Recursive Identification of Binary-Valued Systems Under Uniform Persistent Excitations

Jieming Ke, *Student Member, IEEE*, Ying Wang, *Member, IEEE*, Yanlong Zhao, *Senior Member, IEEE*, and Ji-Feng Zhang, *Fellow, IEEE*

***Abstract*—This article investigates the online identification problem of binary-valued moving average systems. A stochastic approximation-based algorithm without projections or truncations is proposed. To analyze the convergence property of the algorithm, the distribution tail of the parameter estimate is proved to be exponentially convergent through an auxiliary stochastic process. Under uniform persistent excitations, the almost sure and mean square convergence of the algorithm is obtained. When the step-size coefficient is properly selected, the almost sure and mean square convergence rates are proved to reach $O(\sqrt{\ln\ln k/k})$ and $O(1/k)$, respectively, where k is the sample size. A numerical example is given to demonstrate the effectiveness of the proposed algorithm and theoretical results.**

***Index Terms*—Binary-valued systems, recursive identification, stochastic approximation (SA), stochastic systems, uniform persistent excitations.**

I. Introduction

BINARY-VALUED systems emerge widely in practice. For example, in automotive and chemical process applications, oxygen sensors are used for evaluating gas oxygen contents [1], [2], [3]. Inexpensive oxygen sensors are switching types that change their voltage outputs sharply when excess oxygen in the gas is detected. More examples can be seen in genetic association studies [4], [5], radar target recognition [6], and credit scoring [7]. The appearance of the above binary-valued sensors brings forward new requirements for identification theory, which is the focus of this article. There are some important identification algorithms proposed for binary-valued and finite-valued systems [8], [9], [10], [11], [12], [13], [14], many of which are offline. Offline methods take full advantage of the statistical property of the finite-valued outputs, and require fewer assumptions than the online ones. However, in some scenarios, for instance, in adaptive control problems, online identification is of great importance, since online identification methods need less memory and computation complexity, and can update the parameter estimate quickly [15]. The online identification of binary-valued and finite-valued systems has been investigated under different type inputs [2], [16], [17], [18], [19], [20], [21], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34]. For example, the authors in [16], [17], and [18] assume the inputs to be independent and identically distributed (i.i.d.), and propose stochastic approximation (SA) algorithms with expanding truncations for binary-valued systems. You [19] required i.i.d. inputs, and gave a stochastic gradient-based algorithm. These algorithms are all proved to be convergent almost surely. Besides, the authors in [2], [20], [21], [22], [23], [24], [25], [26], [27], [28], and [29] consider periodic inputs, and propose empirical measurement methods. The methods can be applied in infinite impulse response systems and Hammerstein systems with binary-valued observations [23], [28]. The authors in [30] and [31] consider uniformly persistently exciting inputs, and design sign-error type identification algorithms. The authors in [32], [33], and [34] assume the inputs to be persistently exciting, and propose recursive projection methods.

There are two types of sensors considered in the finite-valued system identification problems. One type sensors are adaptive ones, whose thresholds can be adjusted according to historical data [16], [30], [31]. In the adaptive sensor case, the system outputs provide richer information when the thresholds are properly selected. Another type sensors are fixed ones, whose thresholds are time-invariant [2], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [32], [33], [34]. Fixed sensors are more common in practical scenarios. A practical example of the fixed finite-valued sensors is the oxygen sensors in automotive and chemical process [1], [2], [3]. This article focuses on the binary-valued system identification problem under uniform persistent excitations and fixed binary-valued sensors. The problem has been studied in [32], [33], and [34], but these works require that the unknown parameter is located in a known compact set. They design projections according to the a priori information on parameter location to ensure the uniform boundedness of the identification algorithms.

Manuscript received 29 November 2022; revised 21 June 2023, 15 January 2024, and 8 April 2024; accepted 4 May 2024. Date of publication 13 May 2024; date of current version 5 December 2024. This work was supported in part by the National Key R&D Program of China under Grant 2018YFA0703800, in part by the National Natural Science Foundation of China under Grant 62025306, Grant 62303452, and Grant T2293770, in part by CAS Project for Young Scientists in Basic Research under Grant YSBR-008, and in part by China Postdoctoral Science Foundation under Grant 2022M720159. Recommended by Associate Editor X. Bombois. *(Corresponding author: Yanlong Zhao.)*

The authors are with the Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, Chinaand also with the School of Mathematics Sciences, University of Chinese Academy of Sciences, Beijing 100149, China (e-mail: kejieming@amss.ac.cn; wangying96@amss.ac.cn; ylzhao@amss.ac.cn; jif@iss.ac.cn).

Digital Object Identifier 10.1109/TAC.2024.3399968

We consider the case without any a priori information on the parameter location. In this case, the identification algorithm should have the ability to search unknown parameters in the whole space. Therefore, the projections in [32], [33], and [34] should be removed for the convergence properties, which cause the algorithm to lose the uniform boundedness. This makes it difficult to analyze the convergence properties of the algorithm. To overcome the difficulty, in the periodic input case, Zhao et al. [29] calculated the distribution tail of the parameter estimate, which is the probability that the parameter estimate exceeds a certain compact set. In the nonperiodic input case, the distribution of observation sequences does not maintain periodicity and therefore is more complex, which makes the distribution tail of the parameter estimate difficult to be calculated.

To solve the difficulty, this article constructs a stochastic process with averaged observations (SPAO), which builds a bridge between the average of the binary-valued observations and the algorithm. By SPAO, we can utilize the distribution tail of the observation average to estimate the distribution tail of the algorithm.

In this article, an SA-based algorithm without projections or truncations is proposed for the binary-valued moving average (MA) system identification problem. The main contributions of this article are as follows.

1) A new SA-based identification algorithm without projections is proposed for binary-valued MA systems. Using this algorithm, we can recursively obtain the unknown parameter under uniform persistent excitations without any a priori information on the parameter location. It is the first paper where such a property is derived in the fixed finite-level quantizer and nonperiodic deterministic input case.
2) The convergence properties of the SA-based identification algorithm are established. To be specific, the almost sure convergence and mean square convergence are induced through the exponential convergence of the estimation error distribution tail. Besides, when the step-size coefficient is properly selected, the almost sure convergence rate is proved to be $O(\sqrt{\ln\ln k/k})$, which is first achieved among online identification algorithms of stochastic binary-valued systems under nonperiodic inputs. Moreover, the mean square convergence rate is proved to be $O(1/k)$, which is the best mean square convergence rate in theory under binary-valued observations and even accurate ones.
3) A new constructive methodology is developed for the convergence analysis of binary-valued system identification algorithms. Specially, SPAO is constructed to reveal the connection between the average of the binary-valued observations and the convergence properties of the algorithm. Moreover, the methodology is also shown to be practical for a common class of recursive identification algorithms for binary-valued systems, such as the stochastic gradient-based algorithm [19] and the quasi-Newton type algorithms [33], [34].

The rest of this article is organized as follows. Section II formulates the identification problem. Section III proposes an SA-based identification algorithm of binary-valued systems. Section IV gives the convergence analysis. Section IV-A constructs an auxiliary stochastic process named SPAO and discusses its property. Based on SPAO, Section IV-B estimates the distribution tail of the estimation error, and gives the almost sure and mean square convergence. Sections IV-C and IV-D analyze the almost sure and mean square convergence rates, respectively. Section V simulates a numerical example to demonstrate the theoretical results. Finally, Section VI concludes this article.

Notations: In the rest of this article, $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{R}^n$ denote the sets of natural numbers, real numbers, and n-dimensional real vectors, respectively. $I_{\{\cdot\}}$ denotes the indicator function, whose value is 1 if its argument (a formula) is true, and 0, otherwise. $\|x\|$ is the Euclidean norm for vector x. I_n is an $n\times n$ identity matrix. $\lfloor x\rfloor$ is the largest integer that is smaller than or equal to $x\in\mathbb{R}$. The positive part of x is denoted as $x^+=\max\{x,0\}$. For square matrices $A_l,\ldots,A_k$, denote $\prod_{i=l}^{k}A_i=A_k\cdots A_l$ for $k\geq l$. Relations between two series a_k and b_k are defined as follows:

1) $a_k=O(b_k)$ if $a_k=c_kb_k$ for a bounded c_k;
2) $a_k=o(b_k)$ if $a_k=c_kb_k$ for a c_k that converges to 0.

II. Problem Formulation

Consider the MA system

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

where $\phi_k=\phi(u_k,u_{k-1},\ldots,u_{k-\bar{n}+1})\in\mathbb{R}^n$ is a regressed function of inputs u_k for some $\bar{n}>0,\theta\in\mathbb{R}^n$ is the unknown parameter, and d_k is the system noise, respectively. The unobserved system output y_k is measured by a binary-valued sensor with a fixed threshold C, which can be represented by an indicator function

$$s_k=I_{\{y_k\leq C\}}=\begin{cases}1, & y_k\leq C;\\ 0, & y_k>C.\end{cases} \tag{2}$$

Our goal is to identify the unknown parameter θ based on the regressed vector ϕ_k and the binary observation s_k.

Assumption 1: The sequence $\{\phi_k,k\geq 1\}$ is bounded, i.e.,

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

and there exist a positive integer $N\geq n$ and a real number $\delta>0$ such that

$$\frac{1}{N}\sum_{i=k}^{k+N-1}\phi_i\phi_i^\top\geq\delta I_n,\ k\geq 1. \tag{3}$$

Remark 1: The condition (3) is usually called "uniform persistent excitation condition" or "sufficiently rich condition" [32], [35]. Assumption 1 is common in the binary-valued system identification field [31], [32].

Assumption 2: The system noise sequence $\{d_k,k\geq 1\}$ is i.i.d. with zero mean and finite variance, whose distribution and density function are denoted as $F(\cdot)$ and $f(\cdot)$, respectively.

The distribution $F(\cdot)$ is Lipschitz continuous, and the density function $f(\cdot)$ satisfies

$$\inf_{x\in\mathfrak{X}} f(x) > 0 \tag{4}$$

for any bounded open set $\mathfrak{X}$.

For simplicity of notation, denote

$$F_k = F\left(C - \phi_k^\top \theta\right), \quad f_k = f\left(C - \phi_k^\top \theta\right).$$

Then, $\mathbb{E}s_k = \mathbb{P}\{y_k \le C\} = F_k$.

Remark 2: Gaussian noise, Laplacian noise, and t-distribution noise are all examples satisfying Assumption 2. Moreover, if (4) does not hold for the system noise, we can add a dither to the binary sensor [2]. Under Assumption 2, the density function $f(\cdot)$ is bounded because of the Lipschitz continuity of the distribution function $F(\cdot)$.

Remark 3: It will be a more general problem when the noise is Gaussian with an unknown variance σ^2. In this case, we can use the similar technique of [26] to transform the joint identification problem for θ and σ^2 into the identification problem for a new binary-valued system with known noise variance.

III. Identification Algorithm

This section will propose an SA-based algorithm for the MA system (1) with binary observation (2).

In the viewpoint of SA, the identification problem can be treated as the problem to find the roots of

$$\mu_k(\hat{\theta}) = F\left(C - \phi_k^\top \hat{\theta}\right) - F_k, \ k \ge 1.$$

Because θ is unknown, F_k is unavailable. Besides, s_k is available, and its expectation is F_k. We replace F_k with s_k. Then, based on the SA method [36], the identification algorithm is given as

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \rho_k \phi_k \left(F\left(C - \phi_k^\top \hat{\theta}_{k-1}\right) - s_k\right)$$

where $\rho_k \in \mathbb{R}$ is the step-size.

Remark 4: In the algorithm design, F_k can be replaced with s_k because $\{s_k - F_k\}$ is a martingale difference sequence with uniformly bounded variances. When the step-size is properly selected, martingale difference noises with uniformly bounded variances will not influence the convergence of SA-based algorithms [36].

Denote $\hat{F}_k = F(C - \phi_k^\top \hat{\theta}_{k-1})$. Set $\rho_k = \beta/k$, where $\beta > 0$ is a constant coefficient. Then, the SA-based algorithm is given as follows:

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \frac{\beta \phi_k}{k}\left(\hat{F}_k - s_k\right) \quad \forall k > k_0 \tag{5}$$

where $k_0 \in \mathbb{N}$ is the starting point, and the initial value $\hat{\theta}_{k_0}$ can be arbitrarily selected in $\mathbb{R}^n$.

The observation error $\hat{\theta}_k - \theta$ is denoted as $\tilde{\theta}_k$.

Remark 5: Algorithm (5) is similar to the recursive projection algorithm proposed in [32]. The main difference is that we do not introduce any projections or truncations in (5). The difference brings major difficulty in the convergence analysis. The convergence analysis of the recursive projection algorithm relies on the fact that if the search region is constrained in a compact set, then there is a uniform positive lower bound for $-(\hat{F}_k - F_k)/\phi_k^\top \tilde{\theta}_{k-1}$. Without any projection, Algorithm (5) can reach every point in the whole space. Then, the infimum of $-(\hat{F}_k - F_k)/\phi_k^\top \tilde{\theta}_{k-1}$ can be arbitrarily close to 0. To overcome the problem, we should investigate the distribution tail of the algorithm.

Remark 6: The step-size ρ_k that converges to 0 is used to reduce the effect of noise d_k [36]. In the SA method [36], ρ_k should satisfy $\sum_{i=1}^\infty \rho_i = \infty$ and $\sum_{i=1}^\infty \rho_i^2 < \infty$. One of the example is $\rho_k = \beta/k$ that is used in (5). Another example is $\rho_k = \beta/(1 + \sum_{i=1}^k \|\phi_i\|^2)$ that is used in [32].

Remark 7: In Algorithm (5), $\hat{F}_k$ is used to approximate s_k because $\hat{F}_k = \mathbb{E}[s_k(\theta)|\theta = \hat{\theta}_{k-1}]$. Therefore, in the multiple threshold case with threshold number q, Algorithm (5) also works after replacing $\hat{F}_k$ with $\mathbb{E}[s_k^q(\theta)|\theta = \hat{\theta}_{k-1}]$, where s_k^q is the corresponding observation in $\{0, 1, \ldots, q\}$.

IV. Convergence

This section will focus on the convergence analysis of the algorithm, including the distribution tail, almost sure convergence rate, and mean square convergence rate. An auxiliary stochastic process is introduced first to assist in the analysis.

A. Stochastic Process With Averaged Observations

This section will introduce an auxiliary stochastic process satisfying the following:

1) the trajectory of the stochastic process gradually approaches that of the estimation error $\tilde{\theta}_k$;
2) the convergence property of the stochastic process is easy to analyze compared with that of the algorithm.

The construction is inspired by the idea that $\beta\phi_k(F_k - s_k)$ can be replaced by the linear combination of w_k and w_{k-1}, where

$$w_k = \frac{\sum_{i=1}^k \beta \phi_i (F_i - s_i)}{k} \tag{6}$$

i.e.,

$$\begin{aligned}\beta\phi_k(F_k - s_k) &= \sum_{i=1}^k \beta\phi_i(F_i - s_i) - \sum_{i=1}^{k-1} \beta\phi_i(F_i - s_i)\\ &= k\left(w_k - w_{k-1}\right) + w_{k-1}.\end{aligned}$$

Define $\psi_k = \tilde{\theta}_k - w_k$. Then, by the transformation above

$$\begin{aligned}\psi_k = \psi_{k-1} &+ \frac{\beta\phi_k}{k}\left(F\left(C - \phi_k^\top\theta - \phi_k^\top\psi_{k-1} - \phi_k^\top w_{k-1}\right)\right.\\ &\left.- F\left(C - \phi_k^\top\theta\right)\right) + \frac{w_{k-1}}{k}.\end{aligned} \tag{7}$$

The above stochastic process is named as SPAO. With SPAO, the convergence property of the algorithm can be analyzed through that of w_k.

Remark 8: For general SA methods, w_k is also used to verify the robustness of the algorithm ([36], Assumption 2.7.3, Th. 2.7.1).

To analyze the properties of SPAO ψ_k, we should first estimate the distribution tail of w_k.

Lemma 1: Let w_k be defined in (6), and assume that the following holds:

1) $\phi_k \in \mathbb{R}^n$ is bounded;
2) $s_k \in \{0,1\}$ is a binary random variable with expectation F_k, and the sequence $\{s_k, k \geq 1\}$ is independent.

Then, for any $\varepsilon \in (0, \frac{1}{2})$, there exists $m > 0$ such that

$$\mathbb{P}\left\{\sup_{j\geq k} j^{\varepsilon}\|w_j\| > 1\right\} = O\left(\exp(-mk^{1-2\varepsilon})\right).$$

The proof is given in Appendix A.

Next, we give the following lemma to describe the distance between ψ_k and the estimation error $\tilde{\theta}_k$ in three different senses based on Lemma 1.

Lemma 2: Assume that the following holds:

1) system (1) with binary observation (2) satisfies Assumptions 1 and 2;
2) w_k is defined in (6), and $\psi_k = \tilde{\theta}_k - w_k$.

Then, we have the following:

1) for any $\varepsilon \in (0, \frac{1}{2})$, there exists $m > 0$ such that $\mathbb{P}\{\|\tilde{\theta}_k - \psi_k\| > k^{-\varepsilon}\} = O(\exp(-mk^{1-2\varepsilon}))$;
2) $\|\tilde{\theta}_k - \psi_k\| = O(\sqrt{\ln\ln k/k})$, a.s.;
3) $\mathbb{E}\|\tilde{\theta}_k - \psi_k\|^2 = O(1/k)$.

Proof: Since $\tilde{\theta}_k - \psi_k = w_k$, the three parts of the lemma can be obtained immediately from Lemma 1, the law of iterated logarithm ([37], Th. 10.2.1), and $\mathbb{E}\|w_k\|^2 = O(1/k)$, respectively. □

Then, by using Lemmas 1 and 2, the following theorem estimates the distribution tail of SPAO ψ_k.

Theorem 1: Under the condition of Lemma 2, for any $M' > 0$ and $\varepsilon \in (0, \frac{1}{2})$, when k is sufficiently large

$$\left\{\|\psi_k\|^2 < M'\right\} \supseteq \left\{\sup_{j\geq \lfloor k^{1-\varepsilon}\rfloor} j^{\varepsilon}\|w_j\| \leq 1\right\}.$$

Furthermore, there exists $m > 0$ such that

$$\mathbb{P}\left\{\|\psi_k\|^2 \geq M'\right\} = O\left(\exp\left(-mk^{(1-\varepsilon)(1-2\varepsilon)}\right)\right).$$

Proof: Set $k_s = \lfloor k^{1-\varepsilon}\rfloor$ and $k_s' = k - N\lfloor\frac{k-k_s}{N}\rfloor$. It is worth mentioning that $k_s' \in [k_s, k_s + N - 1]$, and $k - k_s'$ is divisible by N. Assume that $\sup_{j\geq k_s} j^{\varepsilon}\|w_j\| \leq 1$ is true in the rest of the proof. Then, it suffices to prove that $\|\psi_k\|^2 < M'$.

We first simplify the recursive formula of $\|\psi_k\|^2$. By (7) and the monotonicity and Lipschitz continuity of $F(\cdot)$, for any positive real number b, we have

$$\begin{aligned}
&\|\psi_k\|^2\\
&\leq \|\psi_{k-1}\|^2 + \frac{2\beta\phi_k^\top\psi_{k-1}}{k}\left(F(C - \phi_k^\top\theta - \phi_k^\top w_{k-1} - \phi_k^\top\psi_{k-1})\right.\\
&\quad\left. - F(C - \phi_k^\top\theta)\right) + \frac{2\psi_{k-1}^\top w_{k-1}}{k} + \frac{(\beta\|\phi_k\| + \|w_{k-1}\|)^2}{k^2}\\
&= \|\psi_{k-1}\|^2 + \frac{2\beta\phi_k^\top\psi_{k-1}}{k}\left(F(C - \phi_k^\top\theta - \phi_k^\top w_{k-1} - \phi_k^\top\psi_{k-1})\right.\\
&\quad\left. - F(C - \phi_k^\top\theta - \phi_k^\top w_{k-1})\right) + O\left(k^{-1-\varepsilon/2}\right)\\
&\leq \|\psi_{k-1}\|^2 + \frac{2\beta\phi_k^\top\psi_{k-1}}{k}\left(F(C - \phi_k^\top\theta - \phi_k^\top w_{k-1} - b)\right.\\
&\quad\left. - F(C - \phi_k^\top\theta - \phi_k^\top w_{k-1})\right) I_{\{\phi_k^\top\psi_{k-1}\geq b\}}\\
&\quad + \frac{2\beta\phi_k^\top\psi_{k-1}}{k}\left(F(C - \phi_k^\top\theta - \phi_k^\top w_{k-1} + b)\right.\\
&\quad\left. - F(C - \phi_k^\top\theta - \phi_k^\top w_{k-1})\right) I_{\{\phi_k^\top\psi_{k-1}\leq -b\}}\\
&\quad + O\left(k^{-1-\varepsilon/2}\right).
\end{aligned} \tag{8}$$

By Assumption 2 and the boundedness of $C - \phi_k^\top\theta - \phi_k^\top w_{k-1}$, there exists $B > 0$ such that

$$\begin{aligned}
&-2\beta\left(F(C - \phi_k^\top\theta - \phi_k^\top w_{k-1} - b)\right.\\
&\qquad\left. - F(C - \phi_k^\top\theta - \phi_k^\top w_{k-1})\right) > B\\
&2\beta\left(F(C - \phi_k^\top\theta - \phi_k^\top w_{k-1} + b)\right.\\
&\qquad\left. - F(C - \phi_k^\top\theta - \phi_k^\top w_{k-1})\right) > B
\end{aligned}$$

which together with (8) implies

$$\|\psi_k\|^2 \leq \|\psi_{k-1}\|^2 - \frac{B|\phi_k^\top\psi_{k-1}|}{k} I_{\{|\phi_k^\top\psi_{k-1}|\geq b\}} + O\left(k^{-1-\varepsilon/2}\right).$$

Set $b = \frac{\sqrt{\delta M'}}{2}$. Then, we have

$$\begin{aligned}
&\|\psi_{k+N}\|^2\\
&\leq \|\psi_k\|^2 - \sum_{i=k+1}^{k+N} \frac{B|\phi_i^\top\psi_{i-1}|}{i} I_{\left\{|\phi_i^\top\psi_{i-1}|\geq\frac{\sqrt{\delta M'}}{2}\right\}}\\
&\quad + \sum_{i=k+1}^{k+N} O\left(i^{-1-\varepsilon/2}\right).
\end{aligned} \tag{9}$$

By (6), (7), and Assumption 1

$$\begin{aligned}
\|\psi_k - \psi_{k-1}\| &\leq \frac{1}{k}\left(2\beta\|\phi_k\| + \|w_{k-1}\|\right)\\
&\leq \frac{1}{k}(2\beta M + 2\beta M) = \frac{4\beta M}{k}.
\end{aligned} \tag{10}$$

Note that $M' > 0$. Then, by Lemma A.2, when k is sufficiently large, there exists $k' \in [k+1, k+N]$ such that

$$|\phi_{k'}^\top\psi_{k'-1}| \geq \sqrt{\frac{\delta}{2}}\|\psi_k\| I_{\left\{\|\psi_k\|^2\geq\frac{M'}{2}\right\}}$$

which implies

$$\begin{aligned}
&\left\{|\phi_{k'}^\top\psi_{k'-1}| \geq \frac{\sqrt{\delta M'}}{2}\right\}\\
&\supseteq \left\{\sqrt{\frac{\delta}{2}}\|\psi_k\| \geq \frac{\sqrt{\delta M'}}{2}\right\} \cap \left\{\|\psi_k\|^2 \geq \frac{M'}{2}\right\}\\
&\supseteq \left\{\|\psi_k\|^2 \geq \frac{M'}{2}\right\}.
\end{aligned} \tag{11}$$

Then, by (9) and (11)

$$\|\psi_{k+N}\|^2$$

$$\leq \|\psi_k\|^2 - \frac{B|\phi_{k'}^\top \psi_{k'-1}|}{k+N} I_{\left\{|\phi_{k'}^\top \psi_{k'-1}| \geq \frac{\sqrt{\delta M'}}{2}\right\}} + O\left(k^{-1-\varepsilon/2}\right)$$

$$\leq \|\psi_k\|^2 - \frac{B\sqrt{\delta}}{\sqrt{2}} \frac{\|\psi_k\|}{k+N} I_{\left\{\|\psi_k\|^2 \geq \frac{M'}{2}\right\}} + O\left(k^{-1-\varepsilon/2}\right). \quad (12)$$

Hence, when $k = k_s' + N(t-1)$, we have

$$\begin{aligned}&\left\|\psi_{k_s'+Nt}\right\|^2 \\ &\leq \left\|\psi_{k_s'+N(t-1)}\right\|^2 - \frac{B\sqrt{\delta}}{\sqrt{2}} \frac{\left\|\psi_{k_s'+N(t-1)}\right\|}{k_s'+Nt} I_{\left\{\|\psi_{k_s'+Nt}\|^2 \geq \frac{M'}{2}\right\}} \\ &\quad + O\left((k_s'+Nt)^{-1-\varepsilon/2}\right). \quad (13)\end{aligned}$$

Since $\lim_{k\to\infty} k_s' = \infty$, we have

$$\lim_{k\to\infty} \sum_{t=1}^{\infty} (k_s'+Nt)^{-1-\varepsilon/2} = 0.$$

Then, by (13) and Lemma A.3 in Appendix A, when k is sufficiently large, we have

$$\|\psi_k\|^2 = \left\|\psi_{k_s'+N\lfloor \frac{k-\lfloor k^{1-\varepsilon}\rfloor}{N}\rfloor}\right\|^2 < \max\left\{M', \Delta_k^2 + \frac{M'}{2}\right\}$$

where

$$\Delta_k = \left(\left\|\psi_{k_s'}\right\| - \frac{B\sqrt{\delta}}{2\sqrt{2}N} \ln\left(\frac{\lfloor \frac{k-\lfloor k^{1-\varepsilon}\rfloor}{N}\rfloor + \frac{k_s'}{N} + 1}{\frac{k_s'}{N}+1}\right)\right)^+.$$

Note that $\ln(\frac{\lfloor (k-\lfloor k^{1-\varepsilon}\rfloor)/N\rfloor + k_s'/N+1}{k_s'/N+1})$ is of the same order as $\ln k$ since $k_s' = O(k^{1-\varepsilon})$. By Corollary A.2 in Appendix A, it holds that $\psi_{k_s'} = O(\sqrt{\ln k})$. Then, when k is sufficiently large, $\Delta_k = 0$ and $\|\psi_k\|^2 < M'$, which proves the theorem. □

Remark 9: The distribution tail estimation of SPAO ψ_k in Theorem 1 can be promoted to Theorem A.1 in Appendix A.

Remark 10: It is worth noting that the constructed SPAO can not only be adapted to Algorithm (5), but also can be extended to a class of identification algorithms of the binary-valued systems. The details are given in Appendix B.

B. Estimate of the Distribution Tail

In this section, the distribution tail of the estimation error will be estimated.

Theorem 2: If system (1) with binary observations (2) satisfies Assumptions 1 and 2, then for any $M' > 0$ and $\varepsilon > 0$, there exists $m > 0$ such that

$$\mathbb{P}\left\{\sup_{j\geq k} \|\tilde{\theta}_j\|^2 \geq M'\right\} = O\left(\exp(-mk^{1-\varepsilon})\right).$$

Proof: Reminding that $\tilde{\theta}_k = \psi_k + w_k$, by Theorem 1, for sufficiently large k, we have

$$\begin{aligned}&\left\{\sup_{j\geq\lfloor k^{1-\varepsilon}\rfloor} j^{\varepsilon}\|w_j\| \leq 1\right\} \\ &\subseteq \left\{\|\psi_k\|^2 < \frac{M'}{4}\right\} \cap \left\{\|w_k\|^2 \leq \frac{M'}{4}\right\} \subseteq \left\{\|\tilde{\theta}_k\|^2 < M'\right\}\end{aligned}$$

and hence

$$\begin{aligned}\left\{\sup_{j\geq k} \|\tilde{\theta}_j\|^2 \geq M'\right\} &\subseteq \bigcup_{j\geq k} \left\{\sup_{j_0\geq\lfloor j^{1-\varepsilon}\rfloor} j_0^{\varepsilon}\|w_{j_0}\| > 1\right\} \\ &= \left\{\sup_{j\geq\lfloor k^{1-\varepsilon}\rfloor} j^{\varepsilon}\|w_j\| > 1\right\}.\end{aligned}$$

So, by Lemma 1

$$\mathbb{P}\left\{\sup_{j\geq k} \|\tilde{\theta}_j\|^2 \geq M'\right\} = O\left(\exp(-mk^{(1-\varepsilon)(1-2\varepsilon)})\right).$$

Thus, the theorem can be proved by the arbitrariness of ε. □

Remark 11: Theorem 2 estimates the distribution tail of the estimation error $\tilde{\theta}_k$. For the convergence analysis of identification algorithms, the existing works are usually interested in the asymptotic properties of the estimation error distribution. For example, the asymptotic normality of $\rho_k^{-1/2}\tilde{\theta}_k$ is given for general SA algorithms under different conditions ([36], Sec. 3.3 and [38]). For the finite-valued system with i.i.d. inputs and designable quantizers, You [19] also analyzed the asymptotic normality of the algorithm. Compared with the asymptotic normality, Theorem 2 weakens the description of the estimate distribution in the neighborhood of θ, but gives a better description on the exponential tail of the estimation error. This helps to obtain the almost sure and mean square convergence of the algorithm.

Theorem 3: Under the condition of Theorem 2, Algorithm (5) converges to θ in both almost sure and mean square sense.

Proof: The almost sure convergence can be immediately obtained by Theorem 2.

By Theorem 2 and Corollary A.1 in Appendix A, for any $M' > 0$ and $\varepsilon > 0$, there exists $m > 0$ such that

$$\begin{aligned}\mathbb{E}\|\tilde{\theta}_k\|^2 &= \int_{\{\|\tilde{\theta}_k^2\| < M'\}} \|\tilde{\theta}_k\|^2 \mathrm{d}\mathbb{P} + \int_{\{\|\tilde{\theta}_k^2\| \geq M'\}} \|\tilde{\theta}_k\|^2 \mathrm{d}\mathbb{P} \\ &< M' + O\left(\ln k \cdot \exp(-mk^{1-\varepsilon})\right) = M' + o(1).\end{aligned}$$

Thus, the mean square convergence can be obtained by the arbitrariness of M'. ■

Remark 12: When the inputs are periodic, the mean square convergence of the empirical measurement method without truncation is also proved by the estimation of the distribution tail [29]. The distribution tail of the estimate is relatively easy to be obtained for the empirical measurement method, because there is a direct connection between the average of the binary-valued observations and the distribution tail. But, in the SA-based algorithm, the relationship is much more complicated. Therefore, SPAO is constructed to reveal the connection.

C. Almost Sure Convergence Rate

This section will estimate the almost sure convergence rate of the SA-based algorithm.

Before the analysis, we define

$$\underline{f}(x) = \sup_{z > M\|\theta\|+x} \inf_{t\in[C-z, C+z]} f(t) > 0 \quad \forall x \geq 0 \quad (14)$$

and

$$\underline{f} = \underline{f}(0). \tag{15}$$

The convergence rate of the algorithm depends on $\underline{f}$.

Remark 13: Under Assumption 2, $\underline{f}$ is the lower bound of $f(C-\phi_k^\top\theta)$ for all possible regressors ϕ_k. The following lemma gives properties of $\underline{f}(\cdot)$ and $\underline{f}$.

Lemma 3: Under Assumption 2, $\underline{f}(\cdot)$ and $\underline{f}$ have the following properties:

a) $\underline{f}(\cdot)$ is nonincreasing and right continuous;
b) $\lim_{x\to 0}\underline{f}(x)=\underline{f}$;
c) $\underline{f}(x)\le \inf_{t\in[C-M\|\theta\|-x,C+M\|\theta\|+x]} f(t)$;
d) if, in addition, $f(\cdot)$ is locally Lipschitz continuous, then so is $\underline{f}(\cdot)$.

The proof is given in Appendix A.

The almost sure convergence rate of the algorithm can be achieved through that of ψ_k.

Theorem 4: Under the condition of Theorem 2, for any $\varepsilon>0$, we have

$$\psi_k=\begin{cases}O\left(\sqrt{\frac{\ln\ln k}{k}}\right), & \eta>\frac{1}{2};\\ O\left(\frac{1}{k^{\eta-\varepsilon}}\right), & \eta\le\frac{1}{2},\end{cases}\quad \text{a.s.} \tag{16}$$

where $\eta=\beta\delta\underline{f}$ with $\underline{f}$ defined in (15). If $f(\cdot)$ is assumed to be locally Lipschitz continuous, then the almost sure convergence rate can be promoted into

$$\psi_k=\begin{cases}O\left(\sqrt{\frac{\ln\ln k}{k}}\right), & \eta>\frac{1}{2};\\ O\left(\ln k\sqrt{\frac{\ln\ln k}{k}}\right), & \eta=\frac{1}{2};\\ O\left(\frac{1}{k^{\eta}}\right), & \eta<\frac{1}{2},\end{cases}\quad \text{a.s.} \tag{17}$$

Proof: The proof is based on Lemma A.5 in Appendix A.

For the proof of (16), we first simplify the recursive formula of $\|\psi_k\|$. By Assumption 2, $F(\cdot)$ is Lipschitz continuous, which implies $\sup_{x\in\mathbb{R}} f(x)<\infty$. Then, in (7), by Lagrange mean value theorem ([39], Th. 5.3.1), there exists ξ_k between $C-\phi_k^\top\theta-\phi_k^\top\psi_{k-1}$ and $C-\phi_k^\top\theta$, and ξ_k' between $C-\phi_k^\top\theta-\phi_k^\top w_{k-1}-\phi_k^\top\psi_{k-1}$ and $C-\phi_k^\top\theta-\phi_k^\top\psi_{k-1}$ such that

$$\begin{aligned}&F(C-\phi_k^\top\theta-\phi_k^\top w_{k-1}-\phi_k^\top\psi_{k-1})-F(C-\phi_k^\top\theta)\\ &=F(C-\phi_k^\top\theta-\phi_k^\top w_{k-1}-\phi_k^\top\psi_{k-1})\\ &\quad-F(C-\phi_k^\top\theta-\phi_k^\top\psi_{k-1})\\ &\quad+F(C-\phi_k^\top\theta-\phi_k^\top\psi_{k-1})-F(C-\phi_k^\top\theta)\\ &=f(\xi_k')\phi_k^\top w_{k-1}+f(\xi_k)\phi_k^\top\psi_{k-1}\\ &=f(\xi_k)\phi_k^\top\psi_{k-1}+O(w_{k-1}).\end{aligned} \tag{18}$$

Then, by the law of iterated logarithm ([37], Th. 10.2.1)

$$\begin{aligned}&F(C-\phi_k^\top\theta-\phi_k^\top w_{k-1}-\phi_k^\top\psi_{k-1})-F_k\\ &=f(\xi_k)\phi_k^\top\psi_{k-1}+O\left(\sqrt{\frac{\ln\ln k}{k}}\right),\ \text{a.s.}\end{aligned}$$

which together with (7) implies

$$\psi_k=\left(I_n-\frac{\beta f(\xi_k)}{k}\phi_k\phi_k^\top\right)\psi_{k-1}+O\left(\sqrt{\frac{\ln\ln k}{k^3}}\right),\quad \text{a.s.}$$

Note that except for the first few steps, we have

$$\begin{aligned}&\left\|\prod_{i=k-N+1}^{k}\left(I_n-\frac{\beta f(\xi_i)}{i}\phi_i\phi_i^\top\right)\right\|\\ &\le\left\|I_n-\frac{\beta}{k}\sum_{i=k-N+1}^{k}f(\xi_i)\phi_i\phi_i^\top\right\|+O\left(\frac{1}{k^2}\right)\\ &\le\left\|I_n-\frac{\beta}{k}\underline{f}\left(\max_{k-N<i\le k}|\phi_i^\top\psi_{i-1}|\right)\sum_{i=k-N+1}^{k}\phi_i\phi_i^\top\right\|+O\left(\frac{1}{k^2}\right)\\ &\le\left(1-\frac{\beta\delta N}{k}\underline{f}\left(\max_{k-N<i\le k}|\phi_i^\top\psi_{i-1}|\right)\right)+O\left(\frac{1}{k^2}\right)\end{aligned} \tag{19}$$

where $\underline{f}(\cdot)$ is defined in (14). Denote

$$\underline{f}_{k|N}=\underline{f}\left(\max_{k-N<i\le k}|\phi_i^\top\psi_{i-1}|\right).$$

Then, we have

$$\|\psi_k\|\le\left(1-\frac{\beta\delta N}{k}\underline{f}_{k|N}\right)\|\psi_{k-N}\|+O\left(\sqrt{\frac{\ln\ln k}{k^3}}\right),\quad\text{a.s.} \tag{20}$$

By (b) of Lemma 3, $\lim_{k\to\infty}\underline{f}_{k|N}=\underline{f}$ almost surely. Then, there exists $\varepsilon_1\in(0,\min\{\eta-\frac{1}{2},\varepsilon\})$, if $\eta>\frac{1}{2}$, and $\varepsilon_1\in(0,\min\{\eta,\varepsilon\})$, otherwise. Therefore, there almost surely exists k_a such that for all $k\ge k_a$, we have $\beta\delta\underline{f}_{k|N}>\beta\delta\underline{f}-\varepsilon_1=\eta-\varepsilon_1$, and thus by (20)

$$\|\psi_k\|\le\left(1-\frac{N}{k}(\eta-\varepsilon_1)\right)\|\psi_{k-N}\|+O\left(\sqrt{\frac{\ln\ln k}{k^3}}\right),\quad\text{a.s.}$$

If $k-k_a$ is divisible by N, then by Lemma A.5 in Appendix A, one can get

$$\begin{aligned}&\|\psi_k\|\\ &\le\prod_{i=1}^{\frac{k-k_a}{N}}\left(1-\frac{N(\eta-\varepsilon_1)}{k_a+iN}\right)\|\psi_{k_a}\|\\ &\quad+O\left(\sum_{l=1}^{\frac{k-k_a}{N}}\prod_{i=l+1}^{\frac{k-k_a}{N}}\left(1-\frac{N(\eta-\varepsilon_1)}{k_a+iN}\right)\sqrt{\frac{\ln\ln(k_a+lN)}{(k_a+lN)^3}}\right)\\ &=O\left(\frac{1}{k^{\eta-\varepsilon_1}}\right)\\ &\quad+O\left(\sum_{l=1}^{\frac{k-k_a}{N}}\prod_{i=l+1}^{\frac{k-k_a}{N}}\left(1-\frac{N(\eta-\varepsilon_1)}{k_a+iN}\right)\sqrt{\frac{\ln\ln(l+2)}{l^3}}\right)\end{aligned}$$

$$= \begin{cases} O\left(\sqrt{\frac{\ln\ln k}{k}}\right), & \eta-\varepsilon_1 > \frac{1}{2}; \\ O\left(\ln k\sqrt{\frac{\ln\ln k}{k}}\right), & \eta-\varepsilon_1 = \frac{1}{2}; \\ O\left(\frac{1}{k^{\eta-\varepsilon_1}}\right), & \eta-\varepsilon_1 < \frac{1}{2}, \end{cases} \quad \text{a.s.}$$

By the settings of ε_1, we have $\varepsilon_1 \le \varepsilon$, and $\eta-\varepsilon_1 > \frac{1}{2}$ if and only if $\eta > \frac{1}{2}$. Therefore

$$\|\psi_k\| = \begin{cases} O\left(\sqrt{\frac{\ln\ln k}{k}}\right), & \eta > \frac{1}{2}; \\ O\left(\frac{1}{k^{\eta-\varepsilon}}\right), & \eta \le \frac{1}{2}, \end{cases} \quad \text{a.s.} \tag{21}$$

If $k-k_a$ is not divisible by N, then there exists an integer $\kappa \in [k-N+1, k]$ such that $\kappa - k_a$ is divisible by N. By (10), $\|\psi_k - \psi_\kappa\| \le \sum_{i=\kappa+1}^{k} \|\psi_i - \psi_{i-1}\| \le \frac{4\beta M(k-\kappa-1)}{k} \le \frac{4\beta MN}{k}$. Hence, by (21)

$$\|\psi_k\| = \|\psi_k - \psi_\kappa\| + \|\psi_\kappa\|$$
$$= \begin{cases} O\left(\sqrt{\frac{\ln\ln k}{k}}\right) + O\left(\frac{1}{k}\right), & \eta > \frac{1}{2}; \\ O\left(\frac{1}{k^{\eta-\varepsilon}}\right) + O\left(\frac{1}{k}\right), & \eta \le \frac{1}{2}, \end{cases} \quad \text{a.s.}$$
$$= \begin{cases} O\left(\sqrt{\frac{\ln\ln k}{k}}\right), & \eta > \frac{1}{2}; \\ O\left(\frac{1}{k^{\eta-\varepsilon}}\right), & \eta \le \frac{1}{2}, \end{cases} \quad \text{a.s.}$$

(16) is thereby proved.

Then, we now prove (17). For sufficiently large k

$$1 - \frac{\beta\delta N}{k}\underline{f}_{k|N} \le \left(1 + \frac{\beta\delta N}{k}\left(\underline{f} - \underline{f}_{k|N}\right)\right)\left(1 - \frac{\beta\delta N}{k}\underline{f}\right) \tag{22}$$

which together with (20) implies

$$\|\psi_k\| \Big/ \prod_{i=N+1}^{k}\left(1 + \frac{\beta\delta N}{i}\left(\underline{f} - \underline{f}_{i|N}\right)\right)$$
$$\le \left(1 - \frac{N\eta}{k}\right)\|\psi_{k-N}\| \Big/ \prod_{i=N+1}^{k-N}\left(1 + \frac{\beta\delta N}{i}\left(\underline{f} - \underline{f}_{i|N}\right)\right)$$
$$+ O\left(\sqrt{\frac{\ln\ln k}{k^3}}\right), \quad \text{a.s.} \tag{23}$$

By (16) and (d) of Lemma 3, $\underline{f} - \underline{f}_{i|N}$ converges to 0 at a polynomial rate. Hence, we have

$$\prod_{i=N+1}^{\infty}\left(1 + \frac{\beta\delta N}{i}\left(\underline{f} - \underline{f}_{i|N}\right)\right) < \infty.$$

Then, (17) can be proved by (23) and Lemma A.5. ∎

Then, the almost sure convergence rate of the algorithm can be obtained by Theorem 4.

Theorem 5: Under the condition of Theorem 2, for any $\varepsilon > 0$

$$\tilde{\theta}_k = \begin{cases} O\left(\sqrt{\frac{\ln\ln k}{k}}\right), & \eta > \frac{1}{2}; \\ O\left(\frac{1}{k^{\eta-\varepsilon}}\right), & \eta \le \frac{1}{2}, \end{cases} \quad \text{a.s.}$$

where $\eta = \beta\delta\underline{f}$ with $\underline{f}$ defined in (15). If the density function $f(\cdot)$ is assumed to be locally Lipschitz continuous, then the almost sure convergence rate can be promoted into

$$\tilde{\theta}_k = \begin{cases} O\left(\sqrt{\frac{\ln\ln k}{k}}\right), & \eta > \frac{1}{2}; \\ O\left(\ln k\sqrt{\frac{\ln\ln k}{k}}\right), & \eta = \frac{1}{2}; \\ O\left(\frac{1}{k^{\eta}}\right), & \eta < \frac{1}{2}, \end{cases} \quad \text{a.s.}$$

Proof: The theorem can be obtained by Lemma 2 and Theorem 4. ∎

Remark 14: By Theorem 5, the algorithm may not achieve the optimal almost sure convergence rate when the coefficient η is smaller than $1/2$. Since $\eta = \beta\delta\underline{f}$, the convergence rate of the algorithm depends on the step-size, the inputs, the noise distribution, and the relationship between the threshold C and $M\|\theta\|$. However, $M\|\theta\|$ relies on the true parameter θ. Thus, the almost sure convergence rate of Algorithm (5) cannot be known without a priori information on θ. The problem can be solved if the step-size is designed as $\rho_k = \beta_k/k$, where

$$\beta_k > 1 \Big/ \left(2\delta \sup_{z > M\|\hat{\theta}_k\|} \inf_{t\in[C-z, C+z]} f(t)\right).$$

The analysis for the modified algorithm is similar to the algorithm with time-invariant β.

Remark 15: For the identification problem of stochastic finite-valued systems, $O(\sqrt{\ln\ln k/k})$ is the best almost sure convergence rate. In the periodic input case, the empirical measurement algorithm in [2] generates a maximum likelihood estimate ([10], Lemma 4). The almost sure convergence rate of the empirical measurement algorithm is $O(\sqrt{\ln\ln k/k})$ [23]. In the nonperiodic input case, Theorem 5 appears to be the first to achieve the almost sure convergence rate of $O(\sqrt{\ln\ln k/k})$ theoretically. Guo and Zhao [32] achieved the almost sure convergence rate of $O(\sqrt{\ln k/k})$ for the recursive projection method. The almost sure convergence rate of SA algorithms with expanding truncations is $O(1/k^{\varepsilon})$ for $\varepsilon \in (0, 1/2)$ [17]. When properly selecting β, the almost sure convergence rate of Algorithm (5) is better than both of them.

D. Mean Square Convergence Rate

This section will estimate the mean square convergence rate of the SA-based algorithm.

Theorem 6: Under the condition of Theorem 2, for any $\varepsilon > 0$

$$\mathbb{E}\|\tilde{\theta}_k\|^2 = \begin{cases} O\left(\frac{1}{k}\right), & \eta > \frac{1}{2} \\ O\left(\frac{1}{k^{2\eta-\varepsilon}}\right), & \eta \le \frac{1}{2} \end{cases} \tag{24}$$

where $\eta = \beta\delta\underline{f}$ with $\underline{f}$ defined in (15). If $f(\cdot)$ is assumed to be locally Lipschitz continuous, then the mean square convergence rate can be promoted into

$$\mathbb{E}\|\tilde{\theta}_k\|^2 = \begin{cases} O\left(\frac{1}{k}\right), & \eta > \frac{1}{2} \\ O\left(\frac{\ln k}{k}\right), & \eta = \frac{1}{2} \\ O\left(\frac{1}{k^{2\eta}}\right), & \eta < \frac{1}{2}. \end{cases} \tag{25}$$

Proof: To prove (24), we first simplify the recursive formula of $\mathbb{E}\|\tilde{\theta}_k\|^2$.

By (5) and the Lagrange mean value theorem ([39], Th. 5.3.1), there exists ζ_k between $C - \phi_k^\top\theta$ and $C - \phi_k^\top\theta - \phi_k^\top\tilde{\theta}_{k-1}$ such that

$$\begin{aligned}\tilde{\theta}_k &= \tilde{\theta}_{k-1} + \frac{\beta\phi_k}{k}\left(\hat{F}_k - F_k\right) + \frac{\beta\phi_k}{k}\left(F_k - s_k\right)\\ &= \left(I_n - \frac{\beta}{k}f(\zeta_k)\phi_k\phi_k^\top\right)\tilde{\theta}_{k-1} + \frac{\beta\phi_k}{k}\left(F_k - s_k\right)\\ &= \prod_{i=k-N+1}^{k}\left(I_n - \frac{\beta}{i}f(\zeta_i)\phi_i\phi_i^\top\right)\tilde{\theta}_{k-N}\\ &\quad + \sum_{l=k-N+1}^{k}\prod_{i=l+1}^{k}\left(I_n - \frac{\beta}{i}f(\zeta_i)\phi_i\phi_i^\top\right)\frac{\beta\phi_l}{l}\left(F_l - s_l\right)\\ &= \prod_{i=k-N+1}^{k}\left(I_n - \frac{\beta}{i}f(\zeta_i)\phi_i\phi_i^\top\right)\tilde{\theta}_{k-N}\\ &\quad + \sum_{l=k-N+1}^{k}\frac{\beta\phi_l}{l}\left(F_l - s_l\right) + O\left(\frac{1}{k^2}\right).\end{aligned}$$

Similar to (19), except for the first few steps, we have

$$\begin{aligned}&\left\|\prod_{i=k-N+1}^{k}\left(I_n - \frac{\beta f(\zeta_i)}{i}\phi_i\phi_i^\top\right)\right\|\\ &\leq \left(1 - \frac{\beta\delta N}{k}\underline{f}'_{k|N}\right) + O\left(\frac{1}{k^2}\right)\end{aligned}$$

where $\underline{f}'_{k|N} = \underline{f}(\max_{k-N<i\leq k}|\phi_i^\top\tilde{\theta}_{i-1}|)$, and $\underline{f}(\cdot)$ is defined in (14). Besides, noticing that $\tilde{\theta}_{k-N}$ is independent of $\sum_{l=N+1}^{k}\frac{\beta\phi_l}{l}(F_l - s_l)$, we have

$$\begin{aligned}&\mathbb{E}\left[\left(\sum_{l=k-N+1}^{k}\frac{\beta\phi_l}{l}\left(F_l - s_l\right)\right)^\top\right.\\ &\qquad\left.\cdot\prod_{i=k-N+1}^{k}\left(I_n - \frac{\beta}{i}f(\zeta_i)\phi_i\phi_i^\top\right)\tilde{\theta}_{k-N}\right]\\ &= \mathbb{E}\left[\left(\sum_{l=k-N+1}^{k}\frac{\beta\phi_l}{l}\left(F_l - s_l\right)\right)^\top\right.\\ &\qquad\left.\cdot\left(\prod_{i=k-N+1}^{k}\left(I_n - \frac{\beta}{i}f(\zeta_i)\phi_i\phi_i^\top\right) - I_n\right)\tilde{\theta}_{k-N}\right]\\ &= O\left(\frac{1}{k^2}\right).\end{aligned}$$

Therefore, for sufficiently large k, one can get

$$\mathbb{E}\|\tilde{\theta}_k\|^2 \leq \mathbb{E}\left[\left(1 - \frac{\beta\delta N}{k}\underline{f}'_{k|N}\right)^2\|\tilde{\theta}_{k-N}\|^2\right] + O\left(\frac{1}{k^2}\right). \tag{26}$$

By Theorem 2 and (b) of Lemma 3, $\mathbb{P}\{\underline{f}'_{k|N} < \underline{f} - \frac{\varepsilon}{2\beta\delta}\} = O(\exp(-mk^{1/2}))$. Hence, by Corollary A.1 in Appendix A, we have

$$\begin{aligned}&\mathbb{E}\left[\left(1 - \frac{\beta\delta N}{k}\underline{f}'_{k|N}\right)^2\|\tilde{\theta}_{k-N}\|^2\right]\\ &\leq \int_{\left\{\underline{f}'_{k|N}\geq\underline{f}-\frac{\varepsilon}{2\beta\delta}\right\}}\left(1 - \frac{N}{k}\left(\eta - \frac{\varepsilon}{2}\right)\right)^2\|\tilde{\theta}_{k-N}\|^2 d\mathbb{P}\\ &\quad + \int_{\left\{\underline{f}'_{k|N}<\underline{f}-\frac{\varepsilon}{2\beta\delta}\right\}}\|\tilde{\theta}_{k-N}\|^2 d\mathbb{P}\\ &= \left(1 - \frac{N}{k}\left(\eta - \frac{\varepsilon}{2}\right)\right)^2\mathbb{E}\|\tilde{\theta}_{k-N}\|^2\\ &\quad + O\left(\ln k\cdot\exp(-mk^{1/2})\right).\end{aligned}$$

Substituting the above estimate into (26) gives

$$\mathbb{E}\|\tilde{\theta}_k\|^2 \leq \left(1 - \frac{N}{k}\left(\eta - \frac{\varepsilon}{2}\right)\right)^2\mathbb{E}\|\tilde{\theta}_{k-N}\|^2 + O\left(\frac{1}{k^2}\right).$$

Thus, (24) can be proved by Lemma A.5 in Appendix A.

Then, we prove (25). Similar to (22), for sufficiently large k

$$1 - \frac{\beta\delta N}{k}\underline{f}'_{k|N} \leq \left(1 + \frac{\beta\delta N}{k}\left(\underline{f} - \underline{f}'_{k|N}\right)\right)\left(1 - \frac{\beta\delta N}{k}\underline{f}\right).$$

Therefore, by (26) and $\eta = \beta\delta\underline{f}$, one can get

$$\begin{aligned}\mathbb{E}\|\tilde{\theta}_k\|^2 &\leq \left(1 - \frac{N\eta}{k}\right)^2\mathbb{E}\left[\left(1 + \frac{\beta\delta N}{k}\left(\underline{f} - \underline{f}'_{k|N}\right)\right)^2\right.\\ &\qquad\left.\cdot\|\tilde{\theta}_{k-N}\|^2\right] + O\left(\frac{1}{k^2}\right).\end{aligned} \tag{27}$$

By (d) of Lemma 3, since $f(\cdot)$ is assumed to be locally Lipschitz continuous here, $\underline{f}(\cdot)$ is also locally Lipschitz continuous. Hence, if $\|\tilde{\theta}_j\| \leq j^{-\varepsilon'}$ for $\varepsilon' > 0$ and all $j = k-N+1,\ldots,k$, then there exists $L > 0$ such that $\underline{f} - \underline{f}'_{k|N} \leq Lk^{-\varepsilon'}$, which together with Corollaries A.1 and A.3 in Appendix A implies that there exist positive numbers m and ε such that

$$\begin{aligned}&\mathbb{E}\left[\left(1 + \frac{\beta\delta N}{k}\left(\underline{f} - \underline{f}'_{k|N}\right)\right)^2\|\tilde{\theta}_{k-N}\|^2\right]\\ &\leq \left(1 + \frac{\beta\delta NL}{k^{1+\varepsilon'}}\right)^2\int_{\cap_{j=k-N+1}^{k}\{\|\tilde{\theta}_j\|\leq j^{-\varepsilon'}\}}\|\tilde{\theta}_{k-N}\|^2 d\mathbb{P}\\ &\quad + O\left(\ln k\cdot\exp(-mk^{1-\varepsilon})\right)\\ &\leq \left(1 + \frac{\beta\delta NL}{k^{1+\varepsilon'}}\right)^2\mathbb{E}\|\tilde{\theta}_{k-N}\|^2 d\mathbb{P} + O\left(\ln k\cdot\exp(-mk^{1-\varepsilon})\right).\end{aligned}$$

Substituting the above estimate into (27) gives

$$\mathbb{E}\|\tilde{\theta}_k\|^2 \leq \left(1 - \frac{N\eta}{k}\right)^2\left(1 + \frac{\beta\delta NL}{k^{1+\varepsilon'}}\right)^2\mathbb{E}\|\tilde{\theta}_{k-N}\|^2 + O\left(\frac{1}{k^2}\right).$$

Therefore, we have

$$\mathbb{E}\|\tilde{\theta}_k\|^2 \Big/ \prod_{i=1}^{k}\left(1 + \frac{\beta\delta NL}{i^{1+\varepsilon'}}\right)^2$$

$$\leq \left(1-\frac{N\eta}{k}\right)^2 \mathbb{E}\|\tilde{\theta}_{k-N}\|^2 \Big/ \prod_{i=1}^{k-N}\left(1+\frac{\beta\delta NL}{i^{1+\varepsilon'}}\right)^2 + O\left(\frac{1}{k^2}\right).$$

Then, by Lemma A.5, one can get

$$\mathbb{E}\|\tilde{\theta}_k\|^2 \Big/ \prod_{i=1}^{k}\left(1+\frac{\beta\delta NL}{i^{1+\varepsilon'}}\right)^2 = \begin{cases} O\left(\frac{1}{k}\right), & \eta > \frac{1}{2} \\ O\left(\frac{\ln k}{k}\right), & \eta = \frac{1}{2} \\ O\left(\frac{1}{k^{2\eta}}\right), & \eta < \frac{1}{2}. \end{cases}$$

Due to the boundedness of $\prod_{i=1}^{\infty}(1+\frac{\beta\delta NL}{i^{1+\varepsilon'}})^2$, (25) is proved. ∎

Remark 16: By Theorem 6, the mean square convergence rate of the SA-based algorithm achieves $O(1/k)$ when properly selecting the coefficient β. By [34], the Cramér–Rao lower bound for estimating θ based on binary observations $s_1, \ldots, s_k$ is

$$\sigma_{CR}^2(s_1, \ldots, s_k) = \left(\sum_{i=1}^{k}\frac{f_i^2}{F_i(1-F_i)}\phi_i\phi_i^\top\right)^{-1} = O\left(\frac{1}{k}\right).$$

Besides, for the identification problem of MA systems with accurate observations and Gaussian noises, the least square algorithm generates a minimum variance estimate ([40], Th. 4.4.2). The mean square convergence rate of the recursive least square algorithm is $O(1/k)$. Therefore, $O(1/k)$ is the best mean square convergence rate in theory of the identification problem of the binary-valued MA systems and even accurate ones.

Remark 17: In the multiple threshold case, when properly selecting the coefficient β, the almost sure and mean square convergence rates of the SA-based algorithm are also $O(\sqrt{\ln\ln k/k})$ and $O(1/k)$, respectively. The analysis is similar to the binary observation case.

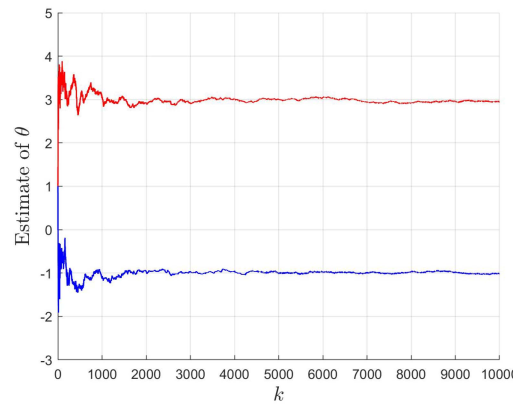
Remark 18: From Theorems 5 and 6, we learn that the almost sure and mean square convergence rates are influenced by the step-size, inputs, and the threshold. Here, we give the following intuitive explanations.

1) The step-size coefficient β influences the convergence rates. If we adopt a small step-size β, then the algorithm updates the estimate at a slow rate.
2) Excitations of $\{\phi_k, k \geq 1\}$ also affect the convergence rates. If δ in (3) is large, then $\{y_k, j \geq 1\}$ provides rich information on θ from every direction, which causes good effectiveness of the algorithm.
3) The threshold C is another factor influencing the convergence rates. If the threshold C is too high or too low, then s_k may have always the same value, which causes poor effectiveness of the algorithm.

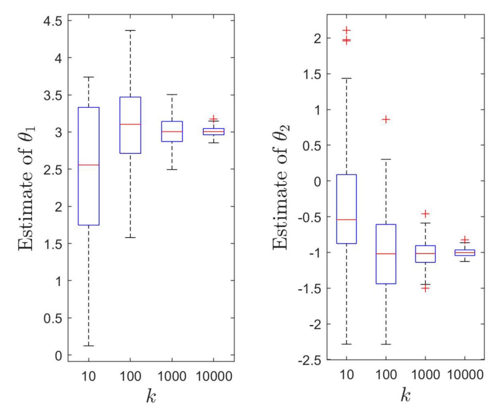
Besides, given $\{\phi_k, k \geq 1\}$, the upper bound M does not influence the actual convergence rate of the algorithm, but only influences the estimation on the convergence rates. If M is too large, then we have less information on $\{\phi_k, k \geq 1\}$ for estimating the convergence rates, which may lead to an unsatisfactory estimation on the convergence rates.

V. Numerical Simulation

A numerical simulation will be performed in this section to verify Theorems 3, 5, and 6.



(a)



(b)

Fig. 1. Convergence of Algorithm (5). (a) Trajectory of $\hat{\theta}_k$. (b) Box-plots of $\hat{\theta}_k$ in 200 repeated experiments.

Consider an MA system $y_k = \phi_k^\top\theta + d_k$ with binary-valued observation

$$s_k = I_{\{y_k \leq C\}} = \begin{cases} 1, & y_k \leq C \\ 0, & y_k > C \end{cases}$$

where the unknown parameter $\theta = [3, -1]^\top$, the threshold $C = 1$, and d_k is i.i.d. Gaussian noise with variance $\sigma^2 = 25$ and zero mean. The regressed function of inputs $\phi_k = [u_k, u_{k-1}]^\top$ is generated by $u_{3i} = -1 + e_{3i}$, $u_{3i+1} = 2 + e_{3i+1}$, $u_{3i+2} = 1 + e_{3i+2}$ for natural number i, where $e_k = 0.1\sin(\ln(k+1))$. It can be verified that the input follows Assumption 1 with $M = 2.38$, $N = 3$, and $\delta = 1.42$.

In the simulation, set $\beta = 20$, $k_0 = 20$, and the initial value $\hat{\theta}_{k_0} = [1, 1]^\top$. Fig. 1(a) shows a trajectory of $\hat{\theta}_k$. Fig. 1(b) gives the box-plots of $\hat{\theta}_k$ in 200 repeated experiments. The figures demonstrate the convergence of Algorithm (5).

Remark 19: We set $\beta = 20$ to have $\eta > 1/2$. When $k_0 = 0$, large β causes large step-sizes in first few steps. Due to the randomness of $\{s_k\}$, the estimate $\hat{\theta}_k$ may run away from the true value θ after first few steps of iterations. Then, it will take much more time to reduce the estimation error. To avoid this situation, we should adjust the starting point k_0 according to the selection of β. In the simulation, we set $k_0 = 20$.

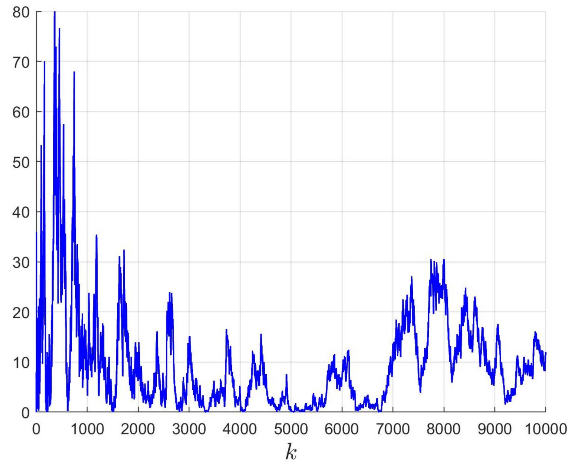


Fig. 2. Trajectory of $k\|\tilde{\theta}_k\|^2/\ln\ln k$.

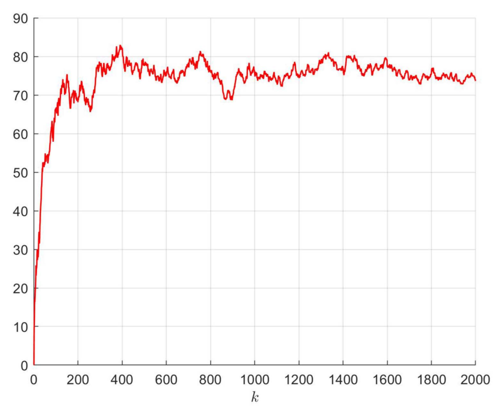


Fig. 3. Trajectory of $k\|\tilde{\theta}_k\|^2$ in 200 repeated experiments.

Note that η is about 0.53. Then, by Theorem 5, Algorithm (5) achieves the almost sure convergence rate of $O(\sqrt{\ln\ln k/k})$. Fig. 2 shows that the trajectory of $k\|\tilde{\theta}_k\|^2/\ln\ln k$ is bounded, which consists of the almost sure convergence rate of $O(\sqrt{\ln\ln k/k})$.

By Theorem 6, Algorithm (5) achieves the mean square convergence rate of $O(1/k)$. Fig. 3 illustrates that the average of the 200 trajectories of $k\|\tilde{\theta}_k\|^2$ is bounded, which consists of the mean square convergence rate of $O(1/k)$.

Besides, by Theorem 6, the step-size β influences the mean square convergence rate. Fig. 4 shows the empirical mean square convergence rate under the case of $\beta = 20$ is faster than that under the case of $\beta = 1$.

It will be an interesting problem to consider the situation where the distribution used in Algorithm (5) is different from the actual noise distribution. When the distribution used in Algorithm (5) is Gaussian with variance 20 and zero mean, but the actual noise variance is 25, Fig. 5 shows that the estimation error is bounded.

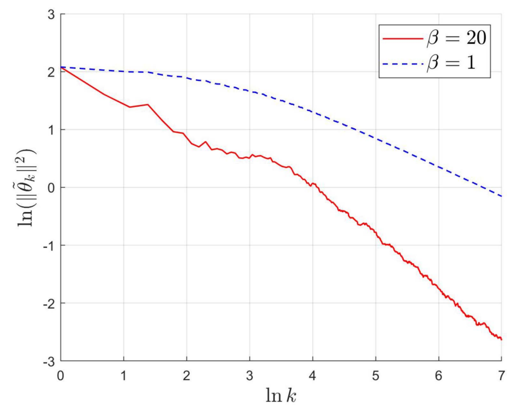


Fig. 4. Empirical mean square convergence rates under different β.

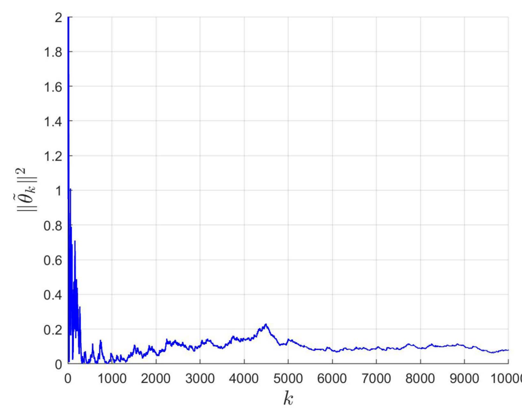


Fig. 5. Trajectory of $\|\tilde{\theta}_k\|^2$ for the wrong variance case.

VI. Conclusion

This article investigates the identification problem of binary-valued MA systems with uniformly persistently exciting inputs. An SA-based algorithm without projection is proposed to identify the unknown parameter. The algorithm appears to be the first online identification method for binary-valued systems whose implementation does not rely on projections or truncations. When properly selecting the coefficients, the almost sure convergence rate of the SA-based algorithm is $O(\sqrt{\ln\ln k/k})$, and the mean square convergence rate is $O(1/k)$. Both the convergence rates are the best for the identification problem of binary-valued systems. Moreover, an auxiliary stochastic process named SPAO is constructed for the effectiveness analysis.

Here, we give some topics for future research. First, the design of the step-size ρ_k is left as an open question. How can we design a dynamic ρ_k to allow the convergence rates to be the best automatically, and how can we design ρ_k to make the identification algorithm achieve the Cramér–Rao lower bound asymptotically? Second, can the algorithm be extended to other general forms of systems, such as the infinite impulse response system? Third, how can we design system control laws to regulate the system performance using the SA-based algorithm?

Appendix A
Lemmas and the Proofs

Proof of Lemma 1: The lemma can be indicated by [41, Th. 5.5.1]. We transfer the problem first.

First, we claim that it is sufficient to prove that there exists $m > 0$ such that $\mathbb{P}\{\|w_k\| > k^{-\varepsilon}\} = O(\exp(-mk^{1-2\varepsilon}))$. This is because $\sum_{j=k}^{\infty} \exp(-mj^{1-2\varepsilon}) = O(k^{2\varepsilon}\exp(-mk^{1-2\varepsilon})) = O(\exp(-mk^{1-2\varepsilon}/2))$.

Second, we claim that it is sufficient to prove that for any $i \in \{1, 2, \ldots, n\}$, $w_{k,i}$ satisfies $\mathbb{P}\{|w_{k,i}| > k^{-\varepsilon}/\sqrt{n}\} = O(\exp(-mk^{1-2\varepsilon}))$, where $w_{k,i}$ is the ith component of w_k. This is because $\{\|w_k\| > k^{-\varepsilon}\} \subseteq \cup_i \{|w_{k,i}| > k^{-\varepsilon}/\sqrt{n}\}$, which implies

$$\mathbb{P}\left\{\|w_k\| > k^{-\varepsilon}\right\} \leq \sum_{i=1}^{n} \mathbb{P}\left\{|w_{k,i}| > k^{-\varepsilon}/\sqrt{n}\right\}.$$

The transformation has been finished. Now, we show that the converted problem is a corollary of [41, Th. 5.5.1].

Lemma A.1 ([41], Th. 5.5.1): Assume that the following holds:

1) $\{X_k, k \geq 1\}$ is a sequence of independent random variables;
2) $\mathbb{E}X_k = 0$ and $|X_k| \leq \bar{X} < \infty$;
3) $S_k = \sum_{i=1}^{k} X_i$, $\sigma_k = \sqrt{\mathrm{var}(S_k)}$.

Then,

$$\mathbb{P}\left\{\frac{S_k}{\sigma_k} > d_k\right\} < \max\left\{\exp\left(-\frac{d_k^2}{4}\right), \exp\left(-\frac{d_k\sigma_k}{4\bar{X}}\right)\right\}.$$

Set $X_{k,i} = \beta\phi_{k,i}(F_k - s_k)$, $S_{k,i} = \sum_{j=1}^{k} X_{j,i}$, $\sigma_{k,i} = \sqrt{\mathrm{var}(S_{k,i})}$, and $d_{k,i} = k^{1-\varepsilon}/\sigma_{k,i}\sqrt{n}$, where $\phi_{k,i}$ is the ith component of ϕ_k. Then, by Lemma A.1

$$\begin{aligned}\mathbb{P}\left\{w_{k,i} > \frac{k^{-\varepsilon}}{\sqrt{n}}\right\} &= \mathbb{P}\left\{\frac{S_{k,i}}{\sigma_{k,i}} > d_{k,i}\right\}\\ &< \max\left\{\exp\left(-\frac{d_{k,i}^2}{4}\right), \exp\left(-\frac{d_{k,i}\sigma_{k,i}}{4\bar{X}}\right)\right\}\\ &= \max\left\{\exp\left(-\frac{k^{2-2\varepsilon}}{4n\sigma_{k,i}^2}\right), \exp\left(-\frac{k^{1-\varepsilon}}{4\bar{X}\sqrt{n}}\right)\right\}.\end{aligned}$$

Noting that

$$\sigma_{k,i}^2 = \mathrm{var}(S_{k,i}) = \sum_{j=1}^{k} \mathrm{var}(X_{j,i}) \leq 4\bar{X}^2 k$$

then $\exp(-k^{2-2\varepsilon}/4n\sigma_{k,i}^2) \leq \exp(-k^{1-2\varepsilon}/16n\bar{X}^2)$. Therefore, there exists $m_+ > 0$ such that $\mathbb{P}\{w_{k,i} > k^{-\varepsilon}/\sqrt{n}\} = O(\exp(-m_+ k^{1-2\varepsilon}))$.

$\mathbb{P}\{w_{k,i} < -k^{-\varepsilon}/\sqrt{n}\}$ can be similarly analyzed.

Combining the two consequences, the converted problem is thereby proved. That is to say, we get Lemma 1. □

Proof of Lemma 3: (a) For $x_1 \geq x_2$, one can get

$$\begin{aligned}\underline{f}(x_1) &= \sup_{z > M\|\theta\| + x_1} \inf_{t \in [C-z, C+z]} f(t)\\ &\leq \sup_{z > M\|\theta\| + x_2} \inf_{t \in [C-z, C+z]} f(t) = \underline{f}(x_2).\end{aligned}$$

Therefore, $\underline{f}(\cdot)$ is nonincreasing.

Due to the monotonicity of $\underline{f}(\cdot)$, $\sup_{x > \chi} \underline{f}(x)$ is the right limit of $\underline{f}(\cdot)$ at the point χ. Then, $\underline{f}(\cdot)$ is right continuous because

$$\begin{aligned}\sup_{x > \chi} \underline{f}(x) &= \sup_{x > \chi} \sup_{z > M\|\theta\| + x} \inf_{t \in [C-z, C+z]} f(t)\\ &= \sup_{z > M\|\theta\| + \chi} \inf_{t \in [C-z, C+z]} f(t) = \underline{f}(\chi).\end{aligned}$$

(b) Since $\underline{f}(\cdot)$ is right continuous and only defined on $[0, \infty)$, we have $\lim_{x \to 0} f(x) = \underline{f}(0) = \underline{f}$.

(c) By (14), we have

$$\begin{aligned}\underline{f}(x) &\leq \sup_{z \geq M\|\theta\| + x} \inf_{t \in [C-z, C+z]} f(t)\\ &= \inf_{t \in [C - M\|\theta\| - x, C + M\|\theta\| + x]} f(t).\end{aligned}$$

(d) We first prove that $g(z) = \inf_{t \in [C-z, C+z]} f(t)$ is locally Lipschitz continuous on $z \geq 0$. Since $f(t)$ is locally Lipschitz continuous, for any given $z_0 \geq 0$, there exist $\delta_1 > 0$ and $K_1 > 0$ such that

$$|f(t_1) - f(t_2)| \leq K_1 |t_1 - t_2| \tag{A.1}$$

for all $t_1, t_2 \in (C + z_0 - \delta_1, C + z_0 + \delta_1)$ and $t_1, t_2 \in (C - z_0 - \delta_1, C - z_0 + \delta_1)$.

Consider $z_1, z_2 \in (z_0 - \delta_1, z_0 + \delta_1) \cap [0, \infty)$.

If $z_1 = z_2$, then $g(z_2) - g(z_1) = 0$.

If $z_1 \neq z_2$, then without the loss of generality, consider $z_1 > z_2$, which implies $g(z_1) = \inf_{t \in [C-z_1, C+z_1]} f(t) \leq \inf_{t \in [C-z_2, C+z_2]} f(t) = g(z_2)$. Hence, $|g(z_1) - g(z_2)| = g(z_2) - g(z_1)$. By the definition of infimum [39], there exists $\tau_1 \in [C - z_1, C + z_1]$ such that

$$g(z_1) = \inf_{t \in [C-z_1, C+z_1]} f(t) \geq f(\tau_1) - (z_1 - z_2). \tag{A.2}$$

When $\tau_1 \in [C - z_2, C + z_2]$

$$g(z_2) - g(z_1) \leq f(\tau_1) - f(\tau_1) + z_1 - z_2 = z_1 - z_2.$$

When $\tau_1 \in [C - z_1, C - z_2)$, set $\tau_2 = C - z_2$. Therefore, $\tau_2 - \tau_1 \leq z_1 - z_2$, and

$$\tau_1, \tau_2 \subseteq [C - z_1, C - z_2] \subseteq (C - z_0 - \delta_1, C - z_0 + \delta_1)$$

which together with (A.1) and (A.2) implies

$$\begin{aligned}g(z_2) - g(z_1) &\leq f(\tau_2) - f(\tau_1) + (z_1 - z_2)\\ &\leq K_1(\tau_2 - \tau_1) + (z_1 - z_2) \leq (K_1 + 1)(z_1 - z_2).\end{aligned} \tag{A.3}$$

When $\tau_1 \in (C + z_2, C + z_1]$, set $\tau_2 = C + z_2$. Then, (A.3) can be obtained similar to the case of $\tau_1 \in [C - z_1, C - z_2)$.

Therefore, $g(z)$ is locally Lipschitz continuous on $z \geq 0$. Now we further prove that $\underline{f}(x) = \sup_{z > M\|\theta\| + x} g(z)$ is also local Lipschitz continuous on $x \geq 0$. Since $g(z)$ is locally Lipschitz continuous, for any given $x_0 \geq 0$, there exist $\delta_2 > 0$ and $K_2 > 0$ such that

$$|g(z_1) - g(z_2)| \leq K_2 |z_1 - z_2| \tag{A.4}$$

for all $z_1, z_2 \in (M\|\theta\| + x_0 - \delta_2, M\|\theta\| + x_0 + \delta_2) \cap [0, \infty)$.

Consider $x_1, x_2 \in (x_0 - \delta_2, x_0 + \delta_2) \cap [0, \infty)$.

If $x_1 = x_2$, then $\underline{f}(x_1) - \underline{f}(x_2) = 0$.

If $x_1 \neq x_2$, then without loss of generality, consider $x_1 > x_2$, which together with (a) of this lemma implies $|\underline{f}(x_1) - \underline{f}(x_2)| = \underline{f}(x_2) - \underline{f}(x_1)$. By the definition of supremum [39], there exists $\upsilon_2 \in (M\|\theta\| + x_2, \infty)$ such that

$$\underline{f}(x_2) = \sup_{z > M\|\theta\| + x_2} g(z) \leq g(\upsilon_2) + (x_1 - x_2). \tag{A.5}$$

When $\upsilon_2 \in (M\|\theta\| + x_1, \infty)$

$$\underline{f}(x_2) - \underline{f}(x_1) \leq g(\upsilon_2) + (x_1 - x_2) - g(\upsilon_2).$$

When $\upsilon_2 \in (M\|\theta\| + x_2, M\|\theta\| + x_1]$, set

$$\upsilon_1 = \min\left\{M\|\theta\| + 2x_1 - x_2, M\|\theta\| + \frac{x_1 + x_0 + \delta_2}{2}\right\}.$$

Therefore, $\upsilon_1 > M\|\theta\| + x_1 \geq \upsilon_2, \upsilon_1 - \upsilon_2 < 2(x_1 - x_2)$, and

$$\upsilon_1, \upsilon_2 \in (M\|\theta\| + x_0 - \delta_2, M\|\theta\| + x_0 + \delta_2) \cap [0, \infty)$$

which together with (A.4) and (A.5) implies

$$\underline{f}(x_2) - \underline{f}(x_1) \leq g(\upsilon_2) + (x_1 - x_2) - g(\upsilon_1)$$
$$\leq K_2(\upsilon_1 - \upsilon_2) + (x_1 - x_2) \leq (2K_2 + 1)(x_1 - x_2).$$

Hence, $\underline{f}(\cdot)$ is local Lipschitz continuous. □

Lemma A.2: Assume that ϕ_k satisfies Assumption 1 and the stochastic process ψ_k satisfies $\|\psi_k - \psi_{k-1}\| \leq \Psi/k$ for some $\Psi > 0$. Then,

$$\delta\|\psi_k\|^2 \leq \frac{1}{N}\sum_{j=k+1}^{k+N}\left(\phi_j^\top \psi_{j-1}\right)^2 + \frac{2NM^2\Psi}{k}\sum_{j=k}^{k+N-1}\|\psi_j\| + \frac{N^2M^2\Psi^2}{k^2}.$$

Furthermore, if $b' > 0$ and k is large enough, then there is $k' \in [k+1, k+N]$ such that $|\phi_{k'}^\top \psi_{k'-1}| \geq \sqrt{\delta/2}\|\psi_k\| I_{\{\|\psi_k\| > b'\}}$.

Proof: The lemma is based on Assumption 1.

Because $\|\psi_k - \psi_{k-1}\| \leq \Psi/k$, $\|\psi_k - \psi_{j-1}\| \leq N\Psi/k$ for any $j \in [k+1, k+N]$. Therefore,

$$\frac{1}{N}\sum_{j=k+1}^{k+N}\left(\phi_j^\top \psi_{j-1}\right)^2$$
$$\geq \frac{1}{N}\sum_{j=k+1}^{k+N}\left(\phi_j^\top \psi_k\right)^2 - \frac{2NM^2\Psi}{k}\sum_{j=k}^{k+N-1}\|\psi_j\| - \frac{N^2M^2\Psi^2}{k^2}.$$

Besides, by Assumption 1

$$\frac{1}{N}\sum_{j=k+1}^{k+N}\left(\phi_j^\top \psi_k\right)^2 = \frac{1}{N}\sum_{j=k+1}^{k+N}\psi_k^\top \phi_j \phi_j^\top \psi_k \geq \delta\|\psi_k\|^2.$$

Thus, the first part of the lemma is proved.

As for the second part, we note that under the condition of the lemma, $\psi_k = O(\ln k)$. Then,

$$\frac{2NM^2\Psi}{k}\sum_{j=k}^{k+N-1}\|\psi_j\| + \frac{N^2M^2\Psi^2}{k^2} = O\left(\frac{\ln k}{k}\right).$$

Hence, if $\|\psi_k\| > b'$ and k is sufficiently large, then one can get

$$\frac{1}{N}\sum_{j=k+1}^{k+N}\left(\phi_j^\top \psi_{j-1}\right)^2 \geq \delta\|\psi_k\|^2 + O\left(\frac{\ln k}{k}\right) > \frac{\delta}{2}\|\psi_k\|^2$$

which implies $\frac{1}{N}\sum_{j=k+1}^{k+N}(\phi_j^\top \psi_{j-1})^2 > \frac{\delta}{2}\|\psi_k\|^2 I_{\{\|\psi_k\| > b'\}}$ for sufficiently large k. Then, there exists $k' \in [k+1, k+N]$ such that $(\phi_{k'}^\top \psi_{k'-1})^2 \geq \frac{\delta}{2}\|\psi_k\|^2 I_{\{\|\psi_k\| > b'\}}$, which verifies the second part of the lemma. □

Lemma A.3: If a sequence $\{a_k\}$ satisfies the recursive function

$$a_k \leq a_{k-1} - \frac{D\sqrt{a_{k-1}}}{k + k_0} I_{\{a_{k-1} \geq \frac{M'}{2}\}} + \nu_k \tag{A.6}$$

where D, k_0, and M' are all positive, and $\sum_{k=1}^{\infty}|\nu_k| < M'/2$, then

$$a_k < \max\left\{M', \left[\left(\sqrt{a_0} - \frac{D}{2}\ln\left(\frac{k + k_0 + 1}{k_0 + 1}\right)\right)^+\right]^2 + \frac{M'}{2}\right\} \tag{A.7}$$

where $x^+ = \max\{0, x\}$.

Proof: If $a_k < M'$, then the lemma is proved. Hence, we can assume that $a_k \geq M'$ in the rest of the proof, which implies

$$a_t \geq a_k - \sum_{i=t+1}^{k}\nu_i \geq a_k - \frac{M'}{2} \geq \frac{M'}{2} \quad \forall t \leq k.$$

Define $a'_0 = a_0$ and $a'_t = a_t - \sum_{i=1}^{t}|\nu_i| > M'/2 - M'/2 = 0$ for $t \geq 1$. Then, we have

$$a'_t = a_t - \sum_{i=1}^{t}|\nu_i| \leq a_{t-1} - \frac{D\sqrt{a_{t-1}}}{t + k_0} + \nu_t - \sum_{i=1}^{t}|\nu_i|$$
$$\leq a_{t-1} - \sum_{i=1}^{t-1}|\nu_i| - \frac{D\sqrt{\left(a_{t-1} - \sum_{i=1}^{t-1}|\nu_i|\right)^+}}{t + k_0}$$
$$= a'_{t-1} - \frac{D\sqrt{a'_{t-1}}}{t + k_0}$$

and hence

$$a'_k < a'_{k-1} - \frac{D\sqrt{a'_{k-1}}}{k + k_0} + \frac{D^2}{4(k + k_0)^2}$$
$$= \left(\sqrt{a'_{k-1}} - \frac{D}{2(k + k_0)}\right)^2$$

which implies $\sqrt{a'_k} < \sqrt{a'_{k-1}} - \frac{D}{2(k + k_0)}$. Therefore, by $x \leq x^+$

$$\sqrt{a'_k} < \sqrt{a'_0} - \sum_{t=1}^{k}\frac{D}{2(t + k_0)}$$
$$\leq \left(\sqrt{a_0} - \frac{D}{2}\ln\left(\frac{k + k_0 + 1}{k_0 + 1}\right)\right)^+.$$

So, we have

$$a_k = a'_k + \sum_{i=1}^{t} |\nu_i| < \left[\left(\sqrt{a_0} - \frac{D}{2} \ln \left(\frac{k+k_0+1}{k_0+1} \right) \right)^+ \right]^2 + \frac{M'}{2}.$$

The lemma is thereby proved. □

Remark A.1: Lemma A.3 ensures the uniform ultimate upper boundedness of the sequence $\{a_k\}$, which satisfies (A.6). Given the initial value a_0

$$\sqrt{a_0} - \frac{D}{2} \ln \left(\frac{k+k_0+1}{k_0+1} \right) < 0$$

when $k > (k_0+1)\exp(2\sqrt{a_0}/D) - k_0 - 1$, which together with (A.7) implies $a_k < M'$.

Lemma A.4: Assume that the following holds:

1) $v(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is a continuously twice differentiable nonnegative function, whose second derivative is bounded;
2) $g_k(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ is uniformly bounded;
3) $\nabla v(x)^\top g_k(x)$ is uniformly upper bounded, where $\nabla v(\cdot)$ is the gradient of $v(\cdot)$;
4) the positive step-size $\rho_k \in \mathbb{R}$ satisfies $\lim_{k\to\infty} \rho_k = 0$;
5) $x_k = x_{k-1} + \rho_k g_k(x_{k-1})$.

Then, $v(x_k) = O(\sum_{i=1}^{k} \rho_i)$.

Proof: From

$$\begin{aligned} v(x_k) &= v(x_{k-1} + \rho_k g_k(x_{k-1})) \\ &= v(x_{k-1}) + \rho_k \nabla v(x_{k-1})^\top g_k(x_{k-1}) + O\left(\rho_k^2\right) \\ &\le v(x_{k-1}) + O(\rho_k) \le O\left(\sum_{i=1}^{k} \rho_i \right) \end{aligned}$$

we get the lemma. ■

Corollary A.1: Under Assumptions 1 and 2, the estimation error of Algorithm (5) satisfies $\tilde{\theta}_k = O(\sqrt{\ln k})$.

Proof: Due to the finite covariance of the noise, by Markov inequality ([37], Th. 5.1.1), when t goes to ∞

$$\begin{aligned} F\left(C - \phi_k^\top \theta - t\right) &= \mathbb{P}\left\{ d_k < C - \phi_k^\top \theta - t \right\} \\ &\le \mathbb{P}\left\{ d_k^2 > \left(C - \phi_k^\top \theta - t\right)^2 \right\} \\ &\le \frac{\mathbb{E} d_k^2}{\left(C - \phi_k^\top \theta - t\right)^2} = O\left(\frac{1}{t^2}\right) \end{aligned} \quad \text{(A.8)}$$

and similarly, when t goes to $-\infty$

$$1 - F\left(C - \phi_k^\top \theta - t\right) = O\left(\frac{1}{t^2}\right). \quad \text{(A.9)}$$

Set $v(x) = x^\top x$. Then, $\nabla v(x) = x$. By (A.8) and (A.9)

$$\begin{aligned} &\nabla v(x)^\top \phi_k \left(F\left(C - \phi_k^\top \theta - \phi_k^\top x\right) - s_k\right) \\ &= \phi_k^\top x \left(F\left(C - \phi_k^\top \theta - \phi_k^\top x\right) - s_k\right) \end{aligned}$$

is uniformly upper bounded. Thus, we get the corollary by Lemma A.4. ■

Corollary A.2: Under the condition of Lemma 2, $\psi_k = O(\sqrt{\ln k})$.

Proof: From $\psi_k = \tilde{\theta}_k - w_k = O(\sqrt{\ln k}) + O(1)$, we get the corollary. ■

Remark A.2: Corollaries A.1 and A.2 estimate the estimation error $\tilde{\theta}_k$ and SPAO ψ_k in the worst case, respectively.

Lemma A.5: For the sequence $\{h_k\}$, assume that the following holds:

1) h_k is positive and monotonically increasing;
2) $\ln h_k = o(\ln k)$.

Then, for nonnegative real numbers i_0, η', and ε, and any positive integer p

$$\sum_{l=1}^{k} \prod_{i=l+1}^{k} \left(1 - \frac{\eta'}{i+i_0}\right)^p \frac{h_l}{l^{1+\varepsilon}} = \begin{cases} O\left(\frac{h_k}{k^\varepsilon}\right), & p\eta' > \varepsilon \\ O\left(\frac{h_k \ln k}{k^\varepsilon}\right), & p\eta' = \varepsilon \\ O\left(\frac{1}{k^{p\eta'}}\right), & p\eta' < \varepsilon. \end{cases}$$

Proof: Since $i_0 \ge 0$, one can get $\frac{l+1+i_0}{l} = 1 + \frac{i_0+1}{l} \le 2 + i_0$ and $\frac{k}{k+i_0} \le 1$ for all positive integers l and k. Then, by [42, Lemma A.2], we have

$$\prod_{i=l+1}^{k} \left(1 - \frac{\eta'}{i+i_0}\right) \le \left(\frac{l+1+i_0}{k+i_0}\right)^{\eta'} \le (2+i)^{\eta'} \left(\frac{l}{k}\right)^{\eta'}$$

which leads to

$$\begin{aligned} &\sum_{l=1}^{k} \prod_{i=l+1}^{k} \left(1 - \frac{\eta'}{i+i_0}\right)^p \frac{h_l}{l^{1+\varepsilon}} \\ &= \sum_{l=1}^{k} \left[\prod_{i=l+1}^{k} \left(1 - \frac{\eta'}{i+i_0}\right)\right]^p \frac{h_l}{l^{1+\varepsilon}} \\ &\le (2+i)^{p\eta'} \sum_{l=1}^{k} \left(\frac{l}{k}\right)^{p\eta'} \frac{h_l}{l^{1+\varepsilon}} = O\left(\frac{1}{k^{p\eta'}} \sum_{l=1}^{k} \frac{h_l}{l^{1+\varepsilon-p\eta'}}\right). \end{aligned}$$

Then, it suffices to estimate $\sum_{l=1}^{k} h_l / l^{1+\varepsilon-p\eta'}$.

First, when $p\eta' < \varepsilon$, by $\ln h_k = o(\ln k)$, we have $h_k < k^{(\varepsilon - p\eta')/2}$ for sufficiently large k, which implies $\sum_{l=1}^{\infty} \frac{h_l}{l^{1+\varepsilon-p\eta'}} < \infty$. So, we can get

$$\sum_{l=1}^{k} \prod_{i=l+1}^{k} \left(1 - \frac{\eta'}{i+i_0}\right)^p \frac{h_l}{l^{1+\varepsilon}} = O\left(\frac{1}{k^{p\eta'}}\right).$$

Second, by the monotonicity of h_k, we have

$$\begin{aligned} \sum_{l=1}^{k} \frac{h_l}{l} &\le \sum_{l=1}^{k} h_l \left(\ln l - \ln(l-1)\right) \\ &\le \sum_{l=1}^{k} \left(h_l \ln l - h_{l-1} \ln(l-1)\right) = h_k \ln k. \end{aligned}$$

Hence, when $p\eta' = \varepsilon$, one can get

$$\sum_{l=1}^{k} \prod_{i=l+1}^{k} \left(1 - \frac{\eta'}{i+i_0}\right)^p \frac{h_l}{l^{1+\varepsilon}} = O\left(\frac{h_k \ln k}{k^\varepsilon}\right).$$

Lastly, when $p\eta' > \varepsilon$, we have

$$\sum_{l=1}^{k} \frac{h_l}{l^{1+\varepsilon-p\eta'}} = O\left(\sum_{l=1}^{k} h_l\left(l^{p\eta'-\varepsilon} - (l-1)^{p\eta'-\varepsilon}\right)\right) \leq O\left(\sum_{l=1}^{k}\left(h_l l^{p\eta'-\varepsilon} - h_{l-1}(l-1)^{p\eta'-\varepsilon}\right)\right) = O\left(h_k k^{p\eta'-\varepsilon}\right)$$

which implies

$$\sum_{l=1}^{k}\prod_{i=l+1}^{k}\left(1-\frac{\eta'}{i+i_0}\right)^p \frac{h_l}{l^{1+\varepsilon}} = O\left(\frac{h_k}{k^{\varepsilon}}\right).$$

Remark A.3: If h_k is constant, $p = 1$ and $i_0 = 0$, then Lemma A.5 implies [29, Lemma 4]. Besides, if $h_k/\ln k$ is assumed to be monotonically decreasing, then the estimate of Lemma A.5 is accurate.

Theorem A.1: Under the condition of Lemma 2, for any $\varepsilon \in (0, 1)$, there exist positive numbers ε' and m such that

$$\mathbb{P}\left\{\|\psi_k\| > k^{-\varepsilon'}\right\} = O\left(\exp\left(-mk^{1-\varepsilon}\right)\right).$$

Proof: The theorem can be proved by verifying that there exists $\varepsilon' > 0$ such that

$$\left\{\|\psi_k\| \leq k^{-\varepsilon'}\right\} \supseteq \left\{\sup_{j\geq\lfloor k^{1-2\varepsilon}\rfloor} j^{\varepsilon}\|w_j\| \leq 1\right\}. \quad \text{(A.10)}$$

By the monotonicity of $\{\sup_{j\geq k} j^{\varepsilon}\|w_j\| \leq 1\}$ and Theorem 1

$$\left\{\sup_{j\geq\lfloor k^{1-\varepsilon}\rfloor}\|\psi_j\|^2 < M'\right\} \supseteq \left\{\sup_{j\geq\lfloor k^{1-2\varepsilon}\rfloor} j^{\varepsilon}\|w_j\| \leq 1\right\}.$$

Therefore, if $\sup_{j\geq\lfloor k^{1-2\varepsilon}\rfloor} j^{\varepsilon}\|w_j\| \leq 1$, then by (18) and (19), for all $j \geq \lfloor k^{1-\varepsilon}\rfloor + N$

$$\|\psi_j\| \leq \left(1 - \frac{\beta\delta N}{j}\underline{f}\left(M\sqrt{M'}\right)\right)\|\psi_{j-N}\| + O\left(\frac{1}{j^{1+\varepsilon}}\right)$$

where $\underline{f}(\cdot)$ is defined in (14). Then, by Corollary A.2 and Lemma A.5, $\|\psi_j\|$ converges at a polynomial rate. Hence, we get (A.10). Then, the theorem can be proved by Lemma 1 and the arbitrariness of ε. ■

Corollary A.3: Under the condition of Theorem 2, for any $\varepsilon > 0$, there exist positive numbers ε' and m such that

$$\mathbb{P}\left\{\|\tilde{\theta}_k\| > k^{-\varepsilon'}\right\} = O\left(\exp\left(-mk^{1-\varepsilon}\right)\right).$$

Proof: By (A.10) and $\tilde{\theta}_k = \psi_k + w_k$, we have

$$\left\{\|\tilde{\theta}_k\|^2 \leq k^{-\varepsilon'} + k^{-\varepsilon}\right\} \supseteq \left\{\|\psi_k\| \leq k^{-\varepsilon'}\right\} \cap \left\{\|w_k\| \leq k^{-\varepsilon}\right\} \supseteq \left\{\sup_{j\geq\lfloor k^{1-2\varepsilon}\rfloor} j^{\varepsilon}\|w_j\| \leq 1\right\}.$$

Then, the corollary can be proved by Lemma 1. ■

Remark A.4: Theorem A.1 and Corollary A.3 are extensions of Theorems 1 and 2, respectively.

Appendix B
Other Application of SPAO

First, the construction of SPAO can be applied to many online identification algorithms of binary-valued systems. For binary-valued systems with threshold C_k, a large number of recursive identification algorithms can be represented as

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \rho_k v_k\left(h(\phi_k, \hat{\theta}_{k-1}) - s_k\right)$$

where $\{\phi_k, k \geq 1\}$ are independent regressed function of inputs, and C_k and v_k are generated by $\{\phi_j, s_{j-1}, j \leq k\}$ [16], [17], [18], [19], [30], [31], [32], [33], [34]. The step-size ρ_k can also be matrices [33], [34].

Define $\psi_k = \tilde{\theta}_k - w_k$, where $\tilde{\theta}_k = \hat{\theta}_k - \theta$ is the estimation error and

$$w_k = \rho_k\left(\sum_{i=1}^{k} v_i\left(\mathbb{E}\left[s_i|\phi_j, s_{j-1}, j \leq i\right] - s_i\right)\right) = \rho_k\left(\sum_{i=1}^{k} v_i\left(F(C_i - \phi_i^{\top}\theta) - s_i\right)\right).$$

Then, one can get

$$\psi_k = \psi_{k-1} + \rho_k\left(\rho_k^{-1} - \rho_{k-1}^{-1}\right)w_{k-1} + \rho_k v_k\left(h(\phi_k, \psi_{k-1} + w_{k-1} + \theta) - F(C_k - \phi_k^{\top}\theta)\right).$$

If there is a good convergence property for w_k, then the trajectory of ψ_k is similar to that of $\tilde{\theta}_k$ and that of the deterministic sequence

$$\overline{\psi}_k = \overline{\psi}_{k-1} + \rho_k v_k\left(h(\phi_k, \overline{\psi}_{k-1} + \theta) - F(C_k - \phi_k^{\top}\theta)\right).$$

Therefore, we can analyze the convergence property of the algorithm through ψ_k.

Second, SPAO technique can be applied in the robustness analysis of Algorithm (5). If the noise distribution used in our algorithm $F(\cdot)$ is different from the true noise distribution $F_{\text{true}}(\cdot)$, then by the SPAO technique, we can prove that under the condition of Theorem 2, the following holds:

$$\overline{\lim_{k\to\infty}}\|\tilde{\theta}_k\|^2 \leq M''(\Delta_F), \text{ a.s.}$$

where $\Delta_F = \sup_{x\in\mathbb{R}}|F(x) - F_{\text{true}}(x)|$, and $M''(\cdot)$ is a positive function satisfying $\lim_{\Delta_F\to 0} M''(\Delta_F) = 0$. The detailed analysis is similar to Theorem 1, and hence, omitted here.

References

[1] 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 *Proc. ASME Int. Mech. Eng. Congr. Expo.*, 2002, pp. 777–785.

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

[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] G. Kang et al., "A robust and powerful set-valued approach to rare variant association analyses of secondary traits in case-control sequencing studies," *Genetics*, vol. 205, no. 3, pp. 1049–1062, 2017.

[5] 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," *Amer. J. Hum. Genet.*, vol. 108, no. 5, pp. 825–839, 2021.

[6] T. Wang, W. Bi, Y. L. Zhao, and W. Xue, “Radar target recognition algorithm based on RCS observation sequence: Set-valued identification method,” *J. Syst. Sci. Complexity*, vol. 29, pp. 573–588, 2016.

[7] X. Wang, M. Hu, Y. L. Zhao, and B. Djehiche, “Credit scoring based on the set-valued identification method,” *J. Syst. Sci. Complexity*, vol. 33, pp. 1297–1309, 2020.

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

[9] E. Colinet and J. Juillard, “A weighted least-squares approach to parameter estimation problems based on binary measurements,” *IEEE Trans. Autom. Control*, vol. 55, no. 1, pp. 148–152, Jan. 2010.

[10] 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.

[11] F. Gustafsson and R. Karlsson, “Statistical results for system identification based on quantized observations,” *Automatica*, vol. 45, no. 12, pp. 2794–2801, 2009.

[12] R. S. Risuleo, G. Bottegal, and H. Hjalmarsson, “Identification of linear models from quantized data: A midpoint-projection approach,” *IEEE Trans. Autom. Control*, vol. 65, no. 7, pp. 2801–2813, Jul. 2020.

[13] X. Shen, P. K. Varshney, and Y. Zhu, “Robust distributed maximum likelihood estimation with dependent quantized data,” *Automatica*, vol. 50, no. 1, pp. 169–174, 2014.

[14] Y. L. Zhao, H. Zhang, T. Wang, and G. Kang, “System identification under saturated precise or set-valued measurements,” *Sci. China Inf. Sci.*, vol. 66, 2023, Art. no. 112204.

[15] L. Ljung and E. T. Söderström, *Theory and Practice of Recursive Identification*, Cambridge, MA, USA: MIT Press, 1983.

[16] B. C. Csóji and E. Weyer, “Recursive estimation of ARX systems using binary sensors with adjustable thresholds,” in *Proc. 16th IFAC Symp. System Identification*, 2012, pp. 1185–1190.

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

[18] Z. Huang and Q. Song, “Identification of linear systems using binary sensors with random thresholds,” *J. Syst. Sci. Complexity*, vol. 37, no. 3, pp. 907–923, 2024.

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

[20] J. D. Diao, J. Guo, and C. Sun, “A compensation method for the packet loss deviation in system identification with event-triggered binary-valued observations,” *Sci. China Inf. Sci.*, vol. 63, no. 12, 2020, Art. no. 229204.

[21] Q. He, G. G. Yin, and L. Y. Wang, “Moderate deviations analysis for system identification under regular and binary observations,” in *Proc. Conf. Control Appl.*, 2013, pp. 51–58.

[22] 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.

[23] H. Mei, L. Y. Wang, 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] A. Moschitta, J. Schoukens, and P. Carbone, “Parametric system identification using quantized data,” *IEEE Trans. Instrum. Meas.*, vol. 64, no. 8, pp. 2312–2322, Aug. 2015.

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

[26] L. Y. Wang, G. G. Yin, and J. F. Zhang, “Joint identification of plant rational models and noise distribution functions using binary-valued observations,” *Automatica*, vol. 42, no. 4, pp. 535–547, 2006.

[27] L. Y. Wang, G. G. Yin, Y. L. Zhao, and J. F. Zhang, “Identification input design for consistent parameter estimation of linear systems with binary-valued output observations,” *IEEE Trans. Autom. Control*, vol. 53, no. 4, pp. 867–880, May 2008.

[28] Y. L. Zhao, J. F. Zhang, L. Y. Wang, and G. G. Yin, “Identification of Hammerstein systems with quantized observations,” *SIAM J. Control Optim.*, vol. 48, no. 7, pp. 4352–4376, 2010.

[29] Y. L. Zhao, T. Wang, and W. Bi, “Consensus protocol for multi-agent systems with undirected topologies and binary-valued communications,” *IEEE Trans. Autom. Control*, vol. 64, no. 1, pp. 206–221, Jan. 2019.

[30] K. Fu, H. F. Chen, and W. X. Zhao, “Distributed system identification for linear stochastic systems with binary sensors,” *Automatica*, vol. 141, 2022, Art. no. 110298.

[31] 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.

[32] 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.

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

[34] 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, May 2021.

[35] H. F. Chen and L. Guo, “Adaptive control via consistent estimation for deterministic systems,” *Int. J. Control*, vol. 45, no. 6, pp. 2183–2202, 1987.

[36] H. F. Chen, *Stochastic Approximation and Its Applications*, vol. 64, New York, USA: Springer, 2006.

[37] Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*, New York, USA: Springer, 1997.

[38] V. Fabian, “On asymptotic normality in stochastic approximation,” *Ann. Math. Stat.*, vol. 39, no. 4, pp. 1327–1332, 1968.

[39] V. A. Zorich, *Mathematical analysis I*. Berlin, Germany: Springer, 2016.

[40] L. Guo, *Introduction to Control Theory: From Basic Concepts to Research Frontier*, Beijing, China: Science Press, 2005.

[41] H. G. Tucker, *A Graduate Course in Probability*, New York, NY, USA: Academic Press, 1967.

[42] J. M. Wang, J. M. Ke, and J. F. Zhang, “Differentially private bipartite consensus over signed networks with time-varying noises,” *IEEE Trans. Autom. Control*, early access, Jan. 09, 2024, doi: 10.1109/TAC.2024.3351869.

Jieming Ke (Student Member, IEEE) received the B.S. degree in mathematics from the University of Chinese Academy of Science, Beijing, China, in 2020. He is currently working toward the Ph.D. degree majoring in system theory with the Academy of Mathematics and Systems Science, Chinese Academy of Science (CAS), Beijing. His research interests include identification and control of quantized systems and the information security problems of control systems.

Ying Wang (Member, IEEE) received the B.S. degree in mathematics from Wuhan University, Wuhan, China, in 2017, 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 2022.

She is currently a Postdoctoral Research Associate with AMSS, CAS. Her research interests include identification and control of quantized systems, and multiagent systems.

Yanlong Zhao (Senior Member, IEEE) 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 full Professor. His research interests include identification and control of quantized systems, and information theory and modeling of financial systems.

Dr. Zhao has been a Deputy Editor-in-Chief of *Journal of Systems Science and Complexity*, an Associate Editor for *Automatica*, *SIAM Journal on Control and Optimization*, and IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS: SYSTEMS. He was a Vice-President of Asian Control Association and a Vice-President of IEEE CSS Beijing Chapter, and is now a Vice General Secretary of Chinese Association of Automation (CAA) and the Chair of Technical Committee on Control Theory, CAA.

Ji-Feng Zhang (Fellow, IEEE) received the B.S. degree in mathematics from Shandong University, Jinan, China, in 1985, and the Ph.D. degree in majoring in operational research and cybernetics from the Institute of Systems Science (ISS), Chinese Academy of Sciences (CAS), Beijing, China, in 1991.

Since 1985, he has been with the ISS, CAS. His current research interests include system modeling, adaptive control, stochastic systems, and multiagent systems.

Dr. Zhang is an IFAC Fellow, CAA Fellow, SIAM Fellow, Member of the European Academy of Sciences and Arts, and Academician of the International Academy for Systems and Cybernetic Sciences. He was the recipient of the Second Prize of the State Natural Science Award of China in 2010 and 2015, respectively. He was a Vice-President of the Chinese Association of Automation, the Chinese Mathematical Society, and the Systems Engineering Society of China. 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. He was the Editor-in-Chief, Deputy Editor-in-Chief, or Associate Editor for more than ten journals, including *Science China Information Sciences*, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and *SIAM Journal on Control and Optimization*, etc.