J Syst Sci Complex (2024) 37(1): 204–229

Threshold Selection and Resource Allocation for Quantized Identification*

WANG Ying · LI Xin · ZHAO Yanlong · ZHANG Ji-Feng

DOI: 10.1007/s11424-024-3369-8
Received: 13 September 2023 / Revised: 25 October 2023

Abstract This paper is concerned with the optimal threshold selection and resource allocation problems of quantized identification, whose aims are improving identification efficiency under limited resources. Firstly, the first-order asymptotically optimal quantized identification theory is extended to the weak persistent excitation condition. Secondly, the characteristics of time and space complexities are established based on the Cramér-Rao lower bound of quantized systems. On these basis, the optimal selection methods of fixed thresholds and adaptive thresholds are established under aperiodic signals, which answer how to achieve the best efficiency of quantized identification under the same time and space complexity. In addition, based on the principle of maximizing the identification efficiency under a given resource, the optimal resource allocation methods of quantized identification are given for the cases of fixed thresholds and adaptive thresholds, respectively, which show how to balance time and space complexity to realize the best identification efficiency of quantized identification.

Keywords Quantized output, resource allocation, system identification, threshold selection.

1 Introduction

1.1 Background and Motivations

System identification is widely applied in various fields, such as engineering[1–3], signal processing[4], biomedicine[5, 6], and so on. Classical identification theory focuses on asymptotic properties[7–9], such as least square and least mean square algorithms, which requires sufficiently numerous data. However, system identification is often carried out under limited

WANG Ying · LI Xin · ZHAO Yanlong (Corresponding author) · ZHANG Ji-Feng
Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190*, China; School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing* 100049*, China.*
Email: wangying96@amss.ac.cn; lixin2020@amss.ac.cn; ylzhao@amss.ac.cn; jif@iss.ac.cn.
*This research was supported by the National Key R&D Program of China under Grant No. 2018YFA0703800, the National Natural Science Foundation of China under Grant Nos. T2293770, 62025306, 62303452, and 122263051, CAS Project for Young Scientists in Basic Research under Grant No. YSBR-008, China Postdoctoral Science Foundation under Grant No. 2022M720159, and Guozhi Xu Postdoctoral Research Foundation.
◇*This paper was recommended for publication by Guest Editor LIN Zongli.*

data resources in daily life. In order to make full use of resources, we must investigate how to select appropriate data to achieve the best identification efficiency under limited resources. It is exactly a kind of resource allocation problem, that is, how to choose the data type and data amount to achieve the best identification efficiency under limited resources.

There are two types of data in daily life: Accurate and quantized data. Quantized data are often seen in medical, engineering and communication fields[10–12], which refers to that it is known that data belongs to a certain or some sets instead of its accurate value[13]. Although quantized data has less information than its accurate value, it is undeniable that quantized data has incomparable advantages in reducing the cost of information transmission and storage. This urges us to investigate how to make quantized data contain as much information as possible, which is also an important part of optimal resource allocation under limited resources.

It is worth noting that the threshold design of quantized data has an important impact on the amount of information contained in the data. For example, data compression ratio can be improved by designing a quantized table in the data compression[14–16]. Specifically, the image compression method proposed by [15] based on adaptive quantized parameters has less loss of image quality and better compression effect, compared with the original JPEG image compression method. The research in [13] has also shown that different quantized intervals contribute differently to the error reduction of the identification algorithm, and even the whole quantized interval has no effect on identification effect for some extreme cases. Therefore, proper threshold designs can make the same type of data contain more information, which helps to achieve the best application effect under the same amount of data. According to the above, this paper considers the optimal threshold selection from the perspective of quantized identification, and then answers how to establish the optimal resource allocation method under limited identification resources.

1.2 Related Works

As known, whether exploring optimal threshold selection problems or optimal resource allocation problems of quantized identification, the basis is to realize the parameter identification based on quantized data. The existing identification methods based on quantized data can be roughly divided into two categories. One is the off-line quantized identification methods, which contain empirical measure (EM) method[10, 17, 18], expectation maximization algorithm[19, 20], maximum likelihood method[21], variational bayesian method[22] and support vector machine[23], etc. The other is the online quantized identification methods, such as recursive projection algorithm[24, 25], stochastic approximation algorithm with extended truncation[26, 27], stochastic gradient algorithm[28], least mean square algorithm[29], Quasi-Newton projection algorithm[30, 31], and so on. All of the above quantized identification methods focus on how to achieve parameter identification under given quantized data, but do not concern with what kind of quantized data could realize a better convergence properties of quantized identification under limited resource.

Actually, there are a few studies involving to select which kinds of quantized data (i.e., threshold selection of quantized data) can achieve better convergence properties of quantized identification, including [32] and [33]. To be specific, in order to describe the influence of data type and amount on quantized identification algorithm, [32] introduced the concept of space complexity (i.e., the number of thresholds determining the types of data) and established the basic property of space complexity, that is, the minimum reachable identification error decreases with the increase of space complexity. However, with the increase of space complexity, the cost of resources required for data transmission and storage also increases. In order to balance space complexity and data transmission cost, [32] explored the optimal threshold selection theory based on the quasi-convex combination estimator based on EM algorithm in [17]. That is, how to design the thresholds to achieve the best identification effect under the same data type and transmission. On this basis, [32] established two kinds of optimal resource allocation criteria for quantized identification. However, due to the limitation of EM algorithm theory, the above threshold selection and resource allocation theory could only be applied to the case of periodic input. Then, [33] constructed a recursive projection algorithm with the time-varying threshold design, gave the upper bound of the estimation error covariance under the weak persisting excitation condition, and established a threshold selection scheme based on the upper bound. [33] is an effective attempt to explore the asymptotic optimal online quantized identification algorithm and optimal threshold selection theory.

Starting from the online quantized identification in [33], this paper explores the optimal threshold design problem in quantized identification under the fixed threshold and time-varying threshold, and studies the resource allocation method in quantized identification on this basis, so as to answer how to balance the amount and type of data to achieve the best identification efficiency of quantized identification under limited resources.

1.3 Contributions

Lightened by [32], this paper investigates the optimal threshold selection and optimal resource allocation problems from the perspective of online quantized identification. The main contributions of this paper is summarized as follows:

- First, the asymptotic optimality of the first-order IBID algorithm (i.e., information based identification algorithm) proposed by [31] is established under more general signal conditions. To be specific, the optimal convergence rate and asymptotic efficiency of the algorithm are proved under weak persisting excitation condition, which lays a theoretical foundation for optimal threshold selection and resource allocation. Compared with [17], the IBID algorithm is appropriate for non-periodic signals.

- This paper designs the optimal selection methods of fixed thresholds and adaptive thresholds based on the principle of minimizing the minimum reachable identification error (i.e., maximizing the identification efficiency), respectively. For this purpose, the influence of the time and space complexity to the identification efficiency are analyzed by taking the norm of Cramér-Rao (CR) lower bound as the metric index. Moreover, this paper not only studies fixed threshold case but also provides the theoretical support of threshold selection by asymptotically optimal algorithm, while [33] designed the adaptive quantization thresholds only by the upper bound of mean square estimation error.

- In addition, based on the principle of minimizing the minimum reachable identification error under limited resources, this paper establishes the optimal resource allocation methods under fixed thresholds and adaptive thresholds, respectively. Compared with [32], this paper establishes the optimal threshold selection methods and resource allocation methods of quantized identification under the more general condition of non-periodic signals.

The rest of this paper is organized as follows. Section 2 describes the problem to be solved. Section 3 demonstrates the identification algorithm and its optimality. Section 4 analyses the influence of time and space complexity to identification efficiency. Section 5 constructs the threshold selection methods for binary sensors under fixed and adaptive threshold cases. Section 6 gives the optimal resource allocation methods. Section 7 gives some concluding remarks and related future works.

Notations In this paper, $\mathbb{Z}$ and $\mathbb{Z}_+$ are the set of integers and positive integers, respectively. $\mathbb{R}^n$ and $\mathbb{R}^{n\times n}$ are the set of n-dimensional real vectors and $n \times n$ dimensional real matrices, respectively. I_n is an n-dimension identity matrix. $\|\cdot\|$ is the Euclidean norm, i.e., $\|x\| = \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}}$ for the vector $x \in \mathbb{R}^n$ and $\|A\| = \left(\lambda_{\max}(AA^{\rm T})\right)^{\frac{1}{2}}$ for the matrix $A \in \mathbb{R}^{n\times n}$. For matrices A_k and B_k, denote $A_k = O\left(\frac{1}{k}\right)$ as $\|A_k\| = O\left(\frac{1}{k}\right)$ and $A_k = o\left(\frac{1}{k}\right)$ as $\|A_k\| = o\left(\frac{1}{k}\right)$. The function $I_{\{\cdot\}}$ denotes the indicator function, whose value is 1 if its argument (a formula) is true, and 0, otherwise.

2 Problem Formulation

2.1 Observation Model

Consider the following dynamic linear system

$$y_k = \phi_k^{\rm T}\theta + d_k, \quad k = 1, 2, \cdots, \tag{1}$$

where k is the time index and $\phi_k \in \mathbb{R}$, $\theta \in \mathbb{R}$, and $d_k \in \mathbb{R}$ are the input, unknown but constant parameter vector, and noise at time k, respectively. The system output y_k only can be measured by a sensor of m thresholds $-\infty < C_1 < C_2 < \cdots < C_m < \infty$. The sensor can be represented by a set of m indicator function, which is given by

$$q_k = \mathcal{Q}(y_k) = \sum_{i=0}^{m} i I_{\{C_i < y_k \le C_{i+1}\}} = \begin{cases} 0, & \text{if } \ y_k \le C_1; \\ 1, & \text{if } \ C_1 < y_k \le C_2; \\ \vdots & \vdots \\ m, & \text{if } \ y_k > C_m, \end{cases} \tag{2}$$

where $C_0 = -\infty$ and $C_{m+1} = \infty$.

2.2 Assumptions

In order to proceed our analysis, we introduce some assumptions concerning the prior information of the unknown parameter, the inputs and the noises.

Assumption 2.1 The prior information on the unknown parameter θ is that $\theta \in \Omega \subset \mathbb{R}$ with Ω being a bounded and closed set. And denote $\overline{\theta} = \sup_{\eta\in\Omega} \|\eta\|$.

Assumption 2.2 The input sequence $\{\phi_k\}$ is supposed to be weak persistent exciting, i.e.,

$$\liminf_{k\to\infty} \frac{1}{k}\sum_{l=1}^{k} \phi_l\phi_l^{\mathrm{T}} > 0 \tag{3}$$

and $\sup_k \|\phi_k\| \leq \overline{\phi} < \infty$.

Assumption 2.3 Suppose that the noise sequence $\{d_k\}$ is a sequence of independent and identically normally distributed variables following $N(0,\sigma^2)$. The distribution and density functions of d_1 are denoted as $F(\cdot)$ and $f(\cdot)$, respectively.

Remark 2.4 Actually, the median μ of the noise could be estimated similarly to [25] when $\mu \neq 0$. Therefore, without loss of generality, we assume that $\mu = 0$ throughout the paper. Assumption 2.1 is common in quantized identification such as [24, 25, 33], which is used to guarantee the boundness of the estimate and good convergence effect in the initial iterative process of the algorithm. Assumption 2.2 is used to keep the identifiability of parameter estimation such as [1, 33], which is weaker than persisting excitation in [24, 25].

This paper discusses threshold selection and resource allocation problems of quantized identification. To solve these, we first answer how to achieve the best efficiency of quantized identification under same data resources. In other words, we will construct the asymptotically optimal online algorithm under more general excitation conditions in the following section.

3 Identification Algorithm and Its Optimality

This section would summarize and expand the key results that support the analysis of the optimal threshold selection problem, namely the convergence and optimality theory of the first-order IBID algorithm proposed by [31].

At first, for simplicity of description, denote

$$F_{i,k} = F\left(C_i - \phi_k^{\mathrm{T}}\theta\right), \quad f_{i,k} = f(C_i - \phi_k^{\mathrm{T}}\theta), \tag{4}$$

and their estimates based on $\widehat{\theta}_{k-1}$ as

$$\widehat{F}_{i,k} = F(C_i - \phi_k^{\mathrm{T}}\widehat{\theta}_{k-1}) \text{ and } \widehat{f}_{i,k} = f(C_i - \phi_k^{\mathrm{T}}\widehat{\theta}_{k-1}), \tag{5}$$

respectively, for $i = 0, 1, \cdots, m+1$. Correspondingly, denote

$$H_{i,k} = F_{i,k} - F_{i-1,k} \text{ and } h_{i,k} = f_{i,k} - f_{i-1,k}, \tag{6}$$

and their estimates as

$$\widehat{H}_{i,k} = \widehat{F}_{i,k} - \widehat{F}_{i-1,k}, \quad \widehat{h}_{i,k} = \widehat{f}_{i,k} - \widehat{f}_{i-1,k}, \tag{7}$$

respectively, for $i = 1, \cdots, m+1$.

Then, the first order form of the asymptotically effective identification algorithm under multi-threshold quantized observations constructed in [31] is given, that is, the first-order IBID algorithm shown in Algorithm 1.

Algorithm 1 First-order IBID algorithm

Beginning with an initial values $\widehat{\theta}_0 \in \Omega$ and a positive definitive matrix $\widehat{P}_0$, the algorithm is recursively defined at any $k \geq 0$ as follows:

1: Update of the adaptive weight coefficients:

$$\widehat{\alpha}_{i,k} = -\frac{\widehat{h}_{i,k}}{\widehat{H}_{i,k}} \quad \text{and} \quad \widehat{\beta}_k = \sum_{i=1}^{m+1} \frac{\widehat{h}_{i,k}^2}{\widehat{H}_{i,k}}, \tag{8}$$

where $\widehat{h}_{i,k}$ and $\widehat{H}_{i,k}$ are defined as (7).

2: Weighted conversion of the quantized observations:

$$s_k = \sum_{i=1}^{m+1} \widehat{\alpha}_{i,k} I_{\{C_{i-1} < y_k \leq C_i\}}. \tag{9}$$

3: Estimation:

$$\widehat{\theta}_k = \Pi_\Omega \left(\widehat{\theta}_{k-1} + \widehat{a}_k \widehat{P}_{k-1} \phi_k \widetilde{s}_k \right), \tag{10}$$

$$\widetilde{s}_k = s_k - \sum_{i=1}^{m+1} \widehat{\alpha}_{i,k} \widehat{H}_{i,k}, \tag{11}$$

$$\widehat{a}_k = \frac{1}{1 + \widehat{\beta}_k \phi_k^{\mathrm{T}} \widehat{P}_{k-1} \phi_k}, \tag{12}$$

$$\widehat{P}_k = \widehat{P}_{k-1} - \widehat{a}_k \widehat{\beta}_k \widehat{P}_{k-1} \phi_k^{\mathrm{T}} \phi_k \widehat{P}_{k-1}, \tag{13}$$

where $\Pi_\Omega(\cdot)$ is the projection mapping defined as $\Pi_\Omega(x) = \arg\min_{z \in \Omega} \|x - z\|, \forall x \in \mathbb{R}^n$.

Next, we will establish the convergence and asymptotic efficiency of the 1-order IBID algorithm under more general condition than that in [31].

Theorem 3.1 *If Assumptions* 2.1–2.3 *hold, then the* 1*-order IBID algorithm is convergent in both mean square and almost sure sense. Besides, its mean square convergence rate is*

$$\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k}\right),$$

where $\widetilde{\theta}_k = \widehat{\theta}_k - \theta$ *is the estimation error.*

Proof The proof of this theorem is supplied in the Appendix. ∎

Then, the CR lower bound of quantized systems (1)–(2) is given as follows.

Proposition 3.2 (see [31]) *For the system* (1) *with quantized observations* (2)*, the CR* *lower bound is*

$$\Delta_k = \left(\sum_{l=1}^{k} \rho_l \phi_l \phi_l^{\mathrm{T}}\right)^{-1}, \tag{14}$$

where $\rho_l = \sum_{i=1}^{m+1} \frac{h_{i,l}^2}{H_{i,l}}$ *with* $h_{i,l}$ *and* $H_{i,l}$ *defined in* (6) *for* $i = 1, \cdots, m+1$.

The following theorem shows the asymptotic efficiency of the first-order IBID algorithm.

Theorem 3.3 *If Assumptions* 2.1–2.3 *hold, then the* 1*-order IBID algorithm is asymptotically efficient, i.e.,*

$$\lim_{k\to\infty} k\left(\mathbb{E}\widetilde{\theta}_k\widetilde{\theta}_k^{\mathrm{T}} - \Delta_k\right) = 0.$$

Moreover, $\widehat{P}_k$ *defined in* (13) *has the following property,*

$$\lim_{k\to\infty} k(\mathbb{E}\widehat{P}_k - \Delta_k) = 0.$$

Based on Theorem 3.1, the proof of Theorem 3.3 is similar to [31, Theorems 4.2 and 4.3], and hence omit here.

Remark 3.4 In the optimal threshold selection and resource allocation problems, the CR lower bound is used to design the measure criterion. Theorem 3.3 shows that $\widehat{P}_k$ in the 1-order IBID algorithm can asymptotically approximate the CR lower bound. It means that we can utilize $\widehat{P}_k$ as the measure criterion to some extent, instead of calculating the CR lower bound by parameter estimates, which is more simple and intuitive.

4 Time and Space Complexity

This section focuses on the characteristics of time and space complexity in order to balance the role of both in subsequent resource allocation problems. Actually, the quantizer threshold number m is regarded as a measure of space complexity, while the data amount K is regarded as a measure of time complexity[32]. It is well-known that both of them influence the identification efficiency (i.e., the minimum reachable identification error). The CR lower bound, the greatest lower bound of the covariance of estimation error, is often used to measure the identification efficiency in parameter estimation. The smaller the CR lower bound, the higher the identification efficiency (i.e., the smaller the minimum reachable identification error). Moreover, it can also be used to analyze the influence of the time and space complexity to the identification efficiency.

When involving communication data transmission, the communication resources for identification are usually limited in order to normally carry out the follow-up control or decision-making. For convenience, the bandwidth resource R of a communication channel is used as the amount of available identification resources in a real-time signal processing. The resource for transmitting K data measured by a quantizer with m thresholds is $K\log_2(m+1)$. For the available resource R of quantized identification, there are usually two ways to allocate it. The first way is to allocate it on space complexity (i.e., to increase the number of thresholds m), and the second way is on time complexity (i.e., to increase the amount of data K). The overall goal of two allocation schemes is to achieve the highest identification efficiency for given identification resources. In this section, we would discuss the effect of space and time complexity on the identification efficiency.

Firstly, we give an index to measure the influence of time K (data amount $\{\phi_k\}_{k=1}^{K}$) and space m (threshold number) on the identification efficiency as follows:

$$\eta(K,m,\theta)=\|\Delta_K(m,\theta)\|, \tag{15}$$

which is based on the CR lower bound. The smaller the index is, the higher the identification efficiency is.

The following theorem states that the quantized identification efficiency improves (i.e., the minimum reachable identification error reduces) with the increase of the data amount.

Theorem 4.1 *Under the same thresholds $\{C_i\}_{i=1}^{m}$, for $K_1\leq K_2$, choose the input subsequence $\{\phi_k\}_{k=1}^{K_1}$ and $\{\phi_k\}_{k=1}^{K_2}$ in the input sequence $\{\phi_k\}_{k=1}^{K}$. Then, the measure index $\eta(K,m,\theta)$ defined in* (15) *has the following property*

$$\eta(K_2,m,\theta)\leq\eta(K_1,m,\theta). \tag{16}$$

Proof From the definition of $h_{i,l}$ and $H_{i,l}$ in (6), we have $\sum_{i=1}^{m+1}\frac{h_{i,l}^2}{H_{i,l}}>0$ and

$$\sum_{l=K_1+1}^{K_2}\sum_{i=1}^{m+1}\frac{h_{i,l}^2}{H_{i,l}}\phi_l\phi_l^{\mathrm{T}}\geq 0.$$

Then, by (14), we get

$$\Delta_{K_2}(m,\theta)=\left(\Delta_{K_1}^{-1}(m,\theta)+\sum_{l=K_1+1}^{K_2}\sum_{i=1}^{m+1}\frac{h_{i,l}^2}{H_{i,l}}\phi_l\phi_l^{\mathrm{T}}\right)^{-1}\leq\Delta_{K_1}(m,\theta),$$

which together with (15) yields (16).

To intuitively demonstrate the monotony of space complexity, we introduce a definition about the way of threshold selection.

Definition 4.2 (see [17]) Let $\mathcal{C}_{m_1}=\{C_1,\cdots,C_{m_1}\}$ and $\mathcal{C}_{m_2}=\{C_1^*,\cdots,C_{m_2}^*\}$ be two placements of sensors, where m_1 and m_2 are two positive integers satisfying $m_1<m_2$, $C_1<C_2<\cdots<C_{m_1}$ and $C_1^*<C_2^*<\cdots<C_{m_2}^*$. Then, we say that $\mathcal{C}_{m_2}$ is a refinement of $\mathcal{C}_{m_1}$ if $\{C_1,\cdots,C_{m_1}\}\subseteq\{C_1^*,\cdots,C_{m_2}^*\}$ holds.

The following theorem states that quantized identification efficiency improves as the number of thresholds increases.

Theorem 4.3 *For the input sequence $\{\phi_k\}_{k=1}^{K}$, assume that $\mathcal{C}_{m_1}$ and $\mathcal{C}_{m_2}$ are two placements of sensor such that $\mathcal{C}_{m_2}$ is a refinement $\mathcal{C}_{m_1}$. Then, the measure index $\eta(K,m,\theta)$ defined in* (15) *satisfies*

$$\eta(K,m_2,\theta)\leq\eta(K,m_1,\theta). \tag{17}$$

Proof Actually, we just need to prove that $\eta(K, m_2, \theta) \leq \eta(K, m_1, \theta)$ when $m_2 = m_1 + 1$. It is because that we get $\eta(K, m_2, \theta) \leq \eta(K, m_1, \theta)$ by recursion based on $\eta(K, m+1, \theta) \leq \eta(K, m, \theta)$ $(m = m_1, \cdots, m_2 - 1)$, which implies (17).

Suppose the one additional threshold added to $\mathcal{C}_{m_1} = \{C_1, \cdots, C_{m_1}\}$ as C, which is between C_i and C_{i+1}. The new placement of sensors is defined as $\mathcal{C}_{m_2}$.

Define the corresponding coefficients $\sum_{i=1}^{m+1} \frac{h_{i,l}^2}{H_{i,l}}$ of $\eta(K, m_2, \theta)$ and $\eta(K, m_1, \theta)$ as $\rho_l^{m_1}$ and $\rho_l^{m_2}$, respectively. From the definitions of $h_{i,l}$ and $H_{i,l}$ in (6), we get

$$\begin{aligned}
&\rho_l^{m_1} - \rho_l^{m_2} \\
&= \sum_{j=1}^{m_1+1} \frac{h_{j,l}^2}{H_{j,l}} - \rho_l^{m_2} \\
&= \frac{(f_{i+1,l} - f_{i,l})^2}{F_{i+1,l} - F_{i,l}} - \frac{\left(f_{i+1,l} - f(C - \phi_l^{\mathrm{T}}\theta)\right)^2}{F_{i+1,l} - F(C - \phi_l^{\mathrm{T}}\theta)} - \frac{\left(f(C - \phi_l^{\mathrm{T}}\theta) - f_{i,l}\right)^2}{F(C - \phi_l^{\mathrm{T}}\theta) - F_{i,l}} \\
&= -\frac{\left((f_{i+1,l} - f_{i,l})(F_{i,l} - F(C - \phi_l^{\mathrm{T}}\theta)) - (f_{i,l} - f(C - \phi_l^{\mathrm{T}}\theta))(F_{i+1,l} - F_{i,l})\right)^2}{(F_{i+1,l} - F_{i,l})(F(C - \phi_l^{\mathrm{T}}\theta) - F_i)(F_{i+1,l} - F(C - \phi_l^{\mathrm{T}}\theta))} \leq 0, \qquad (18)
\end{aligned}$$

where $f_{i,l}$ and $F_{i,l}$ are defined in (4). Then by (14), we have

$$\Delta_K(m_2, \theta) = \left(\sum_{l=1}^{K} \rho_l^{m_2} \phi_l \phi_l^{\mathrm{T}}\right)^{-1} \leq \left(\sum_{l=1}^{K} \rho_l^{m_1} \phi_l \phi_l^{\mathrm{T}}\right)^{-1} = \Delta_K(m_1, \theta),$$

which together with (15) directly yields (17).

According to Theorems 4.1 and 4.3, the growth of both data amounts and threshold numbers can improve the identification efficiency of quantized identification. However, both two contribute the increase of required identification resources, which mean more costs. Therefore, it is not advisable to blindly increase identification resources. We must investigate the balance of time complexity (data amount) and space complexity (threshold number) to achieve optimal resource utilization of quantized identification. At first, we focus on how to design thresholds to achieve the highest identification efficiency under the same time-space complexity in the following section.

5 Threshold Selection for Binary Sensors

This section takes the binary sensor as an example to investigate threshold selection in quantized identification. Actually, an interval $(C_{i-1}, C_i]$ of the output range can offer useful information for quantized identification just when $h_{i,l} \neq 0$[32]. From the definition of $h_{i,l}$ in (6), we learn that the contribution of a sensor interval to error reduction depends on the actual parameter θ, the distribution function $F(\cdot)$ of noises, the threshold C and the inputs $\{\phi_k\}$. In this section, we mainly focus on the influence of threshold on the identification efficiency. An example is given below to show that thresholds have an important influence on error reduction.

Example 5.1 Consider the 1-order system (1), where the unknown parameter $\theta = 5$, the inputs $\{\phi_k\}_{k=1}^{100}$ are randomly chosen in the interval $[-2, 4]$, and the noises follow normal distribution $N(0, 10^2)$ with zero mean and the variance $\sigma^2 = 10^2$. Take the index defined in (15) with the threshold C as

$$\eta_C(100, 1, \theta) = \left(\sum_{l=1}^{100} \frac{f^2(C - \phi_l \theta)}{(1 - F(C - \phi_l \theta))\, F(C - \phi_l \theta)} \phi_l^2 \right)^{-1}.$$

Take C as -30, 20, 40 and 60, respectively. Then,

$$\eta_{-30}(100, 1, 5) = 28.3313; \quad \eta_{20}(100, 1, 5) = 0.4287;$$
$$\eta_{40}(100, 1, 5) = 3.8130; \quad \eta_{60}(100, 1, 5) = 1451.8731.$$

These indicate that the measure index $\eta_C(N, m, \theta)$ of identification efficiency is significantly impacted by the values of the threshold.

Remark 5.2 The significance of investigating the threshold selection problem lies in that proper threshold selection can make the quantized data contain more information of the original data and reduce the required data amount for an identification task. Threshold values greatly affect the identification efficiency of quantized identification, and a proper threshold selection can greatly improve the identification efficiency. Actually, threshold selection is often used in communication fields, that is, we make communication code contain more information through the design of appropriate thresholds (coding protocol). In communication coding, the coding protocol is designed in advance, in which case the thresholds are designable.

5.1 Optimal Design of Fixed Thresholds

This section discusses how to select thresholds to achieve the highest identification efficiency for binary sensors with fixed thresholds. To be specific, we would like to investigate the problem of the threshold selection of fixed thresholds, aiming at the prior information $\theta \in \Omega$, the input sequence $\{\phi_k\}_{k=1}^K$ satisfying persisting excitation (i.e., Assumption 2.2) and normally distributed noise with mean 0 and variance σ^2 (i.e., Assumption 2.3).

When the threshold of binary sensor is C, it can be seen that

$$h_{1,k} = f(C - \phi_k^{\mathrm{T}} \theta); \quad h_{2,k} = -f(C - \phi_k^{\mathrm{T}} \theta);$$
$$H_{1,k} = F(C - \phi_k^{\mathrm{T}} \theta); \quad H_{2,k} = 1 - F(C - \phi_k^{\mathrm{T}} \theta).$$

And then the index defined in (15) can be written as

$$\eta_C(K, 1, \theta) = \left\| \left(\sum_{l=1}^{K} \frac{f^2(C - \phi_l^{\mathrm{T}} \theta)}{\left(1 - F(C - \phi_l^{\mathrm{T}} \theta)\right) F(C - \phi_l^{\mathrm{T}} \theta)} \phi_l \phi_l^{\mathrm{T}} \right)^{-1} \right\|.$$

Denote

$$G(x) = \frac{f^2(x)}{(1 - F(x))\, F(x)}. \tag{19}$$

Then, the index defined in (15) can be transformed into

$$\eta_C(K, 1, \theta) = \left\| \left(\sum_{l=1}^{K} G(C - \phi_l^{\mathrm{T}} \theta) \phi_l \phi_l^{\mathrm{T}} \right)^{-1} \right\|. \tag{20}$$

Since the parameter θ is unknown and only the prior parameter information $\theta \in \Omega$ can be learned, the selection criterion of optimal thresholds is established based on minimizing the minimum reachable identification error (i.e., maximizing identification efficiency) of the worst case in the prior parameter range. Specifically, for the input sequence $\{\phi_k\}_{k=1}^K$ following persisting excitations, the optimal fixed threshold selection method is as follows

$$\eta^*(K,1) = \inf_C \sup_{\theta\in\Omega} \eta_C(K,1,\theta). \tag{21}$$

It is noted that we take the worst case of the prior parameter information in the optimal threshold selection method when the threshold is fixed. In some cases, the performance of the optimal threshold given by this method may be not good because the prior information on parameters is poor. It is noted that the optimal identification algorithm can provide more accurate information about the unknown parameters, which implies that the identification efficiency can be further improved based on it. Next, we will utilize the estimate given by the optimal identification algorithm, i.e., the IBID algorithm, to adaptively design the time-varying threshold C_k to achieve a higher identification efficiency.

5.2 Design of Adaptive Threshold

This part discusses how to select thresholds to realize the highest identification efficiency for binary sensors with time-varying and designable thresholds. Specifically, we would like to study the problem of the threshold selection on time-varying thresholds, aiming at the prior information $\theta \in \Omega$, the input sequence $\{\phi_k\}_{k=1}^K$ satisfying persisting excitation condition (i.e., Assumption 2.2) and normally distributed noise with mean 0 and variance σ^2 (i.e., Assumption 2.3).

When the time-varying threshold of binary sensor is C_k, it can be seen that

$$\begin{aligned}
&h_{1,k} = f(C_k - \phi_k^{\mathrm{T}}\theta); \quad h_{2,k} = -f(C_k - \phi_k^{\mathrm{T}}\theta);\\
&H_{1,k} = F(C_k - \phi_k^{\mathrm{T}}\theta); \quad H_{2,k} = 1 - F(C_k - \phi_k^{\mathrm{T}}\theta).
\end{aligned}$$

Then, the measure index defined in (15) is as follows:

$$\begin{aligned}
\eta_{C_{1:K}}(K,1,\theta) &= \left\| \left(\sum_{l=1}^K \frac{f^2(C_l - \phi_l^{\mathrm{T}}\theta)}{\left(1 - F(C_l - \phi_l^{\mathrm{T}}\theta)\right) F(C_l - \phi_l^{\mathrm{T}}\theta)} \phi_l \phi_l^{\mathrm{T}} \right)^{-1} \right\| \\
&= \left\| \left(\sum_{l=1}^K G(C_l - \phi_l^{\mathrm{T}}\theta) \phi_l \phi_l^{\mathrm{T}} \right)^{-1} \right\|,
\end{aligned} \tag{22}$$

where $C_{1:K} = \{C_1, C_2, \cdots, C_K\}$.

Then, based on the principle of minimizing the minimum reachable identification error (i.e., maximizing the identification efficiency), the selection method of optimal time-varying thresholds is established as

$$\eta^*(K,1,\theta) = \inf_{C_{1:K}} \left\| \left(\sum_{l=1}^K G(C_l - \phi_l^{\mathrm{T}}\theta) \phi_l \phi_l^{\mathrm{T}} \right)^{-1} \right\|. \tag{23}$$

From $\phi_l\phi_l^{\mathrm{T}} \geq 0$ and Weyl's inequality, the selection criterion in (23) can be transformed into

$$\eta^*(K,1,\theta) = \left\| \left(\sum_{l=1}^{K} \sup_{C_l} G(C_l - \phi_l^{\mathrm{T}}\theta)\phi_l\phi_l^{\mathrm{T}} \right)^{-1} \right\|. \tag{24}$$

In order to further analyze the selection of optimal time-varying threshold method, we introduce the following lemma.

Lemma 5.3 (see [17]) *$G(x)$ defined in* (19) *has the following properties,* $G'(x) = \frac{dG(x)}{x} < 0$ *for* $x > 0$ *and* $G'(x) > 0$ *for* $x < 0$.

Then, based on Lemma 5.3, the maximum of $G(C_l - \phi_l^{\mathrm{T}}\theta)$ is got when $C_l = \phi_l^{\mathrm{T}}\theta$, i.e.,

$$\sup_{C_l} G(C_l - \phi_l^{\mathrm{T}}\theta) = G(0) = \frac{(f(0))^2}{F(0)(1-F(0))} = \frac{2}{\pi\sigma^2},$$

which together with (24) gives

$$\eta^*(K,1,\theta) = \left\| \left(\sum_{l=1}^{K} \frac{2}{\pi\sigma^2}\phi_l\phi_l^{\mathrm{T}} \right)^{-1} \right\| = \frac{\pi\sigma^2}{2} \left\| \left(\sum_{l=1}^{K} \phi_l\phi_l^{\mathrm{T}} \right)^{-1} \right\|. \tag{25}$$

However, the optimal time-varying threshold $C_k = \phi_k^{\mathrm{T}}\theta$ is unknown due to containing the unknown parameter. Hence, we use its estimate $\widehat{\theta}_{k-1}$ to design the time-varying threshold C_k instead of the unknown parameter. So the time-varying threshold C_k is adaptively designed as

$$C_k = \phi_k^{\mathrm{T}}\widehat{\theta}_{k-1}, \tag{26}$$

where $\widehat{\theta}_{k-1}$ is given by the 1-order IBID algorithm.

Remark 5.4 From $C_k \in \mathcal{F}_{k-1} = \sigma(d_1, \cdots, d_{k-1})$, the adaptive weight $\widehat{\alpha}_{i,k}$ and $\widehat{\beta}_k$ in the IBID algorithm (8)–(13) still belong to $\mathcal{F}_{k-1}$. Therefore, the convergence and asymptotic efficiency of the IBID algorithm still hold (i.e., Theorems 3.1 and 3.3 hold) under the adaptive threshold design. In other words, the IBID algorithm is an asymptotically optimal quantized identification algorithm with the adaptive threshold (26), and it can be used to design the time-varying thresholds.

Remark 5.5 Compared with [33], this paper provides complete theoretical support for the optimal threshold selection method of time-varying thresholds, instead of utilizing the upper bound of the estimation covariance to design the threshold. Furthermore, this paper also presents the expression of adaptive thresholds. In addition, the design of the adaptive threshold is given by the optimal identification algorithm, without designing additional adaptive threshold iteration. So this design method can reduce the amount of computation and algorithm complexity in contrast with [32].

Note that CR lower bound is $\sigma^2 \left(\sum_{l=1}^{K} \phi_l\phi_l^{\mathrm{T}} \right)^{-1}$ under accurate measurements. Under the design of adaptive threshold (26), the quantized identification efficiency can approximately tend to (25). It means that the minimum reachable identification error under this adaptive threshold of binary sensor is only $\pi/2$ times the one under accurate measurement. Therefore, the identification efficiency under binary-valued data can achieve the approximate performance as the one with accurate data.

Next, an example is given to verify the validity of the optimal fixed threshold selection method and the adaptive threshold selection method.

Example 5.6 Consider a first-order linear system (1), where the unknown parameter is $\theta = 2$ and the prior information is $\theta \in \Omega = [1, 6]$. The input sequence $\{\phi_k\}_{k=1}^{1000}$ is randomly chosen in the range $[3, 6]$. The noises $\{d_k\}$ follow normal distribution $N(0, 15^2)$. It can be verified that the input and noises satisfy Assumptions 2.2 and 2.3, respectively.

Figure 1 shows the measure index $\eta_C(1000, 1) = \sup_{\theta\in\Omega} \eta_C(1000, 1, \theta)$. Then, the optimal fixed threshold is $C = 17.20$, and the corresponding the minimum reachable identification error is $\eta^*(1000, 1) = 0.0218$.

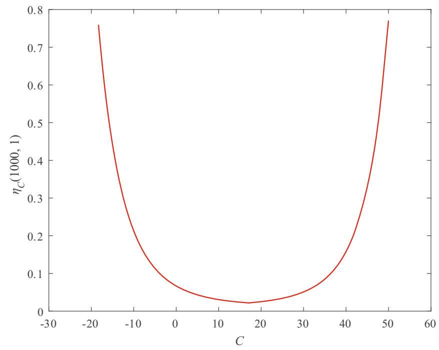

Figure 1 Optimal worst-case threshold selection for $\theta \in [1, 6]$ and $d_k \sim N(0, 15^2)$

Figure 2 demonstrates the convergence properties of the first-order IBID algorithm with non-optimal fixed threshold, optimal fixed threshold and adaptive threshold, respectively. It can be seen that the identification efficiency of the algorithm is significantly improved with the optimal fixed threshold compared with the general non-optimal fixed threshold. Furthermore, the identification efficiency under the adaptive threshold design is better than that under the optimal fixed threshold design.

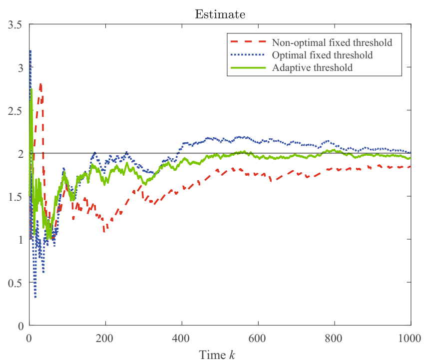

Figure 2 Comparison of identification properties under non-optimal threshold $C = -10$, optimal threshold $C = 17.2$ and adaptive threshold $C_k = \phi_k\widehat{\theta}_{k-1}$

6 Optimal Resource Allocation

This section focuses on the optimal resource allocation problem of quantized identification. That is, in order to achieve the optimal identification efficiency under the given identification resource R, how to balance the time and space complexity (i.e., the amount of data K and the number of thresholds m). We will provide some concrete resource allocation methods for fixed threshold and adaptive threshold cases, respectively.

Fixed threshold case: For the given identification resource R, input sequence $\{\phi_k\}_{k=1}^{K_0}$ ($K_0 \geq R$) and prior information $\theta \in \Omega$, we take the minimizing the minimum reachable identification error (i.e., maximizing the identification efficiency) of the worst case in the prior parameter range as the resource allocation criterion. Then, the minimum reachable identification error of the worst case under the given resource is as follows:

$$\varepsilon(R) = \min_{m \in \mathbb{Z}_+} \max_{\theta \in \Omega} \eta(K, m, \theta) \quad \text{s.t.} \quad K \log_2(m+1) \leq R.$$

Based on it, the optimal resource allocation method for the given resource R is

$$(K, m) = \arg \min_{K \log_2(m+1) \leq R} \varepsilon(R).$$

Consider the following two threshold selection methods to increase space complexity.

The first one is a structured threshold case, i.e., threshold sets are restricted to a pre-specified category. For example, the threshold is selected according to dividing equally the threshold value range $[a, b]$. When $m = 1$, the threshold is set as $C_1^1 = \frac{a+b}{2}$. When $m = 2$, the thresholds are chosen as $C_2^1 = \frac{2a+b}{3}$ and $C_2^2 = \frac{a+2b}{3}$. Select the thresholds by analogy. In this case, when the threshold selection range is given, the threshold selection is one-to-one corresponding to the number of thresholds.

The second one is the non-structured threshold case. For a given number of thresholds m, the threshold in $\mathcal{C}^m = \{C_1^m, \cdots, C_m^m\}$ can be selected within the preset threshold range ℓ_C. The general selection principle is to choose the thresholds and data amount that minimize the identification error under limited resources. In this case, the minimum reachable identification error and the optimal resource allocation method are

$$\begin{aligned}
&\varepsilon(R) = \min_{m \in \mathbb{Z}_+} \min_{\mathcal{C}^m \in \ell_C} \max_{\theta \in \Omega} \eta(K, m, \theta) \quad \text{s.t.} \quad K \log_2(m+1) \leq R, \\
&(K, m) = \arg \min_{K \log_2(m+1) \leq R} \varepsilon(R),
\end{aligned}$$

respectively. In this method, the threshold for $m = 1$ is selected as follows:

$$C = \arg \min_{C \in \ell_C} \max_{\theta \in \Omega} \eta(K, 1, \theta). \tag{27}$$

Adaptive threshold case: For the given identification resource R and input sequence $\{\phi_k\}_{k=1}^{K_0}$ ($K_0 \geq R$), we take the minimizing the minimum reachable identification error (i.e., maximizing the identification efficiency) as the allocation criterion. Then, the minimum reachable identification error and the optimal resource allocation method are as follows:

$$\varepsilon(R) = \min_{m\in\mathbb{Z}_+} \eta_{\mathcal{C}^m_{1:K}}(K, m, \theta) \quad \text{s.t.} \quad K\log_2(m+1) \leq R,$$
$$(K, m) = \arg\min_{K\log_2(m+1)\leq R} \varepsilon(R),$$

where $\mathcal{C}^m_{1:K} = \{C^m_{i,k} = F^{-1}(\frac{i}{m+1}) + \phi_k^{\mathrm{T}}\widehat{\theta}_{k-1}, i = 1, \cdots, m, k = 1, \cdots, K\}$, $F^{-1}(\cdot)$ is the inverse function of the distributed function $F(\cdot)$ of noises and $\widehat{\theta}_{k-1}$ is given by the IBID algorithm.

Remark 6.1 The design idea of the adaptive thresholds is that choosing $C^m_{i,k}$ $(i = 1, 2, \cdots, m)$ make $F(C^m_{i,k} - \phi_k^{\mathrm{T}}\theta)$ divide equally the distribution range $[0, 1]$, which could make each threshold interval contain more information on the unknown parameter. It is noticed that the analytic expression of the adaptive threshold design $C^m_{i,k}$ calculated by the above idea is related to the unknown parameter. Therefore, we use the estimate given by the IBID algorithm to design the adaptive threshold instead of the unknown parameter. To be specific, the design method of the adaptive thresholds is that we can select $C^m_{i,k}$ $(i = 1, 2, \cdots, m)$ to make $F(C^m_{i,k} - \phi_k^{\mathrm{T}}\widehat{\theta}_{k-1}) - F(C^m_{i-1,k} - \phi_k^{\mathrm{T}}\widehat{\theta}_{k-1}) = \frac{1}{m+1}$, where $C^m_{0,k} = -\infty$, $C^m_{m+1,k} = \infty$ and $\widehat{\theta}_{k-1}$ is given by the first-order IBID algorithm. In other words, the adaptive thresholds for the multiple threshold case are $C^m_{i,k} = F^{-1}(\frac{i}{m+1}) + \phi_k^{\mathrm{T}}\widehat{\theta}_{k-1}$ for $i = 1, \cdots, m$. For example, when $m = 1$, the adaptive threshold is designed as $C^1_{1,k} = F^{-1}(\frac{1}{2}) + \phi_k^{\mathrm{T}}\widehat{\theta}_{k-1} = \phi_k^{\mathrm{T}}\widehat{\theta}_{k-1}$, which makes $F(C^1_{1,k} - \phi_k^{\mathrm{T}}\widehat{\theta}_{k-1}) = \frac{1}{2}$.

Example 6.2 For the model in Example 5.6 under a given resource $R = 1000$ bits, we consider the optimal resource allocation problem in the following three cases, respectively.

1) Structured fixed threshold case. The threshold selection range is $[-70, 30]$, and the thresholds are selected according to dividing equally threshold value range. In other words, the threshold is designed as $C^1_1 = -20$ when $m = 1$. In this case, the resource occupied by space complexity is $\log_2(m + 1) = 1$ *bit*, and the corresponding usable time complexity is $K \leq \left\lfloor \frac{R}{\log_2(m+1)} \right\rfloor = 1000$. Then, increase the number of thresholds in sequence. So the thresholds are chosen as $C^2_1 = -\frac{110}{3}$ and $C^2_2 = -\frac{10}{3}$ when $m = 2$. In this case, the resource occupied by space complexity is $\log_2(m + 1) = \log_2 3$ bits, and the corresponding usable time complexity is $K \leq \left\lfloor \frac{R}{\log_2(m+1)} \right\rfloor = 630$. Select the thresholds by analogy.

2) Unstructured fixed threshold case. Choose the thresholds by minimizing the identification error under limited resource. For instance, the threshold is chosen as (27) for $m = 1$.

3) Adaptive threshold case. This method adaptively designs the threshold based on the estimates given by the IBID algorithm, i.e., $C^m_{i,k} = F^{-1}(\frac{1}{m+1}) + \phi_k^{\mathrm{T}}\widehat{\theta}_{k-1}$ for $i = 1, \cdots, m$.

Figures 3–5 give the comparison of the minimum reachable identification error under different space complexity in the case of structured fixed threshold, unstructured fixed optimal threshold and adaptive threshold, respectively. According to Figure 3, the optimal resource allocation is achieved when the number of thresholds is $m = 3$ in the structured threshold case. And Figures 4–5 show that the minimum identification error can be achieved when the number of thresholds is $m = 1$ for both the unstructured fixed optimal threshold and the adaptive threshold cases. Further, comparing Figure 3 and Figures 4–5, we find that it can greatly reduce the identification error and improve the utilization rate of resources by using the optimal fixed threshold and adaptive threshold in the threshold design of binary sensor.

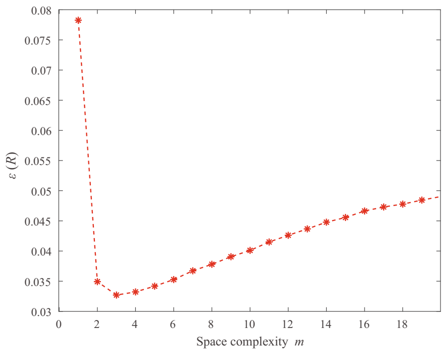

Figure 3 The minimum reachable identification error for the structured thresholds

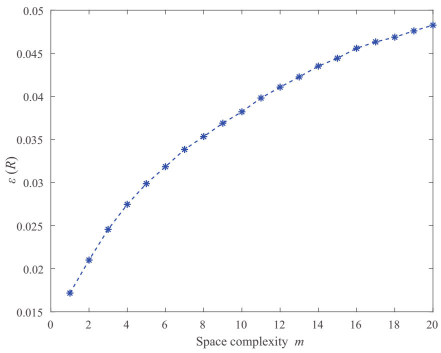

Figure 4 The minimum reachable identification error for the optimal fixed thresholds

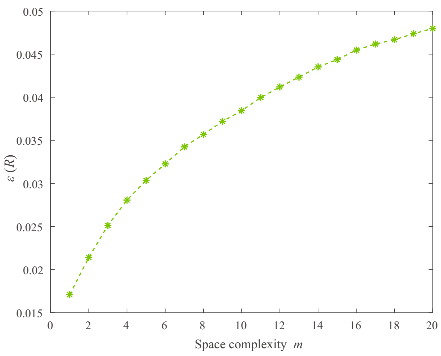

Figure 5 The minimum reachable identification error for the adaptive thresholds

7 Concluding Remarks

This paper investigates the optimal threshold selection problem of quantized identification under aperiodic signals, and answers how to realize quantized identification with higher resource utilization. First, this paper establishes the optimal convergence rate and asymptotic efficiency of the first-order IBID algorithm under general persisting excitation condition. Then, taking the norm of the CR lower bound as the measure index, the influence of space and time complexity on the identification efficiency are analyzed. On this basis, the optimal fixed threshold selection method and adaptive threshold design method are designed to answer how to achieve the highest quantized identification efficiency under the same time-space complexity. In addition, based on the principle of minimizing the minimum reachable identification error under given resources, this paper establishes the optimal resource allocation methods for the fixed threshold and adaptive threshold cases, respectively. And then, this paper shows how to balance the time-space complexity to achieve the best resource utilization.

There are still lots of interesting problems for further research. For example, we can consider the optimal quantized identification algorithm under more general systems and noise conditions, and then based on it we can also consider optimal threshold selection and resource allocation problems of online quantized identification.

Conflict of Interest

ZHAO Yanlong is an editorial board member for Journal of Systems Science & Complexity and was not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

References

[1] Chen H F, *Stochastic Approximation and Its Applications*, Kluwer Academic Publishers, Dordrecht, 2002.

[2] Tan S P, Guo J, Zhao Y L, et al., Adaptive control with saturation-constrainted observations for drag-free satellites — A set-valued identification approach, *Science China Information Sciences*, 2021, **64**: 202202.

[3] Li J, Wu L, Lü W, et al., Lithology classification based on set-valued identification method, *Journal of Systems Science & Complexity*, 2022, **35**(5): 1637–1652.

[4] Zhang X, *Modern Signal Processing*, Boston: De Gruyter, Berlin, 2023.

[5] Kang G L, Bi W J, Zhao Y L, et al., A system identification approach to identifying genetic variants in sequencing studies for a binary phenotype, *Human Heredity*, 2014, **78**: 104–116.

[6] Zhang H, Bi W J, Cui Y, et al., Extreme-value sampling design is cost-benefit only with valid statistical approach for exposure-secondary outcome association analyses, *Statistical Methods in Medical Research*, 2020, **29**(2): 466–480.

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

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

[9] Wang J, Tan J W, and Zhang J F, Differentially private distributed parameter estimation, *Journal of Systems Science & Complexity*, 2023, **36**(1): 187–204.

[10] Wang L Y, Zhang J F, and Yin G, System identification using binary sensors, *IEEE Transactions on Automatic Control*, 2003, **48**(11): 1892–1907.

[11] Bi W J, Kang G L, Zhao Y L, et al., A fast and powerful set-valued system identification approach to identifying rare variants in sequencing studies for ordered categorical traits, *Annals of Human Genetics*, 2015, **79**: 294–309.

[12] Guo J, Jia R, Su R, et al., Identification of FIR systems with binary-valued observations against data tampering attacks, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, **53**(9): 5861–5873.

[13] Wang L Y, Yin G, Zhang J F, et al., *System Identification with Quantized Observations*, Birkhauser, Boston, 2010.

[14] He J, Yang E H, Yang F, et al., Adaptive quantization parameter selection for H.265/HEVC by employing inter-frame dependency, *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, **28**(12): 3424–3436.

[15] Chen X and Wang X, Research on quantization distortion estimation algorithm of JPEG, *Computer Simulation*, 2022, **39**(2): 191–194.

[16] Sun C and Yang E H, An efficient DCT-based image compression system based on Laplacian transparent composite model, *IEEE Transactions on Image Processing*, 2015, **24**(3): 886–900.

[17] Wang L Y and Yin G, Asymptotically efficient parameter estimation using quantized output observations, *Automatica*, 2007, **43**(7): 1178–1191.

[18] Zhao Y L, Zhang H, Wang T, et al., System identification under saturated precise or setvalued measurements, *Science China Information Sciences*, 2023, **66**: 112204.

[19] Godoy B, Goodwin G, Agüero J, et al., On identification of FIR systems having quantized output data, *Automatica*, 2011, **47**(9): 1905–1915.

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

[21] Risuleo R S, Bottegal G, and Hjalmarsson H, Identification of linear models from quantized data: A midpoint-projection approach, *IEEE Transactions on Automatic Control*, 2020, **65**(7): 2801–2813.

[22] Wang X, Li C, Li T, et al, Variational bayesian inference for the identification of FIR systems via quantized output data, *Automatica*, 2021, **132**: 109827.

[23] Goudjil A, Pouliquen M, Pigeon E, et al., Identification of systems using binary sensors via support vector machines, *Proceedings of the 54th IEEE Conference on Decision and Control*, Osaka, 2015, 3385–3390.

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

[25] Wang Y, Zhao Y L, Zhang J F, et al., A unified identification algorithm of FIR systems based on binary observations with time-varying thresholds, *Automatica*, 2022, **135**: 109990.

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

[27] Zhao W, Chen H F, Tempo R, et al., Recursive nonparametric identification of nonlinear systems with adaptive binary sensors, *IEEE Transactions on Automatic Control*, 2017, **62**(8): 3959–3971.

[28] You K, Recursive algorithms for parameter estimation with adaptive quantizer, *Automatica*, 2015, **52**: 192–201.

[29] Jafari K, Juillard J, and Roger M, Convergence analysis of an online approach to parameter

estimation problems based on binary observations, *Automatica*, 2012, **48**(11): 2837–2842.

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

[31] Wang Y, Zhao Y L, and Zhang J F, Asymptotically efficient quasi-newton type identification with quantized observations under bounded persistent excitations, 2023, arXiv: 2309.04984.

[32] Wang L Y, Yin G, Zhang J F, et al., Space and time complexities and sensor threshold selection in quantized identification, *Automatica*, 2008, **44**(12): 3014–3024.

[33] Guo J and Zhao Y L, Identification of the gain system with quantized observations and bounded persistent excitations, *Science China Information Sciences*, 2014, **57**: 012205.

[34] Calamai P H and Moré J J, Projected gradient methods for linearly constrained problems, *Mathematical Programming*, 1987, **39**: 93–116.

Appendix Proof of Theorem 3.1

Based on Lemma 5 and [31, Proposition 2], we have

$$\widehat{\alpha}_{i+1,k} > \widehat{\alpha}_{i,k}, \quad \widehat{\alpha}_{m+1,k} - \widehat{\alpha}_{1,k} \geq \widehat{\alpha} > 0, \tag{A.1}$$

$$|\widehat{\alpha}_{i,k}| \leq \overline{\widehat{\alpha}} \leq \infty, \tag{A.2}$$

$$\underline{\widehat{\beta}} = \inf_k \widehat{\beta}_k > 0, \quad \overline{\widehat{\beta}} = \sup_k \widehat{\beta}_k \leq \infty. \tag{A.3}$$

From (13), we have

$$\widehat{P}_k^{-1} = \widehat{P}_{k-1}^{-1} + \widehat{\beta}_k \phi_k^2 = \widehat{P}_0^{-1} + \sum_{l=1}^{k} \widehat{\beta}_l \phi_l^2.$$

Then, by Assumption 2.2 and (A.3), we have

$$\widehat{P}_k = O\left(\frac{1}{k}\right), \quad \widehat{P}_k^{-1} = O\left(k\right). \tag{A.4}$$

By the definition of $H_{i,k}$ and $\widehat{H}_{i,k}$ in (6) and (7) and the differential mean value theorem, there exists $\check{\theta}_{i,k-1}$ with $\phi_k \check{\theta}_{i,k-1}$ in the interval between $\phi_k^{\mathrm{T}} \theta$ and $\phi_k^{\mathrm{T}} \widehat{\theta}_{k-1}$ such that

$$\begin{aligned}
\mathbb{E}[\widetilde{s}_k | \mathcal{F}_{k-1}] &= \sum_{i=1}^{m+1} \widehat{\alpha}_{i,k} \left(H_{i,k} - \widehat{H}_{i,k}\right) \\
&= \sum_{i=1}^{m} \left(\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}\right) \left(\widehat{F}_{i,k} - F_{i,k}\right) \\
&= -\sum_{i=1}^{m} \left(\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}\right) f(C_i - \phi_k^{\mathrm{T}} \check{\theta}_{i,k-1}) \phi_k^{\mathrm{T}} \widetilde{\theta}_{k-1},
\end{aligned} \tag{A.5}$$

where $\mathcal{F}_{k-1} = \sigma\{d_1, \cdots, d_{k-1}\}$. Define $\check{f}_{i,k} \triangleq f(C_i - \phi_k^{\mathrm{T}} \xi_{i,k})$. Then

$$\mathbb{E}[\widetilde{s}_k | \mathcal{F}_{k-1}] = -\sum_{i=1}^{m} \left(\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}\right) \check{f}_{i,k} \phi_k^{\mathrm{T}} \widetilde{\theta}_{k-1}. \tag{A.6}$$

From (10) and [34, Lemma 2.1], we have

$$\widetilde{\theta}_k^2 \le \widetilde{\theta}_{k-1}^2 + 2\widehat{a}_k \phi_k \widehat{P}_{k-1} \widetilde{\theta}_{k-1} \widetilde{s}_k + \widehat{a}_k^2 \phi_k \widehat{P}_{k-1}^2 \phi_k \widetilde{s}_k^2. \tag{A.7}$$

By (A.2), (A.4)–(A.7) and Assumption 2.2, we get

$$\begin{aligned}
\mathbb{E}\widetilde{\theta}_k^2 &\le \mathbb{E}\widetilde{\theta}_{k-1}^2 + 2\mathbb{E}\widehat{a}_k \phi_k \widehat{P}_{k-1} \widetilde{\theta}_{k-1} \widetilde{s}_k + \mathbb{E}\widehat{a}_k^2 \phi_k \widehat{P}_{k-1}^2 \phi_k \widetilde{s}_k^2 \\
&\le \mathbb{E}\widetilde{\theta}_{k-1}^2 + 2\mathbb{E}\sum_{i=1}^{m+1} \widehat{\alpha}_{i,k} \left(H_{i,k} - \widehat{H}_{i,k}\right) \widehat{a}_k \widehat{P}_{k-1} \phi_k^2 \widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k^2}\right) \\
&\le \mathbb{E}\widetilde{\theta}_{k-1}^2 - 2\mathbb{E}\sum_{i=1}^{m} \left(\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}\right) \check{f}_{i,k} \widehat{a}_k \widehat{P}_{k-1} \phi_k^2 \widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k^2}\right) \\
&\le \mathbb{E}\widetilde{\theta}_{k-1}^2 - 2\mathbb{E}\lambda_k \beta_k a_k P_{k-1} \phi_k^2 \widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k^2}\right),
\end{aligned} \tag{A.8}$$

where P_k is generated by (13) with $\beta_k = \sum_{i=1}^{m+1} \frac{h_{i,k}^2}{H_{i,k}}$, $a_k = \left(1 + \beta_k P_{k-1} \phi_k^2\right)^{-1}$ and

$$\lambda_k = \frac{\sum_{i=1}^{m} \left(\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}\right) \check{f}_{i,k} \widehat{a}_k \widehat{P}_{k-1}}{\beta_k a_k P_{k-1}}. \tag{A.9}$$

From the boundness of $\widehat{\theta}_k$, θ and ϕ_k, (A.1) and the continuity of $f(x)$ and $F(x)$, we learn that λ_k is bounded, i.e.,

$$\underline{\lambda} = \inf_k \lambda_k > 0, \quad \overline{\lambda} = \sup_k \lambda_k < \infty. \tag{A.10}$$

Then, by the boundness of θ and ϕ_k and Assumption 2.3, we have

$$\underline{\beta} = \inf_k \beta_k > 0, \quad \overline{\beta} = \sup_k \beta_k \le \infty, \tag{A.11}$$

which together with $P_k^{-1} = P_0^{-1} + \sum_{l=1}^{k} \beta_l \phi_l^2$ and Assumption 2.2 yields

$$P_k = O\left(\frac{1}{k}\right), \quad P_k^{-1} = O\left(k\right). \tag{A.12}$$

From (13), we get

$$\sum_{l=1}^{k} a_l \beta_l P_{l-1} \phi_l^2 = \sum_{l=1}^{k} \frac{P_l^{-1} - P_{l-1}^{-1}}{P_l^{-1}} \le \sum_{l=1}^{k} \int_{P_{l-1}^{-1}}^{P_l^{-1}} \frac{dx}{x} = \log P_k^{-1} - \log P_0^{-1}. \tag{A.13}$$

Based on (A.12) and (A.13), we get

$$\begin{aligned}
\prod_{l=j+1}^{k} \left(1 - 2\underline{\lambda} a_l \beta_l P_{l-1} \phi_l^2\right) =& \mathrm{e}^{\sum_{l=j+1}^{k} \log\left(1 - 2\underline{\lambda} a_l \beta_l P_{l-1} \phi_l^2\right)} \\
\sim& \mathrm{e}^{-2\underline{\lambda} \sum_{l=j+1}^{k} a_l \beta_l P_{l-1} \phi_l^2}
\end{aligned}$$

$$
\begin{aligned}
&\le \mathrm{e}^{-2\underline{\lambda}\left(\log P_k^{-1}-\log P_j^{-1}\right)}\\
&=\left(\frac{P_j^{-1}}{P_k^{-1}}\right)^{2\underline{\lambda}}\\
&=O\left(\left(\frac{j}{k}\right)^{2\underline{\lambda}}\right). \qquad \text{(A.14)}
\end{aligned}
$$

From (A.8), (A.10) and (A.14), we have

$$
\begin{aligned}
\mathbb{E}\widetilde{\theta}_k^2 &\le \left(1-2\underline{\lambda}\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k^2}\right)\\
&\le \prod_{l=1}^{k}\left(1-2\underline{\lambda}a_l\beta_l P_{l-1}\phi_l^2\right)^2\mathbb{E}\widetilde{\theta}_0^2 + O\left(\sum_{l=1}^{k}\prod_{j=l+1}^{k}\left(1-2\underline{\lambda}a_j\beta_j P_{j-1}\phi_j^2\right)\frac{1}{l^2}\right)\\
&=O\left(\frac{1}{k^{2\underline{\lambda}}}\right)+O\left(\frac{1}{k^{2\underline{\lambda}}}\sum_{l=1}^{k}\frac{1}{l^{2-2\underline{\lambda}}}\right)\\
&=\begin{cases}
O\left(\dfrac{1}{k}\right), & \text{if}\quad 2\underline{\lambda}>1,\\
O\left(\dfrac{\log k}{k}\right), & \text{if}\quad 2\underline{\lambda}=1,\\
O\left(\dfrac{1}{k^{2\underline{\lambda}}}\right), & \text{if}\quad 2\underline{\lambda}<1.
\end{cases} \qquad \text{(A.15)}
\end{aligned}
$$

Noticing that $2\underline{\lambda}>0$, we derive that 1-order IBID algorithm is convergent in mean square sense. By (A.6) and (A.7), we get $\mathbb{E}[\widetilde{\theta}_k^2|\mathcal{F}_{k-1}]\le\widetilde{\theta}_{k-1}^2+O\left(\frac{1}{k^2}\right)$, which together with [1, Lemma 1.2.2] yields that $\widetilde{\theta}_k^2$ converges almost surely to a bounded limit. Then, there is a subsequence of $\widetilde{\theta}_k$ that converges almost surely to 0. Noticing $\mathbb{E}\widetilde{\theta}_k^2$, we learn that $\widetilde{\theta}_k$ almost surely converges to 0.

Similar to (A.7)–(A.15), we can get

$$
\mathbb{E}\widetilde{\theta}_k^{2r}=\begin{cases}
O\left(\dfrac{1}{k^r}\right), & \text{if}\ \ 2\underline{\lambda}>1,\\
O\left(\dfrac{(\log k)^r}{k^r}\right), & \text{if}\quad 2\underline{\lambda}=1,\quad \text{for } r=2,3\cdots.\\
O\left(\dfrac{1}{k^{2r\underline{\lambda}}}\right), & \text{if}\quad 2\underline{\lambda}<1,
\end{cases} \qquad \text{(A.16)}
$$

Next, we will prove $\mathbb{E}\widetilde{\theta}_k^2=O\left(\frac{1}{k}\right)$ by dividing the value of $2\underline{\lambda}$ into different cases, based on the high order moment convergence rate of the estimation error.

Case i $2\underline{\lambda}>1$.

Then, by (A.15), we have $\mathbb{E}\widetilde{\theta}_k^2=O\left(\frac{1}{k}\right)$.

Case ii $2\underline{\lambda}\le 1$.

Denote $\beta(x)=\sum_{i=1}^{m+1}\frac{(f(C_i-x)-f(C_{i-1}-x))^2}{F(C_i-x)-F(C_{i-1}-x)}$. Then, $\beta_k=\beta(\phi_k^{\mathrm{T}}\theta)$ and $\widehat{\beta}_k=\beta(\phi_k^{\mathrm{T}}\widehat{\theta}_{k-1})$.

From $\widehat{\beta}_k = \sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) \widehat{f}_{i,k}$, (A.1), (A.4), (A.9), (A.12), Assumption 2.3 and the continuous differentiability of $f(x)$ and $F(x)$, we have

$$\begin{aligned}
|1-\lambda_k| =& \left|1 - \frac{\sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) \check{f}_{i,k}\widehat{a}_k\widehat{P}_{k-1}}{\beta_k a_k P_{k-1}}\right| \\
\le & \left|1 - \frac{\sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) \check{f}_{i,k}}{\sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) \widehat{f}_{i,k}}\right| + \left|\sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) \left(\frac{\check{f}_{i,k}\widehat{a}_k\widehat{P}_{k-1}}{\widehat{\beta}_k\widehat{a}_k\widehat{P}_{k-1}} - \frac{\check{f}_{i,k}\widehat{a}_k\widehat{P}_{k-1}}{\beta_k a_k P_{k-1}}\right)\right| \\
\le & \left|\frac{\sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) \left(\widehat{f}_{i,k} - \check{f}_{i,k}\right)}{\sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) \widehat{f}_{i,k}}\right| + \left|\sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) \check{f}_{i,k}\widehat{a}_k\widehat{P}_{k-1}\frac{\beta_k - \widehat{\beta}_k}{\widehat{\beta}_k\beta_k\widehat{P}_{k-1}}\right| \\
& + \left|\sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) \check{f}_{i,k}\widehat{a}_k\widehat{P}_{k-1}\left(\frac{\widehat{P}_{k-1}^{-1} - P_{k-1}^{-1}}{\beta_k\widehat{P}_{k-1}^{-1}P_{k-1}^{-1}}\right)\right| \\
\le & \left|\frac{\sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) f^{'}(\xi_{i,k})\phi_k\left(\widehat{\theta}_{k-1} - \check{\theta}_{k-1}\right)}{\sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) \widehat{f}_{i,k}}\right| \\
& + \left|\sum_{i=1}^m (\widehat{\alpha}_{i+1,k} - \widehat{\alpha}_{i,k}) \check{f}_{i,k}\widehat{a}_k\frac{\beta^{'}(\widehat{\zeta}_k)\phi_k\left(\theta - \widehat{\theta}_{k-1}\right)}{\widehat{\beta}_k\beta_k}\right| + O\left(\frac{1}{k^2}\right) \\
=& O\left(|\widetilde{\theta}_k|\right) + O\left(\frac{1}{k^2}\right),
\end{aligned} \tag{A.17}$$

where $\xi_{i,k}$ and $\widehat{\zeta}_{i,k}$ are between $\phi_k^{\mathrm{T}}\widehat{\theta}$ and $\phi_k^{\mathrm{T}}\check{\theta}_{i,k-1}$ and between $\phi_k^{\mathrm{T}}\theta$ and $\phi_k^{\mathrm{T}}\widehat{\theta}_{k-1}$, respectively. And $\check{f}_{i,k}$ and $\check{\theta}_{i,k-1}$ are defined as (A.6).

From Assumption 2.2 and (A.12), taking (A.17) into (A.8) gives

$$\begin{aligned}
\mathbb{E}\widetilde{\theta}_k^2 \le & \mathbb{E}\widetilde{\theta}_{k-1}^2 - 2\mathbb{E}\beta_k a_k P_{k-1}\phi_k^2\widetilde{\theta}_{k-1}^2 \\
& + 2\mathbb{E}(1-\lambda_k)\beta_k a_k P_{k-1}\phi_k^2\widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k^2}\right) \\
\le & \mathbb{E}\widetilde{\theta}_{k-1}^2 - 2\mathbb{E}\beta_k a_k P_{k-1}\phi_k^2\widetilde{\theta}_{k-1}^2 \\
& + O\left(\frac{1}{k}\right)\cdot\mathbb{E}\widetilde{\theta}_{k-1}^3 + O\left(\frac{1}{k^2}\right) \\
\le & \left(1 - 2\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^2 \\
& + O\left(\frac{1}{k}\right)\cdot\sqrt{\mathbb{E}\widetilde{\theta}_{k-1}^2\mathbb{E}\widetilde{\theta}_{k-1}^4} + O\left(\frac{1}{k^2}\right).
\end{aligned} \tag{A.18}$$

Similar to (A.14), we can get

$$\prod_{l=j+1}^k \left(1 - 2a_l\beta_l P_{l-1}\phi_l^2\right) = \left(\frac{P_j^{-1}}{P_k^{-1}}\right)^2 = O\left(\left(\frac{j}{k}\right)^2\right). \tag{A.19}$$

Case ii-0 $2\underline{\lambda} = 1$.

Then, from (A.15) and (A.16), there is $\varepsilon \in \left(0, \frac{1}{3}\right)$ such that $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k^{1-\varepsilon}}\right)$ and $\mathbb{E}\widetilde{\theta}_k^4 = O\left(\frac{1}{k^{2-2\varepsilon}}\right)$. Noticing (A.19), taking the above two into (A.18), which yields

$$\begin{aligned}
\mathbb{E}\widetilde{\theta}_k^2 &\le \left(1 - 2\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k}\right)\cdot\sqrt{O\left(\frac{1}{k^{1-\varepsilon}}\right)\cdot O\left(\frac{1}{k^{2-2\varepsilon}}\right)} + O\left(\frac{1}{k^2}\right)\\
&\le \left(1 - 2\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k^2}\right)\\
&\le \prod_{l=1}^{k}\left(1 - 2a_l\beta_l P_{l-1}\phi_l^2\right)\mathbb{E}\widetilde{\theta}_0^2 + O\left(\sum_{l=1}^{k}\prod_{j=l+1}^{k}\left(1 - a_j\beta_j P_{j-1}\phi_j^2\right)\frac{1}{l^2}\right)\\
&= O\left(\frac{1}{k^2}\right) + O\left(\frac{1}{k^2}\sum_{l=1}^{k}\frac{1}{l^{2-2}}\right)\\
&= O\left(\frac{1}{k}\right).
\end{aligned}$$

Case ii-1 $2\underline{\lambda} < 1$ and $4\underline{\lambda} > 1$.

Then by (A.15) and (A.16), we have $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k^{2\underline{\lambda}}}\right)$ and $\mathbb{E}\widetilde{\theta}_k^4 = O\left(\frac{1}{k^{4\underline{\lambda}}}\right)$. Noting (A.19), substituting the above two into (A.18) which gives

$$\begin{aligned}
\mathbb{E}\widetilde{\theta}_k^2 &\le \left(1 - 2\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k}\right)\cdot\sqrt{O\left(\frac{1}{k^{2\underline{\lambda}}}\right)\cdot O\left(\frac{1}{k^{4\underline{\lambda}}}\right)} + O\left(\frac{1}{k^2}\right)\\
&\le \left(1 - 2\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k^{1+3\underline{\lambda}}}\right) + O\left(\frac{1}{k^2}\right)\\
&\le \prod_{l=1}^{k}\left(1 - 2a_l\beta_l P_{l-1}\phi_l^2\right)\mathbb{E}\widetilde{\theta}_0^2 + O\left(\sum_{l=1}^{k}\prod_{j=l+1}^{k}\left(1 - 2a_j\beta_j P_{j-1}\phi_j^2\right)\frac{1}{l^{1+3\underline{\lambda}}}\right)\\
&\quad + O\left(\sum_{l=1}^{k}\prod_{j=l+1}^{k}\left(1 - 2a_j\beta_j P_{j-1}\phi_j^2\right)\frac{1}{l^2}\right)\\
&= O\left(\frac{1}{k^2}\right) + O\left(\frac{1}{k^2}\sum_{l=1}^{k}\frac{1}{l^{3\underline{\lambda}-1}}\right) + O\left(\frac{1}{k^2}\sum_{l=1}^{k}\frac{1}{l^{2-2}}\right)\\
&= O\left(\frac{1}{k^2}\right) + O\left(\frac{1}{k^2}\sum_{l=1}^{k^{3\underline{\lambda}}}\right) + O\left(\frac{1}{k}\right)\\
&= \begin{cases} O\left(\dfrac{1}{k}\right), & \text{if } \ 3\underline{\lambda} \ge 1,\\[2ex] O\left(\dfrac{1}{k^{3\underline{\lambda}}}\right), & \text{if } \ 3\underline{\lambda} < 1. \end{cases}
\end{aligned}$$

Then, we have $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k}\right)$ when $3\underline{\lambda} \ge 1$. When $3\underline{\lambda} < 1$, taking $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k^{3\underline{\lambda}}}\right)$ and $\mathbb{E}\widetilde{\theta}_k^4 =$

$O\left(\frac{1}{k^{4\underline{\lambda}}}\right)$ into (A.18) which gives

$$\begin{aligned}
\mathbb{E}\widetilde{\theta}_k^2 \leq& \left(1-2\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k}\right)\cdot\sqrt{O\left(\frac{1}{k^{3\underline{\lambda}}}\right)\cdot O\left(\frac{1}{k^{4\underline{\lambda}}}\right)} + O\left(\frac{1}{k^2}\right)\\
\leq& \prod_{l=1}^{k}\left(1-2a_l\beta_l P_{l-1}\phi_l^2\right)\mathbb{E}\widetilde{\theta}_0^2 + O\left(\sum_{l=1}^{k}\prod_{j=l+1}^{k}\left(1-a_j\beta_j P_{j-1}\phi_j^2\right)\frac{1}{l^{1+7\underline{\lambda}/2}}\right)\\
&+O\left(\sum_{l=1}^{k}\prod_{j=l+1}^{k}\left(1-a_j\beta_j P_{j-1}\phi_j^2\right)\frac{1}{l^2}\right)\\
=& O\left(\frac{1}{k^2}\right)+O\left(\frac{1}{k^{7\underline{\lambda}/2}}\right)+O\left(\frac{1}{k}\right)\\
=& \begin{cases} O\left(\frac{1}{k}\right), & \text{if} \quad 7\underline{\lambda}/2 \geq 1,\\ O\left(\frac{1}{k^{7\underline{\lambda}/2}}\right), & \text{if} \quad 7\underline{\lambda}/2 < 1.\end{cases}
\end{aligned}$$

So, we have $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k}\right)$ when $7\underline{\lambda}/2 \geq 1$. When $7\underline{\lambda}/2 < 1$, taking $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k^{7\underline{\lambda}/2}}\right)$ and $\mathbb{E}\widetilde{\theta}_k^4 = O\left(\frac{1}{k^{4\underline{\lambda}}}\right)$ into (A.18) and repeating the above process for p_1 times, we have

$$\mathbb{E}\widetilde{\theta}_k^2 = \begin{cases} O\left(\frac{1}{k}\right), & \text{if } \left(4-2^{-p_1}\right)\underline{\lambda} \geq 1,\\ O\left(k^{-\left(4-2^{-p_1}\right)\underline{\lambda}}\right), & \text{if } \left(4-2^{-p_1}\right)\underline{\lambda} < 1.\end{cases}$$

From $4\underline{\lambda} > 1$, there is p_1 such that $\left(4-2^{-p_1}\right)\underline{\lambda} \geq 1$.

Hence, we have $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k}\right)$.

Case ii-2 $4\underline{\lambda} \leq 1$ and $6\underline{\lambda} > 1$.

By (A.16), we have $\mathbb{E}\widetilde{\theta}_k^4 = O\left(\frac{1}{k^{4\underline{\lambda}}}\right)$ and $\mathbb{E}\widetilde{\theta}_k^6 = O\left(\frac{1}{k^{6\underline{\lambda}}}\right)$. Moreover, from Case ii-1, we have

$$\mathbb{E}\widetilde{\theta}_k^2 = O\left(k^{-\left(4-2^{-p_1}\right)\underline{\lambda}}\right). \tag{A.20}$$

By (A.7) and (A.17), similar to (A.8), we have

$$\begin{aligned}
\mathbb{E}\widetilde{\theta}_k^4 \leq& \mathbb{E}\widetilde{\theta}_{k-1}^4 + 4\mathbb{E}\widehat{a}_k\widehat{P}_{k-1}\phi_k\widetilde{\theta}_{k-1}^3\widetilde{s}_k + O\left(\frac{1}{k^2}\right)\cdot\mathbb{E}\widetilde{\theta}_{k-1}^2\\
\leq& \mathbb{E}\widetilde{\theta}_{k-1}^4 - 4\mathbb{E}\lambda_k\beta_k a_k P_{k-1}\phi_k^2\widetilde{\theta}_{k-1}^4 + O\left(\frac{1}{k^2}\right)\cdot\mathbb{E}\widetilde{\theta}_{k-1}^2\\
\leq& \left(1-4\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^4 + 4\mathbb{E}(1-\lambda_k)\beta_k a_k P_{k-1}\phi_k^2\widetilde{\theta}_{k-1}^4 + O\left(\frac{1}{k^2}\right)\cdot\mathbb{E}\widetilde{\theta}_{k-1}^2\\
\leq& \left(1-4\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^4 + O\left(\frac{1}{k}\right)\cdot\mathbb{E}\widetilde{\theta}_{k-1}^5 + O\left(\frac{1}{k^2}\right)\cdot\mathbb{E}\widetilde{\theta}_{k-1}^2\\
\leq& \left(1-4\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^4 + O\left(\frac{1}{k}\right)\cdot\sqrt{\mathbb{E}\widetilde{\theta}_{k-1}^4\cdot\mathbb{E}\widetilde{\theta}_{k-1}^6} + O\left(\frac{1}{k^2}\right)\cdot\mathbb{E}\widetilde{\theta}_{k-1}^2.
\end{aligned}\tag{A.21}$$

Noticing $4\underline{\lambda} \le 1$ and $6\underline{\lambda} > 1$, taking $\mathbb{E}\widetilde{\theta}_k^4 = O\left(\frac{1}{k^{4\underline{\lambda}}}\right)$ and $\mathbb{E}\widetilde{\theta}_k^6 = O\left(\frac{1}{k^{6\underline{\lambda}}}\right)$ into (A.21) which gives

$$\begin{aligned}
\mathbb{E}\widetilde{\theta}_k^4 \le& \left(1 - 4\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^4 + O\left(\frac{1}{k}\right)\cdot\sqrt{O\left(\frac{1}{k^{4\underline{\lambda}}}\right)\cdot O\left(\frac{1}{k^{6\underline{\lambda}}}\right)} + O\left(\frac{1}{k^{2+2\underline{\lambda}}}\right)\\
\le& \left(1 - 4\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^4 + O\left(\frac{1}{k^{1+5\underline{\lambda}}}\right)\\
\le& \prod_{l=1}^{k}\left(1 - 4a_l\beta_l P_{l-1}\phi_l^2\right)\mathbb{E}\widetilde{\theta}_0^2 + O\left(\sum_{l=1}^{k}\prod_{j=l+1}^{k}\left(1 - 4a_j\beta_j P_{j-1}\phi_j^2\right)\frac{1}{l^{1+5\underline{\lambda}}}\right)\\
=& O\left(\frac{1}{k^4}\right) + O\left(\frac{1}{k^4}\sum_{l=1}^{k}\frac{1}{l^{-3+5\underline{\lambda}}}\right)\\
=& O\left(\frac{1}{k^{5\underline{\lambda}}}\right).
\end{aligned}$$

Then, taking $\mathbb{E}\widetilde{\theta}_k^4 = O\left(\frac{1}{k^{5\underline{\lambda}}}\right)$ and $\mathbb{E}\widetilde{\theta}_k^6 = O\left(\frac{1}{k^{6\underline{\lambda}}}\right)$ into (A.21), which yields $\mathbb{E}\widetilde{\theta}_k^4 = O\left(\frac{1}{k^{11\underline{\lambda}/2}}\right)$, and repeating the above process for p_2 times, which gives

$$\mathbb{E}\widetilde{\theta}_k^4 = O\left(k^{-\left(6-2^{-p_2}\right)\underline{\lambda}}\right). \tag{A.22}$$

Substituting (A.20) and (A.22) into (A.18), which yields

$$\begin{aligned}
\mathbb{E}\widetilde{\theta}_k^2 \le& \left(1 - 2\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k}\right)\cdot\sqrt{O\left(k^{-(5-2^{-p_1-p_2})\underline{\lambda}}\right)} + O\left(\frac{1}{k^2}\right)\\
\le& \left(1 - 2\beta_k a_k P_{k-1}\phi_k^2\right)\mathbb{E}\widetilde{\theta}_{k-1}^2 + O\left(\frac{1}{k^{1+\nu_1}}\right) + O\left(\frac{1}{k^2}\right)\\
\le& \prod_{l=1}^{k}\left(1 - 2a_l\beta_l P_{l-1}\phi_l^2\right)\mathbb{E}\widetilde{\theta}_0^2 + O\left(\sum_{l=1}^{k}\prod_{j=l+1}^{k}\left(1 - 2a_j\beta_j P_{j-1}\phi_j^2\right)\frac{1}{l^{1+\nu_1}}\right)\\
&+ O\left(\sum_{l=1}^{k}\prod_{j=l+1}^{k}\left(1 - 2a_j\beta_j P_{j-1}\phi_j^2\right)\frac{1}{l^2}\right)\\
=& O\left(\frac{1}{k^2}\right) + O\left(\frac{1}{k^2}\sum_{l=1}^{k}\frac{1}{l^{\nu_1-1}}\right) + O\left(\frac{1}{k^2}\sum_{l=1}^{k}\frac{1}{l^{2-2}}\right)\\
=& \begin{cases} O\left(\dfrac{1}{k}\right), & \text{if } \ \nu_1 \ge 1,\\[2ex] O\left(\dfrac{1}{k^{\nu_1}}\right), & \text{if } \ \nu_1 < 1,\end{cases}
\end{aligned}$$

where $\nu_1 = \left(5 - 2^{-(p_1+1)} - 2^{-(p_2+1)}\right)\underline{\lambda}$.

When $\nu_1 \ge 1$, we have $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k}\right)$. Otherwise, taking $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k^{\nu_1}}\right)$ and (A.22) into

(A.18) and computing $\mathbb{E}\widetilde{\theta}_k^2$, then repeating the above process for p_3 times, we can get

$$\mathbb{E}\widetilde{\theta}_k^2 = \begin{cases} O\left(\dfrac{1}{k}\right), & \text{if } \nu_{p_3} \geq 1, \\ O\left(\dfrac{1}{k^{\nu_{p_3}}}\right), & \text{if } \nu_{p_3} < 1, \end{cases}$$

where $\nu_{p_3} = \left(6 - 2^{-p_2} - 2^{-p_3}\left(2 + 2^{-p_1} - 2^{-p_2}\right)\right)\underline{\lambda}$. From $6\underline{\lambda} > 1$, there are p_1, p_2 and p_3 such that $\nu_{p_3} \geq 1$. Therefore, we have $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k}\right)$.

Case ii-*r* $2r\underline{\lambda} \leq 1$ and $2(r+1)\underline{\lambda} > 1$, $\forall r \geq 3$. Similar to Case ii-1 and Case ii-2, we can prove that there exist $r(r+1)/2$ positive integers such that

$$\left(2(r+1) - \varepsilon_{r(r+1)/2}\left(2^{-p_1}, 2^{-p_2}, \cdots, 2^{-p_{r(r+1)/2}}\right)\right)\underline{\lambda} \geq 1,$$

where $\varepsilon_{r(r+1)/2}\left(2^{-p_1}, 2^{-p_2}, \cdots, 2^{-p_{r(r+1)/2}}\right)$ is continuous function. So we have $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k}\right)$.

In summary, it can be seen that $\mathbb{E}\widetilde{\theta}_k^2 = O\left(\frac{1}{k}\right)$, which means the conclusion holds. ∎