RECURSIVE IDENTIFICATION OF SET-VALUED SYSTEMS 
UNDER UNIFORM PERSISTENT EXCITATIONS *

JIEMING KE†, YING WANG†, YANLONG ZHAO†, AND JI-FENG ZHANG†

Abstract. This paper studies the control-oriented identification problem of set-valued moving average systems with uniform persistent excitations and observation noises. The system outputs are measured via a binary sensor with a fixed threshold. Among previous works, the recursive projection method requires a priori information on the unknown parameters, and the empirical measurement method requires periodic inputs for effectiveness analysis. In this paper, a stochastic approximation-based (SA-based) algorithm is proposed. The algorithm overcomes the limitations of the two methods by combining their strengths. To analyze the convergence property of the algorithm, the distribution tail of the estimation error is proved to be exponentially convergent through an auxiliary stochastic process. Based on this key technique, the SA-based algorithm appears to be the first to reach the almost sure convergence rate of $O(\sqrt{\ln \ln k/k})$ theoretically in the non-periodic input case. Meanwhile, the mean square convergence is proved to have a rate of $O(1/k)$, which is the best one even under accurate observations. A numerical example is given to demonstrate the effectiveness of the proposed algorithm and theoretical results.

Key words. Set-valued system, parameter estimation, stochastic approximation, convergence, convergence rate.

AMS subject classifications. 93E12, 93C55, 62L20

1. Introduction. Set-valued systems emerge widely in practical fields, such as genetic association studies [1], ambient assisted living [11], gravity gradient satellites [25], radar target recognition [31], credit scoring [33], and artificial neural networks [19, 30, 37], etc. Therefore, such systems have received much attention in the past two decades. For example, identification problems have been widely researched for such systems since [30]. Adaptive control laws and consensus protocols of set-valued multi-agent systems have also been developed based on identification methods with periodic inputs or projection algorithms [17, 32, 38, 39]. However, there are still limitations on existing control-oriented identification methods. Under this background, we continue the development along the direction of seeking more general control-oriented identification methods for set-valued systems.

There are some excellent identification algorithms proposed for set-valued systems [2, 5, 7, 12, 15, 22, 23], many of which are offline. Offline methods take full advantage of the statistical property of the set-valued outputs, and require fewer assumptions than the online ones. However, recursive forms of the corresponding offline methods are difficult to be constructed and analyzed. And, the computational complexity increases geometrically when offline identification methods are applied to system controls. Therefore, control-oriented identification methods are required to be online.

Moreover, the control-oriented identification methods are supposed to keep a freedom for the input design. Many efficient online identification methods require

*Submitted to the editors December 23, 2021.

Funding: The work is supported by National Key R&D Program of China under Grant 2018YFA0703800, National Natural Science Foundation of China under Grants 62025306 and 61877057.

†Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People’s Republic of China; and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, People’s Republic of China (kejieming@amss.ac.cn, wangying96@amss.ac.cn, ylzhao@amss.ac.cn, jif@iss.ac.cn).

This manuscript is for review purposes only.
strict conditions for inputs. For example, \([8, 24]\) propose stochastic approximation algorithms with expanding truncations for set-valued systems. \([35]\) identifies the parameters with a stochastic gradient based estimation method. Strong consistency is obtained for all these algorithms, but the strong consistency relies on independent and identically distributed (i.i.d.) inputs, which seems not suitable to design control laws. Besides, the empirical measurement method for set-valued systems is proved to be strongly convergent and achieves Cramér-Rao lower bounds asymptotically \([9, 16, 18, 20, 27, 28, 29, 30, 40]\). However, the effectiveness analysis of the empirical measurement method is based on periodic input signals, which cause that the corresponding adaptive control laws waste a lot of information \([17, 38, 39]\).

Identification algorithms with potential in effective control design have been proposed for set-valued output systems \([13, 32, 34, 36, 37]\), but there are still limitations in these works. The identification method proposed in \([36]\) relies on designable quantizers. However, most physical sensors that produce set-valued signals are undesignable and even time-invariant. In the time-invariant quantizer case, \([13, 32, 34, 37]\) raise recursive projection algorithms, which rely on projection operators to restrict the search region in a compact set. But, projection operators require a priori information about the approximate location of the unknown parameters. Therefore, an online control-oriented identification method is desired for the case where the quantizer is undesignable and there is no priori information on the parameters.

The main difficulty of the online algorithm design lies in the trade-off between little information that one set-valued measurement contains and the features of online algorithms. On the one hand, a single observation only contains binary information, hence the identification of set-valued systems requires the accumulation of a number of output signals. On the other hand, when we update online algorithms, only the present or recent several signals can be used. As a result, it is difficult to reveal how the accumulation of the set-valued signals affects the trend of the online algorithm.

In the periodic input case, the solution of the empirical measurement method is based on the online calculation of the observation average \([9, 16, 17, 18, 20, 27, 28, 29, 30, 38, 39, 40]\). However, the method is hard to extend to the general case, because in the general case, the average does not even converge. In the general case, when the quantizer is designable, \([36]\) adopts the Lloyd-Max quantizer, whose quantized error is a white noise when the estimate is sufficiently close to the true value. But, the technique requires a designable quantizer. When the quantizer is fixed, \([13, 32, 34, 37]\) propose a recursive projection method based on stochastic approximation. However, due to the application of the projection, the search region is constrained. So, priori information is required for the recursive projection method to ensure that the parameters are located in the given search region.

To overcome the limitations in these existing works, some ideas of the recursive projection and empirical measurement methods are instructive. Inspired by the recursive projection method, the paper constructs an SA-based algorithm without projection. Then, the uniform persistent excitation condition suffices for the algorithm. And, the effectiveness analysis of the algorithm is inspired by the empirical measurement method. For the mean square convergence analysis of the empirical measurement method without truncation, \([39]\) estimates the distribution tail of the average of binary observations to deal with the low probability and undesired case. Enlightened by the idea, the paper analyzes the effectiveness of the SA-based algorithm based on the distribution tail estimation of the observation average.

A stochastic process with averaged observations (SPAO) is proposed for the realization of the idea. SPAO approaches the estimation error exponentially. Hence, there

This manuscript is for review purposes only.
is a high correlation between the convergence of SPAO and the algorithm. Besides, the update of SPAO depends on the average of observations. Therefore, there is also a high correlation between the distribution tails of SPAO and the observation average. Then, the effectiveness of the SA-based algorithm is verified by the distribution tail of the average as the empirical measurement method.

In the paper, an SA-based algorithm is proposed for set-valued moving average (MA) systems. The quantizer is fixed and no priori information on the unknown parameters is required. The inputs are only assumed to be bounded and uniformly persistently exciting as in the recursive projection method [13, 32], which guarantees persistent outputs and meanwhile keeps freedom for input design. The contributions of the paper are as follows.

i) A new SA-based identification algorithm is proposed for set-valued MA systems. In the non-periodic non-i.i.d. inputs case, the algorithm appears to be the first online identification algorithm of set-valued systems without requiring the priori information of the unknown parameters.

ii) The distribution tail of the estimation error is proved to be exponentially convergent, which induces almost sure and mean square convergence. Besides, when properly selecting coefficients, almost sure and mean square convergence rates are proved to reach $O(\sqrt{\ln \ln k/k})$ and $O(1/k)$, respectively. In the non-periodic input case, our algorithm seems to be the first to reach $O(\sqrt{\ln \ln k/k})$ among online identification algorithms of stochastic set-valued systems. And, $O(1/k)$ is the best mean square convergence rate in theory under set-valued observations and even accurate ones.

iii) A new constructive methodology is developed for the convergence analysis of set-valued system identification algorithms. For the SA-based algorithm, an auxiliary stochastic process named SPAO is constructed to connect the average of the set-valued observations and the convergence of the algorithm. Then, the algorithm can be analyzed through the observation average. The methodology is also shown to be practical for a common class of recursive identification algorithms of set-valued systems.

The rest of the paper is organized as follows. Section 2 formulates the identification problem. Section 3 constructs an SA-based identification algorithm of set-valued systems. The convergence analysis is given in Section 4. Subsection 4.1 constructs an auxiliary stochastic process named SPAO and discusses its property. Based on SPAO, Subsection 4.2 estimates the distribution tail of the estimation error, and gives the almost sure and mean square convergence. Almost sure and mean square convergence rates are estimated in Subsection 4.3 and Subsection 4.4, respectively. A numerical example is simulated in Section 5 to demonstrate the theoretical results. Concluding remarks and future works are given in Section 6.

Notation. In the rest of the paper, $\mathbb{R}$ and $\mathbb{R}^n$ are the sets of real numbers and $n$-dimensional real vectors, respectively. $I_{\{\}}$ denotes the indicator function, whose value is 1 if its argument (a formula) is true, and 0, otherwise. $\|x\| = \|x\|_2$ 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 $A_1, \ldots, A_k$, denote $\prod_{i=1}^{k} A_i = A_k \cdots A_1$ for $k \geq l$ and $\prod_{i=k+1}^{k} A_i = I_n$. Relations between two series $a_k$ and $b_k$ are defined as

i) $a_k = O(b_k)$ if $a_k = c_k b_k$ for an ultimately bounded $c_k$ as $k$ goes to $\infty$;

ii) $a_k = o(b_k)$ if $a_k = c_k b_k$ for a $c_k$ that converges to 0 as $k$ goes to $\infty$. 

This manuscript is for review purposes only.
2. Problem formulation. Consider the MA system:

\[ y_k = \phi_k^T \theta + d_k, \quad k \geq 1, \]

where \( \phi_k = \phi(u_k, u_{k-1}, \ldots, u_{k-n+1}) \in \mathbb{R}^n \) is a regressed function of inputs \( u_k \) for some \( n > 0 \), \( \theta \in \mathbb{R}^n \) 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

\[ s_k = I\{y_k \leq C\} = \begin{cases} 1, & y_k \leq C; \\ 0, & y_k > C. \end{cases} \]

Our goal is to identify the unknown parameter \( \theta \) based on the regressed vector \( \phi_k \) and the binary observation \( s_k \) without priori information on \( \theta \).

**Assumption 2.1.** The sequence \( \{\phi_k, k \geq 1\} \) is uniformly 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

\[ \frac{1}{N} \sum_{i=k}^{k+N-1} \phi_i \phi_i^T \geq \delta I_n, \quad k \geq 1. \]

**Remark 2.2.** The condition (2.4) is usually called “uniform persistent excitation condition” or “sufficiently rich condition” [4, 13]. For accurate and deterministic systems, Assumption 2.1 is required for the exponential convergence of identification methods [4]. For the identification problem of System (2.1) and (2.2), Assumption 2.1 is also adopted in recursive projection algorithms [13, 32].

**Assumption 2.3.** The system noise \( \{d_k, k \geq 1\} \) is a sequence of i.i.d. random variables with zero mean and finite covariance \( \sigma^2 \), whose distribution and density function are supposed to be known and denoted as \( F(\cdot) \) and \( f(\cdot) \), respectively. The distribution \( F(\cdot) \) is Lipschitz continuous, and the density function \( f(\cdot) \) satisfies

\[ \inf_{x \in \mathcal{X}} f(x) > 0 \]

for any bounded open set \( \mathcal{X} \).

For simplicity of notation, denote

\[ F_k = F(C - \phi_k^T \theta), \quad f_k = f(C - \phi_k^T \theta). \]

Then \( \mathbb{E}s_k = P\{y_k \leq C\} = F_k \).

**Remark 2.4.** Under Assumption 2.3, the density function \( f(\cdot) \) is bounded because of the Lipschitz continuity of the distribution function \( F(\cdot) \).

**Remark 2.5.** Gaussian noise, Laplacian noise and \( t \)-distribution noise are all examples satisfying Assumption 2.3. Moreover, Assumption 2.3 can be extended to the unknown but parameterizable noise distribution case. In this case, parameters of the noise distribution and \( \theta \) can be identified jointly in a similar way of identifying \( \theta \) alone [28, 29]. Besides, if (2.5) does not hold for the system noise, we can add a dither to the binary sensor [30].
3. Identification algorithm. In this section, we will give an SA-based algorithm for the MA system (2.1) with binary observation (2.2).

The realization of the SA-based algorithm is inspired by a new viewpoint for the system. Note that $F_k$ contains the entire information of $\hat{\phi}^\top \theta$. Hence, we treat $F_k$ as an accurate but nonlinear output. The rest of the observation, i.e., $s_k - F_k$, is treated as an independent noise. Under Assumptions 2.1 and 2.3, $\hat{\theta} = \theta$ if and only if

$$
\sum_{i=k}^{k+N-1} \phi_i^\top (\theta - \hat{\theta}) \left( F - \phi_i^\top \hat{\theta} - F_i \right) = 0.
$$

Then, the SA-based algorithm is designed as

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

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

where $\rho_k$ is the step size satisfying $\sum_{i=1}^\infty \rho_i = \infty$ and $\sum_{i=1}^\infty \rho_i^2 < \infty$.

Denote

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

Set $\rho_k = \beta/k$, where $\beta > 0$ is a constant coefficient. Then, we have the SA-based algorithm

$$
\hat{\theta}_k = \hat{\theta}_{k-1} + \frac{\beta \phi_k}{k} \left( \hat{F}_k - s_k \right).
$$

Remark 3.1. In addition to $\rho_k = \beta/k$, other types of step sizes can be applied to the SA-based algorithm. For example, $\rho_k = \beta_k/(1 + \sum_{i=1}^k \|\phi_i\|^2)$ for $0 < \beta \leq \beta_k < \infty$ is an alternative step size. Besides, in Algorithm (3.2), $\hat{F}_k$ 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 (3.2) also works after replacing $\hat{F}_k$ with $\mathbb{E}[s_q(\theta)|\theta = \hat{\theta}_{k-1}]$, where $s_q$ is the corresponding observation in $\{0, 1, \ldots, q\}$.

4. 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 firstly to assist in the analysis.

4.1. Stochastic process with averaged observations (SPAO). In this subsection, we will introduce an auxiliary stochastic process satisfying

i) the trajectory of the stochastic process gradually approaches that of the estimation error $\hat{\theta}_k$;

ii) 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 $T_k$ and $T_{k-1}$, where

$$
T_k = \sum_{i=1}^k \beta \phi_i (F_i - s_i) / k.
$$
243 i.e.,
\[ (4.2) \quad \beta \phi_k(F_k - s_k) = \sum_{i=1}^{k} \beta \phi_i(F_i - s_i) - \sum_{i=1}^{k-1} \beta \phi_i(F_i - s_i) = k(T_k - T_{k-1}) + T_{k-1}. \]

Define \( \psi_k = \bar{T}_k - T_k \). Then, by the transformation above,
\[ (4.3) \quad \psi_k = \psi_{k-1} + \frac{\beta \phi_k}{k} \left( F(\bar{C} - \phi_k \bar{\theta} - \phi_k \bar{T}_{k-1}) - F(\bar{C} - \phi_k \bar{\theta}) \right) + \frac{T_{k-1}}{k}. \]

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

**Remark 4.1.** For general stochastic approximation methods, \( T_k \) is also used to verify the robustness of the algorithm (cf. [3], Assumption 2.7.3 and Theorem 2.7.1).

The following lemma estimates the distribution tail of \( T_k \).

**Lemma 4.2.** Let \( T_k \) be defined in (4.1), and assume that
i) \( \phi_k \in \mathbb{R}^n \) is bounded;
ii) \( s_k \in \{0, 1\} \) is an independent binary random variable with expectation \( F_k \);
iii) \( \sup_k F_k < 1 \) and \( \inf_k F_k > 0 \).

Then, for any \( \varepsilon \in (0, \frac{1}{2}) \), there exists \( m > 0 \) such that
\[ \mathbb{P} \left\{ \sup_{j \geq k} j^\varepsilon \|T_j\| > 1 \right\} = O \left( \exp(-mk^{1-2\varepsilon}) \right). \]

The proof is given in Appendix A.1.

Based on Lemma 4.2, we will give properties of \( \psi_k \) in the following propositions.

**Proposition 4.3.** Assume that
i) System (2.1) with binary observation (2.2) satisfies Assumptions 2.1 and 2.3;
ii) \( T_k \) is defined in (4.1), and \( \psi_k = \bar{T}_k - T_k \).

Then, we have
(a) for any \( \varepsilon \in (0, \frac{1}{2}) \), there exists \( m > 0 \) such that \( \mathbb{P} \left\{ \|\bar{T}_k - \psi_k\| > k^{-\varepsilon} \right\} = O \left( \exp (-mk^{1-2\varepsilon}) \right) \);
(b) \( \|\bar{T}_k - \psi_k\| = O \left( \sqrt{\ln \ln k/k} \right) \), a.s.;
(c) \( \mathbb{E} \|\bar{T}_k - \psi_k\|^2 = O(1/k) \).

Proof. Since \( \bar{T}_k - \psi_k = T_k \), the three parts of the proposition can be obtained immediately from Lemma 4.2, the law of the iterated logarithm ([6], Theorem 10.2.1) and \( \mathbb{E} \|T_k\|^2 = O(1/k) \), respectively. \( \square \)

**Remark 4.4.** Proposition 4.3 describes the distance between \( \psi_k \) and the estimation error \( \bar{T}_k \) in three different senses. By Proposition 4.3, the trajectory of \( \psi_k \) is similar to that of \( \bar{T}_k \). Therefore, we can analyze the convergence property of the algorithm through \( \psi_k \).

The following proposition analyzes the convergence of \( \psi_k \).

**Proposition 4.5.** Under the condition of Proposition 4.3, for any \( M' > 0 \) and \( \varepsilon \in (0, \frac{1}{2}) \), when \( k \) is sufficiently large,
\[ \{ \|\psi_k\|^2 < M' \} \supseteq \left\{ \sup_{j \geq \lfloor k^{1-\varepsilon} \rfloor} j^\varepsilon \|T_j\| \leq 1 \right\}. \]
Furthermore, there exists $m > 0$ such that

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

**Proof.** Set $k_0 = \lfloor k^{1-\epsilon} \rfloor$ and $k_0' = k - N \lfloor k^{1-\epsilon}/N \rfloor$. It is worth mentioning that $k_0' \in [k_0, k_0 + N - 1]$, and $k - k_0'$ is divisible by $N$. Assume that $\sup_{j \geq k_0} \|T_j\| \leq 1$ is true in the rest of the proof. Then, it suffices to prove that $\|\psi_k\|^2 < M'$.

We firstly simplify the recursive formula of $\|\psi_k\|^2$. By (4.3) and the monotonicity and Lipschitz continuity of $F(\cdot)$, for any positive real number $b$, we have

$$\|\psi_k\|^2 \leq \|\psi_{k-1}\|^2 + \frac{2\beta \phi_k^T \psi_{k-1}}{k} \left(F(C - \phi_k^T \theta - \phi_k^T T_{k-1} - \phi_k^T \psi_{k-1}) - F(C - \phi_k^T \theta)\right)$$

$$+ \frac{2\psi_{k-1}^T T_{k-1}}{k} + \left(\beta \|\phi_k\| + \|T_{k-1}\|\right)^2/k^2$$

$$= \|\psi_{k-1}\|^2 + \frac{2\beta \phi_k^T \psi_{k-1}}{k} \left(F(C - \phi_k^T \theta - \phi_k^T T_{k-1} - \phi_k^T \psi_{k-1}) - F(C - \phi_k^T \theta - \phi_k^T T_{k-1}) + O\left(k^{-1-\epsilon/2}\right)\right)$$

$$\leq \|\psi_{k-1}\|^2 + \frac{2\beta \phi_k^T \psi_{k-1}}{k} \left(F(C - \phi_k^T \theta - \phi_k^T T_{k-1} - b) - F(C - \phi_k^T \theta - \phi_k^T T_{k-1})\right) I_{\{\phi_k^T \psi_{k-1} \geq b\}}$$

$$+ \frac{2\beta \phi_k^T \psi_{k-1}}{k} \left(F(C - \phi_k^T \theta - \phi_k^T T_{k-1} + b) - F(C - \phi_k^T \theta - \phi_k^T T_{k-1})\right) I_{\{\phi_k^T \psi_{k-1} \leq -b\}} + O\left(k^{-1-\epsilon/2}\right).$$

By Assumption 2.3 and the boundedness of $C - \phi_k^T \theta - \phi_k^T T_{k-1}$, there exists $B > 0$ such that

$$-2\beta \left(F(C - \phi_k^T \theta - \phi_k^T T_{k-1} - b) - F(C - \phi_k^T \theta - \phi_k^T T_{k-1})\right) > B,$$

$$2\beta \left(F(C - \phi_k^T \theta - \phi_k^T T_{k-1} + b) - F(C - \phi_k^T \theta - \phi_k^T T_{k-1})\right) > B,$$

which together with (4.5) implies

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

Set $b = \sqrt{\frac{M'}{2\sqrt{2}}}$. Then, by Lemma A.2, there exists $k' \in [k + 1, k + N]$ such that

$$\|\psi_{k+N}\|^2 \leq \|\psi_k\|^2 - \sum_{i=k+1}^{k+N} \frac{B \phi_i^T \psi_{i-1}}{i} I_{\{\phi_i^T \psi_{i-1} \geq \sqrt{\frac{M'}{2\sqrt{2}}}\}} + \sum_{i=k+1}^{k+N} O\left(i^{-1-\epsilon/2}\right)$$

$$\leq \|\psi_k\|^2 - \frac{B \phi_k^T \psi_{k'-1}}{k+N} I_{\{\phi_k^T \psi_{k'-1} \geq \sqrt{\frac{M'}{2\sqrt{2}}}\}} + O\left(k^{-1-\epsilon/2}\right)$$

$$\leq \|\psi_k\|^2 - \frac{B \sqrt{\delta}}{\sqrt{2}} \|\psi_k\| I_{\{\|\psi_k\| \geq \sqrt{\frac{M'}{2\sqrt{2}}}\}} + O\left(k^{-1-\epsilon/2}\right).$$
Hence, when \( k = k'_s + N(t - 1), \) we have

\[
\begin{aligned}
\|\psi_{k'_s + Nt}\|^2 &\leq \|\psi_{k'_s + N(t-1)}\|^2 - \frac{B\sqrt{\delta}}{2\delta} \|\psi_{k'_s + N(t-1)}\| I\{\psi_{k'_s + Nt} \| \geq M'\} \\
&+ O \left( (k'_s + Nt)^{-1-\varepsilon}/2 \right).
\end{aligned}
\]

(4.6)

Since \( \lim_{k \to \infty} k'_s = \infty, \) we have \( \lim_{k \to \infty} \sum_{t=1}^\infty \theta \left((k'_s + Nt)^{-1-\varepsilon}/2 \right) = 0. \) Then, by (4.6) and Lemma A.3, when \( k \) is sufficiently large, we have

\[
\|\psi_k\|^2 = \|\psi_{k'_s + N\left(\frac{k-1-\varepsilon}{N}\right)}\|^2
\]

< \max \left\{ M', \left[ \left\|\psi_{k'_s}\right\| - \frac{B\sqrt{\delta}}{22\delta} \ln \left( \left(\frac{k-1-\varepsilon}{N} + \frac{k'_s}{N} + 1\right) \right) \right]^+ \right\}^2 + \frac{M'}{2}.
\]

Recalling that \( k'_s = O(1)\), we have \( \frac{B\sqrt{\delta}}{22\delta} \ln \left( \left(\frac{k-1-\varepsilon}{N} + \frac{k'_s}{N} + 1\right) \right) \) is of the same order as \( \ln k \). By Corollary A.7, \( \psi_{k'_s} = O(\sqrt{\ln k} \) \). Therefore, when \( k \) is sufficiently large, \( \|\psi_k\|^2 < M', \) which proves the proposition.

**Remark 4.6.** Proposition 4.5 estimates the distribution tail of \( \psi_k \). The estimation can be promoted as in Proposition A.11 (cf. Appendix A.2). In addition, due to the arbitrariness of \( M' \), the almost sure and mean square convergence of \( \psi_k \) can be immediately obtained by Proposition 4.5. Notice that by Proposition 4.3, the convergence property of the estimation error \( \hat{\theta}_k \) is similar to that of \( \psi_k \). Therefore, we can analyze the convergence of the SA-based algorithm through Proposition 4.5.

It is worth noting that SPAO can be extended to a class of identification algorithms of the set-valued systems. The details are given in Appendix B.

### 4.2. Estimate of the distribution tail

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

**Theorem 4.7.** If System (2.1) with binary observations (2.2) satisfies Assumptions 2.1 and 2.3, then for any \( M' > 0 \) and \( \varepsilon > 0 \), there exists \( m > 0 \) such that

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

**Proof.** Reminding that \( \hat{\theta}_k = \psi_k + T_k \), by Proposition 4.5, for sufficiently large \( k \), we have

\[
\left\{ \sup_{j \geq k} \right\|\hat{\theta}_j\right\|^2 - M' \right\} \subseteq \left\{ \right\|\psi_k\right\|^2 < \frac{M'}{4} \right\} \cap \left\{ \right\|T_k\right\|^2 < \frac{M'}{4} \right\} \subseteq \left\{ \right\|\hat{\theta}_k\right\|^2 < M' \right\},
\]

and hence,

\[
\left\{ \sup_{j \geq k} \right\|\hat{\theta}_j\right\|^2 - M' \right\} \subseteq \bigcup_{j \geq k} \left\{ \sup_{j \geq k} \right\|\hat{\theta}_j\right\|^2 - M' \right\} = \left\{ \sup_{j \geq k} \right\|\hat{\theta}_j\right\|^2 - M' \right\}.
\]

So, by Lemma 4.2, \( \mathbb{P}\{ \sup_{j \geq k} \right\|\hat{\theta}_j\right\|^2 - M' \right\} = O \left( \exp(-m k^{1-\varepsilon(1-2\varepsilon)}) \right). \) Thus, the theorem can be proved by the arbitrariness of \( \varepsilon \).
Remark 4.8. Theorem 4.7 estimates the distribution tail of the estimation error \( \tilde{\theta}_k \). For the convergence analysis of identification algorithms, previous 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 stochastic approximation algorithms under different conditions (cf. [3], Section 3.3 and [10]). For the set-valued system with i.i.d. inputs and designable quantizer, [35] also analyzes the asymptotic normality of the algorithm. Compared with the asymptotic normality, Theorem 4.7 weakens the description of the distribution in the neighborhood of \( \theta \), but gives a better description on the exponential tail. This helps to obtain the almost sure and mean square convergence of the algorithm.

Theorem 4.9. Under the condition of Theorem 4.7, Algorithm (3.2) converges to \( \theta \) in both almost sure and mean square sense.

Proof. The almost sure convergence can be immediately obtained by Theorem 4.7. By Theorem 4.7 and Corollary A.6, for any \( M' > 0 \) and \( \varepsilon > 0 \), there exists \( m > 0 \) such that

\[
E\|\tilde{\theta}_k\|^2 = \int_{\{\|\tilde{\theta}_k\| < M'\}} \|\tilde{\theta}_k\|^2dP + \int_{\{\|\tilde{\theta}_k\| \geq M'\}} \|\tilde{\theta}_k\|^2dP < M' + O(\ln k \cdot \exp(-mk^{1-\varepsilon})) = M' + o(1).
\]

Thus, the mean square convergence can be obtained by the arbitrariness of \( M' \).

4.3. Almost sure convergence rate. In this subsection we will estimate the almost sure convergence rate of the SA-based algorithm.

Before the analysis, we define

\[
(4.8) \quad f(x) = \lim_{z \to (M\|\theta\|+x)^+} \inf_{t \in [C-z,C+z]} f(t) > 0, \quad \forall x \geq 0,
\]

and

\[
(4.9) \quad \underline{f}(x) = \inf_{t \in [C-M\|\theta\|,C+M\|\theta\|]} f(t).
\]

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

Remark 4.10. Under Assumption 2.1, \( f \) is the lower bound of \( f(C - \phi_k^\top \theta) \) for all possible \( \phi_k \), which motivates us to analyze \( \overline{f}(\cdot) \) and \( \underline{f}(\cdot) \). By (4.8), we give properties of \( \overline{f}(\cdot) \) and \( f \).

(a) \( \overline{f}(\cdot) \) is monotonically decreasing and right continuous.

(b) \( \lim_{x \to 0} f(x) = \inf f(t) = 0 \) and \( \lim_{x \to \infty} f(x) = 0 \). Then, by Proposition 4.5,

\[
\lim_{k \to \infty} f(k) = \lim_{k \to \infty} f(\phi_k^\top \psi_k) = f, \quad \text{a.s.}
\]

(c) The following inequality holds for \( f(\cdot) \).

\[
\overline{f}(x) \leq \inf_{t \in [C-M\|\theta\|+x,C+M\|\theta\|+x]} f(t),
\]

\[
\underline{f}(x) \leq \inf_{t \in [C-M\|\theta\|,C+M\|\theta\|]} f(t).
\]

The equality is achieved when \( f(\cdot) \) is continuous.

(d) If \( f(\cdot) \) is locally Lipschitz continuous, so is \( \underline{f}(\cdot) \).
The almost sure convergence rate of the algorithm can be achieved through that of \( \psi_k \).

**Proposition 4.11.** Under the condition of Proposition 4.3, for any \( \varepsilon > 0 \), we have

\[
\psi_k = \begin{cases} 
O \left( \sqrt{\frac{\ln \ln k}{k}} \right), & \eta > \frac{1}{2}; \\
O \left( \frac{1}{k^{\frac{1}{2}\eta}} \right), & \eta \leq \frac{1}{2}, 
\end{cases}
\]

where \( \eta = \beta \delta f \) with \( f \) defined in (4.9). If \( f(\cdot) \) is assumed to be locally Lipschitz continuous, then the almost sure convergence rate can be promoted into

\[
\psi_k = \begin{cases} 
O \left( \frac{\ln \ln k}{k} \right), & \eta > \frac{1}{2}; \\
O \left( \frac{1}{k^{\frac{1}{2}\eta}} \right), & \eta = \frac{1}{2}; \\
O \left( \frac{1}{k^{\eta}} \right), & \eta < \frac{1}{2}, 
\end{cases}
\]

**Proof.** The proof is based on Lemma A.9.

For the proof of (4.10), we firstly simplify the recursive formula of \( k \). In (4.3), by the Lagrange mean value theorem ([41], Theorem 5.3.1), there exists \( \xi_k \) between

\[
(C - \phi_k^T \theta - \phi_k^T \psi_{k-1} - C - \phi_k^T \theta),
\]

(4.12) \( F(C - \phi_k^T \theta - \phi_k^T T_{k-1} - \phi_k^T \psi_{k-1} - F(C - \phi_k^T \theta) = f(\xi_k)\phi_k^T \psi_{k-1} + O(T_{k-1}) \).

Then, by the law of the iterated logarithm ([6], Theorem 10.2.1),

\( F(C - \phi_k^T \theta - \phi_k^T T_{k-1} - \phi_k^T \psi_{k-1} - F_k = f(\xi_k)\phi_k^T \psi_{k-1} + O \left( \sqrt{\frac{\ln k}{k}} \right), \) a.s.,

which together with (4.3) implies

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

Note that except for the first few steps, we have

\[
\left\| \prod_{i=k-N+1}^k \left( I_n - \frac{\beta f(\xi_i)}{k} \phi_i \phi_i^T \right) \right\| 
\leq \left\| I_n - \frac{\beta f(\xi)}{k} \sum_{i=k-N+1}^k f(\xi_i) \phi_i \phi_i^T \right\| + O \left( \frac{1}{k^2} \right) 
\leq \left\| I_n - \frac{\beta}{k} \sum_{i=k-N+1}^k f(\xi_i) \phi_i \phi_i^T \right\| + O \left( \frac{1}{k^2} \right) 
\leq \left( 1 - \frac{\beta \delta N}{k} \right) F \left( \max_{k-N < i \leq k} | \phi_i^T \psi_{i-1} | \right) + O \left( \frac{1}{k^2} \right),
\]

(4.13) where \( f(\cdot) \) is defined in (4.8). Denote

\[
f_{k|N} = f \left( \max_{k-N < i \leq k} | \phi_i^T \psi_{i-1} | \right).
\]

This manuscript is for review purposes only.
Then, we have

\[ \| \psi_k \| \leq \left( 1 - \frac{\beta \delta N}{k} f_{k|N} \right) \| \psi_{k-N} \| + O \left( \sqrt{\frac{\ln \ln k}{k^3}} \right), \quad \text{a.s.} \]  

By Item (b) of Remark 4.10, \( \lim_{k \to \infty} f_{k|N} = f \) almost surely. Therefore, for any \( \varepsilon \in (0, \beta \delta f) \), there almost surely exists a \( k_0 \) such that for all \( k \geq k_0 \), we have \( \beta \delta f_{k|N} > \beta \delta f - \varepsilon \), and thus, by (4.14),

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

Without loss of generality, assume that \( k - k_a \) is divisible by \( N \). Then, by Lemma A.9, one can get

\[ \| \psi_k \| \leq \prod_{i=1}^{k-k_a} \left( 1 - \frac{N(\eta - \varepsilon)}{k_a + iN} \right) \| \psi_{k_a} \| \]

\[ + O \left( \sum_{i=1}^{k-k_a} \prod_{i=l+1}^{k-k_a} \left( 1 - \frac{N(\eta - \varepsilon)}{k_a + iN} \right) \sqrt{\frac{\ln \ln (k_a + iN)}{(k_a + iN)^3}} \right) \]

\[ = O \left( \frac{1}{k^{\eta - \varepsilon}} \right) + O \left( \sum_{i=1}^{k-k_a} \prod_{i=l+1}^{k-k_a} \left( 1 - \frac{N(\eta - \varepsilon)}{k_a + iN} \right) \sqrt{\frac{\ln(l + 2)}{l^3}} \right) \]

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

Notice that \( \varepsilon \) can be arbitrarily chosen, which implies that when \( \eta > 1/2 \), we can choose \( \varepsilon > 0 \) such that \( \eta - \varepsilon > 1/2 \). Hence, (4.10) is proved.

Then, we prove (4.11). For sufficiently large \( k \),

\[ 1 - \frac{\beta \delta N}{k} f_{k|N} \leq \left( 1 + \frac{\beta \delta N}{k} \left( f - f_{k|N} \right) \right) \left( 1 - \frac{\beta \delta N}{k} f \right). \]

which together with (4.14) implies

\[ \| \psi_k \| \int_{i=N+1}^{k} \left( 1 + \frac{\beta \delta N}{i} \left( f - f_{i|N} \right) \right) \]

\[ \leq \left( 1 - \frac{N\eta}{k} \right) \| \psi_{k-N} \| \int_{i=N+1}^{k-N} \left( 1 + \frac{\beta \delta N}{i} \left( f - f_{i|N} \right) \right) + O \left( \sqrt{\frac{\ln \ln k}{k^3}} \right), \quad \text{a.s.} \]

By (4.10) and Item (d) of Remark 4.10, \( f - f_{i|N} \) converges to 0 at a polynomial rate. Hence, we have \( \prod_{i=N+1}^{\infty} \left( 1 + \frac{\beta \delta N}{i} \left( f - f_{i|N} \right) \right) < \infty \). Then, (4.11) can be proved by (4.16) and Lemma A.9.
Remark 4.12. Proposition 4.11 gives the almost sure convergence rate of $\psi_k$. Since the optimal almost sure convergence rate of $\psi_k$ is $O(\sqrt{\ln \ln k/k})$, by Proposition 4.3, the SA-based algorithm can achieve the almost sure convergence rate of $\psi_k$.

**Theorem 4.13.** Under the conditions of Theorem 4.7, for any $\varepsilon > 0$,

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

where $\eta = \beta \delta f$ with $f$ defined in (4.9). If the density function $f(\cdot)$ is assumed to be locally Lipschitz continuous, then the almost sure convergence rate can be promoted into

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

**Proof.** The theorem can be obtained by Propositions 4.3 and 4.11.

Remark 4.14. By Theorem 4.13, the algorithm may not achieve the optimal almost sure convergence rate when the coefficient $\eta$ is smaller than $1/2$. Since $\eta = \beta \delta f$, 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 value $\theta$. Thus, the almost sure convergence rate of Algorithm (3.2) cannot be known without priori information on $\theta$. The problem can be solved if the step size is designed as $\rho_k = \beta_k/k$, where

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

The analysis for the modified algorithm is consistent with the algorithm with time-invariant $\beta$.

Remark 4.15. For the identification problem of stochastic set-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 [30] generates a maximum likelihood estimate (cf. [12], Lemma 4). The almost sure convergence rate of the empirical measurement algorithm is $O(\sqrt{\ln \ln k/k})$ [18]. In the non-periodic input case, Theorem 4.13 appears to be the first to achieve the almost sure convergence rate of $O(\sqrt{\ln \ln k/k})$ theoretically. [13] achieves the almost sure convergence rate of $O(\sqrt{\ln k/k})$ for the recursive projection method. And, the almost sure convergence rate of stochastic approximation algorithms with expanding truncations is $O(1/k^\varepsilon)$ for $\varepsilon \in (0,1/2)$ [24]. When properly selecting $\beta$, the almost sure convergence rate of Algorithm (3.2) is better than both of them.

**4.4. Mean square convergence rate.** This subsection will estimate the mean square convergence rate of the SA-based algorithm.

**Theorem 4.16.** Under the conditions of Theorem 4.7, for any $\varepsilon > 0$,

$$\mathbb{E}\|\tilde{\theta}_k\|^2 = \begin{cases} O\left(\frac{1}{k}\right), & \eta > \frac{1}{2}; \\ O\left(\frac{1}{k^{1-\varepsilon}}\right), & \eta \leq \frac{1}{2}. \end{cases}$$
where \( \eta = \beta \delta f \) with \( f \) defined in (4.9). If \( f(\cdot) \) is assumed to be locally Lipschitz continuous, then the mean square convergence rate can be promoted into

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

**Proof.** To prove (4.19), we first simplify the recursive formula of \( \mathbb{E}\|\tilde{\theta}_k\|^2 \).

By (3.2) and the Lagrange mean value theorem ([41], Theorem 5.3.1), there exists \( \zeta_k \) between \( C - \phi_k^T \theta \) and \( C - \phi_k^T \theta - \phi_k^T \hat{\theta}_{k-1} \) such that

\[
\tilde{\theta}_k = \tilde{\theta}_{k-1} + \frac{\beta \phi_k}{k} \big( \tilde{F}_k - F_k \big) + \frac{\beta \phi_k}{k} (F_k - s_k)
= \left(I_n - \frac{\beta}{l} f(\zeta_k) \phi_i \phi_i^T \right) \tilde{\theta}_{k-1} + \frac{\beta \phi_k}{k} (F_k - s_k)
= \prod_{i=k-N+1}^k \left( I_n - \frac{\beta}{l} f(\zeta_i) \phi_i \phi_i^T \right) \tilde{\theta}_{k-N}
+ \sum_{l=k-N+1}^k \prod_{i=l+1}^k \left( I_n - \frac{\beta}{l} f(\zeta_i) \phi_i \phi_i^T \right) \frac{\beta \phi_i}{l} (F_i - s_i)
= \prod_{i=k-N+1}^k \left( I_n - \frac{\beta}{l} f(\zeta_i) \phi_i \phi_i^T \right) \tilde{\theta}_{k-N} + \sum_{l=k-N+1}^k \frac{\beta \phi_i}{l} (F_i - s_i) + O\left(\frac{1}{k^2}\right).
\]

Similar to (4.13), except for the first few steps, we have

\[
\left\| \prod_{i=k-N+1}^k \left( I_n - \frac{\beta}{l} f(\zeta_i) \phi_i \phi_i^T \right) \right\| \leq \left( 1 - \frac{f'_{k|N}}{\mathbb{E} f'_{k|N}} \right) + O\left(\frac{1}{k^2}\right),
\]

where \( f'_{k|N} = \frac{\beta \Delta N}{k} \mathbb{E} \left[ \max_{k-N \leq i \leq k} | \phi_i^T (\hat{\theta}_{i-1}) | \right] \), and \( f(\cdot) \) is defined in (4.8). Besides, noticing that \( \tilde{\theta}_{k-N} \) is independent of \( \sum_{l=k-N+1}^k \frac{\beta \phi_i}{l} (F_i - s_i) \), we have

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

Therefore, for sufficiently large \( k \), one can get

\[
\mathbb{E}\|\tilde{\theta}_k\|^2 \leq \mathbb{E} \left[ \left( 1 - \frac{\beta \Delta N}{k} \mathbb{E} f'_{k|N} \right)^2 \|\tilde{\theta}_{k-N}\|^2 \right] + O\left(\frac{1}{k^2}\right).
\]
By Theorem 4.7 and the property of $f(\cdot)$, $\mathbb{P}\{\varepsilon_k < f - \frac{k}{2\sqrt{N}}\} = O(\exp(-mk^{1/2}))$.

Hence, by Corollary A.6, we have

$$\mathbb{E}\left[\left(1 - \frac{\beta\delta N}{k} f_{|k|N}\right)^2 \|\hat{\theta}_{k-N}\|^2\right]$$

$$\leq \int_{\{\varepsilon_k \geq \frac{1}{2k}\}} \left(1 - \frac{N}{k} \left(\eta - \frac{\varepsilon}{2}\right)\right)^2 \|\hat{\theta}_{k-N}\|^2 d\mathbb{P} + \int_{\{\varepsilon_k < \frac{1}{2k}\}} \|\hat{\theta}_{k-N}\|^2 d\mathbb{P}$$

$$= \left(1 - \frac{N}{k} \left(\eta - \frac{\varepsilon}{2}\right)\right)^2 \mathbb{E}\|\hat{\theta}_{k-N}\|^2 + O\left(\ln k \cdot \exp(-mk^{1/2})\right).$$

Substituting the above estimate into (4.22) gives

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

Thus, (4.19) can be proved by Lemma A.9.

Then, we prove (4.20). Similar to (4.15), for sufficiently large $k$,

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

Therefore, by (4.21) and $\eta = \beta\delta f$, one can get

$$\mathbb{E}\|\hat{\theta}_{k}\|^2 \leq \left(1 - \frac{N\eta}{k}\right)^2 \mathbb{E}\left[\left(1 + \frac{\beta\delta N}{k} (f - f_{|k|N})\right)^2 \|\hat{\theta}_{k-N}\|^2\right] + O\left(\frac{1}{k^2}\right).$$

By Item (d) of Remark 4.10, since $f(\cdot)$ is assumed to be locally Lipschitz continuous here, $f(\cdot)$ is also locally Lipschitz continuous. Hence, if $\|\hat{\theta}_j\| \leq j^{-\varepsilon'}$ for a $\varepsilon' > 0$

and all $j = k - N + 1, \ldots, k$, then there exists $L > 0$ such that $f - f_{|k|N} \leq Lk^{-\varepsilon'}$, which together with Corollaries A.6 and A.12 implies that there exist positive numbers $m$ and $\varepsilon$ such that

$$\mathbb{E}\left[\left(1 + \frac{\beta\delta N L}{k^{1-\varepsilon'}}\right)^2 \|\hat{\theta}_{k-N}\|^2\right]$$

$$\leq \left(1 + \frac{\beta\delta N L}{k^{1-\varepsilon'}}\right)^2 \int_{\|j\| \leq \frac{1}{2k}} \|\hat{\theta}_{k-N}\|^2 d\mathbb{P} + O\left(\ln k \cdot \exp(-mk^{1-\varepsilon})\right)$$

$$\leq \left(1 + \frac{\beta\delta N L}{k^{1-\varepsilon'}}\right)^2 \mathbb{E}\|\hat{\theta}_{k-N}\|^2 d\mathbb{P} + O\left(\ln k \cdot \exp(-mk^{1-\varepsilon})\right).$$

Substituting the above estimate into (4.22) gives

$$\mathbb{E}\|\hat{\theta}_{k}\|^2 \leq \left(1 - \frac{N\eta}{k}\right)^2 \left(1 + \frac{\beta\delta N L}{k^{1-\varepsilon'}}\right)^2 \mathbb{E}\|\hat{\theta}_{k-N}\|^2 + O\left(\frac{1}{k^2}\right).$$
Therefore, we have
\[
E\|\hat{\theta}_k\|^2 \left/ \prod_{i=1}^{k} \left(1 + \frac{\beta\delta NL}{i^{1+\epsilon'}}\right)^2 \right. \\
\leq \left(1 - \frac{N\eta}{k}\right)^2 E\|\hat{\theta}_{k-N}\|^2 \left/ \prod_{i=1}^{k-N} \left(1 + \frac{\beta\delta NL}{i^{1+\epsilon'}}\right)^2 \right. + O\left(\frac{1}{k^2}\right).
\]

Then, by Lemma A.9, one can get
\[
E\|\hat{\theta}_k\|^2 \left/ \prod_{i=1}^{k} \left(1 + \frac{\beta\delta NL}{i^{1+\epsilon'}}\right)^2 \right. = \begin{cases} O\left(\frac{1}{k}\right), \quad \eta > \frac{1}{2}; \\
O\left(\frac{\ln k}{k}\right), \quad \eta = \frac{1}{2}; \\
O\left(\frac{1}{k^{2\eta}}\right), \quad \eta < \frac{1}{2}. \end{cases}
\]

Due to the boundedness of \(\prod_{i=1}^{\infty} \left(1 + \frac{\beta\delta NL}{i^{1+\epsilon'}}\right)^2\), (4.20) is proved.

**Remark 4.17.** By Theorem 4.16, the mean square convergence rate of the SA-based algorithm achieves \(O(1/k)\) when properly selecting the coefficient \(\beta\). By [37], the Cramér-Rao lower bound for estimating \(\theta\) based on binary observations \(s_1, \ldots, s_k\) is
\[
\sigma^2_{CR}(s_1, \ldots, s_k) = \left(\sum_{i=1}^{k} \frac{f_i^2}{F_i(1-F_i)} \phi_i\phi_i^\top\right)^{-1} = O\left(\frac{1}{k}\right).
\]

Besides, for the identification problem of MA systems with accurate observations and Gaussian noise, the recursive least square algorithm generates a minimum variance estimate ([14], Theorem 4.4.2). And, 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 set-valued MA systems and even accurate ones.

**Remark 4.18.** In the multiple threshold case, when properly selecting the coefficient \(\beta\), 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.

5. **Numerical simulation.** A numerical simulation will be performed in the section to verify Theorems 4.9, 4.13, and 4.16.

Consider an MA system \(y_k = \phi_k^\top \theta + d_k\) with binary 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 \(u_3i = -1 + e_{3i}, u_{3i+1} = 2 + e_{3i+1}, u_{3i+2} = e_{3i+2}\) for natural number \(i\), where \(e_k\) is randomly chosen in the interval \([-0.1, 0.1]\). It can be verified that the input follows Assumption 2.1.

To achieve a better simulation result, we adjust Algorithm (3.2) as
\[
\hat{\theta}_k = \hat{\theta}_{k-1} + \frac{\beta \phi_k}{k} \left(\hat{F}_k - s_k\right), \quad k \geq k_0.
\]
In the simulation, set $\beta = 20$, $k_0 = 20$, and the initial value $\hat{\theta}_{k_0} = [1, 1]^T$. Figure 1 shows a trajectory of $\hat{\theta}_k$, which verifies the convergence of the SA-based algorithm. In Figure 2, the trajectory of $k\|\hat{\theta}_k\|^2/\ln \ln k$ is shown to be bounded, which verifies that the SA-based algorithm can achieve the almost sure convergence rate of $O(\sqrt{\ln \ln k/k})$.

The empirical variance of $\hat{\theta}_k$ is obtained through 200 repeated experiments with the same inputs. The average of the 200 trajectories of $k\|\hat{\theta}_k\|^2$ is shown to be bounded in Figure 3, which verifies that the SA-based algorithm can achieve the mean square convergence rate of $O(1/k)$.

**Fig. 1.** Convergence of Algorithm (3.2).

**Fig. 2.** The trajectory of $k\|\hat{\theta}_k\|^2/\ln \ln k$.

6. Conclusion. The paper investigates MA systems with uniformly persistently exciting and bounded inputs and binary-valued observations. The binary-valued sensor is fixed. An SA-based algorithm without projection is proposed to identify the parameters. The algorithm appears to be the first online identification method for such systems without requiring *priori* information on the unknown parameters. The
convergence property of the algorithm is shown by the distribution tail and the convergence rates in both the almost sure and mean square sense. The distribution tail converges exponentially. 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 $O(1/k)$. For stochastic MA systems with binary observations and periodic inputs, almost sure convergence rate of the maximum likelihood estimate is $O(\sqrt{\ln \ln k/k})$. And, for MA systems with accurate observations and Gaussian noise, the mean square convergence rate of the minimum variance estimate is $O(1/k)$. Hence, both the convergence rates are the best in theory. The analysis is based on an auxiliary stochastic process named SPAO. The methodology can be extended to a large number of identification algorithms of set-valued systems (cf. Appendix B).

Here we give some topics for future research. Firstly, the design of the step-size $\rho_k$ is left as an open question. How can we design a dynamic $\rho_k$ to allow the convergence rates to be the best automatically, and how can we design $\rho_k$ to make the identification algorithm achieve the Cramér-Rao lower bound asymptotically? Secondly, can the algorithm be extended to other forms of systems, e.g., nonlinear systems or systems with other kinds of nonlinear observations? And thirdly, how can we design system control laws to regulate the system performance using the SA-based algorithm?

Appendix A. Lemmas and the proofs.

A.1. Proof of Lemma 4.2. The lemma can be indicated by Theorem 5.5.1 of [26]. We transfer the problem first.

Firstly, we claim that it is equivalent to proving that there exists $m > 0$ such that $P \left\{ \|T_k\| > k^{-\epsilon} \right\} = O(\exp(-mk^{1-2\epsilon}))$. This is because $\sum_{j=k}^{\infty} \exp(-mj^{1-2\epsilon}) = O(k^{2\epsilon} \exp(-mk^{1-2\epsilon})) = O(\exp(-mk^{1-2\epsilon}/2))$.

Secondly, we claim that it is equivalent to prove that for any $i \in \{1, 2, \ldots, n\}$, the $i$-th component of $T_k$, $T_{k,i}$ satisfies $P \left\{ |T_{k,i}| > k^{-\epsilon}/\sqrt{n} \right\} = O(\exp(-mk^{1-2\epsilon}))$. This is because $\{\|T_k\| > k^{-\epsilon}\} \subseteq \bigcup_i \{ |T_{k,i}| > k^{-\epsilon}/\sqrt{n} \}$, which implies

$$P \left\{ \|T_k\| > k^{-\epsilon} \right\} \leq \sum_{i=1}^{n} P \left\{ |T_{k,i}| > k^{-\epsilon}/\sqrt{n} \right\}.$$
Now we have finished the transformation. The converted problem is a corollary of Theorem 5.5.1 of [26].

**Lemma A.1** ([26], Theorem 5.5.1). Assume that

i) \(\{X_k, k \geq 1\}\) is a sequence of independent random variables;

ii) \(E X_k = 0\) and \(|X_k| \leq X < \infty\);

iii) \(S_k = \sum_{i=1}^{k} X_i, \sigma_k = \sqrt{\text{var}(S_k)}\).

Then

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

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

\[
\mathbb{P}\left\{ T_{k,i} > \frac{k^{-\varepsilon}}{\sqrt{n}} \right\} = \mathbb{P}\left\{ \frac{S_{k,i}}{\sigma_{k,i}} > d_{k,i} \right\} < \max \left\{ \exp\left( -\frac{d_{k,i}^2}{4} \right), \exp\left( -\frac{d_{k,i} \sigma_{k,i}}{4X} \right) \right\}
\]

\[
= \max \left\{ \exp\left( -\frac{k^{2-2\varepsilon}}{4n\sigma_{k,i}^2} \right), \exp\left( -\frac{k^{1-\varepsilon}}{4X \sqrt{n}} \right) \right\}.
\]

Noting that

\[
\sigma_{k,i}^2 = \text{var}(S_{k,i}) = \sum_{j=1}^{k} \text{var}(X_{j,i}) \leq 4X^2 k,
\]

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

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

Combining the two consequences, the converted problem is thereby proved. That is to say, we get Lemma 4.2.

### A.2. Lemmata in Section 4.

**Lemma A.2.** Assume that \(\phi_k\) satisfies Assumption 2.1 and the stochastic process \(\psi_k\) satisfies \(\|\psi_k - \psi_{k-1}\| \leq \Psi/k\) for some \(\Psi > 0\). Then,

\[
\delta \|\psi_k\|^2 \leq \frac{1}{N} \sum_{j=k+1}^{k+N} (\phi_j^T \psi_{j-1})^2 + \frac{2NM^2\Psi}{k} \sum_{j=k}^{k+N-1} \|\psi_j\| + \frac{N^2M^2\Psi^2}{k^2}.
\]

Furthermore, if \(b' > 0\) and \(k\) is large enough, then there is \(k' \in [k + 1, k + N]\) such that \(|\phi_{k'}^T \psi_{k'-1}| \geq \sqrt{\delta/2} \|\psi_k\| I_{\{\|\psi_k\| > b'\}}\).

**Proof.** The lemma is based on Assumption 2.1. 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,

\[
\frac{1}{N} \sum_{j=k+1}^{k+N} (\phi_j^T \psi_{j-1})^2 \geq \frac{1}{N} \sum_{j=k+1}^{k+N} (\phi_j^T \psi_{j-1})^2 - \frac{2NM^2\Psi}{k} \sum_{j=k}^{k+N-1} \|\psi_j\| - \frac{N^2M^2\Psi^2}{k^2}.
\]
Besides, by Assumption 2.1,
\[
\frac{1}{N} \sum_{j=k+1}^{k+N} (\phi_j^\top \psi_k)^2 = \frac{1}{N} \sum_{j=k+1}^{k+N} \psi_k^\top \phi_j \phi_j^\top \psi_k \geq \delta \|\psi_k\|^2.
\]

Thus, the first part of the lemma is proved.

As for the second part, we note that under the condition of the lemma, \( \psi_k = O(1) \). So,
\[
\frac{2NM^2\Psi}{k} \sum_{j=k}^{k+N-1} \|\psi_j\| + \frac{N^2M^2\Psi^2}{k^2} = O\left(\frac{\ln k}{k}\right).
\]

Hence, if \( \|\psi_k\| > b' \) and \( k \) is sufficiently large, then one can get
\[
\frac{1}{N} \sum_{j=k+1}^{k+N} (\phi_j^\top \psi_{j-1})^2 \geq \delta \|\psi_k\|^2 + O\left(\frac{\ln k}{k}\right) > \frac{\delta}{2} \|\psi_k\|^2,
\]
which implies
\[
\frac{1}{N} \sum_{j=k+1}^{k+N} (\phi_j^\top \psi_{j-1})^2 > \frac{\delta}{2} \|\psi_k\|^2 \mathbb{I}_{\{\|\psi_k\| > b'\}}
\]
for sufficiently large \( k \).

Then, there exists \( k' \in [k+1, k+N] \) such that
\[
(\phi_k^\top \psi_{k'-1})^2 \geq \frac{\delta}{2} \|\psi_k\|^2 \mathbb{I}_{\{\|\psi_k\| > b'\}},
\]
which verifies the second part of the lemma.

**Lemma A.3.** If a sequence \( a_k \) satisfies the recursive function
\[
(A.1) \quad a_k \leq a_{k-1} - \frac{D\sqrt{a_{k-1}}}{k+k_0} \mathbb{I}_{\{a_{k-1} > M'/2\}} + d_k,
\]
where \( D, k_0 \) and \( M' \) are all positive, and \( \sum_{k=1}^{\infty} |d_k| < M'/2 \), then
\[
(A.2) \quad a_k < \max\left\{ M', \left[ \left( \sqrt{a_0} - \frac{D}{2} \ln \left( \frac{k+k_0+1}{k_0+1} \right) \right)^+ + \frac{M'}{2} \right] \right\},
\]
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
\[
a_t \geq a_k - \sum_{i=t+1}^k d_i \geq a_k - \frac{M'}{2} \geq \frac{M'}{2}, \quad \forall t \leq k.
\]

Define \( a'_0 = a_0 \) and \( a'_t = a_t - \sum_{i=1}^t |d_i| > M'/2 - M'/2 = 0 \) for \( t \geq 1 \). Then, we have
\[
a'_t = a_t - \sum_{i=1}^t |d_i| \leq a_{t-1} - \frac{D\sqrt{a_{t-1}}}{t+k_0} + d_t - \sum_{i=1}^t |d_i| \leq a_{t-1} - \frac{\sum_{i=1}^{t-1} |d_i|}{t+k_0} = a'_{t-1} - \frac{D\sqrt{a'_{t-1}}}{t+k_0},
\]

This manuscript is for review purposes only.
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 \( x \leq x^+ \)
\[ \sqrt{a'_k} < \sqrt{a'_0} - \sum_{i=1}^{k} \frac{D}{2(t + k_0)} \leq \left( \sqrt{a'_0} - \frac{D}{2} \ln \left( \frac{k + k_0 + 1}{k_0 + 1} \right) \right)^+ . \]
So,
\[ a_k = a'_k + \sum_{i=1}^{k} |d_i| < \left[ \left( \sqrt{a'_0} - \frac{D}{2} \ln \left( \frac{k + k_0 + 1}{k_0 + 1} \right) \right)^+ \right]^2 + M', \]
which proves the lemma.

Remark A.4. Lemma A.3 proves the uniform ultimate upper boundedness of the sequence \( a_k \) which satisfies (A.1). 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.2) implies \( a_k < M' \).

Lemma A.5. Assume that
i) \( v(.) : \mathbb{R}^n \rightarrow \mathbb{R} \) is a continuously twice differentiable non-negative function, whose second derivative is bounded;
ii) \( g_k(.) : \mathbb{R}^n \rightarrow \mathbb{R}^n \) is uniformly bounded;
iii) \( \nabla v(x)^\top g_k(x) \) is uniformly upper bounded, where \( \nabla v(.) \) is the gradient of \( v(.) \);
iv) the positive step size \( \rho_k \in \mathbb{R} \) satisfies \( \lim_{k \rightarrow \infty} \rho_k = 0 \);
v) \( x_k = x_{k-1} + \rho_k g_k(x_{k-1}) \).

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

Proof. From
\[ v(x_k) = v(x_{k-1} + \rho_k g_k(x_{k-1})) = v(x_{k-1}) + \rho_k \nabla v(x_{k-1})^\top g_k(x_{k-1}) + O \left( \rho_k^2 \right) \]
\[ \leq v(x_{k-1}) + O(\rho_k) \leq O \left( \sum_{i=1}^{k} \rho_i \right) , \]
we get the lemma.

Corollary A.6. Under Assumptions 2.1 and 2.3, the estimation error of Algorithm (3.2) satisfies \( \bar{\theta}_k = O \left( \sqrt{\ln k} \right) \).

Proof. Due to the finite covariance of the noise, by Markov inequality ([6], Theorem 5.1.1), when \( t \) goes to \( \infty \),
\[ F \left( C - \phi_k^\top \theta - t \right) = \mathbb{P} \left\{ \| d_k \| < C - \phi_k^\top \theta - t \right\} \leq \mathbb{P} \left\{ d_k^2 > (C - \phi_k^\top \theta - t)^2 \right\} \]
\[ \leq \frac{\mathbb{E}d_k^2}{(C - \phi_k^\top \theta - t)^2} = O \left( \frac{1}{t^2} \right) , \]
(A.3)
Then, it suffices to estimate $P(C - \phi_k^T \theta - t) = O\left(\frac{1}{t^2}\right)$.

Set $v(x) = x^T x$. Then, $\nabla v(x) = x$. By (A.3) and (11),

$$\nabla v(x)^T \phi_k \left(F \left(C - \phi_k^T \theta - \phi_k^T x\right) - s_k\right) = \phi_k^T x \left(F \left(C - \phi_k^T \theta - \phi_k^T x\right) - s_k\right)$$

is uniformly upper bounded. Thus, we get the corollary by Lemma A.5.

**Corollary A.7.** Under the condition of Proposition 4.3, $\psi_k = O(\sqrt{\ln k})$.

**Proof.** From $\psi_k = \hat{\theta}_k - T_k = O(\sqrt{\ln k}) + O(1)$, we get the corollary.

**Remark A.8.** Corollaries A.6 and A.7 estimate the estimation error $\tilde{\theta}_k$ and SPAO $\psi_k$ in the worst case, respectively.

**Lemma A.9.** For the sequence $h_k$, assume that

i) $h_k$ is positive and monotonically increasing;

ii) $\ln h_k = O(\ln k)$.

Then, for real numbers $d$, $\eta$ and $\varepsilon$, and any positive integer $p$,

$$\sum_{l=1}^{k} \prod_{i=l}^{k} \left(1 - \frac{\eta}{i + d}\right)^p h_l^{\frac{1}{l^{1+\varepsilon}}} = O\left(\left(\frac{h_k}{k}\right)^{\eta}; \quad p\eta > \varepsilon;\right)$$

$$O\left(\frac{h_k \ln k}{k^{\varepsilon}}\right), \quad p\eta = \varepsilon;\right)$$

$$O\left(\frac{1}{k^p}\right), \quad p\eta < \varepsilon.$$ 

**Proof.** By Lemma 4 in [39], we have

$$\prod_{i=l}^{k} \left(1 - \frac{\eta}{i + d}\right) \leq \prod_{i=l+|d|+2}^{k+|d|+1} \left(1 - \frac{\eta}{i}\right) = O\left(\left(\frac{l}{k}\right)^\eta\right),$$

which leads to

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

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

Firstly, 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 $\sum_{l=1}^{\infty} h_l^{\frac{1}{l^{1+\varepsilon-p\eta}}} < \infty$. So, we can get

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

Secondly, by the monotonicity of $h_k$, we have

$$\sum_{l=1}^{k} h_l \leq \sum_{l=1}^{k} h_l (\ln l - \ln(l - 1)) \leq \sum_{l=1}^{k} (h_l \ln l - h_{l-1} \ln(l - 1)) = h_k \ln k.$$ 

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

$$\sum_{l=1}^{k} \prod_{i=l}^{k} \left(1 - \frac{\eta}{i + d}\right)^p h_l^{\frac{1}{l^{1+\varepsilon}}} = O\left(\frac{h_k \ln k}{k^{\varepsilon}}\right).$$
Lastly, when \( p \eta > \varepsilon \), we have
\[
\sum_{l=1}^{k} \frac{h_l}{l^{1+\varepsilon-p\eta}} = O \left( \sum_{l=1}^{k} \left( l^{p\eta-\varepsilon} - (l-1)^{p\eta-\varepsilon} \right) \right)
\]
\[
\leq O \left( \sum_{l=1}^{k} \left( h_l l^{p\eta-\varepsilon} - h_{l-1} (l-1)^{p\eta-\varepsilon} \right) \right) = O \left( h_k k^{p\eta-\varepsilon} \right),
\]
which implies
\[
\sum_{l=1}^{k} \prod_{i=l+1}^{k} \left( 1 - \frac{\eta}{i + d} \right)^{p} \frac{h_l}{l^{1+\varepsilon}} = O \left( \frac{h_k}{k^{\varepsilon}} \right).
\]

**Remark A.10.** If \( h_k \) is constant, \( p = 1 \) and \( d = 0 \), then Lemma A.9 implies Lemma 4 in [39]. Besides, if \( h_k/\ln k \) is assumed to be monotonically decreasing, then the estimate of Lemma A.9 is accurate.

**Proposition A.11.** Under the condition of Proposition 4.3, for any \( \varepsilon \in (0, 1) \), there exist positive numbers \( \varepsilon' \) and \( m \) such that
\[
P \left\{ \| \psi_k \| > k^{-\varepsilon'} \right\} = O \left( \exp \left( -m k^{1-\varepsilon} \right) \right).
\]

**Proof.** The proposition can be proved by verifying that there exists \( \varepsilon' > 0 \) such that
\[
\left( \| \psi_k \| \leq k^{-\varepsilon'} \right) \supseteq \left\{ \sup_{j \geq k^{1-2\varepsilon}} \| T_j \| \leq 1 \right\}.
\]

By the monotonicity of \( \{ \sup_{j \geq k^{1-2\varepsilon}} \| T_j \| \leq 1 \} \) and Proposition 4.5,
\[
\left\{ \sup_{j \geq k^{1-2\varepsilon}} \| \psi_j \| < M' \right\} \supseteq \left\{ \sup_{j \geq k^{1-2\varepsilon}} j^{\varepsilon} \| T_j \| \leq 1 \right\}.
\]

Therefore, if \( \sup_{j \geq k^{1-2\varepsilon}} j^{\varepsilon} \| T_j \| \leq 1 \), then by (4.12) and (4.13),
\[
\| \psi_j \| \leq \left( 1 - \frac{\beta \delta N}{j} f \left( M \sqrt{M'} \right) \right) \| \psi_{j-N} \| + O \left( \frac{1}{j^{1+\varepsilon}} \right), \quad \forall j \geq k^{1-\varepsilon} + N,
\]
where \( f(\cdot) \) is defined in (4.8). Then, by Corollary A.7 and Lemma A.9, \( \| \psi_j \| \) converges at a polynomial rate. Hence, we get (A.5). Then, the proposition can be proved by Lemma 4.2 and the arbitrariness of \( \varepsilon \).

**Corollary A.12.** Under the condition of Theorem 4.7, for any \( \varepsilon > 0 \), there exist positive numbers \( \varepsilon' \) and \( m \) such that
\[
P \left\{ \| \tilde{\psi}_k \| > k^{-\varepsilon'} \right\} = O \left( \exp \left( -m k^{1-\varepsilon} \right) \right).
\]

**Proof.** By (A.5) and \( \tilde{\psi}_k = \psi_k + T_k \), we have
\[
\left\{ \| \tilde{\psi}_k \|^2 \leq k^{-\varepsilon'} + k^{-\varepsilon} \right\} \supseteq \left\{ \| \psi_k \| \leq k^{-\varepsilon'} \right\} \cap \left\{ \| T_k \| \leq k^{-\varepsilon} \right\}
\]
\[
\sup_{j \geq k^{1-2\varepsilon}} j^{\varepsilon} \| T_j \| \leq 1 \right\}.
\]
Then, the corollary can be proved by Lemma 4.2.
Remark A.13. Proposition A.11 and Corollary A.12 are extensions of Proposition 4.5 and Theorem 4.7, respectively.

Appendix B. SPAO for other algorithms.

The construction of SPAO can be applied to many online identification algorithms of set-valued systems. For set-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, $C_k$ and $v_k$ are generated by $\{\phi_j, s_j, j \leq k\}$ $[8, 13, 21, 24, 32, 34, 35, 36, 37]$. The step size $\rho_k$ can also be matrices $[21, 34, 37]$.

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

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

Then, one can get

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

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

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

Therefore, we can analyze the convergence property of the algorithm through $\psi_k$.

REFERENCES


