}^{k}\|\tilde{\theta}_{l-1}\|^2 + O(\ln k). \end{aligned} \tag{38}$$

In addition, by (13) we have

$$\begin{aligned} \mathbb{E}\Big[&s_k p\left(s_k(\phi_k^\top\hat{\theta}_{k-1} - C)\right) - q\left(\phi_k^\top\hat{\theta}_{k-1} - C\right) \times \\ &\left(F\left(\phi_k^\top\theta - C\right) - F\left(\phi_k^\top\hat{\theta}_{k-1} - C\right)\right)|\mathscr{F}_{k-1}\Big] = 0, \end{aligned}$$

and when $k > k_a$,

$$\begin{aligned} \Big|&s_k p\left(s_k(\phi_k^\top\hat{\theta}_{k-1} - C)\right) - q\left(\phi_k^\top\hat{\theta}_{k-1} - C\right) \times \\ &\left(F\left(\phi_k^\top\theta - C\right) - F\left(\phi_k^\top\hat{\theta}_{k-1} - C\right)\right)\Big| < \infty. \end{aligned}$$

Then by Theorem 1.3.10 in [39], for any $\lambda > \frac{1}{2}$,

$$\begin{aligned} &\sum_{l=k_a+1}^{k} \phi_l^\top\tilde{\theta}_{l-1}\Big\{s_l p\left(s_l(\phi_l^\top\hat{\theta}_{l-1} - C)\right) - \\ &q\left(\phi_l^\top\hat{\theta}_{l-1} - C\right)\left(F\left(\phi_l^\top\theta - C\right) - F\left(\phi_l^\top\hat{\theta}_{l-1} - C\right)\right)\Big\} \\ =&O\left(\sqrt{\sum_{l=k_a+1}^{k}\left(\phi_l^\top\tilde{\theta}_{l-1}\right)^2}\left(\log\sqrt{\sum_{l=k_a+1}^{k}\left(\phi_l^\top\tilde{\theta}_{l-1}\right)^2}\right)^{\lambda}\right) \\ =&O(1) + o\left(\sum_{l=k_a+1}^{k}\|\tilde{\theta}_{l-1}\|^2\right). \end{aligned} \tag{39}$$

Finally, combining (37) with (38) and (39), we can get

$$\begin{aligned} & k\|\tilde{\theta}_k\|^2 - k_a\|\tilde{\theta}_{k_a}\|^2 \\ \leq & \sum_{l=k_a+1}^{k} \|\tilde{\theta}_{l-1}\|^2 - 2\beta\gamma\delta \sum_{l=k_a+1}^{k} \|\tilde{\theta}_{l-1}\|^2 \\ & + o\left(\sum_{l=k_a+1}^{k} \|\tilde{\theta}_{l-1}\|^2\right) + O(\ln k) \\ = & (1 + o(1) - 2\beta\gamma\delta) \sum_{l=k_a+1}^{k} \|\tilde{\theta}_{l-1}\|^2 + O(\ln k). \end{aligned} \tag{40}$$

Thus, we have $\|\tilde{\theta}_k\|^2 = O\left(\frac{\ln k}{k}\right)$ when $\beta > \frac{1}{2\gamma\delta}$. The proof is completed. ■

## V. SIMULATIONS

In this section, two numerical simulations are given to demonstrate the theoretical results.

### A. Verification of algorithm convergence

Consider the following FIR system with binary-valued observations

$$\begin{cases} y_k = \phi_k^\top \theta + d_k, k \geq 1, \\ s_k = \begin{cases} 1, & y_k \geq C; \\ -1, & y_k < C, \end{cases} \end{cases}$$

where $\theta = [0.2, -0.5, 0.7]^\top$ is the unknown parameter, the threshold $C = 0$, and $d_k \sim N(0, 1)$ is i.i.d. noise. The input $\phi_k = [u_k, u_{k-1}, u_{k-2}]^\top$ and

$$u_k = \begin{cases} -0.4 + e_k, & k \mod 4 = 0; \\ 0.4 + e_k, & \text{else}, \end{cases} \tag{41}$$

where $e_k = 0.01\sin(k)$.

It can be verified that the input satisfies Assumption 2. By setting $N = 4$, we have $\delta = 0.153$, $M = 0.705$, $\|\theta\| = \sqrt{0.78}$, $b = 1.246$, $B = 0.235$ and $\gamma = 0.111$. Choose the step size $\beta = 30$ to satisfy $\beta > 1/2\gamma\delta$, $k_0 = 30$ and the initial value $\hat{\theta}_{k_0} = [0, 0, 0]^\top$ for subsequent simulations.

*Remark 6:* A larger $\beta$ will result in larger step sizes in the very first few steps of the algorithm (8), which may cause the estimated value $\hat{\theta}_k$ to deviate from the true value $\theta$ after the first few iterations. It will require more time to reduce estimation error. To avoid excessive initial errors caused by large $\beta$, we start the algorithm at $k_0 = \beta$, which also applies to other selected $\beta$ in Fig. 3.

Fig. 2 shows that the estimates $\hat{\theta}_k$ given by the proposed algorithm (8) converges to the true parameter $[0.2, -0.5, 0.7]^\top$, which is consistent with Theorem 1.

Fig. 3 depicts the average trajectories of $k\|\tilde{\theta}_k\|^2$ from 200 repeated experiments to estimate the empirical variance of $\tilde{\theta}_k$ with $\hat{\theta}_{k_0} = [0, 0, 0]^\top$ and different step sizes $\beta$. It demonstrates that the proposed recursive algorithm achieves the mean square convergence rate of $O(1/k)$ when $\beta = 30$, as stated in Theorem 2. Besides, it can also be seen from Theorem 2 that the step size condition is a sufficient but unnecessary condition. When $2\beta\gamma\delta > 1$, the mean square convergence rate can definitely achieve $O(1/k)$. However, when $\beta$ is small, such as when $\beta = 2$ or 4 in Fig. 3, it cannot guarantee a convergence rate of $O(1/k)$. Fig. 3 validates the conclusions of Theorem 2.

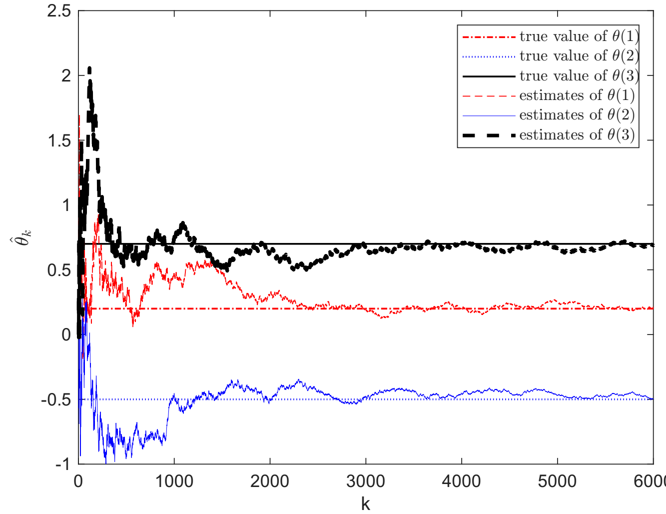


Fig. 2. Convergence of the proposed algorithm.

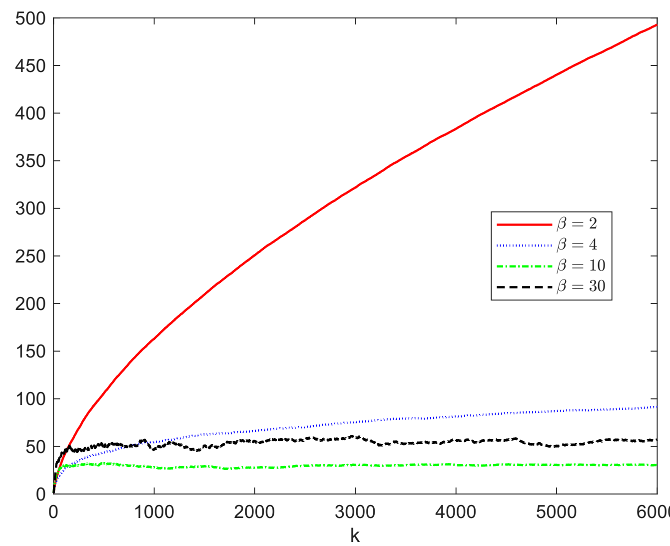


Fig. 3. Average trajectories of $k\|\tilde{\theta}_k\|^2$ from 200 experiments with $\hat{\theta}_{k_0} = [0, 0, 0]^\top$ and different $\beta$.

In Fig. 4, a bounded trajectory of $k\|\tilde{\theta}_k\|^2/\ln k$ is presented to illustrate that the almost sure convergence rate of the proposed algorithm (8) is $O\left(\sqrt{\ln k/k}\right)$, as established in Theorem 3.

### B. Comparison with Existing Algorithm

This simulation aims to compare the performance of the existing algorithm (5) and the proposed algorithm (8), to verify the claim of Section IV-B.

We maintain the same parameters as those in Subsection V-A, with $\theta = [0.2, -0.5, 0.7]^\top$, $C = 0$, $\phi_k = [u_k, u_{k-1}, u_{k-2}]^\top$, $u_k$ is defined as (41), and i.i.d. noise

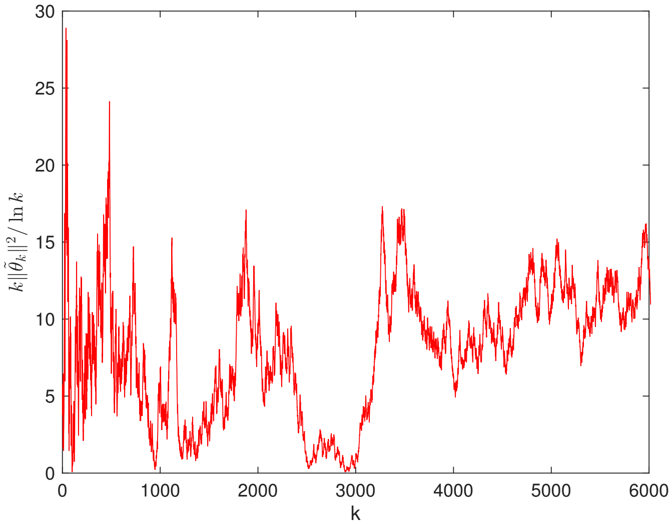


Fig. 4. Trajectory of $\boldsymbol{k}\|\tilde{\boldsymbol{\theta}}_{\boldsymbol{k}}\|^2/\ln \boldsymbol{k}$ with $\hat{\boldsymbol{\theta}}_{\boldsymbol{k_0}} = [\mathbf{0},\mathbf{0},\mathbf{0}]^\top$.

$d_k \sim N(0,1)$. Let the step size $\beta = 30$ and start from $k_0 = 30$ in both algorithms.

Starting with a initial estimate $\hat{\theta}_{k_0} = [-10,-10,-10]^\top$, Fig. 5 shows the average trajectories of $\|\tilde{\theta}_k\|^2$ from 200 repeated experiments for algorithms (5) and (8) under identical scenario. While both algorithms are convergent, the proposed algorithm (8) achieves markedly accelerated convergence during initial stage when parameter estimates contain large errors, compared to algorithm (5). More clearly, it can be seen from Fig. 6, which plots the average trajectories of $k\|\tilde{\theta}_k\|^2$ from 200 repeated experiments. The boundedness of both trajectories confirms that their mean square convergence rate can reach $O(1/k)$. Crucially, the proposed algorithm (8) exhibits better transient estimation performance and has an accelerating effect when estimates deviate far from the true value, thus validating the claim in Section IV-B.

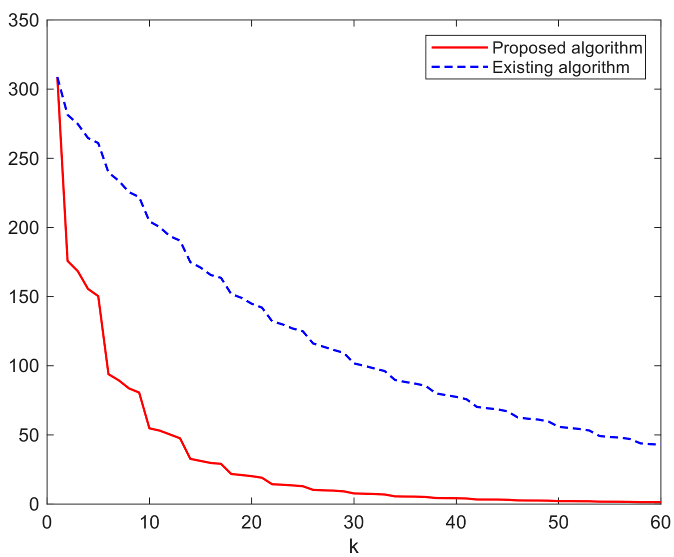


Fig. 5. Average trajectories of $\|\tilde{\boldsymbol{\theta}}_{\boldsymbol{k}}\|^2$ from 200 experiments with $\hat{\boldsymbol{\theta}}_{\boldsymbol{k_0}} = [-\mathbf{10},-\mathbf{10},-\mathbf{10}]^\top$.

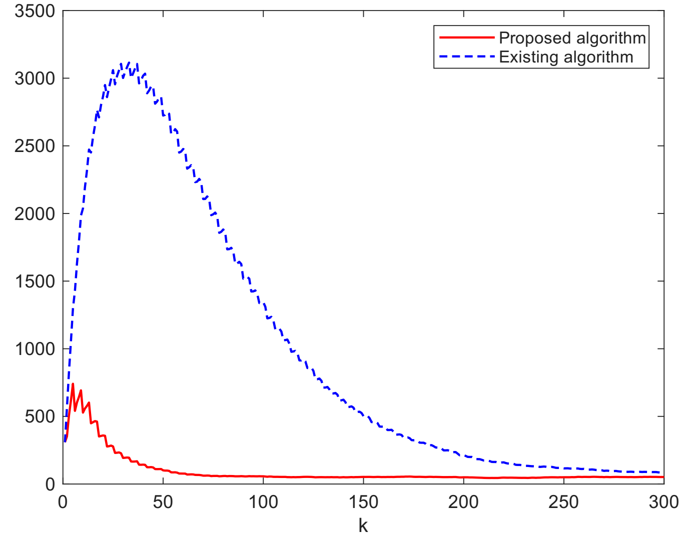


Fig. 6. Average trajectories of $\boldsymbol{k}\|\tilde{\boldsymbol{\theta}}_{\boldsymbol{k}}\|^2$ from 200 experiments with $\hat{\boldsymbol{\theta}}_{\boldsymbol{k_0}} = [-\mathbf{10},-\mathbf{10},-\mathbf{10}]^\top$.

## VI. Concluding Remarks

This paper studies the recursive identification of FIR systems with binary-valued observations under fixed threshold. A novel recursive algorithm is proposed using the statistical property of system noises and observations. The gradient of the proposed algorithm is derived from the local likelihood function, which has not been previously considered. Compared to the existing sign-error type algorithm proposed in [33] based on time-varying thresholds, a weighting approach on the binary-valued observations in the proposed algorithm makes the sign-error type algorithm applicable to fixed threshold scenarios, thus avoiding the complexity caused by time-varying thresholds for quantizers.

The proposed algorithm is proved to be convergent in both almost sure and mean square sense under bounded persistent excitations. Furthermore, the almost sure and mean square convergence rates are also established, achieving $O(\sqrt{\ln k/k})$ and $O(1/k)$, respectively. Compared to the existing recursive algorithm in [31] for the identification with binary-valued observations under fixed threshold, the main advantage of the proposed algorithm in this paper is that the adaptive recursive weight has an accelerating effect when the estimated value deviates far from the true value. Two simulations are conducted to demonstrate the effectiveness of the algorithm and advantage of convergence rate over existing algorithm.

There are three topics for future research. Firstly, relaxing the input assumptions is a critical consideration. Simplifying these requirements could make the algorithm more accessible for control-oriented problems. Secondly, the design of step size $a_k$ in Remark 2 is worth further exploring. Two noteworthy papers [41], [42] provide valuable insights on designing the step size for stochastic approximation method. Thirdly, it is also possible to consider extending to other more complex models, such as quantized identification for ARMA models [43].

# Appendix I
# The proofs for Lemmas

## A. Proof of Lemma 2

*Proof:*

i. The first is clearly true because

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right),$$

and

$$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t.$$

ii. When $x>0$, we have

$$F(x)=\frac{1}{2}+\int_0^x \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{t^2}{2\sigma^2}\right)\mathrm{d}t,$$

then

$$\begin{aligned}&F(x)[1-F(x)]\\=&\left(\frac{1}{2}+\int_0^x \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{t^2}{2\sigma^2}\right)\mathrm{d}t\right)\times\\&\left(\frac{1}{2}-\int_0^x \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{t^2}{2\sigma^2}\right)\mathrm{d}t\right)\\=&\frac{1}{4}-\frac{1}{2\pi\sigma^2}\int_0^x\int_0^x \exp\left(-\frac{t_1^2+t_2^2}{2\sigma^2}\right)\mathrm{d}t_1\mathrm{d}t_2.\end{aligned}$$

Let $t_1=r\cos\alpha, t_2=r\sin\alpha$. Then, we get

$$\begin{aligned}&F(x)[1-F(x)]\\\leq&\frac{1}{4}-\frac{1}{2\pi\sigma^2}\int_0^x\int_0^{\frac{\pi}{2}}\exp\left(-\frac{r^2}{2\sigma^2}\right)r\mathrm{d}\alpha\mathrm{d}r\\=&\frac{1}{4}-\frac{1}{4\sigma^2}\int_0^x\exp\left(-\frac{r^2}{2\sigma^2}\right)r\mathrm{d}r\\=&\frac{1}{4}\exp\left(-\frac{x^2}{2\sigma^2}\right),\end{aligned}$$

and then

$$\frac{f(x)}{F(x)\left[1-F(x)\right]}\geq\frac{4}{\sqrt{2\pi}\sigma}.$$

iii. By Corollary 1, $p(x)\leq\frac{\sqrt{2}}{\sqrt{\pi}\sigma}+\frac{|x|}{\sigma^2}$. When $x>0$,

$$\frac{f(x)}{F(x)\left[1-F(x)\right]}=\frac{p(-x)}{F(x)}\leq\frac{1}{F(x)}\left(\frac{1}{\sigma}\sqrt{\frac{2}{\pi}}+\frac{x}{\sigma^2}\right).$$

Therefore, if

$$\frac{1}{F(x)}\left(\frac{1}{\sigma}\sqrt{\frac{2}{\pi}}+\frac{x}{\sigma^2}\right)<\frac{2}{\sigma}\sqrt{\frac{2}{\pi}}+\frac{x}{\sigma^2}\tag{42}$$

holds, then $q(x)<\frac{2}{\sigma}\sqrt{\frac{2}{\pi}}+\frac{x}{\sigma^2}$ holds.

Let

$$h_1(x)=(F(x)-1)\frac{x}{\sigma^2}+(2F(x)-1)\frac{1}{\sigma}\sqrt{\frac{2}{\pi}},$$

then its derivative function is

$$h_1'(x)=f(x)\frac{x}{\sigma^2}+\frac{F(x)-1}{\sigma^2}+\frac{2f(x)}{\sigma}\sqrt{\frac{2}{\pi}},$$

where $h_1'(0)=\frac{2}{\pi\sigma^2}-\frac{1}{2\sigma^2}>0$. To obtain the properties of $h_1'(x)$, take its derivative, yielding

$$\begin{aligned}h_1''(x)=&-\frac{x^2}{\sigma^4}f(x)+\frac{2f(x)}{\sigma^2}-\frac{2x}{\sigma^3}f(x)\sqrt{\frac{2}{\pi}}\\=&f(x)\left(-\frac{x^2}{\sigma^4}-\frac{2x}{\sigma^3}\sqrt{\frac{2}{\pi}}+\frac{2}{\sigma^2}\right).\end{aligned}$$

By analysing $h_1''(x)$, we can get that the function $h_1'(x)$ increases on the interval $\left[0,\sigma\left(\sqrt{\frac{2}{\pi}+2}-\sqrt{\frac{2}{\pi}}\right)\right]$, and decreases on the interval $\left[\sigma\left(\sqrt{\frac{2}{\pi}+2}-\sqrt{\frac{2}{\pi}}\right),+\infty\right)$. Since $\lim_{x\to+\infty}h_1'(x)=0$, $h_1'(x)>0$ always holds in the interval of $(0,+\infty)$. Due to $h_1(0)=0$, $h_1(x)>0$ always holds in the interval of $(0,+\infty)$. Hence, (42) holds, and then $q(x)<\frac{2}{\sigma}\sqrt{\frac{2}{\pi}}+\frac{x}{\sigma^2}$ holds.

iv. When $x>0$, for the function $h_2(x)=\sigma^2 f(x)-xF(-x)$, we have

$$h_2(0)=\frac{\sigma}{\sqrt{2\pi}}>0,$$

and

$$\lim_{x\to+\infty}h_2(x)=\lim_{x\to+\infty}-\frac{F(-x)}{\frac{1}{x}}=\lim_{x\to+\infty}\frac{f(x)}{-\frac{1}{x^2}}=0.$$

Additionally,

$$\frac{\mathrm{d}h_2(x)}{\mathrm{d}x}=-xf(x)-F(-x)+xf(x)=-F(-x)\leq 0.$$

Therefore, when $x>0$, $\sigma^2 f(x)\geq xF(-x)$ and

$$1-F(x)\leq\frac{\sigma^2 f(x)}{x},$$

which is part of Mills' Ratio inequality [44].
Hence, we obtain

$$\frac{f(x)}{F(x)\left[1-F(x)\right]}\geq\frac{x}{\sigma^2}, \text{when } x>0.$$

The proof is completed. ■

## B. Proof of Lemma 3

*Proof:* Using $p(x)\leq\frac{\sqrt{2}}{\sqrt{\pi}\sigma}+\frac{|x|}{\sigma^2}$ of Corollary 1, we have

$$\begin{aligned}&\|\tilde{\theta}_k-\tilde{\theta}_{k-1}\|\\\leq&\frac{\beta M}{k}p\left(s_k(\phi_k^\top\hat{\theta}_{k-1}-C)\right)\\\leq&\frac{\beta M}{k}\left(\frac{\sqrt{2}}{\sqrt{\pi}\sigma}+\frac{|(\phi_k^\top\tilde{\theta}_{k-1}-C+\phi_k^\top\theta)|}{\sigma^2}\right)\\\leq&\frac{\beta M}{k}\left(\frac{|\phi_k^\top\tilde{\theta}_{k-1}|}{\sigma^2}+\frac{\sqrt{2}}{\sqrt{\pi}\sigma}+\frac{M\|\theta\|+|C|}{\sigma^2}\right)\\\leq&\frac{\beta M|\phi_k^\top\tilde{\theta}_{k-1}|}{k\sigma^2}+O\left(\frac{1}{k}\right),\end{aligned}$$

and further

$$\|\tilde{\theta}_k - \tilde{\theta}_{k-1}\|^2 \leq \frac{2\beta^2 M^2 |\phi_k^\top \tilde{\theta}_{k-1}|^2}{k^2\sigma^4} + O\left(\frac{1}{k^2}\right).$$

The proof is completed. ■

### C. Proof of Lemma 4

*Proof:* For the function $h_3(x) = F(x) - F(x+\alpha)$, it is easy to get $\frac{\mathrm{d}}{\mathrm{d}x}h_3(x) = f(x) - f(x+\alpha)$. Then, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}h_3(x) < 0, x \in (-\infty, -\frac{\alpha}{2}),$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x}h_3(x) > 0, x \in (-\frac{\alpha}{2}, +\infty).$$

When $\alpha > b$, $h_3(x)$ strictly increases on the interval $\left[-\frac{b}{2}, \frac{b}{2}\right] \subseteq \left(-\frac{\alpha}{2}, +\infty\right)$, and

$$h_3(x) \leq F\left(\frac{b}{2}\right) - F\left(\frac{b}{2}+\alpha\right) \leq -\left(F\left(\frac{3b}{2}\right) - F\left(\frac{b}{2}\right)\right).$$

Similarly, when $\alpha < -b$, $h_3(x)$ strictly decreases on the interval $\left[-\frac{b}{2}, \frac{b}{2}\right] \subseteq \left(-\infty, -\frac{\alpha}{2}\right)$, and

$$h_3(x) \geq F\left(\frac{b}{2}\right) - F\left(\frac{b}{2}+\alpha\right) \geq F\left(\frac{b}{2}\right) - F\left(-\frac{b}{2}\right).$$

Furthermore, when $x \in (0, +\infty)$, due to $3f(3x) - 3f(x) < 0$, which is the derivative of $F(3x) - 2F(x) + F(-x)$, we have $F(3x) - 2F(x) + F(-x) < 0$ and

$$F\left(\frac{3b}{2}\right) - F\left(\frac{b}{2}\right) < F\left(\frac{b}{2}\right) - F\left(-\frac{b}{2}\right).$$

Therefore, taking $B = F\left(\frac{3b}{2}\right) - F\left(\frac{b}{2}\right)$ completes the proof. ■

## References


[1] L. Y. Wang, J. F. Zhang, and G. G. Yin, "System identification using binary sensors," *IEEE Trans. Automa. Control*, vol. 48, no. 11, pp. 1892–1907, 2003.

[2] Y. L. Zhao, L. Y. Wang, G. G. Yin, and J. F. Zhang, "Identification of Wiener systems with binary-valued output observations," *Automatica*, vol. 43, no. 10, pp. 1752-1765, 2007.

[3] L. Y. Wang, G. G. Yin, J. F. Zhang, and Y. L. Zhao, *System identification with quantized observations*, Boston, MA, USA: Birkhäuser, 2010.

[4] W. Bi et al., "SVSI: Fast and powerful set-valued system identification approach to identifying rare variants in sequencing studies for ordered categorical traits," *Ann. Hum. Genet.*, vol. 79, no. 4, pp. 294-309, 2015.

[5] J. P. Bradfield et al., "A genome-wide association meta-analysis identifies new childhood obesity loci," *Nat. Genet.*, vol. 44, pp. 526–531, 2012.

[6] W. Bi, W. Zhou, R. Dey, B. Mukherjee, J. N. Sampson, and S. Lee, "Efficient mixed model approach for large-scale genome-wide association studies of ordinal categorical phenotypes," *Am. J. Hum. Genet.*, vol. 108, no. 5, pp. 825-839, 2021.

[7] T. Wang, W. Bi, Y. L. Zhao, and W. C. Xue, "Radar target recognition algorithm based on RCS observation sequence – set-valued identification method," *J. Syst. Sci. Complex.*, vol. 29, no. 3, pp. 573-588, 2016.

[8] T. M. Cover, *Elements of information theory*, John Wiley & Sons, 1999.

[9] J. He, E. H. Yang, F. Z. Yang, and K. H. Yang, "Adaptive quantization parameter selection for H.265/HEVC by employing inter-frame dependency," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 12, pp. 3424-3436, 2018.

[10] L. Y. Wang, Y. W. Kim, and J. Sun, "Prediction of oxygen storage capacity and stored NOx by HEGO sensors for improved LNT control strategies," In *ASME International Mechanical Engineering Congress and Exposition*, 2002, pp. 777-785.

[11] G. Gagliardi, D. Mari, F. Tedesco, and A. Casavola, "An air-to-fuel ratio estimation strategy for turbocharged spark-ignition engines based on sparse binary HEGO sensor measures and hybrid linear observers," *Control Engineering Practice*, vol. 107, 2021, Art. no. 104694.

[12] R. Alejandro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case," *IEEE Trans. Signal Process*, vol. 54, no. 3, pp. 1131-1143, 2006.

[13] R. Alejandro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part II: unknown probability density function," *IEEE Trans. Signal Process*, vol. 54, no. 7, pp. 2784-2796, 2006.

[14] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 978-1002, 2001.

[15] J. J. Xiao and Z. Q. Luo, "Universal decentralized detection in a bandwidth-constrained sensor network," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2617-2624, 2005.

[16] Z. Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210-2219, 2005.

[17] L. Y. Wang and G. G. Yin, "Asymptotically efficient parameter estimation using quantized output observations," *Automatica*, vol. 43, no. 7, pp. 1178-1191, 2007.

[18] E. Colinet and J. Juillard, "A weighted least-squares approach to parameter estimation problems based on binary measurements," *IEEE Trans. Automa. Control*, vol. 55, no. 1, pp. 148-152, 2010.

[19] Y. L. Zhao, W. Bi, and T. Wang, "Iterative parameter estimate with batched binary-valued observations," *Sci. China Inf. Sci.*, vol. 59, no. 5, 2016, Art. no. 052201.

[20] X. Li, Z. Xu, J. Cui, and L. Zhang, "Suboptimal adaptive tracking control for FIR systems with binary-valued observations," *Sci. China Inf. Sci.*, vol. 64, 2021, Art. no. 172202.

[21] B. I. Godoy, G. C. Goodwin, J. C. Agüero, D. Marelli, and T. Wigren, "On identification of FIR systems having quantized output data," *Automatica*, vol. 47, no.9, pp. 1905-1915, 2011.

[22] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005-2018, 2016.

[23] H. Mei and G. Yin, "Almost sure convergence rates for system identification using binary, quantized, and regular sensors," *Automatica*, vol. 50, no. 8, pp. 2120-2127, 2014.

[24] L. Ljung and T. Söderström, *Theory and practice of recursive identification*, Cambridge, MA: MIT Press, 1983.

[25] K. Jafari, J. Juillard, and E. Colinet, "A recursive system identification method based on binary measurements," in *Proceeding of 49th IEEE Conference on Decision and Control*, 2010, pp. 1154-1158.

[26] K. Jafari, J. Juillard, and M. Roger, "Convergence analysis of an online approach to parameter estimation problems based on binary observations," *Automatica*, vol. 48, no. 11, pp. 2837-2842, 2012.

[27] M. Pouliquen, T. Menard, E. Pigeon, O. Gehan, and A. Goudjil, "Recursive system identification algorithm using binary measurements," in *Proceeding of 2016 European Control Conference*, 2016, pp. 1353–1358.

[28] J. Guo and Y. L. Zhao, "Recursive projection algorithm on FIR system identification with binary-valued observations," *Automatica*, vol. 49, no. 11, pp. 3396-3401, 2013.

[29] T. Wang, M. Hu, and Y. L. Zhao, "Convergence properties of recursive projection algorithm for system identification with binary-valued observations," in *Proceedings of 2018 Chinese Automation Congress*, 2018, pp. 2961-2966.

[30] H. Zhang, T. Wang, and Y. L. Zhao, "Asymptotically efficient recursive identification of FIR systems with binary-valued observations," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 5, pp. 2687-2700, 2021.

[31] J. Ke, Y. Wang, Y. L. Zhao, and J. F. Zhang, "Recursive identification of binary-valued systems under uniform persistent excitations," *IEEE Trans. Automa. Control*, vol. 69, no. 12, pp. 8204-8219, 2024.

[32] J. Guo and Y. L. Zhao, "Identification of the gain system with quantized observations and bounded persistent excitations," *Sci. China Inf. Sci.*, vol. 57, pp. 1-15, 2014.

[33] Y. Wang, Y. L. Zhao, J. F. Zhang, and J. Guo, "A unified identification algorithm of FIR systems based on binary observations with time-varying thresholds," *Automatica*, vol. 135, 2022, Art. no. 109990.

[34] K. You, "Recursive algorithms for parameter estimation with adaptive quantizer," *Automatica*, vol. 52, pp.192–201, 2015.

[35] W. Zhao, H. F. Chen, R. Tempo and F. Dabbene, "Recursive nonparametric Identification of nonlinear systems with adaptive binary sensors," *IEEE Trans. Automa. Control*, vol. 62, no. 8, pp. 3959-3971, 2017.

[36] Y. Wang, Y. L. Zhao, and J. F. Zhang, "Distributed recursive projection identification with binary-valued observations," *J. Syst. Sci. Complex.*, vol. 34, no. 5, pp. 2048-2068, 2021.

[37] H. F. Chen, *Stochastic approximation and its application*. Dordrecht: Kluwer Academic Publishers, 2002.

[38] L. Ljung, *System Identification: Theory for the User (2nd ed.)*, Prentice Hall, 1999.

[39] L. Guo, *Time-Varying Stochastic Systems: Stability and Adaptive Theory*. Beijing: Science Press, 2020.

[40] V. A. Zorich, *Mathematical Analysis I*, Berlin, Germany: Universitext, Springer-Verlag, 2016.

[41] D. Shen and J. X. Xu, "An iterative learning control algorithm with gain adaptation for stochastic systems," *IEEE Trans. Automa. Control*, vol. 65, no. 3, pp. 1280-1287, 2020.

[42] D. Shen, N. Huo, and S. S. Saab, "A probabilistically quantized learning control framework for networked linear systems," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 12, pp. 7559-7573, 2022.

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

[44] R. D. Gordon, "Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument," *The Annals of Mathematical Statistics*, vol. 12, no. 3, pp. 364-366, 1941.


**Xin Li** received the B.S. degree in mathematics from Shandong University, Jinan, China, in 2020, and the Ph.D. degree from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2025. He is currently a Postdoctoral Fellow with the AMSS, CAS.

His research interests include the identification and control of quantized systems, information theory.

**Mingjie Shao** received the B.S. degree from the Xidian University, Xi'an, China, in 2015 and Ph.D. degree from the Chinese University of Hong Kong (CUHK) in 2020. He was a Postdoctoral Fellow with the Department of Electronic Engineering, CUHK from 2020 to 2023. He was a Professor (Qilu Young Scholar) in the School of Information Science and Engineering from 2023 to 2024, Shandong University, Qingdao, China. He is now an Associate Professor in the Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS). He was the recipient of the Hong Kong PhD Fellowship Scheme (HKPFS) from August 2015. He was listed in the Student Best Paper Finalists in ICASSP 2017.

His research interests focus on signal processing, convex and nonconvex optimization, and machine learning for wireless communication.

**Ji-Feng Zhang** received the B.S. degree in mathematics from Shandong University, China, in 1985 and the Ph.D. degree from the Institute of Systems Science (ISS), Chinese Academy of Sciences (CAS), China, in 1991. Since 1985, he has been with the ISS, CAS. Now he is also with the School of Automation and Electrical Engineering, Zhongyuan University of Technology. His current research interests include system modeling, adaptive control, stochastic systems, and multi-agent systems.

He is an IEEE Fellow, IFAC Fellow, CAA Fellow, CSIAM Fellow, member of the European Academy of Sciences and Arts, and Academician of the International Academy for Systems and Cybernetic Sciences. He received the second prize of the State Natural Science Award of China in 2010 and 2015, respectively. He was a Vice-Chair of the IFAC Technical Board, member of the Board of Governors, IEEE Control Systems Society; Convenor of Systems Science Discipline, Academic Degree Committee of the State Council of China; Vice-President of the Systems Engineering Society of China, the Chinese Mathematical Society, and the Chinese Association of Automation. He has served as Editor-in-Chief, Deputy Editor-in-Chief, Senior Editor or Associate Editor for more than 10 journals, including *Science China Information Sciences*, *National Science Review*, *IEEE Transactions on Automatic Control*, and *SIAM Journal on Control and Optimization* etc.

**Yanlong Zhao** received the B.S. degree in mathematics from Shandong University, Jinan, China, in 2002, and the Ph.D. degree in systems theory from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2007. Since 2007, he has been with the AMSS, CAS, where he is currently a Professor. His research interests include identification and control of quantized systems, and modeling of communication systems, etc.

He served as a Vice-President of IEEE CSS Beijing Chapter and a Vice-President of Asian Control Association, and is now a Vice President of Chinese Association of Automation (CAA) and the Chair of Technical Committee on Control Theory (TCCT), CAA. He has been a Deputy Editor-in-Chief of the Journal of Systems Science and Complexity, an Associate Editor of Automatica, SIAM Journal on Control and Optimization, and IEEE Transactions on Systems, Man and Cybernetics: Systems.