Automatica 166 (2024) 111722

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

# Asymptotically efficient Quasi-Newton type identification with quantized observations under bounded persistent excitations☆

Ying Wang [a], Yanlong Zhao [a,b,*], Ji-Feng Zhang [a,b]

[a] Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, PR China
[b] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, PR China

ARTICLE INFO

*Article history:*
Received 31 July 2023
Received in revised form 27 November 2023
Accepted 28 March 2024
Available online xxxx

*Keywords:*
System identification
Quantized observations
Cramér–Rao lower bound
Asymptotic efficiency
Quasi-Newton type algorithm

ABSTRACT

This paper is concerned with the optimal identification problem of dynamical systems in which only quantized output observations are available under the assumption of fixed thresholds and bounded persistent excitations. Based on a time-varying projection, a weighted Quasi-Newton type projection (WQNP) algorithm is proposed. With some mild conditions on the weight coefficients, the algorithm is proved to be mean square and almost surely convergent, and the convergence rate can be the reciprocal of the number of observations, which is the same order as the optimal estimate under accurate measurements. Furthermore, inspired by the structure of the Cramér–Rao lower bound, an information-based identification (IBID) algorithm is constructed with an adaptive design about weight coefficients of the WQNP algorithm, where the weight coefficients are related to the parameter estimates which leads to the essential difficulty of algorithm analysis. Beyond the convergence properties, this paper demonstrates that the IBID algorithm tends asymptotically to the Cramér–Rao lower bound, and hence is asymptotically efficient. A numerical example is simulated to show the effectiveness of the proposed algorithms.



## 1. Introduction

### 1.1. Background and motivations

Along with the modern science and technology rapid development, quantized systems have been widely applied in practical fields such as industrial systems, networked systems and even biological systems. For example, (i) industrial systems (Auber et al., 2018; Gagliardi et al., 2021; Tan et al., 2021): usually quantized sensors are more cost effective than regular sensors. In many applications, they are the only ones available during real-time operations. There are numerous examples of quantized observations such as switching sensors for exhaust gas oxygen, ABS (anti-lock braking systems), and shift-by-wire; photoelectric sensors for positions, gravity gradiometers with saturation constraints; traffic condition indicators in the asynchronous transmission mode networks; and gas content sensors (CO, $CO_2$, $H_2$, etc.) in gas and oil industry. (ii) Networked systems (Dargie & Poellabauer, 2010; Sohraby & Znati, 2007): thousands, even millions, of sensors are interconnected using a heterogeneous network of wireless systems. On account of limitations of the sensor power or communication bandwidth, the information from each sensor turns out to be quantized observations with a finite bit or even 1 bit. (iii) Biological systems (Ghysen, 2003; Wang et al., 2010, 2003): only two states of information, "excitation" or "inhibition", are detected from outside of the neuron. When the potential is bigger than the potential threshold, the neuron shows the excitation state, otherwise shows the inhibition state.

Due to the widespread adoption of systems with quantized observations, lots of researches related to the identification of such systems have emerged in the literature (Carbone et al., 2020; Casini et al., 2011; Godoy et al., 2011; Risuleo et al., 2020; Wang et al., 2010; Zhao et al., 2023). In addition, numerous methods are proposed to achieve identification with quantized observations such as empirical measure method (Wang & Yin, 2007; Wang et al., 2003), expectation maximization method (Godoy et al., 2011; Zhao et al., 2016), sign-error type algorithm (Csáji & Weyer,

☆ The work is supported by National Key R&D Program of China under Grant 2018YFA0703800, National Natural Science Foundation of China under Grants 62025306, 62303452 and T2293770, CAS Project for Young Scientists in Basic Research under Grant YSBR-008, China Postdoctoral Science Foundation under Grant 2022M720159, and Guozhi Xu Postdoctoral Research Foundation. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Gianluigi Pillonetto under the direction of Editor Alessandro Chiuso.

* Corresponding author at: Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, PR China.

*E-mail addresses:* wangying96@amss.ac.cn (Y. Wang), ylzhao@amss.ac.cn (Y. Zhao), jif@iss.ac.cn (J.-F. Zhang).

https://doi.org/10.1016/j.automatica.2024.111722

2012; Wang et al., 2022), stochastic approximation type algorithm (Guo & Zhao, 2013; Song, 2018), and stochastic gradient type algorithm (Guo & Zhao, 2014; Zhang et al., 2021). The emergence of these algorithms prompts us to explore how to achieve better identification effect by use of algorithm designs. Moreover, the study of the optimal quantized identification algorithms not only could achieve the improvement of identification theory, but also is helpful to improve the resource utilization with the limited communication bandwidth resources in the communication fields.

It is worth noticing that there is usually no explicit solution for the log-likelihood function of quantized systems due to the strong nonlinearity of quantized observations (Zhao et al., 2016), which makes it extremely hard to design the optimal identification algorithm by minimizing the objective function. Fortunately, it is known that the Cramér–Rao (CR) lower bound is a measure that the system data contains the amount of information of unknown parameters and can be used as a criterion to check the effectiveness of a procedure. In other words, the corresponding identification algorithm is termed efficient if the CR lower bound is achieved. Therefore, this paper investigates the optimal identification under quantized observations from the point of the CR lower bound.

### 1.2. Related literature

Actually, there are some interesting discussions about the CR lower bound of quantized systems (Guo & Zhao, 2014; Gustafsson & Karlsson, 2009; Wu et al., 2013). For example, Gustafsson and Karlsson (2009) and Wu et al. (2013) investigated a detailed study on the CR lower bound and derived its expression under different quantized measurements. Moreover, some results have also appeared for asymptotically efficient algorithms under quantized observations in the past two decades (Guo & Diao, 2020; Guo et al., 2015; Wang et al., 2018; Wang & Yin, 2007; Wang et al., 2003; Yang & Fang, 2014; You, 2015; Zhang et al., 2021). For example, Wang et al. (2018, 2003) established the asymptotical efficiency properties of empirical measure method and non-truncated empirical measure method for FIR systems under binary-valued observations and periodic inputs, respectively. Based on empirical measure method, Wang and Yin (2007) proposed a quasi-convex combination estimator for multi-threshold sensors and established its strong consistency and asymptotical optimality under periodic inputs, Guo et al. (2015) and Guo and Diao (2020) investigated asymptotically efficient algorithms for the systems with general quantized periodic inputs under various cases. Apart from the off-line algorithms mentioned above, there are also some discussions on online algorithms. Yang and Fang (2014) presented a recursive identification method for FIR systems with quantized measurements based on the stochastic approximation algorithm with expanding truncation bounds, and proved its asymptotic efficiency under independent and identically distributed (i.i.d) two-valued random inputs. You (2015) developed a stochastic approximation type recursive estimator with adaptive binary observations and i.i.d. input signals, and demonstrated it asymptotically approached the CR lower bound. Zhang et al. (2021) proposed a stochastic gradient-based recursive algorithm under binary-valued observations, and shown its convergence and asymptotic efficiency under bounded persistent excitations for first-order FIR systems.

However, almost all of the existing investigations on asymptotically efficient quantized identification algorithms suffer from some fundamental limitations. Most of these researches are based on the empirical measure algorithm, which is off-line and thus is difficult to apply to feedback controls. On the other hand, the conditions required are strict in the almost all of the online ones, such as the periodic or two-valued random or i.i.d. inputs, the adaptive and designable thresholds and so on.

Therefore, the goal of this paper is to develop an asymptotically efficient online algorithm, which could relax or remove the above-mentioned limitations. It is our hope that the approach of this paper will open up new avenues for further studies in the area of integrated design of identification and control with quantized constraints.

### 1.3. Main contributions

This paper investigates the asymptotically efficient recursive identification of the systems under quantized observations with multiple thresholds. The main contributions of this paper can be summarized as follows:

- Inspired by a time-varying projection in Zhang et al. (2022), a novel weighted Quasi-Newton type projection (WQNP) algorithm is proposed under quantized observations with multiple thresholds. With some mild conditions, the WQNP algorithm is proved to be convergent in both mean square and almost sure sense under bounded persistent excitations with the help of a scalar type Lyapunov function. Besides, the convergence rate can achieve the reciprocal of the number of observations under a proper requirement of weight coefficients, which is the same order as that under accurate measurements.
- This paper gives the CR lower bound of the system with multiple-threshold quantized observations. Then, based on the recursive form of its CR lower bound to design the weight coefficients of the WQNP algorithm, an information-based identification (IBID) algorithm is constructed, whose adaptive weight coefficients depend on the parameter estimates. Besides, the convergence rate is proved to reach the reciprocal of the time step by combining the scalar type and matrix type Lyapunov function methods. Moreover, the IBID algorithm is shown to be asymptotically efficient under bounded persistent excitations. In contrast with Wang and Yin (2007), the algorithm is an asymptotically efficient online algorithm under non-periodic or non-independent signals.
- The theoretical analysis method is different from the existing quantized identification algorithms. This paper adopts an idea of higher moment acceleration to solute the strong coupling between the weighted coefficients and the estimates of the IBID algorithm in the matrix type Lyapunov function method. It is worth mentioning that Markov inequality and the higher moments of estimation errors are used to establish the convergence rate of the matrix type Lyapunov function.

The rest of this paper is organized as follows. Section 2 describes the identification problem under multiple sensor thresholds. Section 3 presents the WQNP algorithm, and demonstrates its convergence properties. Section 4 constructs the IBID algorithm based on the CR lower bound, and establishes its convergence properties and asymptotic efficiency. All of the proofs of the main results are uniformly provided in Section 5. Section 6 supplies a numerical example to show the main results. Section 7 gives the concluding remarks and related future works.

*Notation.* In this paper, $\mathbb{R}^n$ and $\mathbb{R}^{n\times n}$ are the sets 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 $\|A\| = \sqrt{\lambda_{\max}(AA^T)}$ for the matrix $A \in \mathbb{R}^{n\times n}$. Besides, the trace of the matrix $A$ is $\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$. For the matrix $A_k$, 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

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

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

$$q_k = \begin{cases} 0, & \text{if } y_k \le C_1, \\ 1, & \text{if } C_1 < y_k \le C_2, \\ \vdots & \vdots \\ m, & \text{if } y_k > C_m, \end{cases} \tag{2}$$

which can also be represented as $q_k = \sum_{i=0}^{m} i I_{\{C_i < y_k \le C_{i+1}\}}$, where $C_0 = -\infty$ and $C_{m+1} = \infty$.

### 2.2. Assumptions

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

**Assumption 2.1.** The prior information on the unknown parameter $\theta$ is that $\theta \in \Omega \subset \mathbb{R}^n$ with $\Omega$ being a bounded convex set. And denote $\bar{\theta} = \sup_{\theta \in \Omega} \|\theta\|$.

**Assumption 2.2.** The vector sequence $\{\phi_k\}$ is supposed to be bounded persistently exciting, i.e.,

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

and $\sup_k \|\phi_k\| \le \bar{\phi} < \infty$.

**Assumption 2.3.** Assume that $\{d_k\}$ 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 2.1.** Actually, the median $\mu$ of the noise could be estimated similarly to Wang et al. (2022) when $\mu \neq 0$. Thus, without loss of generality, we assume that $\mu = 0$ throughout the paper. Moreover, Assumption 2.3 can be extended to the unknown but parameterizable noise distribution case, such as normal distribution with unknown mean value and variance. In this case, the parameters of the noise distribution and the unknown parameter $\theta$ can be jointly identified by the same way of identifying the unknown parameter $\theta$ alone (Wang et al., 2006). Furthermore, the noise under Assumption 2.3 also can be generalized to the one that the second derivation of the logarithm density function is less than zero (i.e., $\frac{d^2 \ln f(x)}{dx^2} < 0$), and the density function of the noise satisfies $\min_{\substack{1 \le i \le m \\ x \in [C_i - \bar{\phi}\bar{\theta}, C_i + \bar{\phi}\bar{\theta}]}} f(x) > 0$.

The goal of this paper is to develop an online asymptotically efficient algorithm to estimate the unknown parameter $\theta$ based on the information from input $\phi_k$, quantized observation $q_k$, and the stochastic property of the system noise $d_k$ under bounded persistent excitations.

## 3. The WQNP algorithm

This section will construct a Quasi-Newton type identification algorithm under quantized observations, and establish its convergence properties.

### 3.1. Algorithm design

For the simplicity of description, denote $F_i(x) = F(C_i - x)$, $f_i(x) = f(C_i - x)$, for $i = 0, \ldots, m+1$, and $H_i(x) = F_i(x) - F_{i-1}(x)$, $h_i(x) = f_i(x) - f_{i-1}(x)$ for $i = 1, \ldots, m+1$. Moreover, denote

$$F_{i,k} = F_i\left(\phi_k^T \theta\right), \quad f_{i,k} = f_i(\phi_k^T \theta), \tag{4}$$

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

$$\hat{F}_{i,k} = F_i(\phi_k^T \hat{\theta}_{k-1}), \quad \hat{f}_{i,k} = f_i(\phi_k^T \hat{\theta}_{k-1}), \tag{5}$$

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

$$H_{i,k} = H_i(\phi_k^T \theta), \quad h_{i,k} = h_i(\phi_k^T \theta), \tag{6}$$

and their estimates as

$$\hat{H}_{i,k} = H_i(\phi_k^T \hat{\theta}_{k-1})\ \hat{h}_{i,k} = h_i(\phi_k^T \hat{\theta}_{k-1}), \tag{7}$$

for $i = 1, \ldots, m+1$. Hence, $\mathbb{E} q_k = \sum_{i=1}^{m+1} (i-1) H_{i,k}$.

Next, we would like to introduce the idea of the Quasi-Newton type identification algorithm under quantized observations. Actually, the identification problem of unknown parameter $\theta$ is to find the roots of

$$u_k(\hat{\theta}) = \sum_{i=1}^{m+1} (i-1) H_{i,k} - \sum_{i=1}^{m+1} (i-1) H_i(\phi_k^T \hat{\theta}),$$

for all $k \ge 0$. Note $\sum_{i=1}^{m+1} (i-1) H_{i,k}$ is unavailable due to the existence of unknown parameter $\theta$, and $q_k$ is available with its expectation $\sum_{i=1}^{m+1} (i-1) H_{i,k}$. Therefore, we replaced $\sum_{i=1}^{m+1} (i-1) H_{i,k}$ with $q_k$ in $u_k(\hat{\theta})$. By instrumental variable method (Ljung & Söderström, 1983), we use $\phi_k$-s instrumental variable to define the vector-valued scores

$$U_k(\hat{\theta}) = -\sum_{l=1}^{k} \left( q_l - \sum_{i=1}^{m+1} (i-1) H_i(\phi_l^T \hat{\theta}) \right) \phi_l, \tag{8}$$

whose Jacobian matrix is used to construct the Newton-type step. Then, we calculate $\frac{\partial U_k(\hat{\theta})}{\partial \hat{\theta}}$ as

$$\frac{\partial U_k(\hat{\theta})}{\partial \hat{\theta}} = -\sum_{l=1}^{k} \sum_{i=1}^{m+1} (i-1) h_i(\phi_l^T \hat{\theta}) \phi_l \phi_l^T. \tag{9}$$

We generalize the above calculated Newton step as

$$P_k = \sum_{l=1}^{k} \beta_l \phi_l \phi_l^T. \tag{10}$$

Then, based on the idea of recursive least squares, we construct the identification algorithm as

$$\hat{\theta}_k = \hat{\theta}_{k-1} + a_k P_{k-1} \phi_k \left( q_k - \sum_{i=1}^{m+1} (i-1) \hat{H}_{i,k} \right),$$

$$a_k = \frac{1}{1 + \beta_k \phi_k^T P_{k-1} \phi_k},$$

$$P_k = P_{k-1} - a_k \beta_k P_{k-1} \phi_k \phi_k^T P_{k-1}.$$

Then, we design the weight coefficients $\varepsilon_{i,k}$ on the quantized observation $q_k$ to adjust the performance of the identification algorithm, i.e.,

$$s_k = \sum_{i=1}^{m+1} \varepsilon_{i,k} I_{\{C_{i-1} < y_k \le C_i\}}.$$

Moreover, we utilize the specific time-varying projection operator in Zhang et al. (2022) to guarantee the boundness of estimates, which is also helpful in the convergence analysis of the scalar type Lyapunov function method. Based on the above idea, a weighted Quasi-Newton type projection (WQNP) algorithm is constructed as Algorithm 1.

**Algorithm 1** The WQNP Algorithm

Beginning with an initial value $\hat{}_0 \in$ and a positive definitive matrix $P_0 \in \mathbb{R}^{n\times n}$, the algorithm is recursively defined at any $k \geq 0$ as follows:

1: Weighted conversion of the quantized observations:

$$s_k = \sum_{i=1}^{m+1} {}_{i;k} I_{\{C_{i-1}<y_k\leq C_i\}}. \tag{11}$$

2: Estimation:

$$\hat{}_k = {}_{P_k^{-1}} \hat{}_{k-1} + a_k P_{k-1} {}_k \tilde{s}_k ; \tag{12}$$

$$\tilde{s}_k = s_k - \sum_{i=1}^{m+1} {}_{i;k} \hat{H}_{i;k}; \tag{13}$$

$$a_k = \frac{1}{1 + {}_k{}_k^T P_{k-1}{}_k}; \tag{14}$$

$$P_k = P_{k-1} - a_k {}_k P_{k-1} {}_k {}_k^T P_{k-1}; \tag{15}$$

where $\hat{H}_{i;k}$ are defined in (7). Besides, ${}_Q(\cdot)$ is the projection mapping defined as

$${}_Q(x) = \arg\min_{z\in} \|x - z\|_Q, \forall x \in \mathbb{R}^n; \tag{16}$$

where $\|\cdot\|_Q$ is defined as $\| \ \|_Q = \sqrt{{}^T Q}, \forall \in \mathbb{R}^n$ and Q is a positive definite matrix.

**Remark 3.1.** It is worth noticing that when the quantized output is binary-valued observation (i.e., $m = 1$) and the dimension of the unknown parameter is one (i.e., $n = 1$), the WQNP algorithm can degrade into the unified stochastic gradient-based recursive algorithm in Zhang et al. (2021). More specifically, the innovation of the quantized observation in (13) can be rewritten as $\tilde{s}_k = ({}_{2;k} - {}_{1;k})\left(F(C_1 - {}_k^T\hat{}_{k-1}) - I_{\{y\leq C_1\}}\right)$. Therefore, the WQNP algorithm is a general extension of the algorithm in Zhang et al. (2021) from binary-valued observations to multiple sensor threshold observations.

*3.2. Convergence properties*

Before establishing the convergence, the following assumption about the weight coefficient is given.

**Assumption 3.1.** The weight coefficients ${}_{i;k}$ $(i = 1, \ldots, m+1)$ and ${}_k$ are scalars satisfying $-\infty < \underline{} \leq {}_{1;k} < {}_{2;k} < \cdots < {}_{m+1;k} \leq \bar{} < \infty$ with ${}_{m+1;k} - {}_{1;k} \geq > 0$ and $0 < \underline{} \leq {}_k \leq \bar{} < \infty$, respectively. Besides, the weight coefficients satisfy $\frac{2}{} \cdot \min_{1\leq i\leq m} {}_{x\in[C_i-\bar{}, C_i+\bar{}]} f(x) > 1 - \frac{1}{n}$.

**Theorem 3.1.** *If Assumptions 2.1–2.3 and 3.1 hold, then the WQNP algorithm is convergent both in mean square and high rank square, i.e.,*

$$\lim_{k\to\infty} \mathrm{E}\,\tilde{}_k^T\tilde{}_k = 0 \quad \text{and} \quad \lim_{k\to\infty} \mathrm{E}\|\tilde{}_k\|^{2r} = 0; \tag{17}$$

*and there exists a positive real number $< \infty$ such that*

$$\mathrm{E}\|\tilde{}_k\|^2 = O\left(\frac{1}{k}\right) \quad \text{and} \quad \mathrm{E}\|\tilde{}_k\|^{2r} = O\left(\frac{1}{k^r}\right); \tag{18}$$

*for $r = 2, 3, \ldots$ Besides, the WQNP algorithm is also convergent almost surely, i.e.,*

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

*where $\tilde{}_k = \hat{}_k -$ is the estimation error.*

The proof of Theorem 3.1 is supplied in Section 5.1.

**Remark 3.2.** Theorem 3.1 establishes the convergence properties of the WQNP algorithm for high-order parameter systems with quantized observations while Zhang et al. (2021) show the convergence properties for 1-order parameter systems. The key difficulty of the proof is how to guarantee the compression coefficient less than 1, which is related to dealing with the non-commutative matrices. Two techniques are applied in this part. First, a time-varying projection operator is introduced to deal with the product of the non-commutative matrix $P_k {}_k {}_k^T$ and keep the boundness of estimates. Besides, the boundness of estimates and regressor is used to ensure $\underline{f} = \min_{1\leq i\leq m} \min_{x\in[C_i-\bar{}, C_i+\bar{}]} f(x) > 0$, and then make the compression factor $1 - \frac{2\underline{f}}{}$ in (37) less than 1.

Besides the convergence, the convergence rate is another major problem that should be made clear.

**Theorem 3.2.** *Under the conditions of Theorem 3.1, if the condition (3) in Assumption 2.2 is enhanced as there exist a positive integer h and positive number $> 0$ such that $\frac{1}{h}\sum_{l=k+1}^{k+h} {}_l{}_l^T \geq {}^2 I_n$, and*

$$= > 2\inf_k \min_{1\leq i\leq m} \min_{\#\in} f(C_i - {}_k^T\#)^{-1}; \tag{19}$$

*then the WQNP algorithm has a mean square convergence rate as $O\left(\frac{1}{k}\right)$, i.e.,*

$$\mathrm{E}\|\tilde{}_k\|^2 = O\left(\frac{1}{k}\right);$$

*where and $\bar{}$ are defined in Assumption 3.1.*

The proof of Theorem 3.2 is put in Section 5.2.

**Remark 3.3.** Theorem 3.2 describes the fact that even under quantized observations, the convergence rate of $O\left(\frac{1}{k}\right)$ can be achieved with a suitable design of weight coefficients in the WQNP algorithm (12)–(15), which is the same rate as the case with accurate measurements.

Similar to the proof of Theorem 3.1, the following corollary can be derived directly, which is concerned with high rank square convergence rate.

**Corollary 3.1.** *Under the condition of Theorem 3.2, we have $\mathrm{E}\|\tilde{}_k\|^{2r} = o\left(\frac{1}{k}\right)$ for $r = 2, 3, \ldots$*

## 4. Asymptotically efficient algorithm

This section focuses on how to design and analyze the optimal identification algorithm under quantized observations. To realize it, we give a criterion, the CR lower bound under quantized observations, based on which an asymptotically efficient algorithm is constructed.

*4.1. Cramér–Rao lower bound*

Aiming at the system (1) with quantized observations (2), the following proposition establishes the CR lower bound of parameter estimates.

**Proposition 1.** *For the system (1) with quantized observations (2), the CR lower bound is*

$$_k = \left( \sum_{l=1}^{k} {}_l \, {}_l \, {}_l^T \right)^{-1}; \quad (20)$$

*where*

$$_l = \sum_{i=1}^{m+1} \frac{h_{i;l}^2}{H_{i;l}}; \quad (21)$$

*with $h_{i;l}$ and $H_{i;l}$ defined in (6) for $i = 1; \ldots; m+1$.*

The proof of Proposition 1 is supplied in Section 5.3. To understand the relationship between identification under quantized observations and the one under accurate observations, the following proposition is given.

**Proposition 2.** *Under Assumption 2.3, $_l$ defined in (21) satisfies $\lim_{\max_{i=1;\ldots;m+1}(C_i - C_{i-1}) \to 0} {}_l = \frac{1}{^2}$ where $^2$ is the covariance of $d_l$.*

The proof of Proposition 2 is supplied in Section 5.4.

**Remark 4.1.** The CR lower bound of the system (1) with accurate observations is $\frac{1}{^2} \left( \sum_{l=1}^{k} {}_l \, {}_l^T \right)^{-1}$. Combined it with Proposition 2, we find that the influence of quantized observations on the identification effect can be represented by the CR lower bound to some extent.

*4.2. The IBID algorithm*

This part will construct an asymptotically efficient algorithm with a proper design of weight coefficients on the WQNP algorithm, which is based on CR lower bound.

By the structure of CR lower bound, $_k$ defined in (20) can be written recursively as

$$_k = {}_{k-1} - \frac{_k \, {}_{k-1} \, {}_k \, {}_k^T \, {}_{k-1}}{1 + {}_k \, {}_k^T \, {}_{k-1} \, {}_k}. \quad (22)$$

Since $_k$ depends on the unknown parameter , we estimate it by use of $\hat{}_{k-1}$ as $\hat{}_k = \sum_{i=1}^{m+1} \frac{\hat{h}_{i;k}^2}{\hat{H}_{i;k}}$, where $\hat{H}_{i;k}$ and $\hat{h}_{i;k}$ are defined in (7).

Note that $P_k$ in recursive least square algorithm could represent the covariance of the estimation error to some extent, which enlightens us to design its weight coefficient as the estimate of CR lower bound coefficients, i.e.,

$$_k = \hat{}_k = \sum_{i=1}^{m+1} \frac{\hat{h}_{i;k}^2}{\hat{H}_{i;k}} \, , \, \hat{}_k. \quad (23)$$

Moreover, noticing (9) and (10) during the structure process of Newton step, we have $_k = -\sum_{i=1}^{m} {}_{i;k}\hat{h}_{i;k}$. Therefore, the weight coefficient of the weighted conversion is designed as

$$_{i;k} = \hat{}_{i;k} \, , -\frac{\hat{h}_{i;k}}{\hat{H}_{i;k}}; \quad i = 1; \ldots; m+1. \quad (24)$$

From Lemma 5.8 in Section 5.5 and the boundness of the estimate $\hat{}_k$ and the regressor $_k$, the following proposition can be established directly to illustrate the properties of $\hat{}_{i;k} (i = 1; \ldots; m+1)$ and $\hat{}_k$.

**Proposition 3.** *Denote*

$$\hat{} = \inf_k \min_{x \in} \left( \frac{f_m(x)}{1 - F_m(x)} + \frac{f_1(x)}{F_1(x)} \right).$$

*Then, $\hat{}_k$ and $\hat{}_{i;k}$ defined by (23)–(24) satisfy $0 < \hat{}_k < \infty$, $-\infty < \hat{}_{1;k} < \cdots < \hat{}_{m+1;k} < \infty$ and $\hat{}_{m+1;k} - \hat{}_{1;k} \geq \hat{} > 0$.*

Based on the WQNP algorithm and the weight coefficients in (23)–(24), an information-based identification (IBID) algorithm is constructed as Algorithm 2.

---

**Algorithm 2** The IBID Algorithm

---

Beginning with an initial value $\hat{}_0 \in$ and a positive definitive matrix $\hat{P}_0 \in \mathbb{R}^{n \times n}$, the algorithm is recursively defined at any $k \geq 0$ as follows:

1: Update of the adaptive weight coefficients:

$$\hat{}_{i;k} = -\frac{\hat{h}_{i;k}}{\hat{H}_{i;k}} \quad \text{and} \quad \hat{}_k = \sum_{i=1}^{m+1} \frac{\hat{h}_{i;k}^2}{\hat{H}_{i;k}}; \quad (25)$$

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

2: Weighted conversion of the quantized observations:

$$s_k = \sum_{i=1}^{m+1} \hat{}_{i;k} I_{\{C_{i-1} < y_k \leq C_i\}}. \quad (26)$$

3: Estimation:

$$\hat{}_k = {}_{\hat{P}_k^{-1}} \left( \hat{}_{k-1} + \hat{a}_k \hat{P}_{k-1} \, {}_k \tilde{s}_k \right); \quad (27)$$

$$\tilde{s}_k = s_k - \sum_{i=1}^{m+1} \hat{}_{i;k} \hat{H}_{i;k}; \quad (28)$$

$$\hat{a}_k = \frac{1}{1 + \hat{}_k \, {}_k^T \hat{P}_{k-1} \, {}_k}; \quad (29)$$

$$\hat{P}_k = \hat{P}_{k-1} - \hat{a}_k \hat{}_k \hat{P}_{k-1} \, {}_k^T \, {}_k \hat{P}_{k-1}. \quad (30)$$

---

**Remark 4.2.** Different from the WQNP algorithm and the weighted least square algorithm, the weight coefficients $_{i;k}$ and $_k$ of the IBID algorithm are related to the estimates. This leads to the essential difficulty of algorithm analysis since the properties of the adaptive weight coefficients and the convergence of the estimate are interdependent, which make the scalar type Lyapunov function method no longer applicable. Therefore, we introduce a matrix type Lyapunov function method to analyze the convergence rate of the IBID algorithm.

*4.3. Convergence properties*

The following theorem shows the convergence and the optimal convergence rate of the IBID algorithm.

**Theorem 4.1.** *If Assumptions 2.1–2.3 hold and the noise density function satisfies*

$$2 \min_{x \in [C_i - \,;\, C_i + \,]} f(x) \geq \max_{x \in [C_i - \,;\, C_i + \,]} f(x); \quad (31)$$

*for $i = 1; \ldots; m$, then the IBID algorithm is convergent in both mean square and almost sure sense, i.e., $\lim_{k \to \infty} \mathrm{E} \tilde{}_k^T \tilde{}_k = 0$ and $\lim_{k \to \infty} \tilde{}_k = 0$; a.s. Besides, the mean square convergence rate is*

$$\mathrm{E}\|\tilde{}_k\|^2 = O\left(\frac{1}{k}\right).$$

The proof of Theorem 4.1 is supplied in Section 5.6.

**Remark 4.3.** The noise condition (31) is mainly used to guarantee that the convergence of the scalar type Lyapunov function. This keeps in essence $\frac{f(C_i - {}_k^T \grave{}_{i;k-1})}{f(C_i - {}_k^T \hat{}_{i;k-1})} > \frac{1}{2}$ in (56) hold for $i = 1; \ldots; m$, where $\grave{}_{i;k-1}$ with ${}_k^T \, {}_{i;k-1}$ in the interval between ${}_k^T$ and ${}_k^T \hat{}_{k-1}$. This point is also the key difficulty in the convergence analysis of the IBID algorithm. This will be left as an open question. A possibly effective way in the authors' view is removing the limitation of the projection and using the covariance matrix of estimation error to analyze the convergence analysis of the IBID algorithm.

According to the proof of Theorems 3.1 and 4.1, the following corollary is derived directly, which is on the high-rank square convergence rate of the IBID algorithm.

**Corollary 4.1.** *Under the condition of Theorem 4.1, the IBID algorithm is convergent in high rank square with* $\mathrm{E}\|\tilde{}_k\|^{2r} = o\ \frac{1}{k}$ *, for* $r = 2; 3; \ldots$

*4.4. Asymptotical efficiency*

The following theorem shows $\hat{P}_k$ of IBID algorithm represents the covariance of estimation error to some extent.

**Theorem 4.2.** *If Assumptions 2.1–2.3 and (31) hold, then* $\hat{P}_k$ *defined in (30) has the following property,*

$$\lim_{k\to\infty} k(\mathrm{E}\hat{P}_k - \ _k) = 0.$$

The proof of Theorem 4.2 is supplied in Section 5.7. The following theorem demonstrates that the IBID algorithm can achieve the CR lower bound asymptotically, which implies that the IBID algorithm is asymptotically efficient and optimal.

**Theorem 4.3.** *If Assumptions 2.1–2.3 and (31) hold, then the IBID algorithm is asymptotically efficient, i.e.,*

$$\lim_{k\to\infty} k\ \mathrm{E}\,\tilde{}_k\tilde{}_k^T - \ _k = 0.$$

The proof of Theorem 4.3 is put in Section 5.8.

## 5. Proofs of the main results

*5.1. Proof of Theorem 3.1*

Before proving the convergence of the WQNP algorithm, some lemmas are collected and established, which are frequently used in the analysis of convergence.

**Lemma 5.1** (Calamai & Moré, 1987)**.** *For the bounded convex set , the projection is defined as* $_Q(x) = \arg\min_{z\in}\ \|x - z\|_Q$ *for all* $x \in \mathrm{R}^n$*, where Q is a positive definite matrix. Then, for all* $x \in \mathrm{R}$ *and* $x^* \in$ *, it holds* $_Q(x) - x^*\ _Q \le \|x - x^*\|_Q$.

**Lemma 5.2** (Zhang et al., 2022)**.** *Let* $X_1, X_2, \ldots$ *be any bounded sequence of vectors in* $\mathrm{R}^n (n \ge 1)$*. Denote* $A_k = A_0 + \sum_{i=1}^k X_i X_i^T$ *with* $A_0 > 0$*. Then, it holds that* $\sum_{k=1}^\infty (X_k^T A_k^{-1} X_k)^2 < \infty$.

**Lemma 5.3** (Chen, 2002)**.** *Let* $(v_k, F_k)$, $(w_k, F_k)$ *be two nonnegative adapted sequences. If* $\mathrm{E}(v_{k+1}|F_k) \le v_k + w_k$ *and* $\mathrm{E}\ \sum_{k=1}^\infty w_k < \infty$*, then* $v_k$ *converges a.s. to a finite limit.*

**Lemma 5.4** (Zhang et al., 2021)**.** *For any given positive integer l and* $a, b \in \mathrm{R}$*, the following results hold*

$$\prod_{i=l+1} 1 - \frac{a}{i} = O\ \frac{l}{k}^{\ a};$$

$$\sum_{l=1}\prod_{i=l+1} 1 - \frac{a}{i}\ \frac{1}{l^{1+b}} = \begin{cases} O\ \frac{1}{k^a} & ; a < b; \\ O\ \frac{\ln k}{k^a} & ; a = b; \\ O\ \frac{1}{k^b} & ; a > b. \end{cases}$$

**Lemma 5.5.** *Under Assumption 2.2,* $P_k$ *defined in (15) has the following properties: (i) the inverse of* $P_k$ *follows* $P_k^{-1} = P_{k-1}^{-1} + \ _k\ _k\ _k^T$*; (ii) For any initial* $P_0 > 0$,

$$0 \le P_k \le P_{k-1} \quad \text{and} \quad P_k = O\ \frac{1}{k}\ .$$

**Proof.** From (15), we have $P_k^{-1} = P_{k-1}^{-1} + \ _k\ _k\ _k^T$. Then, by $P_{k-1}P_k^{-1} = I_n + P_{k-1}\ _k\ _k\ _k^T$ and iterating the right parts of last equation, one can get $P_k^{-1} = P_0^{-1} + \sum_{i=1}^k\ _i\ _i\ _i^T$. Consequently, by $_i \ge\ \_ > 0$ and Assumption 2.2, the conclusion is true.

**Lemma 5.6.** *If Assumptions 2.2 and 3.1 hold, then*

$$\tilde{}_{k+j} - \tilde{}_k \le j(m+1)|\bar{}|\bar{}\ \|P_k\|; j \ge 0.$$

**Proof.** If $j = 0$, then the conclusion is true. Otherwise,

$$\|\tilde{}_{k+j} - \tilde{}_k\| = \|\hat{}_{k+j} - \hat{}_k\| \le \sum_{l=k+1}^{k+j} \|\hat{}_l - \hat{}_{l-1}\|. \quad (32)$$

By Lemma 5.1 and (12), we have

$$\|\hat{}_l - \hat{}_{l-1}\|^2_{P_l^{-1}} = \|\ _{P_l^{-1}}\ \hat{}_{l-1} + a_l P_{l-1}\ _l\tilde{s}_l\ - \hat{}_{l-1}\|^2_{P_l^{-1}}$$
$$\le \|a_l P_{l-1}\ _l\tilde{s}_l\|^2_{P_l^{-1}} = a_l^2\ _l^T P_{l-1}(P_{l-1}^{-1} + \ _l\ _l\ _l^T)P_{l-1}\ _l\tilde{s}_l^2$$
$$= a_l^2\ _l^T P_{l-1}\ _l(1 + \ _l\ _l^T P_{l-1}\ _l)\tilde{s}_l^2 = a_l\ _l^T P_{l-1}\ _l\tilde{s}_l^2.$$

Noting $P_l > 0$, we have $\|P_l\| = \ _{\max}(P_l) = \ _{\min}^{-1}(P_l^{-1})$. By Assumptions 2.2 and 3.1, $0 < a_l \le 1$, $\|\hat{}_l - \hat{}_{l-1}\|^2 \le \|\hat{}_l - \hat{}_{l-1}\|^2_{P_l^{-1}} /\ _{\min}(P_l^{-1})$, and Lemma 5.5, we can get

$$\|\hat{}_l - \hat{}_{l-1}\| \le \sqrt{a_l\ _l^T P_{l-1}\ _l\tilde{s}_l^2 /\ _{\min}(P_l^{-1})}$$
$$\le \sqrt{a_l\ _l^T P_{l-1}\ _l}\,|\tilde{s}_l|\,\|P_l\|^{\frac{1}{2}} \le 2|\bar{}|\bar{}\ \|P_{l-1}\| \le 2|\bar{}|\bar{}\ \|P_k\|.$$

Then, taking it into (32) yields this lemma.

**Proof of Theorem 3.1.** The proof is based on a scalar type Lyapunov function method, divided into the following three parts.

**Part I**: The mean square convergence properties.

Denote a scalar type Lyapunov function as $V_k = \tilde{}_k^T P_k^{-1}\tilde{}_k$. From (12), (14) and Lemma 5.1, we have

$$V_k = \ _{P_k^{-1}}\ \hat{}_{k-1} + a_k P_{k-1}\ _k\tilde{s}_k\ - \ ^T P_k^{-1}$$
$$\cdot\ _{P_k^{-1}}\ \hat{}_{k-1} + a_k P_{k-1}\ _k\tilde{s}_k\ -$$
$$\le\ \tilde{}_{k-1} + a_k P_{k-1}\ _k\tilde{s}_k\ ^T P_k^{-1}\ \tilde{}_{k-1} + a_k P_{k-1}\ _k\tilde{s}_k$$
$$\le \tilde{}_{k-1}^T P_{k-1}^{-1}\tilde{}_{k-1} + \ _k\tilde{}_{k-1}^T\ _k\ _k^T\tilde{}_{k-1} + 2a_k\ _k^T\tilde{}_{k-1}\tilde{s}_k$$
$$+ a_k^2\ _k^T P_{k-1}\ _k\tilde{s}_k^2 + 2\ _k a_k\ _k^T P_{k-1}\ _k\ _k^T\tilde{}_{k-1}\tilde{s}_k$$
$$+ a_k^2\ _k\ _k^T P_{k-1}\ _k\ _k^T P_{k-1}\ _k\tilde{s}_k^2$$
$$\le V_{k-1} + \ _k\tilde{}_{k-1}^T\ _k\ _k^T\tilde{}_{k-1} + 2\ _k^T\tilde{}_{k-1}\tilde{s}_k$$
$$+ a_k\ _k^T P_{k-1}\ _k\tilde{s}_k^2. \quad (33)$$

By $\mathrm{E}s_k = \sum_{i=1}^m\ _{i,k}H_{i,k}$, $\hat{F}_{0,k} = F_{0,k} = 0$, $\hat{F}_{m+1,k} = F_{m+1,k} = 1$ and the differential mean value theorem,

$$\mathrm{E}[\tilde{s}_k|F_{k-1}] = \sum_{i=1}^{m+1}\ _{i,k}\ H_{i,k} - \hat{H}_{i,k}$$
$$= \sum_{i=1}^{k}(\ _{i+1,k} - \ _{i,k})\ F(C_i - \ _k^T\hat{}_{k-1}) - F\ C_i - \ _k^T$$

$$= - \sum_{i=1}^{m} ( \quad_{i+1;k} - \quad_{i;k}) f(C_i - \quad_k^T \quad_{i;k}) \quad_k^T \tilde{\quad}_{k-1}$$

$$= - \sum_{i=1}^{m} ( \quad_{i+1;k} - \quad_{i;k}) \check{f}_{i;k} \quad_k^T \tilde{\quad}_{k-1}; \tag{34}$$

where $\quad_{i;k}$ with $\quad_k^T \quad_{i;k}$ in the interval between $\quad_k^T$ and $\quad_k^T \hat{\quad}_{k-1}$ such that $F(C_i - \quad_k^T \hat{\quad}_{k-1}) - F(C_i - \quad_k^T \quad) = -f(C_i - \quad_k^T \quad_{i;k}) \quad_k^T \tilde{\quad}_{k-1}$, and $\check{f}_{i;k} \; , \; f(C_i - \quad_k^T \quad_{i;k})$. Then from (33)–(34) and $|\tilde{s}_k| \le 2^-$, we have

$$EV_k \le EV_{k-1} + E \quad_k \tilde{\quad}_{k-1}^T \quad_k \quad_k^T \tilde{\quad}_{k-1} + Ea_k \quad_k^T P_{k-1} \quad_k \tilde{s}_k^2$$

$$- 2 \sum_{i=1}^{m} E( \quad_{i+1;k} - \quad_{i;k}) \check{f}_{i;k} \tilde{\quad}_{k-1}^T \quad_k \quad_k^T \tilde{\quad}_{k-1}$$

$$\le EV_{k-1} + E \left( 1 - \frac{2 \sum_{i=1}^{m} ( \quad_{i+1;k} - \quad_{i;k}) \check{f}_{i;k}}{\quad_k} \right)$$

$$\cdot \quad_k \tilde{\quad}_{k-1}^T \quad_k \quad_k^T \tilde{\quad}_{k-1} + 4^{-2} E a_k \quad_k^T P_{k-1} \quad_k$$

$$\le E \tilde{\quad}_{k-1}^T P_{k-1}^{-\frac{1}{2}} \left( I_n + \left( 1 - \frac{2 \sum_{i=1}^{m} ( \quad_{i+1;k} - \quad_{i;k}) \check{f}_{i;k}}{\quad_k} \right) P_{k-1}^{\frac{1}{2}} \right.$$

$$\left. \cdot \quad_k \quad_k \quad_k^T P_{k-1}^{\frac{1}{2}} \right) P_{k-1}^{-\frac{1}{2}} \tilde{\quad}_{k-1} + 4^{-2} a_k \quad_k^T P_{k-1} \quad_k. \tag{35}$$

Denote

$$\underline{f} = \min_{1 \le i \le m} \min_{x \in [C_i - \quad ; C_i + \quad]} f(x). \tag{36}$$

From Assumption 2.3, we have $\underline{f} > 0$ and $\check{f}_{i;k} \ge \underline{f}$. By Assumption 3.1 and (35), we get

$$EV_k \le E \tilde{\quad}_{k-1}^T P_{k-1}^{-\frac{1}{2}} \left( I_n + \left( 1 - \frac{2 \underline{f}}{\quad} \right) P_{k-1}^{\frac{1}{2}} \quad_k \quad_k \quad_k^T P_{k-1}^{\frac{1}{2}} \right) P_{k-1}^{-\frac{1}{2}} \tilde{\quad}_{k-1}$$

$$+ 4^{-2} a_k \quad_k^T P_{k-1} \quad_k; \tag{37}$$

where $1 - 2 \underline{f} / \quad \le 1/n$ from Assumption 3.1 and (36).

Next, we show the mean square convergence of WQNP algorithm in two cases, $1 - \frac{2 \underline{f}}{\quad} \le 0$ and $1 - \frac{2 \underline{f}}{\quad} > 0$.

**Case I-1**: $1 - \frac{2 \underline{f}}{\quad} \le 0$.

Noticing $P_k^{-1} = P_{k-1}^{-1} \left( I_n + \quad_k P_{k-1} \quad_k \quad_k^T \right)$, we have $a_k \quad_k \quad_k^T P_{k-1} \quad_k = P_k^{-1} - P_{k-1}^{-1} = P_k^{-1}$ and $P_k^{-1} = P_{k-1}^{-1} \cdot (1 + \quad_k \quad_k^T P_{k-1} \quad_k)$. Then,

$$\sum_{l=1} a_l \quad_l \quad_l^T P_{l-1} \quad_l = \sum_{l=1} \frac{P_l^{-1} - P_{l-1}^{-1}}{P_l^{-1}} \le \sum_{l=1} \int_{P_{l-1}^{-1}}^{P_l^{-1}} \frac{dx}{x}$$

$$\le \log P_k^{-1} - \log P_0^{-1} . \tag{38}$$

From Lemma 5.2, we have

$$\sum_{l=1} \quad_l \quad_l^T P_{l-1} \quad_l^2 < \infty. \tag{39}$$

And by (37) and (38), we get

$$EV_k \le EV_{k-1} + 4^{-2} a_k \quad_k^T P_{k-1} \quad_k$$

$$\le EV_0 + \sum_{l=1} \frac{4^{-2}}{\quad} \sum_{l=1} a_l \quad_k \quad_l^T P_{l-1} \quad_l = O \left( \log P_k^{-1} \right) .$$

Then, combining Lemma 5.5 gives

$$E \tilde{\quad}_k^T \tilde{\quad}_k \le EV_k / \min P_k^{-1} = O ( \log k / k ) . \tag{40}$$

**Case I-2**: $1 - \frac{2 \underline{f}}{\quad} > 0$. In this case, from (37) we have

$$EV_k \le E \tilde{\quad}_{k-1}^T P_{k-1}^{-\frac{1}{2}} \left( 1 + \left( 1 - \frac{2 \underline{f}}{\quad} \right) \quad_k \quad_k^T P_{k-1} \quad_k \right) P_{k-1}^{-\frac{1}{2}} \tilde{\quad}_{k-1}$$

$$+ 4^{-2} a_k \quad_k^T P_{k-1} \quad_k$$

$$\le \left( 1 + \left( 1 - \frac{2 \underline{f}}{\quad} \right) \quad_k \quad_k^T P_{k-1} \quad_k \right) EV_{k-1} + 4^{-2} a_k \quad_k^T P_{k-1} \quad_k$$

$$\le \prod_{l=1} \left( 1 + \left( 1 - \frac{2 \underline{f}}{\quad} \right) \quad_l \quad_l^T P_{l-1} \quad_l \right) EV_0$$

$$+ 4^{-2} \sum_{l=1} \prod_{i=l+1} \left( 1 + \left( 1 - \frac{2 \underline{f}}{\quad} \right) \quad_i \quad_i^T P_{i-1} \quad_i \right) a_l \quad_l^T P_{l-1} \quad_l . \tag{41}$$

First, we estimate the first item on the right side of (41) by (38), (39) and Lemma 5.5. By $0 < a_k \le 1$, we have

$$\prod_{l=1} \left( 1 + \left( 1 - \frac{2 \underline{f}}{\quad} \right) \quad_l \quad_l^T P_{l-1} \quad_l \right)$$

$$= e^{\sum_{l=1}^{k} \log \left( 1 + (1 - 2 \underline{f} / \quad) \quad_l \quad_l^T P_{l-1} \quad_l \right)}$$

$$\sim e^{(1 - 2 \underline{f} / \quad) \sum_{l=1}^{k} \quad_l \quad_l^T P_{l-1} \quad_l}$$

$$= e^{\left( 1 - \frac{2 \underline{f}}{\quad} \right) \sum_{l=1}^{k} a_l \quad_l \quad_l^T P_{l-1} \quad_l} \cdot e^{\left( 1 - \frac{2 \underline{f}}{\quad} \right) \sum_{l=1}^{k} a_l ( \quad_l \quad_l^T P_{l-1} \quad_l )^2}$$

$$\le e^{(1 - 2 \underline{f} / \quad) \left( \log P_k^{-1} - \log P_0^{-1} \right)} \cdot M$$

$$= M \left( P_k^{-1} = P_0^{-1} \right)^{(1 - 2 \underline{f} / \quad)}; \tag{42}$$

where $M$ is a constant related to (39).

Then, we estimate the second item on the right side of (41). Noticing (39) and (42), we have

$$4^{-2} \sum_{l=1} \prod_{i=l+1} \left( 1 + \left( 1 - \frac{2 \underline{f}}{\quad} \right) \quad_i \quad_i^T P_{i-1} \quad_i \right) a_l \quad_l^T P_{l-1} \quad_l$$

$$\le \frac{4^{-2}}{\quad} \sum_{l=1} \prod_{i=l+1} \left( 1 + \left( 1 - \frac{2 \underline{f}}{\quad} \right) \quad_i \quad_i^T P_{i-1} \quad_i \right) a_l \quad_l \quad_l^T P_{l-1} \quad_l$$

$$\le \frac{4M^{-2}}{\quad} P_k^{-1 \; 1 - \frac{2 \underline{f}}{\quad}} \sum_{l=1} \frac{P_l^{-1} - P_{l-1}^{-1}}{P_l^{-1 \; 2 - 2 \underline{f} / \quad}}$$

$$\le \frac{4M^{-2}}{\quad \left( 1 - \frac{2 \underline{f}}{\quad} \right)} \left( \frac{P_k^{-1}}{P_0^{-1}} \right)^{1 - 2 \underline{f} / \quad} . \tag{43}$$

Then, taking (42) and (43) into (41) gives

$$EV_k = O \left( P_k^{-1 \; 1 - 2 \underline{f} / \quad} \right) .$$

Hence, for $1 - 2 \underline{f} / \quad > 0$, combining Lemma 5.5 gives

$$E \tilde{\quad}_k^T \tilde{\quad}_k \le E \frac{V_k}{\min P_k^{-1}} = O \left( k^{n \cdot (1 - 2 \underline{f} / \quad) - 1} \right); \tag{44}$$

where Assumption 3.1 assures $n \left( 1 - 2 \underline{f} / \quad \right) - 1 < 0$.

Therefore, combining (40) and (44) yields

$$E \tilde{\quad}_k^T \tilde{\quad}_k = \begin{cases} O \left( \frac{\log k}{k} \right); & \text{if } \frac{2 \underline{f}}{\quad} \ge 1; \\ O \left( k^{n \cdot (1 - 2 \underline{f} / \quad) - 1} \right); & \text{if } \frac{2 \underline{f}}{\quad} < 1. \end{cases} \tag{45}$$

**Part II**: This part focuses on the convergence property of the WQNP algorithm in the high rank square.

When $r = 2$, from (33), we have

$$V_k^2 \leq V_{k-1} + ( {}_k^T \tilde{}_{k-1})^2 + 2 {}_k^T \tilde{}_{k-1}\tilde{s}_k + a_k {}_k^T P_{k-1} {}_k \tilde{s}_k^2 {}^2$$
$$\leq V_{k-1}^2 + ( {}_k \tilde{}_{k-1}^T {}_k {}_k^T \tilde{}_{k-1})^2 + 2( {}_k^T \tilde{}_{k-1})^2 \tilde{s}_k^2$$
$$+ a_k^2 ( {}_k^T P_{k-1} {}_k)^2 \tilde{s}_k^4 + 4 {}_k(\tilde{}_{k-1}^T {}_k)^3 \tilde{s}_k$$
$$+ 2V_{k-1} {}_k \tilde{}_{k-1}^T {}_k {}_k^T \tilde{}_{k-1} + 2 {}_k^T \tilde{}_{k-1}\tilde{s}_k$$
$$+ 2V_{k-1} a_k {}_k^T P_{k-1} {}_k \tilde{s}_k^2$$
$$+ a_k {}_k^T P_{k-1} {}_k \tilde{s}_k^2 {}_k(\tilde{}_{k-1}^T {}_k)^2 + 2 {}_k^T \tilde{}_{k-1}\tilde{s}_k :$$

Noticing $|\tilde{s}_k| \leq (m+1)|^-|$, $\| {}_k\| \leq {}^-$, (36), (45), Assumption 2.1 and Lemma 5.5, we have

$$EV_k^2 \leq EV_{k-1}^2 + 2EV_{k-1}( {}_k - 2 \sum_{i=1} ( {}_{i+1;k} - {}_{i;k})\check{f}_{i;k})\tilde{}_{k-1}^T {}_k {}_k^T \tilde{}_{k-1}$$
$$+ O\ E\|\tilde{}_{k-1}\|^2$$
$$\leq EV_{k-1}^2 + 2\ 1 - 2\ \underline{f}=^-\ EV_{k-1} {}_k \tilde{}_{k-1}^T {}_k {}_k^T \tilde{}_{k-1}$$
$$+ O\ E\|\tilde{}_{k-1}\|^2 : \tag{46}$$

Next, we consider this problem from the following two cases, i.e., $1 - 2\ \underline{f}=^- \leq 0$ and $1 - 2\ \underline{f}=^- > 0$.

**Case II-1**: $1 - 2\ \underline{f}=^- \leq 0$. From (45) and (46), we have

$$EV_k^2 \leq EV_{k-1}^2 + O\ E\|\tilde{}_{k-1}\|^2 \leq EV_{k-1}^2 + O \frac{\log k}{k}$$
$$= EV_0 + O \sum_{l=1} \frac{\log l}{l} ! = O\ \log^2 k ;$$

which together with Lemma 5.5 yields

$$E\|\tilde{}_k\|^4 \leq \frac{EV_k^2}{{}^2_{\min} P_k^{-1}} = O \frac{\log^2 k}{k^2} : \tag{47}$$

**Case II-2**: $1 - 2\ \underline{f}=^- > 0$. By (44) and (46), we have

$$EV_k^2 \leq 1 + 2\ 1 - 2\ \underline{f}=^- {}_k {}_k^T P_{k-1} {}_k\ EV_{k-1}^2$$
$$+ O\ k^{n.\,1-2\ \underline{f}=^-/-1}$$
$$\leq \prod_{l=1} 1 + 2\ 1 - \frac{2\ \underline{f}}{=} {}_l {}_l^T P_{l-1} {}_l\ EV_0^2$$
$$+ O \sum_{1=1} \prod_{i=l+1} 1 + 2\ 1 - \frac{2\ \underline{f}}{=} {}_i {}_i^T P_{i-1} {}_i$$
$$\cdot l^{n.\,1-2\ \underline{f}=^-/-1} = O\ k^{2n.\,1-2\ \underline{f}=^-/} : \tag{48}$$

For $1 - 2\ \underline{f}=^- > 0$, combining Lemma 5.5 and (48) gives

$$E\|\tilde{}_k\|^4 \leq \frac{EV_k}{{}^2_{\min} P_k^{-1}} = O\ k^{2n.\,1-2\ \underline{f}=^-/-2} : \tag{49}$$

Therefore, from (47) and (49), we have

$$E\|\tilde{}_k\|^2 = \begin{cases} O \frac{\log^2 k}{k^2} ; & \text{if } \frac{2\ \underline{f}}{=} \geq 1; \\ O\ k^{2n.\,1-2\ \underline{f}=^-/-2} ; & \text{if } \frac{2\ \underline{f}}{=} < 1: \end{cases}$$

Similarly, for any $r \geq 3$, we can get

$$E\|\tilde{}_k\|^{2r} = \begin{cases} O \frac{(\log k)^r}{k^r} ; & \text{if } \frac{2\ \underline{f}}{=} = 1; \\ O\ k^{rn.\,1-2\ \underline{f}=^-/-r} ; & \text{if } \frac{2\ \underline{f}}{=} < 1; \end{cases}$$

where Assumption 3.1 keeps $rn\ 1 - 2\ \underline{f}=^- - r < 0$.

In summary, there exists $< \infty$ such that (18) holds for any $r \geq 1$, which implies (17).

**Part III**: The almost sure convergence of WQNP algorithm is considered in this part. Denote $\bar{V}_k = \frac{V_k}{{}_{\min}(P_k^{-1})}$. By (37) and Lemma 5.5, we have

$$E[\bar{V}_k|F_{k-1}] \leq \bar{V}_{k-1} + 2^{-2} a_k {}_k^T P_{k-1} {}_k = {}_{\min}(P_{k-1}^{-1})$$
$$\leq \bar{V}_{k-1} + O\ 1=k^2 ; \text{ for } 1 - 2\ \underline{f}=^- \leq 0;$$
$$E[\bar{V}_k|F_{k-1}] \leq \bar{V}_{k-1} + 1 - \frac{2\ \underline{f}}{=} \frac{\tilde{}_{k-1}^T {}_k {}_k {}_k^T \tilde{}_{k-1}}{{}_{\min}(P_k^{-1})}$$
$$+ 2^{-2} a_k {}_k^T P_{k-1} {}_k = {}_{\min}(P_{k-1}^{-1})$$
$$\leq \bar{V}_{k-1} + 1 - 2\ \underline{f}=^- {}^{-\,-2}\|\tilde{}_{k-1}\|^2 = k$$
$$+ O\ 1=k^2 ; \text{ for } 1 - 2\ \underline{f}=^- > 0:$$

From (45), we have $E\frac{\|\tilde{}_{k-1}\|^2}{k} = O\ k^{-2+n.\,1-2\ \underline{f}=^-/}$ when $1 - 2\ \underline{f}=^- > 0$. From $\sum_{k=1}^{\infty} k^{-2+n.\,1-2\ \underline{f}=^-/} < \infty$, $\sum_{k=1}^{\infty} 1=k^2 < \infty$ and Lemma 5.3, $\bar{V}_k$ converges almost surely to a bounded limit. From (40) and (44), we have $E\bar{V}_k \to 0; k \to \infty$. Then, there is a subsequence of $\bar{V}_k$ that converges almost surely to 0. Noticing $\|\tilde{}_k\|^2 \leq \bar{V}_k$, $\tilde{}_k$ almost surely converges to 0.

### 5.2. Proof of Theorem 3.2

Since $\check{f}_{i;k} \geq \inf_k \min_{1 \leq i \leq m} \min_{\# \in} f(C_i - {}_k^T \#)$ , $\underline{f}$, noticing (35) we have

$$EV_k \leq E\tilde{}_{k-1}^T P_{k-1}^{-\frac{1}{2}} I_n + 1 - 2\ \underline{f}=^- P_{k-1}^{\frac{1}{2}} {}_k {}_k {}_k^T P_{k-1}^{\frac{1}{2}} P_{k-1}^{-\frac{1}{2}} \tilde{}_{k-1}$$
$$+ 4^{-2} a_k {}_k^T P_{k-1} {}_k$$
$$\leq EV_{k-h} - 2\ \underline{f}=^- - 1 \sum_{l=k-h}^{k-1} E\tilde{}_l^T {}_{l+1} {}_{l+1} {}_{l+1}^T \tilde{}_l$$
$$+ \sum_{l=k-h}^{k-1} 4^{-2} a_{l+1} {}_{l+1}^T P_l {}_{l+1}; \tag{50}$$

where $2\ \underline{f}=^- - 1 > 0$ by (19). From Assumptions 2.1 and 2.2, we have $\|\tilde{}_l\| \leq 2^-$ and $\| {}_l\| \leq {}^-$. For $l = k-h; \dots; k-1$, using Lemmas 5.5 and 5.6 give

$$- \tilde{}_l^T {}_{l+1} {}_{l+1} {}_{l+1}^T \tilde{}_l$$
$$= - \tilde{}_{k-h}^T {}_{l+1} {}_{l+1} {}_{l+1}^T \tilde{}_{k-h} + 2\tilde{}_l^T {}_{l+1} {}_{l+1} {}_{l+1}^T(\tilde{}_l - \tilde{}_{k-h})$$
$$- (\tilde{}_l - \tilde{}_{k-h})^T {}_{l+1} {}_{l+1} {}_{l+1}^T(\tilde{}_{l-1} - \tilde{}_{k-h})$$
$$\leq - \tilde{}_{k-h}^T {}_{l+1} {}_{l+1} {}_{l+1}^T \tilde{}_{k-h} + 2\tilde{}_l^T {}_{l+1} {}_{l+1} {}_{l+1}^T(\tilde{}_l - \tilde{}_{k-h})$$
$$= - \tilde{}_{k-h}^T {}_{l+1} {}_{l+1} {}_{l+1}^T \tilde{}_{k-h} + O\ .1=(k-h)/ : \tag{51}$$

By Assumption 3.1, we have

$$\sum_{l=k-h}^{k-1} {}_{l+1} {}_{l+1} {}_{l+1}^T \geq \underline{h}\ {}^2 I_n \geq \frac{\underline{h}\ {}^2}{\|P_0^{-1}\| + {}^{-\,-2}\ k}$$
$$\cdot\ P_0^{-1} + \sum_{l=1} {}_l {}_l {}_l^T ! \geq \frac{\underline{h}\ {}^2 P_{k-h}^{-1}}{\|P_0^{-1}\| + {}^{-\,-2}\ k} : \tag{52}$$

Denote $= \frac{{}^2}{\|P_0^{-1}\| + {}^{-\,-2}} \frac{2\ \underline{f}}{=} - 1 > 0$. By Lemmas 5.4 and 5.5, substituting (51) and (52) into (50) gives

$$EV_k \leq EV_{k-h} - 2\ \underline{f}=^- - 1\ \underline{h}\ {}^2 = \|P_0^{-1}\| + {}^{-\,-2}\ k$$
$$\cdot E\tilde{}_{k-h}^T P_{k-h}^{-1} \tilde{}_{k-h} + O\ .1=(k-h)/$$

$$= \left(1 - h/k\right) EV_{k-h} + O\left(1/(k-h)\right)$$
$$= \prod_{l=1}^{\lfloor \frac{k}{h} \rfloor -1}\left(1 - \frac{h}{k-lh}\right) EV_{k-\lfloor \frac{k}{h} \rfloor h} + O\left(\sum_{l=1}^{\lfloor \frac{k}{h} \rfloor -1}\prod_{q=0}^{l-1}\left(1-\frac{h}{k-qh}\right)\frac{1}{k-lh}\right)$$
$$= O(1/k) + O(1/) = O(1/).$$

Then, by Lemma 5.5, we have

$E\tilde{}_k^T\tilde{}_k \le EV_k/_{\min}(P_k^{-1}) = O(1/k).$

Thus, the WQNP algorithm has a mean square convergence rate as $O\left(\frac{1}{k}\right)$.

### 5.3. Proof of Proposition 1

Since the noises $\{d_k\}$ are i.i.d., we have

$$P(s_1, s_2, \ldots, s_k|\ ) = \prod_{l=1} P(s_l|\ ) = \prod_{l=1}\prod_{i=1}^{m+1} H_{i,l} I_{\{s_l = \ _{i,l}\}}.$$

Denote the log-likelihood function as

$$l_k(\ ) = \log P(s_1, s_2, \ldots, s_k|\ ) = \sum_{l=1} \log P(s_l|\ )$$
$$= \sum_{l=1}\sum_{i=1}^{m+1} \log(H_{i,l}) I_{\{s_l = \ _{i,l}\}}.$$

Noticing that $\frac{\partial \log H_{i,l}}{\partial} = -\frac{h_{i,l}}{H_{i,l}}\ _l$, and continuing the partial process, we have $\frac{\partial^2 \log H_{i,l}}{\partial\ ^2} = \frac{h'_{i,l}H_{i,l} - h_{i,l}^2}{H_{i,l}^2}\ _l\ _l^T$, where $h'_{i,l} = f'(C_i - \ _l^T\ ) - f'(C_{i-1} - \ _l^T\ )$ for $i = 2, \ldots, m$ and $h'_{1,l} = f'(C_1 - \ _l^T\ )$, $h'_{m+1,l} = -f'(C_m - \ _l^T\ )$ with $f'(x) = \partial f(x)/\partial x$. Hence, $\sum_{i=1}^{m+1} h'_{i,l} = 0$ and

$$\frac{\partial^2 l_k}{\partial\ \partial\ } = \sum_{l=1}\sum_{i=1}^{m+1}\left[\frac{h'_{i,l}H_{i,l} - h_{i,l}^2}{H_{i,l}^2} I_{\{s_l = \ _{i,l}\}}\right]\ _l\ _l^T;$$

together with $EI_{\{s_l = \ _{i,l}\}} = H_{i,l}$, the CR lower bound is

$$_k = \left(-E\frac{\partial^2 l_k}{\partial\ ^2}\right)^{-1}$$
$$= \left(-\sum_{l=1}\sum_{i=1}^{m+1}\left(\frac{h'_{i,l}H_{i,l} - h_{i,l}^2}{H_{i,l}^2} EI_{\{s_l = \ _{i,l}\}}\right)\ _l\ _l^T\right)^{-1}$$
$$= \left(-\sum_{l=1}\sum_{i=1}^{m+1}\frac{h'_{i,l}H_{i,l} - h_{i,l}^2}{H_{i,l}}\ _l\ _l^T\right)^{-1}$$
$$= \left(\left(-\sum_{l=1}\sum_{i=1}^{m+1} h'_{i,l} + \sum_{l=1}\sum_{i=1}^{m+1}\frac{h_{i,l}^2}{H_{i,l}}\right)\ _l\ _l^T\right)^{-1}$$
$$= \left(\sum_{l=1}\sum_{i=1}^{m+1}\frac{h_{i,l}^2}{H_{i,l}}\ _l\ _l^T\right)^{-1}.$$

### 5.4. Proof of Proposition 2

Since $f'(x) = -\frac{x}{\ ^2}f(x)$ for the normally density function $f(x)$ with covariance $\ ^2$, we have

$$\lim_{C\to 0}\ _l = \lim_{\max_{i=1,\ldots,m+1}(C_i - C_{i-1})\to 0}\sum_{i=1}^{m+1}\frac{h_{i,l}^2}{H_{i,l}}$$
$$= \int_{-\infty}^{\infty}\left(\frac{f'(x)}{f(x)}\right)^2 dx = \int_{-\infty}^{\infty}\left(-\frac{x}{\ ^2}\right)^2 f(x)dx = \frac{1}{\ ^2};$$

Hence, Proposition 2 holds.

### 5.5. Proof of Proposition 3

Before proving Proposition 3, we give the following lemma to analyze the properties of $h_{i,k}$ and $H_{i,k}$.

**Lemma 5.7.** *Let* $g(x, y) = \begin{cases}\frac{f(x)-f(y)}{F(x)-F(y)}, & \text{if } x \neq y,\\ -\frac{y}{\ ^2}, & \text{if } x = y.\end{cases}$ *Then,* $g_x(x, y) < 0$ *when* $x \neq y$.

**Proof.** Denote $\bar{g}(x, y) = \frac{x}{\ ^2}(F(x) - F(y)) + (f(x) - f(y))$. Noticing that $\bar{g}(y, y) = 0$ and $\bar{g}'_x(x, y) = F(x) - F(y)/\ ^2$, we have $\bar{g}(x, y) > 0$ when $x \neq y$. Since $f'(x) = -xf(x)/\ ^2$, we get

$$g'_x(x, y) = ((f(x) - f(y))/(F(x) - F(y)))'_x$$
$$= \frac{-\frac{x}{\ ^2}f(x)(F(x) - F(y)) - f(x)(f(x) - f(y))}{(F(x) - F(y))^2}$$
$$= -f(x)\bar{g}(x, y)/(F(x) - F(y))^2.$$

So, we have $g'_x(x, y) < 0$ when $x \neq y$.

Based on Lemma 5.7, we give the following lemma, which can lead to Proposition 3 directly.

**Lemma 5.8.** *For* $x \in (-\infty, \infty)$ *and* $i = 1, \ldots, m+1$, *denote* $h_i(x) = f(C_i - x) - f(C_{i-1} - x)$ *and* $H_i(x) = F(C_i - x) - F(C_{i-1} - x)$. *Then, for* $i = 2, \ldots, m+1$,

$$\frac{h_i(x)}{H_i(x)} < \frac{h_{i-1}(x)}{H_{i-1}(x)}. \tag{53}$$

**Proof.** From Lemma 5.7 and $C_i > C_{i-1} > C_{i-2}$ for $i = 2, \ldots, m+1$, we have $\frac{f(C_i - x) - f(C_{i-1}-x)}{F(C_i-x)-F(C_{i-1}-x)} < \frac{f(C_{i-2}-x)-f(C_{i-1}-x)}{F(C_{i-2}-x)-F(C_{i-1}-x)}$, which is equivalent to (53).

### 5.6. Proof of Theorem 4.1

From the definition of $H_{i,k}$ and $\hat{H}_{i,k}$ in (6) and (7), there exists $\grave{}_{i,k-1}$ with $\ _k^T\grave{}_{i,k-1}$ in the interval between $\ _k^T$ and $\ _k^T\hat{}_{k-1}$ such that

$$E[\tilde{s}_k|F_{k-1}] = \sum_{i=1}^{m+1}\hat{}_{i,k}\left(H_{i,k} - \hat{H}_{i,k}\right)$$
$$= \sum_{i=1}^{m}\left(\hat{}_{i+1,k} - \hat{}_{i,k}\right)\left(\hat{F}_{i,k} - F_{i,k}\right)$$
$$= -\sum_{i=1}^{m}\left(\hat{}_{i+1,k} - \hat{}_{i,k}\right)\grave{f}_{i,k}\ _k^T\tilde{}_{k-1}, \tag{54}$$

where $\grave{f}_{i,k} \triangleq f(C_i - \ _k^T\grave{}_{i,k-1}) \geq \min_{x\in[C_i - \ ,C_i + \ ]} f(x)$. Denote

$$_k = \sum_{i=1}^{m}\left(\hat{}_{i+1,k} - \hat{}_{i,k}\right)\grave{f}_{i,k}/\hat{}_k. \tag{55}$$

By the continuity of $f(x)$ and $F(x)$, $\ _k$ and $\hat{}_k$ are bounded. From (7), (25), (31) and (55),

$$_k = \frac{\sum_{i=1}^{m}\left(\hat{}_{i+1,k} - \hat{}_{i,k}\right)\grave{f}_{i,k}}{\sum_{i=1}^{m+1}\frac{\hat{h}_{i,k}^2}{\hat{H}_{i,k}}} = \frac{\sum_{i=1}^{m}\left(\hat{}_{i+1,k} - \hat{}_{i,k}\right)\grave{f}_{i,k}}{\sum_{i=1}^{m}\left(\hat{}_{i+1,k} - \hat{}_{i,k}\right)\hat{f}_{i,k}}$$

$$\geq \frac{\sum_{i=1}^{m} (\hat{\ }_{i+1;k} - \hat{\ }_{i;k}) \min_{x\in[C_i-\bar{\ };C_i+\bar{\ }]} f(x)}{\sum_{i=1}^{m} (\hat{\ }_{i+1;k} - \hat{\ }_{i;k}) \max_{x\in[C_i-\bar{\ };C_i+\bar{\ }]} f(x)} \geq \frac{1}{2}. \quad (56)$$

Let

$$\underline{\ } = \inf_k \ _k, \bar{\ } = \sup_k \ _k, \bar{\hat{\ }} = \sup_k \max_{i=1,\dots,m+1} |\hat{\ }_{i;k}|, \quad (57)$$

$$\underline{\hat{\ }} = \inf_k \hat{\ }_k, \bar{\hat{\ }} = \sup_k \hat{\ }_k. \quad (58)$$

Then, it can be seen that $\underline{\ } > 1/2$, $\underline{\hat{\ }} > 0$, $\bar{\ } < \infty$ and $\bar{\hat{\ }} < \infty$ from the boundness of $\hat{\ }_k$ and $\ _k$.

Let $\hat{V}_k = \tilde{\ }_k^T \hat{P}_k^{-1} \tilde{\ }_k$. Similar to (33), we have

$$\hat{V}_k \leq \hat{V}_{k-1} + \hat{\ }_k(\ _k^T \tilde{\ }_{k-1})^2 + 2\ _k^T \tilde{\ }_{k-1}\tilde{s}_k + \hat{a}_k\ _k^T \hat{P}_{k-1}\ _k \tilde{s}_k^2. \quad (59)$$

By (54)–(59), we have

$$\begin{aligned} E\hat{V}_k \leq & E\hat{V}_{k-1} + E\hat{\ }_k \tilde{\ }_{k-1}^T\ _k\ _k^T \tilde{\ }_{k-1} + E\hat{a}_k\ _k^T \hat{P}_{k-1}\ _k \tilde{s}_k^2 \\ & + 2E \sum_{i=1}^{m+1} \hat{\ }_{i;k} (H_{i;k} - \hat{H}_{i;k})\ _k^T \tilde{\ }_{k-1} \\ \leq & E\hat{V}_{k-1} + E(1-2\ _k)\hat{\ }_k \tilde{\ }_{k-1}^T\ _k\ _k^T \tilde{\ }_{k-1} + E\hat{a}_k\ _k^T \hat{P}_{k-1}\ _k \tilde{s}_k^2 \\ \leq & E\hat{V}_{k-1} + E(1-2\underline{\ })\hat{\ }_k \tilde{\ }_{k-1}^T\ _k\ _k^T \tilde{\ }_{k-1} \\ & + E\hat{a}_k\ _k^T \hat{P}_{k-1}\ _k \tilde{s}_k^2, \end{aligned} \quad (60)$$

where $1 - 2\underline{\ } \leq 0$.

Next, we discuss the convergence rate based on the higher moments and covariance of estimation errors.

First, we show the mean square convergence rate of IBID algorithm can reach $O\left(\frac{\log k}{k}\right)$. Similar to (38), we have

$$\sum_{l=1} \hat{a}_l \hat{\ }_l\ _l^T \hat{P}_{l-1}\ _l \leq \log\left(\hat{P}_k^{-1}\right) - \log\left(\hat{P}_0^{-1}\right). \quad (61)$$

Noticing $|\tilde{s}_k| \leq 2\bar{\hat{\ }}$, (60) and (61), we have

$$\begin{aligned} E\hat{V}_k \leq & E\hat{V}_{k-1} + E\hat{a}_k\ _k^T \hat{P}_{k-1}\ _k \tilde{s}_k^2 \\ \leq & E\hat{V}_0 + \frac{4\bar{\hat{\ }}^2}{\underline{\hat{\ }}} \sum_{l=1} E\hat{a}_l \hat{\ }_k\ _l^T \hat{P}_{l-1}\ _l = O\left(\log E\left(\hat{P}_k^{-1}\right)\right). \end{aligned}$$

From (58) and Assumption 2.2, we get

$$\hat{P}_k = O(1/k) \text{ and } \hat{P}_k^{-1} = O(k). \quad (62)$$

From (62), we have

$$E\tilde{\ }_k^T \tilde{\ }_k \leq E\hat{V}_k / \ _{\min}\left(\hat{P}_k^{-1}\right) = O(\log k / k). \quad (63)$$

Second, we establish the higher moments convergence rate of estimation errors (i.e., $E\|\tilde{\ }_k\|^{2r}, r \geq 2$) similarly to Part II in the proof of Theorem 3.1.

Based on (59), (60) and (63), similar to (46) we can get

$$\begin{aligned} E\hat{V}_k^2 \leq & E\hat{V}_{k-1}^2 + 2E\hat{V}_{k-1}(\ _k - 2\sum_{i=1} (\hat{\ }_{i+1;k} - \hat{\ }_{i;k})\hat{f}_{i;k})\tilde{\ }_{k-1}^T\ _k\ _k^T \tilde{\ }_{k-1} \\ & + O\left(E\|\tilde{\ }_{k-1}\|^2\right) \\ \leq & E\hat{V}_{k-1}^2 + 2(1-2\underline{\ })E\hat{V}_{k-1}\tilde{\ }_{k-1}^T\ _k\ _k^T \tilde{\ }_{k-1} + O\left(E\|\tilde{\ }_{k-1}\|^2\right) \\ \leq & E\hat{V}_{k-1}^2 + O\left(\frac{\log k}{k}\right) \leq E\hat{V}_0^2 + O\left(\sum_{l=1} \frac{\log l}{l}\right) \\ = & O\left(\log^2 k\right), \end{aligned}$$

which together with Lemma 5.5 yields

$$E\|\tilde{\ }_k\|^4 \leq E\hat{V}_k^2 / \ _{\min}^2\left(P_k^{-1}\right) = O\left(\log^2 k / k^2\right).$$

Similar, for any $r \geq 1$, we can get

$$E\|\tilde{\ }_k\|^{2r} = O\left((\log k)^r / k^r\right), \forall r = 1, 2, 3 \dots \quad (64)$$

Third, we construct a matrix type Lyapunov function by the covariance $E\tilde{\ }_k\tilde{\ }_k^T$ of the estimation errors to prove that the mean square convergence rate of the IBID algorithm reaches $O\left(\frac{1}{k}\right)$. Let $\ _k = \hat{\ }_{k-1} + \hat{a}_k\hat{P}_{k-1}\ _k\tilde{s}_k$ and $\bar{\ }_k = \ _k - \ $. Then, $\hat{\ }_k = \ _{P_k^{-1}}(\ _k)$ and

$$\bar{\ }_k = \tilde{\ }_{k-1} + \hat{a}_k\hat{P}_{k-1}\ _k\tilde{s}_k. \quad (65)$$

Based on (62), (64) and (65), we have

$$E\|\bar{\ }_k\|^{2r} = O\left((\log k)^r / k^r\right), r = 1, 2, 3 \dots \quad (66)$$

Without loss of generality, we assume $\ \in \ - @\ $, where $@\ $ is the edge set of $\ $. Denote $\underline{!} = \min_{! \in @\ } \|! - \ \| > 0$. Then by Markov inequality,

$$\begin{aligned} P(\ _k \in \ ) \leq & P\left(\|\ _k - \ \| \geq \underline{!}\right) = P\left(\|\bar{\ }_k\| \geq \underline{!}\right) \\ = & P\left(\|\bar{\ }_k\|^{2r} \geq \underline{!}^{2r}\right) \leq E\|\bar{\ }_k\|^{2r} / \underline{!}^{2r}. \end{aligned} \quad (67)$$

Noticing $\|\bar{\ }_k - \tilde{\ }_k\| = 0$ when $\ _k \in \ $, and $\|\bar{\ }_k - \tilde{\ }_k\| \leq \|\hat{a}_k\hat{P}_{k-1}\ _k\tilde{s}_k\| = O\left(\frac{1}{k}\right)$ when $\ _k \in \ $, we have

$$\begin{aligned} E(\bar{\ }_k - \tilde{\ }_k)(\bar{\ }_k - \tilde{\ }_k)^T \leq & E\|\bar{\ }_k - \tilde{\ }_k\|^2 I_n \\ \leq & O\left(1/k^2\right) \cdot P(\ _k \in \ ). \end{aligned} \quad (68)$$

For $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^n$, we have $ab^T + ba^T \leq 2\sqrt{a^T a b^T b} I_n$. Then, from (66), (67) and (68), we have

$$\begin{aligned} E\tilde{\ }_k\tilde{\ }_k^T = & E\bar{\ }_k\bar{\ }_k^T + E(\tilde{\ }_k - \bar{\ }_k)\bar{\ }_k^T + E\bar{\ }_k(\tilde{\ }_k - \bar{\ }_k)^T \\ & + E(\bar{\ }_k - \tilde{\ }_k)(\bar{\ }_k - \tilde{\ }_k)^T \\ \leq & E\bar{\ }_k\bar{\ }_k^T + 2\sqrt{E\bar{\ }_k^T\bar{\ }_k \cdot E(\bar{\ }_k - \tilde{\ }_k)^T(\bar{\ }_k - \tilde{\ }_k)} I_n \\ \leq & E\bar{\ }_k\bar{\ }_k^T + O\left(1/k^{3/2}\right)\sqrt{P(\ _k \in \ )} + O\left(1/k^2\right) \cdot P(\ _k \in \ ) \\ = & E\bar{\ }_k\bar{\ }_k^T + o\left(1/k^2\right). \end{aligned} \quad (69)$$

By (7) and (24), we have $E\left[\tilde{s}_k^2 | \mathcal{F}_{k-1}\right] = \sum_{i=1}^{m+1} \hat{\ }_{i;k}^2 H_{i;k}$ and $\sum_{i=1}^{m+1} \hat{\ }_{i;k}\hat{H}_{i;k} = 0$. Then, by (54), (62), (65), (69), $E[\tilde{s}_k|\mathcal{F}_{k-1}] = \sum_{i=1}^{m+1} \hat{\ }_{i;k}H_{i;k}$ and Assumptions 2.1–2.2,

$$\begin{aligned} E\tilde{\ }_k\tilde{\ }_k^T \leq & E\tilde{\ }_{k-1}\tilde{\ }_{k-1}^T + E\sum_{i=1}^{m+1} \hat{\ }_{i;k}^2 H_{i;k}\hat{a}_k^2\hat{P}_{k-1}\ _k\ _k^T\hat{P}_{k-1} \\ & + E\sum_{i=1}^{m+1} \hat{\ }_{i;k}(H_{i;k} - \hat{H}_{i;k})\hat{a}_k\tilde{\ }_{k-1}\ _k^T\hat{P}_{k-1} \\ & + E\sum_{i=1}^{m+1} \hat{\ }_{i;k}(H_{i;k} - \hat{H}_{i;k})\hat{a}_k\hat{P}_{k-1}\ _k\tilde{\ }_{k-1}^T + o\left(\frac{1}{k^2}\right) \\ \leq & E\tilde{\ }_{k-1}\tilde{\ }_{k-1}^T - E\tilde{\ }_{k-1}\tilde{\ }_{k-1}^T a_k\ _k\ _k\ _k^T P_{k-1} \\ & - Ea_kP_{k-1}\ _k\ _k\ _k^T\tilde{\ }_{k-1}\tilde{\ }_{k-1}^T + E\tilde{\ }_{k-1}\tilde{\ }_{k-1}^T\ _k\ _k^T \\ & \cdot\left(\ _k a_k P_{k-1} - \sum_{i=1} (\hat{\ }_{i+1;k} - \hat{\ }_{i;k})\hat{f}_{i;k}\hat{a}_k\hat{P}_{k-1}\right) \\ & + E\left(\ _k a_k P_{k-1} - \sum_{i=1} (\hat{\ }_{i+1;k} - \hat{\ }_{i;k})\hat{f}_{i;k}\hat{a}_k\hat{P}_{k-1}\right) \\ & \cdot\ _k\ _k^T\tilde{\ }_{k-1}\tilde{\ }_{k-1}^T + O\left(1/k^2\right), \end{aligned} \quad (70)$$

where $P_k$ is generated by (15) with $\ _k = \sum_{i=1}^{m+1} \frac{h_{i;k}^2}{H_{i;k}}$, $\ _{i;k} = -\frac{h_{i;k}}{H_{i;k}}$ and $a_k = \left(1 + \ _k\ _k^T P_{k-1}\ _k\right)^{-1}$. Then, by $f_{m+1;k} = f_{0;k} = 0$ and (6), we have $\ _k = -\sum_{i=1}^{m+1} \ _{i;k}(f_{i;k} - f_{i-1;k}) = \sum_{i=1}^{m} (\ _{i+1;k} - \ _{i;k})$

$f_{i,k}$. From Assumption 2.2, we have

$$P_k = O(1/k) \text{ and } P_k^{-1} = O(k). \quad (71)$$

Denote $_i(x) = -\frac{f(C_i-x)-f(C_{i-1}-x)}{F(C_i-x)-F(C_{i-1}-x)}$. Then, $_{i,k} = {}_i(\ _k\ )$, $\hat{}_{i,k} = {}_i(\ _k \hat{}_{k-1})$. From the continuous differentiability of $f(\cdot)$ and $F(\cdot)$, we get $_i(\cdot)$ is the continuous differentiable. From (34), (62), (71) and $a_k, \hat{a}_k \in (0, 1)$,

$$
\begin{aligned}
& {}_k a_k P_{k-1} - \sum_{i=1}^{m} \left(\hat{}_{i+1,k} - \hat{}_{i,k}\right) \grave{f}_{i,k} \hat{a}_k \hat{P}_{k-1} \\
=& O\left(\frac{1}{k}\right) \cdot \left( \sum_{i=1}^{m} \left( _{i+1,k} - {}_{i,k} \right) f_{i,k} - \sum_{i=1}^{m} \left(\hat{}_{i+1,k} - \hat{}_{i,k}\right) \grave{f}_{i,k} \right) \\
=& O\left(\frac{1}{k}\right) \cdot \sum_{i=1}^{m} \left( _{i+1,k} - {}_{i,k} \right) \left( f_{i,k} - \grave{f}_{i,k} \right) \\
&+ O\left(\frac{1}{k}\right) \cdot \sum_{i=1}^{m+1} \left( _{i,k} - \hat{}_{i,k} \right) \left( \grave{f}_{i,k} - \grave{f}_{i-1,k} \right) \\
=& O\left(\frac{1}{k}\right) \cdot \sum_{i=1}^{m+1} \left( _{i+1,k} - {}_{i,k} \right) f'(\grave{}_{i,k})\ _k^T(\breve{}_{k-1} - \ ) \\
&+ O\left(\frac{1}{k}\right) \cdot \sum_{i=1}^{m+1} {}'_i(\hat{}_{i,k})\ _k^T \tilde{}_k \left( \grave{f}_{i,k} - \grave{f}_{i-1,k} \right) \\
=& O(1/k) \cdot \|\tilde{}_k\|,
\end{aligned} \quad (72)
$$

where $\hat{}_{i,k}$ is between $_k^T$ and $_k^T \hat{}_{k-1}$, $\grave{}_{i,k}$ is between $C_i - {}_k^T \grave{}_{i,k-1}$ and $C_i - {}_k^T$, $\grave{f}_{i,k}$ and $\grave{}_{i,k-1}$ are denoted as (54). Then, based on (66), we have

$$
\begin{aligned}
& \mathrm{E}\left( \left( {}_k a_k P_{k-1} - \sum_{i=1}^{m} \left(\hat{}_{i+1,k} - \hat{}_{i,k}\right) \grave{f}_{i,k} \hat{a}_k \hat{P}_{k-1} \right) \tilde{}_{k-1} \tilde{}_{k-1}^T\ _k\ _k^T \right) \\
&+ \mathrm{E}\left( {}_k\ _k^T \tilde{}_{k-1} \tilde{}_{k-1}^T \left( {}_k a_k P_{k-1} - \sum_{i=1}^{m} \left(\hat{}_{i+1,k} - \hat{}_{i,k}\right) \grave{f}_{i,k} \hat{a}_k \hat{P}_{k-1} \right) \right) \\
\le& O\left(\frac{1}{k}\right) \cdot \mathrm{E}\|\tilde{}_k\|^3 \le O\left(\frac{1}{k}\right) \cdot \sqrt{\mathrm{E}\|\tilde{}_k\|^2 \cdot \mathrm{E}\|\tilde{}_k\|^4} \\
=& O(1/k) \cdot \sqrt{O\left(\log^3 k/k^3\right)} \\
=& o\left(1/k^2\right).
\end{aligned} \quad (73)
$$

By (71) and $P_k = P_{k-1} - a_k P_{k-1}\ _k\ _k\ _k^T P_{k-1}$, taking (73) into (70) yields

$$
\begin{aligned}
\mathrm{E}\tilde{}_k \tilde{}_k^T \le& \mathrm{E}\tilde{}_{k-1}\tilde{}_{k-1}^T - \mathrm{E}\tilde{}_{k-1}\tilde{}_{k-1}^T a_k\ _k\ _k\ _k^T P_{k-1} \\
&- \mathrm{E} a_k P_{k-1}\ _k\ _k\ _k^T \tilde{}_{k-1}\tilde{}_{k-1}^T + O\left(1/k^2\right) \\
\le& (I_n - a_k P_{k-1}\ _k\ _k\ _k^T)\mathrm{E}\tilde{}_{k-1}\tilde{}_{k-1}^T \\
&\cdot (I_n - a_k\ _k\ _k\ _k^T P_{k-1}) + O\left(1/k^2\right) \\
\le& P_k P_{k-1}^{-1} \mathrm{E}\tilde{}_{k-1}\tilde{}_{k-1}^T P_{k-1}^{-1} P_k + O\left(1/k^2\right) \\
=& P_k P_0^{-1} \mathrm{E}\tilde{}_0 \tilde{}_0^T P_0^{-1} P_k + O\left( \sum_{l=1}^{k} P_k P_l^{-1} \frac{1}{l^2} P_l^{-1} P_k \right) \\
=& O(1/k).
\end{aligned}
$$

Therefore, $\mathrm{E}\|\tilde{}_k\|^2 = \mathrm{tr}(\mathrm{E}\tilde{}_k \tilde{}_k^T) = O\left(\frac{1}{k}\right)$.

**Remark 5.1.** The key of this proof is the following three point. First, we introduce high-order moments of estimation errors and the scalar type Lyapunov function following Zhang et al. (2021) to overcome the difficulty that the weight coefficients $_{i,k}$ and $_k$ are stochastic and coupled with estimates. However, this method can only reach the convergence rate of $O\left(\frac{\log k}{k}\right)$ due to the loss of matrix scaling to constant coefficients. In order to solve it, a matrix type Lyapunov function (i.e., the covariance of estimation errors) is constructed to prove the convergence rate can be $O\left(\frac{1}{k}\right)$. Third, it is noticing that the projection operator makes it unable to directly iterate the covariance of estimation errors. Therefore, we calculate the difference between the projection and no-projection values by Markov inequality, and then, estimate the covariance of estimation errors with the no-projection value.

### 5.7. Proof of Theorem 4.2

By Assumption 2.2 and (20), $_k = \left( \sum_{l=1}^{k}\ _l\ _l\ _l^T \right)^{-1} = O\left(\frac{1}{k}\right)$. Denote $(x) = \sum_{i=1}^{m+1} \frac{(f(C_i-x)-f(C_{i-1}-x))^2}{F(C_i-x)-F(C_{i-1}-x)}$. Then, $_k = (\ _k^T\ )$ and $\hat{}_k = (\ _k^T \hat{}_{k-1})$. By Assumption 2.2 and the continuity of $f(x)$ and $F(x)$, there exists $\hat{}_k$ that is between $_k^T \hat{}_{k-1}$ and $_k^T \hat{}_k$ such that

$$
\begin{aligned}
\hat{}_k - {}_k =& (\ _k^T \hat{}_{k-1}) - (\ _k^T\ ) \\
=& {}'(\hat{}_k)\ _k^T \tilde{}_{k-1} = O\left(\|\tilde{}_{k-1}\|\right).
\end{aligned} \quad (74)
$$

From Theorem 4.1, we have $\mathrm{E}\|\tilde{}_{l-1}\| \le \sqrt{\mathrm{E}\|\tilde{}_{l-1}\|^2} = O\left(\frac{1}{\sqrt{l}}\right)$. Noticing $\|\ _k\| \le \bar{}$ and $_k = O(1/k)$, we have $_k^{\frac{1}{2}} \hat{P}_k\ _k^{\frac{1}{2}} = o(1)$ and $_k^{\frac{1}{2}} \sum_{l=1}^{k} O\left(\mathrm{E}\|\tilde{}_{l-1}\|\right)\ _l\ _l^T\ _k^{\frac{1}{2}} = o(1)$. And then, by $_k = {}_k$ and (74), we have

$$
\begin{aligned}
\mathrm{E}k\hat{P}_k =& \mathrm{E}k\left( {}_k^{-1} + \sum_{l=1}^{k} (\hat{}_l - {}_l)\ _l\ _l^T + \hat{P}_0 \right)^{-1} \\
=& \mathrm{E}k\ _k^{\frac{1}{2}}\left( I + {}_k^{\frac{1}{2}} \sum_{l=1}^{k} O\left(\|\tilde{}_{l-1}\|\right)\ _l\ _l^T\ _k^{\frac{1}{2}} + {}_k^{\frac{1}{2}} \hat{P}_0\ _k^{\frac{1}{2}} \right)^{-1} {}_k^{\frac{1}{2}} \\
=& \mathrm{E}k\ _k^{\frac{1}{2}}\left( I - {}_k^{\frac{1}{2}} \sum_{l=1}^{k} O\left(\|\tilde{}_{l-1}\|\right)\ _l\ _l^T\ _k^{\frac{1}{2}} - {}_k^{\frac{1}{2}} \hat{P}_0\ _k^{\frac{1}{2}} \right. \\
&\left. + \sum_{i=2}^{\infty} (-1)^i \left( {}_k^{\frac{1}{2}} \sum_{l=1}^{k} O\left(\|\tilde{}_{l-1}\|\right)\ _l\ _l^T\ _k^{\frac{1}{2}} + {}_k^{\frac{1}{2}} \hat{P}_0\ _k^{\frac{1}{2}} \right)^i \right) {}_k^{\frac{1}{2}} \\
=& k\ _k + O\left( k\ _k \left( \sum_{l=1}^{k} O\left(\mathrm{E}\|\tilde{}_{l-1}\|\right)\ _l\ _l^T \right) {}_k + k\ _k \hat{P}_0\ _k \right) \\
=& k\ _k + o(1) = k\ _k,
\end{aligned}
$$

where the fifth equality is got by Taylor expansion of the symmetric matrix, i.e., $(I + A)^{-1} = I + \sum_{k=1}^{\infty}(-1)^k A^k$ for the symmetric matrix $A$, and the sixth equality is got by Lyapunov inequality.

Therefore, $\lim_{k\to\infty} k(\mathrm{E}\hat{P}_k - {}_k) = 0$.

### 5.8. Proof of Theorem 4.3

Based on Theorem 4.1 and (70), we have

$$
\begin{aligned}
\mathrm{E}\tilde{}_k \tilde{}_k^T \le& \mathrm{E}\tilde{}_{k-1}\tilde{}_{k-1}^T - \mathrm{E}\tilde{}_{k-1}\tilde{}_{k-1}^T a_k\ _k\ _k\ _k^T P_{k-1} \\
&- \mathrm{E} a_k P_{k-1}\ _k\ _k\ _k^T \tilde{}_{k-1}\tilde{}_{k-1}^T + \mathrm{E}\tilde{}_{k-1}\tilde{}_{k-1}^T\ _k\ _k^T \\
&\cdot \left( {}_k a_k P_{k-1} - \sum_{i=1}^{m} \left(\hat{}_{i+1,k} - \hat{}_{i,k}\right) \grave{f}_{i,k} \hat{a}_k \hat{P}_{k-1} \right) \\
&+ \mathrm{E}\left( {}_k a_k P_{k-1} - \sum_{i=1}^{m} \left(\hat{}_{i+1,k} - \hat{}_{i,k}\right) \grave{f}_{i,k} \hat{a}_k \hat{P}_{k-1} \right) \\
&\cdot {}_k\ _k^T \tilde{}_{k-1}\tilde{}_{k-1}^T + \mathrm{E} \sum_{i=1}^{m+1} H_{i,k}\ \hat{}_{i,k}^2 \hat{a}_k^2 \hat{P}_{k-1}\ _k\ _k^T \hat{P}_{k-1}
\end{aligned}
$$

$$- \ {}^{2}_{i;k} P_{k-1} \ {}_{k} \ {}^{T}_{k} P_{k-1} + \sum_{i=1}^{m+1} \frac{h_{i;k}^2}{H_{i;k}} P_{k-1} \ {}_{k} \ {}^{T}_{k} P_{k-1} + o\left(\frac{1}{k^2}\right); \quad (75)$$

where $P_k$, ${}_k$, ${}_{i;k}$ and $a_k$ are defined in the proof of Theorem 4.1. Then, from (62), (71) and (72), we have

$$\sum_{i=1}^{m+1} H_{i;k} \left( \hat{}^{2}_{i;k} \hat{a}_k^2 \hat{P}_{k-1} \ {}_{k} \ {}^{T}_{k} \hat{P}_{k-1} - {}^{2}_{i;k} P_{k-1} \ {}_{k} \ {}^{T}_{k} P_{k-1} \right)$$
$$= O\left(\frac{1}{k^2}\right) \sum_{i=1}^{m+1} |\hat{}_{i;k} - {}_{i;k}| = O\left(\frac{1}{k^2}\right) \sum_{i=1}^{m+1} |{}'_i(\hat{}_{i;k}) \ {}^{T}_{k} \tilde{}_k|$$
$$= O\left(1/k^2\right) \cdot \|\tilde{}_k\|.$$

where $\hat{}_{i;k}$ is between ${}^{T}_{k}$ and ${}^{T}_{k}\hat{}_{k-1}$. From Theorem 4.1 and $E\|\tilde{}_k\| \leq \sqrt{E\|\tilde{}_k\|^2} = O\left(\frac{1}{\sqrt{k}}\right)$, we have

$$E \sum_{i=1}^{m+1} H_{i;k} \left( \hat{}^{2}_{i;k} \hat{a}_k^2 \hat{P}_{k-1} \ {}_{k} \ {}^{T}_{k} \hat{P}_{k-1} - {}^{2}_{i;k} P_{k-1} \ {}_{k} \ {}^{T}_{k} P_{k-1} \right)$$
$$= O\left(1/k^2\right) \cdot E\|\tilde{}_k\| = o\left(1/k^2\right). \quad (76)$$

Next, we will show $P_{k-1} - P_k = O\left(\frac{1}{k^2}\right)$. Noticing $P_k = P_{k-1} - a_k \ {}_k P_{k-1} \ {}_k \ {}^{T}_{k} P_{k-1}$, where ${}_k = \ ( \ {}^{T}_{k} \ )$ is bounded and positive, we have

$$\|P_{k-1} - P_k\| \leq \ {}_k^{-2} \|P_{k-1}\|^2 = O\left(1/k^2\right). \quad (77)$$

From Theorem 4.1, substituting (73), (76) and (77) into (75) gives

$$E\tilde{}_k \tilde{}_k^T = E\tilde{}_{k-1}\tilde{}_{k-1}^T - E\tilde{}_{k-1}\tilde{}_{k-1}^T a_k \ {}_k \ {}_k \ {}^{T}_{k} P_{k-1}$$
$$- E a_k P_{k-1} \ {}_k \ {}_k \ {}^{T}_{k} \tilde{}_{k-1} \tilde{}_{k-1}^T$$
$$+ \sum_{i=1}^{m+1} \frac{h_{i;k}^2}{H_{i;k}} P_k \ {}_k \ {}^{T}_{k} P_k + o\left(\frac{1}{k^2}\right)$$
$$= P_k P_{k-1}^{-1} E\tilde{}_{k-1}\tilde{}_{k-1}^T P_{k-1}^{-1} P_k + \sum_{i=1}^{m+1} \frac{h_{i;k}^2}{H_{i;k}} P_k \ {}_k \ {}^{T}_{k} P_k + o\left(\frac{1}{k^2}\right)$$
$$= P_k P_0^{-1} E\tilde{}_0 \tilde{}_0^T P_0^{-1} P_k + o\left( \sum_{l=1} P_k P_l^{-1} \frac{1}{l^2} P_l^{-1} P_k \right)$$
$$+ \sum_{l=1} \sum_{i=1}^{m+1} \frac{h_{i;l}^2}{H_{i;l}} P_k P_l^{-1} P_l \ {}_l \ {}^{T}_{l} P_l P_l^{-1} P_k$$
$$= O\left(\frac{1}{k^2}\right) + P_k \ {}_k^{-1} P_k + o\left( \sum_{l=1} P_k P_l^{-1} \frac{1}{l^2} P_l^{-1} P_k \right)$$
$$= o(1/k) + P_k \ {}_k^{-1} P_k = \ {}_k + o(1/k),$$

which implies the conclusion.

## 6. Numerical example

**Example 1.** Consider an in-orbit estimation problem of drag-free satellite mass (Tan et al., 2021), in which the relation between the residual acceleration $a_k$, the thrust $P_k$ and the unknown satellite mass $M$ is described as

$$a_k = \frac{P_k}{M} - \frac{C \ S}{2M} v_k^2, \quad (78)$$

where $v_k$ is the speed of the satellite along the tangent direction. The unknown parameters $C$, , $S$ are the atmospheric drag coefficient, atmospheric density and windward area, respectively. The measurement of the residual acceleration $a_k$ can be modeled as a

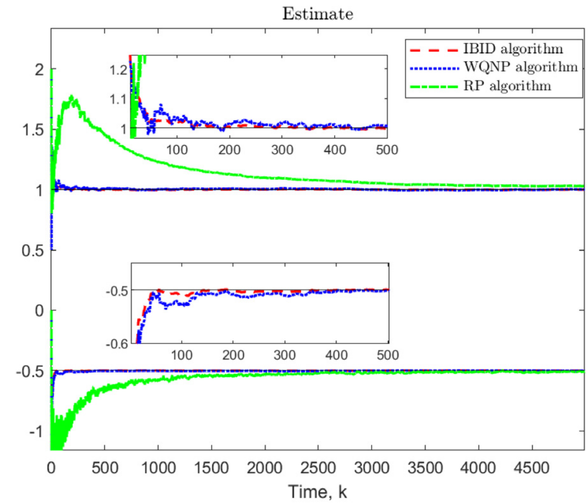


**Fig. 1.** Convergence of the IBID algorithm, the WQNP algorithm and the RP algorithm.

quantized observation $s_k = \sum_{i=0}^{m} i I_{\{C_i < a_k + d_k \leq C_{i+1}\}}$, where $d_k$ and $C_i$ are the measurement noise and the thresholds, respectively.

Set the unknown parameters $M = 1 \times 10^3$ kg and $C \ S = 1 \times 10^{-9}$ kg/m. The thrust $P_k$ and the speed $v_k$ follow the uniform distributions of the intervals $[1 \times 10^{-3}, 2 \times 10^{-3}]$ and $[1 \times 10^{3}, 3 \times 10^{3}]$, respectively. The measurement noise follows $N(0, 2^2 \times 10^{-12})$, and the thresholds are $[C_1, C_2, C_3] = [-3, 0, 3]$. To avoid round-off error caused by the computer, we multiply (78) by $10^6$ and set

$$= \begin{bmatrix} \frac{1}{M} \times 10^3 \\ -\frac{C \ S}{2M} \times 10^{12} \end{bmatrix} = \begin{bmatrix} 1 \\ -0.5 \end{bmatrix}, \quad {}_k = \begin{bmatrix} P_k \times 10^3 \\ v_k^2 \times 10^{-6} \end{bmatrix}.$$

And the prior information is $\ \in \ = [0, 2] \times [-2, 0]$.

In this example, we compare the efficiency of the IBID algorithm with other algorithms (including the WQNP algorithm with $[\ {}_{1,k}, \ {}_{2,k}, \ {}_{3,k}, \ {}_{4,k}, \ {}_k] = [-5, 0, 5, 10, 0.5]$ and the recursive projection (RP) algorithm in Tan et al. (2021) with ${}_1 = \ {}_2 = \ {}_3 = 50$). Here we repeat the simulation 500 times under the same initial values $\hat{}_0 = [2, 0]^T$ and $\hat{P}_0 = P_0 = 3I_2$ to establish the empirical variance of estimation errors representing the mean square errors.

From Figs. 1–3, it can be seen that the IBID algorithm performs better than the RP algorithm in Tan et al. (2021) and the WQNP algorithm, even if the convergence rate of the other two algorithms can also reach $O\left(\frac{1}{k}\right)$ under the appropriate weight design. Moreover, the covariance of the IBID algorithm tends to the CR lower bound, which shows its asymptotical efficiency.

## 7. Concluding remarks

This paper focuses on how to design an optimal identification algorithm under quantized observations. First, a weighted Quasi-Newton type projection algorithm is proposed to identify dynamical systems with quantized observations under bounded persistent excitations. Then, based on the adaptive design on the weight coefficients of the WQNP algorithm via the structure of CR lower bound, an IBID algorithm is constructed. And the mean square convergence rate of the algorithm can reach the reciprocal of the number of observations. Moreover, the asymptotic efficiency of the IBID algorithm is established, which means its optimality.

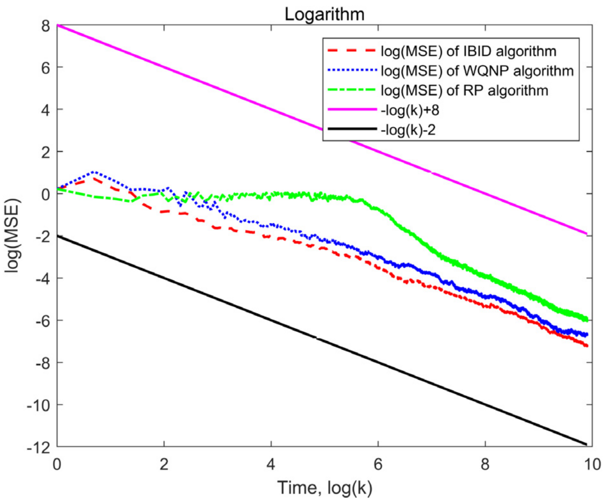


**Fig. 2.** Convergence rate of the IBID algorithm, the WQNP algorithm and the RP algorithm.

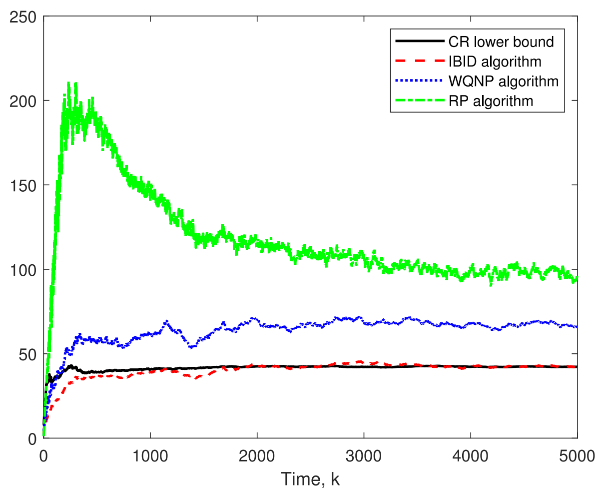


**Fig. 3.** Comparison between the empirical variance ($k\tilde{\ }_k^T\tilde{\ }_k$) of the IBID algorithm, the WQNP algorithm, the RP algorithm and the CR lower bound ($k\,\mathrm{tr}(\ _k)$).

These optimality results lay a foundation for designing appropriate communication protocol (threshold choice) and communication times to achieve the best identification performance under same communication resources. Correspondingly, future work is directed at studying sensor threshold selection to achieve optimal utility of communication bandwidth resources in enhancing identification accuracy.

## References


Auber, R., Pouliquen, M., Pigeon, E., Chapon, P. A., & Moussay, S. (2018). Activity recognition from binary data. In *2018 UKaCC 12th international conference on control* (pp. 158–162).

Calamai, P. H., & Moré, J. J. (1987). Projected gradient methods for linearly constrained problems. *Mathematical Programming*, *39*, 93–116.

Carbone, P., Schoukens, J., & Moschitta, A. (2020). Quick estimation of periodic signal parameters from 1-bit measurements. *IEEE Transactions on Instrumentation and Measurement*, *69*, 339–353.

Casini, M., Garulli, A., & Vicino, A. (2011). Input design in worst-case system identification using binary sensors. *IEEE Transactions on Automatic Control*, *56*, 1186–1191.

Chen, H. F. (2002). *Stochastic approximation and its applications*. Dordrecht: Kluwer Academic Publishers.

Csáji, B. C., & Weyer, E. (2012). Recursive estimation of ARX systems using binary sensors with adjustable thresholds. *IFAC Proceedings Volumes*, *45*, 1185–1190.

Dargie, W., & Poellabauer, C. (2010). *Fundamentals of wireless sensor networks: theory and practice*. Hoboken, NJ, USA: Wiley.

Gagliardi, G., Mari, D., Tedesco, F., & Casavola, A. (2021). An Air-to-Fuel ratio estimation strategy for turbocharged spark-ignition engines based on sparse binary hego sensor measures and hybrid linear observers. *Control Engineering Practice*, *107*, Article 104694.

Ghysen, A. (2003). The origin and evolution of the nervous system. *Control Engineering Practice*, *47*, 555–562.

Godoy, B., Goodwin, G., Agüero, J., Marelli, D., & Wigren, T. (2011). On identification of FIR systems having quantized output data. *Automatica*, *47*, 1905–1915.

Guo, J., & Diao, J. (2020). Prediction-based event-triggered identification of quantized input FIR systems with quantized output observations. *Science China. Information Sciences*, *63*, Article 112201.

Guo, J., Wang, L. Y., Yin, G., Zhao, Y. L., & Zhang, J. F. (2015). Asymptotically efficient identification of FIR systems with quantized observations and general quantized inputs. *Automatica*, *57*, 113–122.

Guo, J., & Zhao, Y. L. (2013). Recursive projection algorithm on FIR system identification with binary-valued observations. *Automatica*, *49*, 3396–3401.

Guo, J., & Zhao, Y. L. (2014). Identification of the gain system with quantized observations and bounded persistent excitations. *Science China. Information Sciences*, *57*, Article 012205.

Gustafsson, F., & Karlsson, R. (2009). Statistical results for system identification based on quantized observations. *Automatica*, *45*, 2794–2801.

Ljung, L., & Söderström, T. (1983). *Theory and practice of recursive identification*. London: The MIT Press.

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

Sohraby, D. M. K., & Znati, T. (2007). *Wireless sensor networks: technology, protocols, and applications*. John Wiley and Sons.

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

Tan, S., Guo, J., Zhao, Y. L., & Zhang, J. F. (2021). Adaptive control with saturation-constrainted observations for drag-free satellites – a set-valued identification approach. *Science China. Information Sciences*, *64*, Article 202202.

Wang, T., Tan, J., & Zhao, Y. L. (2018). Asymptotically efficient non-truncated identification for FIR systems with binary-valued outputs. *Science China. Information Sciences*, *61*, Article 129208.

Wang, L. Y., & Yin, G. (2007). Asymptotically efficient parameter estimation using quantized output observations. *Automatica*, *43*, 1178–1191.

Wang, L. Y., Yin, G., & Zhang, J. F. (2006). Joint identification of plant rational models and noise distribution functions using binary-valued observations. *Automatica*, *42*, 535–547.

Wang, L. Y., Yin, G., Zhang, J. F., & Zhao, Y. (2010). *System identification with quantized observations*. Birkhauser Boston.

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

Wang, Y., Zhao, Y. L., Zhang, J. F., & Guo, J. (2022). A unified identification algorithm of FIR systems based on binary observations with time-varying thresholds. *Automatica*, *135*, Article 109990.

Wu, H., Wang, W., & Ye, H. (2013). Lower bounds in parameter estimation based on quantized measurements. In *52nd IEEE conference on decision and control* (pp. 6341–6346).

Yang, X., & Fang, H. T. (2014). Asymptotically efficient recursive identification method for FIR system with quantized observations. In *Proceedings of the 33rd Chinese control conference* (pp. 6832–6837).

You, K. (2015). Recursive algorithms for parameter estimation with adaptive quantizer. *Automatica*, *52*, 192–201.

Zhang, H., Wang, T., & Zhao, Y. L. (2021). Asymptotically efficient recursive identification of FIR systems with binary-valued observations. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *51*, 2687–2700.

Zhang, L., Zhao, Y. L., & Guo, L. (2022). Identification and adaptation with binary-valued observations under non-persistent excitation condition. *Automatica*, *138*, Article 110158.

Zhao, Y. L., Bi, W. J., & Wang, T. (2016). Iterative parameter estimate with batched binary-valued observations. *Science China. Information Sciences*, *59*, Article 052201.

Zhao, Y. L., Zhang, H., Wang, T., & Kang, G. (2023). System identification under saturated precise or set-valued measurements. *Science China. Information Sciences*, *66*, Article 112204.

**Ying Wang** received the B.S. degree in Mathematics from Wuhan University, Wuhan, China, in 2017, and the Ph.D. degree in system theory from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Science (CAS), Beijing, China, in 2022. Now she is a postdoctoral fellow at the AMSS, CAS.

Her research interests include identification and control of quantized systems, and multiagent systems.

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

His research interests include identification and control of quantized systems, and modeling of financial systems.

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

**Ji-Feng Zhang** received the B.S. degree in mathematics from Shandong University, China, in 1985, and the Ph.D. degree from the Institute of Systems Science (ISS), Chinese Academy of Sciences (CAS), China, in 1991. Since 1985, he has been with the ISS, CAS.

His current research interests include system modeling, adaptive control, stochastic systems, and multi-agent systems.

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