Automatica 57 (2015) 113–122

Contents lists available at ScienceDirect

Automatica

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

Asymptotically efficient identification of FIR systems with quantized observations and general quantized inputs[☆]

Jin Guo[a], Le Yi Wang[b,1], George Yin[c], Yanlong Zhao[d], Ji-Feng Zhang[d]

[a] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, PR China
[b] Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202, USA
[c] Department of Mathematics, Wayne State University, Detroit, MI 48202, USA
[d] Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, PR China

ARTICLE INFO

Article history:
Received 20 September 2014
Received in revised form
2 February 2015
Accepted 25 March 2015
Available online 16 May 2015

Keywords:
System identification
Quantization
Non-periodic input
Asymptotic efficiency
Input design

ABSTRACT

This paper introduces identification algorithms for finite impulse response systems under quantized output observations and general quantized inputs. While asymptotically efficient algorithms for quantized identification under periodic inputs are available, their counterpart under general inputs has encountered technical difficulties and evaded satisfactory resolutions. Under quantized inputs, this paper resolves this issue with constructive solutions. A two-step algorithm is developed, which demonstrates desired convergence properties including strong convergence, mean-square convergence, convergence rates, asymptotic normality, and asymptotical efficiency in terms of the Cramér–Rao lower bound. Some essential conditions on input excitation are derived that ensure identifiability and convergence. It is shown that by a suitable selection of the algorithm's weighting matrix, the estimates become asymptotically efficient. The strong and mean-square convergence rates are obtained. Optimal input design is given. Also the joint identification of noise distribution functions and system parameters is investigated. Numerical examples are included to illustrate the main results of this paper.

1. Introduction

To reduce costs of sensors and accommodate communication system limitations (Akyildiz, Su, Sankarasubramaniam, & Cayirci, 2002; Li, Zhang, Cui, Fan, & Athanasios, 2014), system identification under quantized observation has drawn great research effort and experienced substantial advancement during the past decade (Agüero, Goodwin, & Yuz, 2007; Casini, Garulli, & Vicino, 2011; Wang, Yin, Zhang, & Zhao, 2010; Wang, Zhang, & Yin, 2003). Compared with conventional system identification, quantized observations provide very limited information on system output signals, and consequently introduce essential difficulties in system identification. Fundamental progress has been achieved in methodology development, identification algorithms, essential convergence properties, and applications (Casini, Garulli, & Vicino, 2012; Cerone, Piga, & Regruto, 2013; Chen, Zhao, & Ljung, 2012; Colinet & Juillard, 2010; Gustafsson & Karlsson, 2009; You, Xie, Sun, & Xiao, 2011).

Under full rank periodic inputs, for both linear systems and nonlinear Wiener and Hammerstein systems under stochastic and bounded noises, quantized identification algorithms and their key convergence properties have been obtained, including strong and mean-square convergence, convergence rates, asymptotic normality, asymptotic efficiency, and large deviation principles (Agüero et al., 2007; Casini et al., 2011; He, Wang, & Yin, 2013; Mei, Yin, & Wang, 2014; Wang & Yin, 2007; Zhao, Wang, Yin, & Zhang, 2007). Also, related time and space complexities, identification accuracy with respect to the disturbances and unmodeled dynamics, and optimal input design have been investigated comprehensively. For example, Agüero et al. (2007) studied quantized identification of linear systems with colored

☆ This research was supported in part by the Army Research Office under grant W911NF-12-1-0223 and in part by the National Natural Science Foundation of China (61174042, 61403027, 612279002). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Alessandro Chiuso under the direction of Editor Torsten Söderström.

E-mail addresses: guojin@amss.ac.cn (J. Guo), lywang@wayne.edu (L.Y. Wang), gyin@math.wayne.edu (G. Yin), ylzhao@amss.ac.cn (Y. Zhao), jif@iss.ac.cn (J.-F. Zhang).

[1] Tel.: +1 313 577 4715; fax: +1 313 577 1101.

http://dx.doi.org/10.1016/j.automatica.2015.04.009
0005-1098/

noises based on multi-sine input signals and quantized data. Casini et al. (2011) proposed a method for designing optimal input sequences to minimize time complexity on parameter estimation using binary sensors under bounded disturbances. Wang et al. (2003) gave a strong consistent identification algorithm with binary-valued observations. For general quantization thresholds under full-rank periodic inputs, Wang and Yin (2007) introduced the optimal quasi-convex combination estimator (QCCE) which was shown to achieve the Cramér–Rao (CR) lower bound asymptotically.

However, commonly encountered persistent excitation inputs are not necessarily periodic. Input signals often cannot be arbitrarily selected to be periodic (Ljung, 1987), and in adaptive control the control input is adjusted in real time and is usually non-periodic (Guo, 1993). Great effort has been made to resolve this issue. While it is well understood that the maximum likelihood (ML) estimates are efficient algorithms, they are not constructive. One main direction of research is to approximate the ML functions, such as expectation maximization (EM) methods, under quantized observations and general inputs. Another approach is to construct directly identification algorithms of stochastic approximation type by using innovation of quantized observations. Based on the ML criterion, Godoy, Goodwin, Agüero, Marelli, and Wigren (2011) introduced an iterative batch algorithm for identifying finite impulse response (FIR) systems with quantized output data and persistent excitations, and proved that the ML criterion was achieved as the iterative step goes to infinity. Marelli, You, and Fu (2013) extended the work in Godoy et al. (2011) to ARMA models with intermittent quantized output observations, where the joint effect of finite-level quantization and random packet dropouts on identification accuracy was characterized. Under a regularity assumption on the parametric model describing the data, Chiuso (2008) showed that the ML estimator can be found from quantized data. Guo and Zhao (2013) proposed a recursive projection algorithm for FIR systems with binary-valued observations, proved its strong and mean-square convergence, and obtained convergence rates under sufficiently rich inputs. For gain systems and time-varying thresholds, Guo and Zhao (2014) discussed the quantized identification problem with general persistent excitation inputs, and provided an optimal scheme of selecting thresholds and quantization values.

Despite comprehensive progress on quantized system identification under both periodic and general inputs, asymptotically efficient algorithms for quantized identification under general inputs have encountered technical difficulties and evaded satisfactory resolutions. Under quantized inputs, this paper resolves this issue with constructive solutions. Focusing on FIR systems, we consider quantized system identification under quantized inputs. By classifying the regressor sequence into distinct pattern sets according to the values of the regressor vector, we show that the information on the system can be divided, without losing any information, on the basis of both quantized output observations and input regressor patterns. A modified optimal QCCE (Wang & Yin, 2007) is employed together with a weighted least-squares optimization to combine optimally the set of information to derive an estimate of the unknown parameters. By adjusting the weighting matrix, different asymptotic behavior of the estimate can be obtained. It is shown that the algorithms can achieve the CR lower bound asymptotically by a suitable design of the weighting matrix, and hence, is asymptotically efficient. Construction of the optimal weighting matrix is presented. Other related and desired convergence properties of the algorithms are established, including strong convergence, mean-square convergence, convergence rates, and asymptotic normality. Some essential conditions on inputs for ensuring identifiability and strong convergence are derived. The scaled sequence of estimation errors is shown to be asymptotically normal. Optimal input design problems are also studied, which aim to increase convergence rates by selecting input patterns and adjusting their frequencies of occurrence. Furthermore, we also investigate the case where the noise distribution functions are unknown. The identifiability, algorithm design and convergence analysis are discussed.

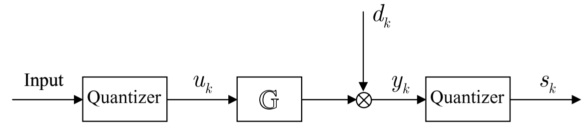


Fig. 1. System configuration.

The rest of the paper is arranged as follows. Section 2 formulates the problem with system structures, noise characterizations, and other related conditions. Section 3 reviews some key techniques on quantized system identification under periodic inputs that are used in this paper for quick reference. Section 4 designs the identification algorithms based on the optimal QCCE and weighted least-squares optimization. Input excitation conditions are introduced. Section 5 analyzes convergence properties, including strong and mean-square convergence, convergence rates, asymptotical normality, and asymptotic efficiency. Section 6 discusses optimal input design problems. Section 7 investigates the joint identification of noise distribution functions and system parameters. Numerical examples are presented in Section 8 to demonstrate the main results of this paper. Finally, findings of the paper are summarized in Section 9, together with some potential future directions.

2. Problem formulation

Consider a single-input–single-output linear time-invariant stable discrete-time system $\mathbb{G}$ represented by $y_k = \mathbb{G}u_k + d_k$, $k = 1, 2, \ldots$, where u_k is the input and d_k is the system noise. For simplicity, in this paper, $\mathbb{G}$ is an FIR (finite impulse response) system. In this case,

$$y_k = a_1 u_k + \cdots + a_n u_{k-n+1} + d_k = \phi_k' \theta + d_k, \tag{1}$$

where $\phi_k' = [u_k, \ldots, u_{k-n+1}]$ is the regressor and $\theta = [a_1, \ldots, a_n]'$ is the parameter vector to be identified and z' denotes the transpose.

The system structure is shown in Fig. 1, in which the input is finitely quantized with r possible values, $u_k \in \mathcal{U} = \{\mu_1, \ldots, \mu_r\}$. The output y_k is measured by a sensor of m thresholds $-\infty < C_1 < \cdots < C_m < \infty$. The output sensor can be represented by a set of m indicator functions $s_k = [s_k^1, \ldots, s_k^m]'$, where $s_k^i = I_{\{-\infty < y_k \le C_i\}}$ for $i = 1, \ldots, m$, and $I_{\{y_k \in A\}} = \begin{cases} 1, & \text{if } y_k \in A, \\ 0, & \text{otherwise.} \end{cases}$ An alternative representation of $I_{\{y_k \in A\}}$ is by defining $\widetilde{s}_k^i = I_{\{C_{i-1} < y_k \le C_i\}}$ and $\widetilde{s}_k = \sum_{i=1}^{m+1} i \widetilde{s}_k^i$. Hence, $\widetilde{s}_k = i$, for $i = 1, \ldots, m+1$, implying that $y_k \in (C_{i-1}, C_i]$ with $C_0 = -\infty$ and $C_{m+1} = \infty$ (with the interval (C_m, ∞)). This representation will be used in deriving the Cramér–Rao lower bound. This paper aims to design an asymptotically efficient algorithm to estimate the unknown parameter vector θ based on $\{u_k\}$ and $\{s_k\}$.

Assumption 1. $\{d_k\}$ is a sequence of i.i.d. (independent and identically distributed) random variables. The accumulative distribution function $F(\cdot)$ of d_1 is invertible and its inverse function, denoted by $F^{-1}(\cdot) = G(\cdot)$, is twice continuously differentiable. The moment generating function of d_1 exists.

For convenience, we denote $F_i(x) = F(C_i - x)$ and define the matrix functions

$$U(x) = \text{diag}\left[\left(\frac{\partial F_1(x)}{\partial x} \right)^{-1}, \ldots, \left(\frac{\partial F_m(x)}{\partial x} \right)^{-1} \right] \tag{2}$$

$$P(x)=\begin{bmatrix} F_1(x)-F_1^2(x) & F_1(x)-F_1(x)F_2(x) & \cdots & F_1(x)-F_1(x)F_m(x) \\ F_1(x)-F_1(x)F_2(x) & F_2(x)-F_2^2(x) & \cdots & F_2(x)-F_2(x)F_m(x) \\ \vdots & & \ddots & \vdots \\ F_1(x)-F_m(x)F_1(x) & F_2(x)-F_m(x)F_2(x) & \cdots & F_m(x)-F_m^2(x) \end{bmatrix} \quad (3)$$

Box I.

and $P(x)$ (given by (3) in Box I) for any $x \in \mathbb{R}$, where diag[...] represents a diagonal matrix.

3. Review on quantized system identification

This section summarizes a few related key results in quantized system identification, including the QCCE and its convergence properties. The main content comes from Wang et al. (2010), Wang and Yin (2007), and Mei et al. (2014).

Consider the special case of a gain system: $y_k = u_k\theta + d_k$. Choose u_k to be a constant. Without loss of generality, assume $u_k \equiv 1$. Then

$$y_k = \theta + d_k. \quad (4)$$

Under Assumption 1, $\{y_k\}$ is an i.i.d. sequence that has the accumulative distribution function $F(\cdot - \theta)$. For the system (4), the probability of $\{s_k^i = 1\}$ is

$$p_i = \Pr\{-\infty < y_k \le C_i\} = F(C_i - \theta) = F_i(\theta). \quad (5)$$

We begin with estimation of p_i in (5). Take N measurements on s_k. Then for $i \in \{1, \dots, m\}$, $\xi_N^i = \frac{1}{N}\sum_{k=1}^N s_k^i$ is the sample relative frequency of y_k taking values in $(-\infty, C_i]$ and an unbiased estimator of p_i for each N, i.e., $E\xi_N^i = p_i$.

An estimator θ_N^i of θ can be derived from $\xi_N^i = F_i(\theta_N^i)$. Consequently, $\theta_N^i = F_i^{-1}(\xi_N^i) = G_i(\xi_N^i)$ is an estimator for θ; and θ_N^i, $i = 1, \dots, m$, are m asymptotically unbiased estimators of θ based on samples of size N. Denote the estimation errors by $e_N = [e_N^1, \dots, e_N^m]'$ with $e_N^i = \theta_N^i - \theta$, and use the notation $\Theta_N = [\theta_N^1, \dots, \theta_N^m]'$, $\mathbb{1} = [1, 1, \dots, 1]'$ of compatible dimension. It is readily seen that $e_N = \Theta_N - \theta\mathbb{1}$. Define $V_N(\theta) = Ee_Ne_N'$. Since $Ee_N \to 0$ as $N \to \infty$, $V_N(\theta)$ is a covariance matrix of e_N, and is positive semi-definite.

Define $\gamma = [\gamma_1, \dots, \gamma_m]'$ such that $\gamma_1 + \cdots + \gamma_m = 1$. One can construct an estimator of θ by $\widehat{a}_N = \sum_{i=1}^m \gamma_i\theta_N^i = \gamma'\Theta_N$. $\widehat{a}_N$ is called a *Quasi-Convex Combination Estimator* (QCCE). The term "quasi-convex" is used since γ_i may not be nonnegative. Since θ_N^i is asymptotically unbiased,

$$E\widehat{a}_N = \gamma' E\Theta_N \to \gamma'\theta\mathbb{1} = \theta \quad \text{as } N \to \infty. \quad (6)$$

Hence, $\widehat{a}_N$ is an asymptotically unbiased estimator of θ. Moreover, the variance of the estimation error $\widehat{a}_N - \theta$ is given by $\overline{\sigma}_N^2 := E(\gamma'\Theta_N - \theta)^2\gamma' Ee_Ne_N'\gamma = \gamma' V_N(\theta)\gamma$. That is, the variance is a quadratic form with respect to the vector γ. The optimal QCCE minimizes $\overline{\sigma}_N^2$ which is obtained from $\sigma_N^2 = \min_{\gamma,\gamma'\mathbb{1}=1}\overline{\sigma}_N^2 = \min_{\gamma,\gamma'\mathbb{1}=1}\gamma' V_N(\theta)\gamma$.

Theorem 1 (*Mei et al., 2014; Wang & Yin, 2007*). *If Assumption 1 holds, then the QCCE $\widehat{a}_N$ converges strongly to θ, i.e., $\widehat{a}_N \to \theta$ w.p.1, and has the strong and mean-square convergence rates $\widehat{a}_N - \theta = O\left(N^{-\frac{1}{2}}\sqrt{\log\log N}\right)$ w.p.1, $N\overline{\sigma}_N^2 \to \gamma' U(\theta)P(\theta)U(\theta)\gamma$ as $N \to \infty$, where $U(\cdot)$ and $P(\cdot)$ are given by (2) and (3). Also the centered and scaled sequence of $\widehat{a}_N$ is asymptotically normal in the sense that $\sqrt{N}(\widehat{a}_N - \theta) \xrightarrow{d} \mathcal{N}\left(0, \gamma' U(\theta)P(\theta)U(\theta)\gamma\right)$, where $\xrightarrow{d}$ denotes convergence in distribution.*

Theorem 2 (*Wang & Yin, 2007*). *Suppose that Assumption 1 holds and $V_N(\theta)$ is positive definite. Then the optimal QCCE can be obtained by choosing $\gamma^* = \frac{V_N^{-1}(\theta)\mathbb{1}}{\mathbb{1}'V_N^{-1}(\theta)\mathbb{1}}$, $\widehat{a}_N = (\gamma^*)'\Theta_N$, and the minimal variance is $\sigma_N^2 = \frac{1}{\mathbb{1}'V_N^{-1}(\theta)\mathbb{1}}$.*

Theorem 3 (*Wang & Yin, 2007*). *The Cramér–Rao Lower Bound for estimating θ based on observations $\{s_k, 1 \le k \le N\}$ is $\sigma_{CR}^2(N, m) = \left(N\sum_{i=1}^{m+1}\frac{\widetilde{h}_i^2}{\widetilde{p}_i}\right)^{-1}$ where $\widetilde{p}_i = F_i(\theta) - F_{i-1}(\theta)$ and $\widetilde{h}_i = \partial\widetilde{p}_i/\partial\theta$, $i = 1, \dots, m+1$.*

Theorem 4 (*Wang & Yin, 2007*). *The optimal QCCE is asymptotically efficient in the sense that $N\sigma_N^2 - N\sigma_{CR}^2(N, m) \to 0$ as $N \to \infty$.*

4. Identification algorithms

Suppose that $u = \{u_k, k = 1, 2, \dots\}$ is an arbitrary input sequence taking quantized values in $\mathcal{U} = \{\mu_1, \dots, \mu_r\}$. The input u generates a regressor sequence $\{\phi_k'\}$ that takes values in $l = r^n$ possible (row vector) patterns denoted by $\mathcal{P} = \{\pi_1, \dots, \pi_l\}$. For example, $\pi_1 = [\mu_1, \dots, \mu_1, \mu_1]$, $\pi_2 = [\mu_1, \dots, \mu_1, \mu_2]$, etc. For $k = n+1, \dots, n+N$, we partition the regressor set $Q_N = \{\phi_{n+1}', \dots, \phi_{n+N}'\}$ according to their patterns π_j. Assume that Q_N contains N_j of pattern π_j, and note that N_j may be zero and $\sum_{j=1}^l N_j = N$. The input pattern set is $M_N = \{\pi_j : N_j \ne 0\}$, i.e., M_N is the collection of all π_j's that have appeared in Q_N.

Assumption 2. There exists $\beta_j \ge 0$ such that $\lim_{N\to\infty} N_j/N = \beta_j$ for $j \in L = \{1, \dots, l\}$.

Definition 1. We use the following notion throughout the paper.

(1) The pattern π_j is said to be *persistent* if $\beta_j > 0$. Without loss of generality, suppose that $\beta_j \ne 0$ for $j \in L_0 = \{1, \dots, l_0\}$ and $\beta_j = 0$ for $j \in L_0^- = \{l_0+1, \dots, l\}$.
(2) $P_u = \{\pi_1, \dots, \pi_{l_0}\}$ is called the *persistent pattern set* of u.
(3) The persistent pattern set P_u of u is said to be *full rank* if the matrix

$$\Psi = (\pi_1', \dots, \pi_{l_0}')' \in \mathbb{R}^{l_0 \times n} \quad (7)$$

is full column rank.
(4) The input u is said to be *persistently exciting* if P_u is full rank.

Remark 1. Non-persistent patterns may exist or appear even infinitely many times. But they do not have impact on asymptotic behavior of the estimates.

We introduce the following two-step estimation algorithms based on the QCCE and weighted least-squares optimization.

Identification algorithm:

(1) At N, if $\pi_j \in M_N$, the observation equations under π_j are

$$y_k^j = w^j + d_k^j, \quad k = 1, \dots, N_j, \quad (8)$$

where $w^j = \pi_j\theta$, is unknown and must be estimated. Let

$$\gamma^j = [\gamma_1^j, \dots, \gamma_m^j]' \in \mathbb{R}^m \quad \text{such that } \gamma_1^j + \cdots + \gamma_m^j = 1, \quad (9)$$

and the corresponding QCCE estimate of w^j be denoted by $\widehat{w}^j_{N_j}$ with estimation error e_{N_j}, which gives

$$\widehat{w}^j_{N_j} = w^j + e_{N_j}. \tag{10}$$

(2) The weighted vector-valued estimate $\widehat{W}_N$ is defined as

$$\widehat{W}_N = \begin{bmatrix} \sqrt{N_1}w^1 \\ \vdots \\ \sqrt{N_l}w^l \end{bmatrix} + \begin{bmatrix} \sqrt{N_1}e_{N_1} \\ \vdots \\ \sqrt{N_l}e_{N_l} \end{bmatrix} = \Phi_N\theta + D_N, \tag{11}$$

where

$$\Phi_N = \begin{bmatrix} \sqrt{N_1}\pi_1 \\ \vdots \\ \sqrt{N_l}\pi_l \end{bmatrix} \quad \text{and} \quad D_N = \begin{bmatrix} \sqrt{N_1}e_{N_1} \\ \vdots \\ \sqrt{N_l}e_{N_l} \end{bmatrix}$$

are the weighted regression matrix and the scaled estimation error vector, respectively.

Suppose that $\Phi_N'\Lambda\Phi_N$ is full rank, where $\Lambda = \text{diag}[\lambda_1, \dots, \lambda_l] > 0$ will be designed later for improving asymptotic properties and convergence rates. Then, the estimate of θ is

$$\widehat{\theta}_N = \left(\frac{1}{N}\Phi_N'\Lambda\Phi_N\right)^{-1}\frac{1}{N}\Phi_N'\Lambda\widehat{W}_N. \tag{12}$$

Remark 2. To highlight the dependence of N_j on l, we write $N_j = N_j(l)$, which depends on the actual input and frequency of occurrence of a given pattern. Typically, a large l generates more data to estimate θ because the dimension of $[\widehat{w}^1_{N_1(\bar{l})}, \dots, \widehat{w}^{\bar{l}}_{N_{\bar{l}}(\bar{l})}]'$ is higher than that of $[\widehat{w}^1_{N_1(l)}, \dots, \widehat{w}^l_{N_l(l)}]'$. From Theorems 6 and 7 it can be seen that $\widehat{\theta}_N - \theta = O\left(N^{-1/2}\sqrt{\log\log N}\right)$ and $E(\widehat{\theta}_N - \theta)(\widehat{\theta}_N - \theta)' = O(1/N)$, which means that the order of the convergence rate of the estimation error only depends on N, and has little to do with l.

It is noted that the total number l of "possible" patterns can be very big. However, in practical usage, the actual complexity depends on how many "active patterns" and "persistent patterns" are actually contained in the given input. The algorithm creates a new pattern only when it appears, and within these active patterns, only the "persistent patterns" contribute to asymptotic convergence properties. The number of "active patterns" can be far less than l, depending on the input design. In other words, in implementing the algorithms, the input should be designed with both convergence and complexity in mind.

5. Convergence properties

This section establishes convergence properties of the identification algorithms under certain persistent excitation conditions on the input. We will consider strong and mean-square convergence, strong and mean-square convergence rates, asymptotic normality, and asymptotic efficiency.

$\Sigma(N)$ denotes the covariance matrix of the estimation error, i.e., $\Sigma(N) = E(\widehat{\theta}_N - \theta)(\widehat{\theta}_N - \theta)'$, and

$$H_1 = \text{diag}[\lambda_1\beta_1, \dots, \lambda_{l_0}\beta_{l_0}] \tag{13}$$

where $\lambda_j, j \in L_0$, is the jth diagonal element of Λ and β_j is given by Assumption 2.

5.1. Convergence, convergence rate and asymptotic normality

Theorem 5. *For system (1) with quantized observations $I_{\{y_k\in A\}}$, if Assumption 1 holds and the input u is persistently exciting, then $\widehat{\theta}_N$ from (12) converges strongly to the true value, $\widehat{\theta}_N \to \theta$ w.p.1 as $N \to \infty$.*

Proof. For $j \in L_0$, by Theorem 1 we have $\widehat{w}^j_{N_j} \to \pi_j\theta$ w.p.1 as $N \to \infty$. From this and (11), when $N \to \infty$ it can be verified that

$$\frac{1}{N}\Phi_N'\Lambda\Phi_N = \sum_{j=1}^{l}\lambda_j\frac{N_j}{N}\pi_j'\pi_j \to \sum_{j=1}^{l_0}\lambda_j\beta_j\pi_j'\pi_j = \Psi'H_1\Psi \tag{14}$$

and

$$\frac{1}{N}\Phi_N'\Lambda\widehat{W}_N = \sum_{j=1}^{l}\lambda_j\frac{N_j}{N}\pi_j'\widehat{w}^j_{N_j} \to \sum_{j=1}^{l_0}\lambda_j\beta_j\pi_j'\pi_j\theta = \Psi'H_1\Psi\theta \tag{15}$$

where Ψ and H_1 are given by (7) and (13). Since $\Lambda > 0$ and $\beta_j > 0$ for $j \in L_0$, we know that $H_1 > 0$ and $\Psi'H_1\Psi > 0$ due to the persistent excitation condition. From (14)–(15) and (12), the theorem follows. □

Corollary 1. *Under the conditions of Theorem 5, $\widehat{\theta}_N$ is an asymptotically unbiased estimator of θ, i.e., $E\widehat{\theta}_N \to \theta$ as $N \to \infty$.*

Proof. Since $E\widehat{w}^j_{N_j} \to \pi_j\theta$ as $N \to \infty$ by (6), in view of (14) and (15) we obtain the desired result. □

Theorem 6. *Under the conditions of Theorem 5, the algorithm (12) has the convergence rate $\widehat{\theta}_N - \theta = O\left(\sqrt{\frac{\log\log N}{N}}\right)$ w.p.1 as $N \to \infty$.*

Proof. In light of (11) and (12), by (14) and $\Psi'H_1\Psi > 0$ we have

$$\begin{aligned}\widehat{\theta}_N - \theta &= \left(\frac{1}{N}\Phi_N'\Lambda\Phi_N\right)^{-1}\frac{1}{N}\Phi_N'\Lambda(\Phi_N\theta + D_N) - \theta \\ &= \left(\frac{1}{N}\Phi_N'\Lambda\Phi_N\right)^{-1}\frac{1}{N}\Phi_N'\Lambda D_N \end{aligned}\tag{16}$$

for sufficiently large N. Let $\widetilde{L}_0 = \{j \in L : N_j \to \infty \text{ as } N \to \infty\}$ and $\widetilde{L}_0^- = L/\widetilde{L}_0 = \{j \in L : N_j \text{ is bounded}\}$. According to Theorem 1, it follows that $e_{N_j} = O\left(N_j^{-1/2}\sqrt{\log\log N_j}\right)$ for $j \in \widetilde{L}_0$ and $N_j/N = O(1/N)$ for $j \in \widetilde{L}_0^-$, which implies

$$\begin{aligned}&\frac{1}{N}\Phi_N'\Lambda D_N \\ &= \sum_{j\in\widetilde{L}_0}\lambda_j\frac{N_j}{N}\pi_j'e_{N_j} + \sum_{j\in\widetilde{L}_0^-}\lambda_j\frac{N_j}{N}\pi_j'e_{N_j} \\ &= \sum_{j\in\widetilde{L}_0}\lambda_j\frac{N_j}{N}\pi_j'O\left(\sqrt{\frac{\log\log N_j}{N_j}}\right) + \sum_{j\in\widetilde{L}_0^-}\lambda_jO\left(\frac{1}{N}\right)\pi_j'e_{N_j} \\ &= O\left(\sqrt{\frac{\log\log N}{N}}\right) \quad \text{w.p.1 as } N \to \infty. \end{aligned}\tag{17}$$

This and (16) prove the theorem. □

Remark 3. For the strong convergence and convergence rate, Assumption 2 can be relaxed. If $\beta_j := \liminf_{N\to\infty} N_j/N$ and Definition 1 remains the same, then under the conditions of

Theorem 5 it can be seen that

$$\liminf_{N\to\infty}\frac{1}{N}\Phi_N'\Lambda\Phi_N \ge \sum_{j=1}^{l_0}\lambda_j\left(\liminf_{N\to\infty}N_j/N\right)\pi_j'\pi_j = \sum_{j=1}^{l_0}\lambda_j\beta_j\pi_j'\pi_j = \Psi'H_1\Psi > 0.$$

From this and (16)–(17), we still have $\widehat{\theta}_N \xrightarrow{w.p.1} \theta$ and $\widehat{\theta}_N - \theta = O\left(N^{-1/2}\sqrt{\log\log N}\right)$ as $N \to \infty$.

Theorem 7. *Under the conditions of Theorem 5, the algorithm (12) has the mean-square convergence rate*

$$N\Sigma(N) \to \left(\Psi'H_1\Psi\right)^{-1}\Psi'H_2\Psi\left(\Psi'H_1\Psi\right)^{-1} \quad as\ N\to\infty, \tag{18}$$

where $H_2 = [\beta_1\lambda_1^2\sigma_1^2, \ldots, \beta_{l_0}\lambda_{l_0}^2\sigma_{l_0}^2]$ *and* $\sigma_j^2 = (\gamma^j)'U(\pi_j\theta)P(\pi_j\theta)U(\pi_j\theta)\gamma^j$ *with* $U(\cdot)$, $P(\cdot)$ *and* γ^j *being given by (2), (3) and (9).*

Proof. With (16) and $\Delta_N := \frac{1}{N}\Phi_N'\Lambda\left(ED_ND_N'\right)\Lambda\Phi_N$, we have

$$N\Sigma(N) = \left(\frac{1}{N}\Phi_N'\Lambda\Phi_N\right)^{-1}\Delta_N\left(\frac{1}{N}\Phi_N'\Lambda\Phi_N\right)^{-1}. \tag{19}$$

For $j \in L_0$, by virtue of Theorem 1 we have

$$N_jEe_{N_j}^2 \to \sigma_j^2 \quad \text{and} \quad \sqrt{N_j}Ee_{N_j} \to 0 \quad \text{as } N\to\infty. \tag{20}$$

As a result, $ED_ND_N' \to \begin{bmatrix}\text{diag}\{\sigma_1^2,\ldots,\sigma_{l_0}^2\} & 0\\ 0 & *\end{bmatrix}$. Since

$$\frac{1}{\sqrt{N}}\Phi_N'\Lambda = \frac{1}{\sqrt{N}}\left(\sqrt{N_1}\lambda_1\pi_1',\ldots,\sqrt{N_l}\lambda_l\pi_l'\right) \to \left(\sqrt{\beta_1}\lambda_1\pi_1',\ldots,\sqrt{\beta_{l_0}}\lambda_{l_0}\pi_{l_0}',0,\ldots,0\right), \tag{21}$$

we obtain

$$\Delta_N = \left(\frac{1}{\sqrt{N}}\Phi_N'\Lambda\right)\left(ED_ND_N'\right)\left(\frac{1}{\sqrt{N}}\Phi_N'\Lambda\right)' \to \left[\sqrt{\beta_1}\lambda_1\pi_1',\ldots,\sqrt{\beta_{l_0}}\lambda_{l_0}\pi_{l_0}'\right]\text{diag}\{\sigma_1^2,\ldots,\sigma_{l_0}^2\} = \Psi'H_2\Psi$$

which together with (14) and (19) completes the proof. □

Theorem 8. *Under the conditions of Theorem 5, the centered and scaled sequence of* $\widehat{\theta}_N$ *is asymptotically normal, i.e.,*

$$\sqrt{N}\left(\widehat{\theta}_N - \theta\right) \xrightarrow{d} \mathcal{N}\left(0,\ \left(\Psi'H_1\Psi\right)^{-1}\Psi'H_2\Psi\left(\Psi'H_1\Psi\right)^{-1}\right)$$

as $N\to\infty$.

Proof. By (16), one can derive

$$\sqrt{N}\left(\widehat{\theta}_N - \theta\right) = \left(\frac{1}{N}\Phi_N'\Lambda\Phi_N\right)^{-1}\frac{1}{\sqrt{N}}\Phi_N'\Lambda D_N. \tag{22}$$

Theorem 1 implies that $\sqrt{N_j}e_{N_j} \xrightarrow{d} \mathcal{N}(0,\sigma_j^2)$ for $j \in L_0$. Consequently, as $N\to\infty$, $D_N \xrightarrow{d} \begin{bmatrix}\mathcal{N}\left(0,\text{diag}[\sigma_1^2,\ldots,\sigma_{l_0}^2]\right) & *\\ * & *\end{bmatrix}$ which together with (14), (21) and (22) indicates the theorem. □

Theorems 5–8 establish some convergence properties of the algorithm (12). The key technical methods are to classify the quantized observations according to the patterns and deal with them respectively. For $j \in L_0$, the results in Section 3 are fully used. For $j \in L_0^-$, we mainly employ the properties of the patterns' frequency of occurrence. The following subsection will construct the optimal weighting matrix such that the algorithm (12) can achieve the CR lower bound asymptotically.

5.2. Asymptotic efficiency

Theorem 9. *The Cramér–Rao Lower Bound for estimating* θ *based on observations* $\{s_k, 1\le k\le N\}$ *is*

$\Sigma_{CR}(N) = \left(\sum_{j=1}^{l}N_j\pi_j'\pi_j\sum_{i=1}^{m+1}\frac{\eta_{i,j}^2}{\zeta_{i,j}}\right)^{-1}$ *where* $\zeta_{i,j}(\theta) = F_i(\pi_j\theta) - F_{i-1}(\pi_j\theta)$ *and* $\eta_{i,j}(\theta) = \partial\zeta_{i,j}/\partial(\pi_j\theta)$, $i = 1,\ldots,m+1$, $j\in L$.

Proof. Augment s_k to $v_k = [s_k', 1]' \in \mathbb{R}^{m+1}$, where the added element represents $1 = \Pr\{-\infty < y_k < \infty\}$. Let x_k be some possible sample value of v_k. Noting that x_k always takes the form of $[0,\ldots,0,1,1,\ldots,1]'$, we have $\Pr\{v_k = x_k;\theta\} = \Pr\{C_{i_0(k)-1} < y_k \le C_{i_0(k)}\} = F_{i_0(k)}(\phi_k'\theta) - F_{i_0(k)-1}(\phi_k'\theta)$, where $i_0(k)$ is the index of the first 1 in x_k. Since $\{d_k\}$ is i.i.d, the likelihood function of $v_1,\ldots,v_N$ taking values $x_1,\ldots,x_N$, conditioned on θ, is $\ell(x_1,\ldots,x_N;\theta) = \Pr\{v_1 = x_1,\ldots,v_N = x_N;\theta\} = \prod_{k=1}^{N}\left[F_{i_0(k)}(\phi_k'\theta) - F_{i_0(k)-1}(\phi_k'\theta)\right]$.

Replace the particular realizations x_k by their corresponding random variables v_k, and denote the resulting quantity by $\ell = \ell(v_1,\ldots,v_N;\theta)$. Set $M_N^j = \{k : \phi_{n+k}' = \pi_j, 1\le k\le N\}$ and $\chi_{i,j} = \frac{1}{N_j}\sum_{k\in M_N^j}\widetilde{s}_k^i$. It is apparent that M_N^j has N_j elements and $E\chi_{i,j} = \zeta_{i,j}$. Then, we have $\ell = \prod_{j=1}^{l}\prod_{k\in M_N^j}\left[F_{i_0(k)}(\pi_j\theta) - F_{i_0(k)-1}(\pi_j\theta)\right] = \prod_{j=1}^{l}\prod_{i=1}^{m+1}\zeta_{i,j}^{N_j\chi_{i,j}}$. Consequently, one can get $\frac{\partial}{\partial\theta}\log\ell = \sum_{j=1}^{l}\sum_{i=1}^{m+1}N_j\chi_{i,j}\frac{\eta_{i,j}}{\zeta_{i,j}}\pi_j$ and

$$\frac{\partial^2}{\partial\theta^2}\log\ell = \sum_{j=1}^{l}N_j\sum_{i=1}^{m+1}\chi_{i,j}\frac{\frac{\partial}{\partial\theta}\eta_{i,j}}{\zeta_{i,j}}\pi_j - \sum_{j=1}^{l}N_j\pi_j'\pi_j\sum_{i=1}^{m+1}\chi_{i,j}\frac{\eta_{i,j}^2}{\zeta_{i,j}^2}. \tag{23}$$

Noticing that $\sum_{i=1}^{m+1}\zeta_{i,j} = 1$, we have

$$E\sum_{j=1}^{l}N_j\sum_{i=1}^{m+1}\chi_{i,j}\frac{\frac{\partial}{\partial\theta}\eta_{i,j}}{\zeta_{i,j}}\pi_j = \sum_{j=1}^{l}N_j\left[\frac{\partial^2}{\partial\theta^2}\sum_{i=1}^{m+1}\zeta_{i,j}\right]\pi_j = 0,$$

which together with (23) and $\Sigma_{CR}(N) = -\left(E\frac{\partial^2}{\partial\theta^2}\log\ell\right)^{-1}$ implies the theorem. □

Theorem 10. *Under the conditions of Theorem 5, if the optimal QCCE is used in (8)–(10) and*

$$\Lambda = \Lambda^* = \text{diag}\left[\sum_{i=1}^{m+1}\frac{\eta_{i,1}^2}{\zeta_{i,1}},\ldots,\sum_{i=1}^{m+1}\frac{\eta_{i,l_0}^2}{\zeta_{i,l_0}},\lambda_{l_0+1},\ldots,\lambda_l\right], \tag{24}$$

then $\widehat{\theta}_N$ *from (12) is asymptotically efficient in the sense that* $N\Sigma(N) - N\Sigma_{CR}(N) \to 0$ *as* $N\to\infty$.

Proof. Under hypothesis, by Theorems 3 and 4 we have $\sigma_j^2 = \left(\sum_{i=1}^{m+1}\frac{\eta_{i,j}^2}{\zeta_{i,j}}\right)^{-1}$ for $j\in L_0$. Furthermore, by (24) it can be verified that

$$H_1 = H_2 = \text{diag}\left[\beta_1\sum_{i=1}^{m+1}\frac{\eta_{i,1}^2}{\zeta_{i,1}},\ldots,\beta_{l_0}\sum_{i=1}^{m+1}\frac{\eta_{i,l_0}^2}{\zeta_{i,l_0}}\right] := H^*.$$

From (18), we have

$$N\Sigma(N) \to \left(\Psi' H^* \Psi\right)^{-1} \quad \text{as } N \to \infty. \tag{25}$$

On the other hand, according to Theorem 9

$$\begin{aligned} N\Sigma_{CR}(N) &= \left(\sum_{j=1}^{l} \frac{N_j}{N}\pi_j'\pi_j \sum_{i=1}^{m+1} \frac{\eta_{i,j}^2}{\zeta_{i,j}}\right)^{-1} \\ &\to \left(\sum_{j=1}^{l_0} \beta_j \pi_j'\pi_j \sum_{i=1}^{m+1} \frac{\eta_{i,j}^2}{\zeta_{i,j}}\right)^{-1} = \left(\Psi' H^* \Psi\right)^{-1}. \end{aligned} \tag{26}$$

Then, the theorem follows from (25) and (26). □

Since θ is unknown, the optimal QCCE is not implementable and Λ^* is also unknown. Hence, Theorem 10 is not a constructive result, and the corresponding asymptotically efficient algorithm cannot be implemented directly. Two implementable algorithms have been developed to approximate the optimal QCCE in Wang and Yin (2007). The next subsection will focus on the implementation of Λ^*.

5.3. Algorithm implementation

Assume that the optimal QCCE is used in (8)–(10). We use an estimate of Λ^* to substitute for its real value to yield an adaptive approximate asymptotically efficient estimation algorithm. It can be recursively expressed by

$$\widehat{\theta}_N = \left(\frac{1}{N}\Phi_N' \Lambda_N \Phi_N\right)^{-1} \frac{1}{N}\Phi_N' \Lambda_N \widehat{W}_N \tag{27}$$

$$\Lambda_N = \text{diag}[\widehat{\sigma}_1^{-2}(N), \ldots, \widehat{\sigma}_l^{-2}(N)] \tag{28}$$

$$\widehat{\sigma}_j^2(N) = \left(\sum_{i=1}^{m+1} \left(\eta_{i,j}(\widehat{\theta}_{N-1})\right)^2 \left(\zeta_{i,j}(\widehat{\theta}_{N-1})\right)^{-1}\right)^{-1}. \tag{29}$$

Similar to (16), by (11) it is known that $\widehat{\theta}_N - \theta = \left(\frac{1}{N}\Phi_N'\Lambda_N\Phi_N\right)^{-1} \frac{1}{N}\Phi_N'\Lambda_N D_N$ and

$$\frac{1}{N}\Phi_N'\Lambda_N D_N = \sum_{j\in L_0} \widehat{\sigma}_j^{-2}(N_j)\frac{N_j}{N}\pi_j' e_{N_j} + \sum_{j\in L_0^-} \lambda_j \frac{N_j}{N}\pi_j' e_{N_j}.$$

In addition, we have $e_{N_j} \to 0$ for $j \in L_0$ and $N_j/N \to 0$ for $j \in L_0^-$, as $N \to \infty$. Thus, $\widehat{\theta}_N \to \theta$ w.p.1, and it follows that

$$\Lambda_N \to \text{diag}\left[\sum_{i=1}^{m+1} \frac{\eta_{i,1}^2}{\zeta_{i,1}}, \ldots, \sum_{i=1}^{m+1} \frac{\eta_{i,l}^2}{\zeta_{i,l}}\right] \quad \text{w.p.1 as } N \to \infty. \tag{30}$$

Another possible construction is to use the sample covariance in place of Λ^*, which gives

$$\overline{W}_N = \sum_{i=1}^{N} \widehat{W}_i / N$$

$$\Lambda_N = \text{Diag}\left(\frac{1}{N-1}\sum_{i=1}^{N}\left(\widehat{W}_i - \overline{W}_i\right)\left(\widehat{W}_i - \overline{W}_i\right)'\right)$$

$$\widehat{\theta}_N = \left(\frac{1}{N}\Phi_N'\Lambda_N\Phi_N\right)^{-1}\frac{1}{N}\Phi_N'\Lambda_N\widehat{W}_N$$

where Diag(B) means a diagonal matrix whose elements come from the diagonal elements of matrix B. By (20), one can get

$$\Lambda_N \to \text{diag}\left[\sum_{i=1}^{m+1} \frac{\eta_{i,1}^2}{\zeta_{i,1}}, \ldots, \sum_{i=1}^{m+1} \frac{\eta_{i,l_0}^2}{\zeta_{i,l_0}}, *, \ldots, *\right] \quad \text{w.p.1} \tag{31}$$

as $N \to \infty$.

In view of (24)–(26), we know that only the first l_0 diagonal elements of Λ^* contribute to the asymptotical efficiency of the algorithms. Thus, (30) and (31) constitute a constructive implementation of Λ^* with the desired asymptotic efficiency.

6. Optimal input design

This section discusses how to design the input optimally for improving mean-square convergence rates. The input design problem is formulated as a max–min optimization problem, which can be solved explicitly in some cases. In particular, when the numbers of output thresholds are large, limiting cases of optimal input design are obtained. Some properties of the solutions are given and an adaptive realization process of the optimal input sequence is provided.

From (25) and (26), it can be seen that

$$\lim_{N\to\infty} N\Sigma(N) = \lim_{N\to\infty} N\Sigma_{CR}(N) = \left(\Psi' H^* \Psi\right)^{-1},$$

which implies that the smaller the size $\| \left(\Psi' H^* \Psi\right)^{-1} \|$ is, the faster the convergence rate of the estimation algorithm (12) becomes. Thus, we can increase the convergence rate by selecting suitable input patterns π_j's from $\mathcal{P}$, the set of all patterns, and adjusting their frequencies of occurrences to minimize $\| \left(\Psi' H^* \Psi\right)^{-1} \|$. In this sense, the optimal input design problem can be stated as a constrained minimization problem

$$\begin{aligned} &\min_{l_0,\pi_1,\ldots,\pi_{l_0},\beta_1,\ldots,\beta_{l_0}} \| \left(\Psi' H^* \Psi\right)^{-1} \| \\ &\text{s.t. } \Psi \text{ is full column rank } (u \text{ is persistently exciting}) \\ &\sum_{j=1}^{l_0} \beta_j = 1, \quad \beta_j > 0,\ \pi_j \in \mathcal{P},\ j \in L_0 \end{aligned} \tag{32}$$

where $\| \cdot \|$ is the spectral norm of a matrix and "s.t." denotes "subject to".

Since l_0 is the number of nonzero β_j $(1 \le j \le l)$ and $\| \left(\Psi' H^* \Psi\right)^{-1} \| = \sigma_{\max}(\left(\Psi' H^* \Psi\right)^{-1}) = (\sigma_{\min}\left(\Psi' H^* \Psi\right))^{-1}$, the optimization problem (32) can be reformulated as

$$\begin{aligned} &\max_{\beta_1,\ldots,\beta_l} Q(\beta_1, \ldots, \beta_l) \\ &\text{s.t. } Q(\beta_1, \ldots, \beta_l) > 0, \\ &\sum_{j=1}^{l} \beta_j = 1, \quad \beta_j \ge 0,\ j \in L, \end{aligned} \tag{33}$$

where $Q(\beta_1, \ldots, \beta_l) := \sigma_{\min}\left(\bar{\Psi}' H \bar{\Psi}\right)$ with

$$\bar{\Psi} = \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_l \end{bmatrix}, \qquad H = \text{diag}\left[\beta_1 \sum_{i=1}^{m+1} \frac{\eta_{i,1}^2}{\zeta_{i,1}}, \ldots, \beta_l \sum_{i=1}^{m+1} \frac{\eta_{i,l}^2}{\zeta_{i,l}}\right],$$

and $\sigma_{\max}(\cdot)$ and $\sigma_{\min}(\cdot)$ represent the maximum and minimum eigenvalues, respectively.

In the case of large numbers of output thresholds, this problem is simplified. As we can see, $\Psi' H^* \Psi$ is dependent on the system input, output thresholds, system parameters, and the distribution and density functions of the disturbance. When the number of output thresholds goes to infinity, the following lemma demonstrates that $\Psi' H^* \Psi$ will converge to a limit that depends only on the input and variance of the disturbance.

Lemma 1. *If d_k is a sequence of i.i.d. Gaussian random variables with zero mean and variance σ^2, and $C_1 \to -\infty$, $C_m \to \infty$, $\max_{1\le i\le m} |C_i - C_{i-1}| \to 0$ as $m \to \infty$, then $\lim_{m\to\infty} \bar{\Psi}' H \bar{\Psi} = \bar{\Psi}'\text{diag}[\beta_1, \ldots, \beta_l]\bar{\Psi}/\sigma^2$.*

Proof. By Theorem 13 in Wang and Yin (2007), we have $\lim_{m\to\infty}\sum_{i=1}^{m+1}\frac{\eta_{i,j}^2}{\zeta_{i,j}}=\frac{1}{\sigma^2}$. Thus, as $m\to\infty$,

$$\begin{aligned}\bar{\Psi}'H\bar{\Psi} &= \sum_{j=1}^{l}\beta_j\pi_j'\pi_j\sum_{i=1}^{m+1}\frac{\eta_{i,j}^2}{\zeta_{i,j}}\\ &\to \frac{1}{\sigma^2}\sum_{j=1}^{l}\beta_j\pi_j'\pi_j=\frac{1}{\sigma^2}\Psi'\mathrm{diag}[\beta_1,\ldots,\beta_l]\Psi,\end{aligned}$$

which gives the lemma. □

Under the conditions of Lemma 1, the optimization problem (33) is reduced to

$$\begin{aligned}&\max_{\beta_1,\ldots,\beta_l}\ \overline{Q}(\beta_1,\ldots,\beta_l)\\ &\text{s.t. } \overline{Q}(\beta_1,\ldots,\beta_l)>0\\ &\sum_{j=1}^{l}\beta_j=1,\quad \beta_j\ge 0,\ j\in L,\end{aligned}\tag{34}$$

with $\overline{Q}(\beta_1,\ldots,\beta_l)=\sigma_{\min}\left(\bar{\Psi}'\mathrm{diag}[\beta_1,\ldots,\beta_l]\bar{\Psi}\right)$.

To solve (33), one can diagonalize $\bar{\Psi}'H\bar{\Psi}$. Since $\left(\bar{\Psi}'H\bar{\Psi}\right)'=\bar{\Psi}'H\bar{\Psi}$, there exists an invertible matrix V such that $V^{-1}\bar{\Psi}'H\bar{\Psi}V$ is a diagonal matrix, denoted as $\mathrm{diag}[a_1,\ldots,a_n]$, where a_i is a function of $\beta_1,\ldots,\beta_l$. Then, the problem becomes $\max_{\beta_1,\ldots,\beta_l}\min\{a_1,\ldots,a_n\}$. This is a max–min optimization problem. We use $\boldsymbol{\beta}^*=[\beta_1^*,\ldots,\beta_l^*]'$ to represent a solution of (33). To reflect the relationship between $\boldsymbol{\beta}^*$ and θ, we also write $\boldsymbol{\beta}^*(\theta)=[\beta_1^*(\theta),\ldots,\beta_l^*(\theta)]'$. Its properties are given by the following proposition.

Proposition 1. *The following assertions hold.*

(i) $\boldsymbol{\beta}^*$ *may be non-unique.*
(ii) $\boldsymbol{\beta}^*$ *has at least n nonzero elements.*
(iii) $Q(\beta_1^*,\ldots,\beta_l^*)\le\max_{1\le j\le l}\sum_{i=1}^{m+1}\frac{\eta_{i,j}^2}{\zeta_{i,j}}\pi_j\pi_j'$.

Proof. (i) This will be shown by Example 2.

(ii) This follows from the fact that Ψ is full column rank under $\boldsymbol{\beta}^*$.

(iii) For any $\beta_1,\ldots,\beta_l$, by

$$\Phi'H\Phi=\sum_{j=1}^{l}\beta_j\sum_{i=1}^{m+1}\frac{\eta_{i,j}^2}{\zeta_{i,j}}\pi_j'\pi_j\tag{35}$$

and $0\le\beta_j\le 1$ we have

$$\begin{aligned}&Q(\beta_1,\ldots,\beta_l)\\ &\le\sum_{j=1}^{l}\sigma_{\max}\left(\beta_j\sum_{i=1}^{m+1}\frac{\eta_{i,j}^2}{\zeta_{i,j}}\pi_j'\pi_j\right)\\ &=\sum_{j=1}^{l}\beta_j\sum_{i=1}^{m+1}\frac{\eta_{i,j}^2}{\zeta_{i,j}}\pi_j\pi_j'\le\sum_{j=1}^{l}\beta_j\max_{1\le j\le l}\left\{\sum_{i=1}^{m+1}\frac{\eta_{i,j}^2}{\zeta_{i,j}}\pi_j\pi_j'\right\}\end{aligned}$$

which together with $\sum_{j=1}^{l}\beta_j=1$ indicates that $Q(\beta_1,\ldots,\beta_l)\le\max_{1\le j\le l}\sum_{i=1}^{m+1}\frac{\eta_{i,j}^2}{\zeta_{i,j}}\pi_j\pi_j'$. Hence, (iii) is true. □

Consequently, a question arises naturally on the realization of $\boldsymbol{\beta}^*$, i.e., how to design an input sequence such that $\lim_{N\to\infty}[N_1,\ldots,N_l]'/N=\boldsymbol{\beta}^*$. If $\boldsymbol{\beta}^*$ is known, one can define a sequence $\{x_k,k\ge 1\}$ of i.i.d. discrete random variables with $\Pr(x_1=j)=\beta_j^*$ for $j\in L$, which is also denoted as $x_1=x_1(1:l;\boldsymbol{\beta}^*)$. At N, let $\phi_{n+1}'=\pi_{x_1},\ldots,\phi_{n+N}'=\pi_{x_N}$. Then, $N_j=\sum_{i=1}^{N}I_{\{x_i=j\}}$, and by the laws of large numbers,

$$\lim_{N\to\infty}\frac{N_j}{N}=\frac{1}{N}\sum_{i=1}^{N}I_{\{x_i=j\}}=EI_{\{x_1=j\}}=\Pr(x_1=j)=\beta_j^*.$$

Thus, $[N_1,\ldots,N_l]'/N\to\boldsymbol{\beta}^*$ as $N\to\infty$.

However, $\boldsymbol{\beta}^*$ is unknown since θ is unknown. Suppose that $\boldsymbol{\beta}^*=\boldsymbol{\beta}^*(\theta)=[\beta_1^*(\theta),\ldots,\beta_{l_0}^*(\theta),\ 0,\ldots,0]'$ and there exists a constant $\underline{\beta}_j>0$ such that $\beta_j^*\ge\underline{\beta}_j$ for $j\in L_0$. For any $Y=[y_1,\ldots,y_{l_0}]'\in\mathbb{R}^{l_0}$, define $\psi(Y)=[z_1,\ldots,z_{l_0}]'\in\mathbb{R}^{l_0}$, where $[z_1,\ldots,z_{l_0}]=[y_1,\ldots,y_{l_0}]$ if $y_j\ge\underline{\beta}_j$ for $j\in L_0$; otherwise,

$$z_j=\begin{cases}\underline{\beta}_j, & \text{if } y_j\le\underline{\beta}_j;\\ \underline{\beta}_j+\upsilon_j(y_j-\underline{\beta}_j), & \text{if } y_j>\underline{\beta}_j,\end{cases}$$

where $\upsilon_j\ge 0$ satisfy $\sum_{j:\,y_j>\underline{\beta}_j}\upsilon_j(y_j-\underline{\beta}_j)=1-\sum_{j=1}^{l_0}\underline{\beta}_j$. It can be verified that

$$\sum_{j=1}^{l_0}z_j=\sum_{j:\,y_j\le\underline{\beta}_j}\underline{\beta}_j+\sum_{j:\,y_j>\underline{\beta}_j}\left(\underline{\beta}_j+\upsilon_j(y_j-\underline{\beta}_j)\right)=1.$$

Given $\theta_0\in\mathbb{R}^n$, let $Y_0=[\beta_1^*(\theta_0),\ldots,\beta_{l_0}^*(\theta_0)]'$, $x_1=x_1(1:l_0;\psi(Y_0))$ and $\phi_{n+1}'=\pi_{x_1}$. By (12), $\widehat{\theta}_1$ can be calculated. Then, let $Y_1=[\beta_1^*(\widehat{\theta}_1),\ldots,\beta_{l_0}^*(\widehat{\theta}_1)]'$, $x_2=x_2(1:l_0;\psi(Y_1))$, $\phi_{n+2}'=\pi_{x_2}$. From ϕ_2 and ϕ_1, we can get $\widehat{\theta}_2$. In general, let $Y_{N-1}=[\beta_1^*(\widehat{\theta}_{N-1}),\ldots,\beta_{l_0}^*(\widehat{\theta}_{N-1})]'$, $x_N=x_N(1:l_0;\psi(Y_{N-1}))$, $\phi_{n+N}'=\pi_{x_N}$, and by (12) calculate $\widehat{\theta}_N$ from $\phi_{n+N},\phi_{n+N-1},\ldots,\phi_{n+1}$. Generate a sequence $\{\phi_{n+1},\phi_{n+2},\ldots\}$ by repeating the process. It can be seen that $\liminf_{N\to\infty}N_j/N\ge\underline{\beta}_j$ for $j\in L_0$. According to Remark 3, we have $\widehat{\theta}_N\to\theta$ as $N\to\infty$. Furthermore, we have $\Pr(x_N=j)=\beta_j^*(\widehat{\theta}_{N-1})\to\beta_j^*(\theta)=\beta_j^*$ and $N_j/N=\frac{1}{N}\sum_{i=1}^{N}I_{\{x_i=j\}}\to\beta_j^*$ for $j\in L_0$. Hence, $[N_1,\ldots,N_{l_0},0\ldots,0]'/N\to\boldsymbol{\beta}^*$ as $N\to\infty$.

7. Joint identification of noise distribution functions and system parameters

The developments above rely on the knowledge of the distribution function $F(\cdot)$ and its inverse. However, in most applications, the noise distributions are not known, or only limited information is available. On the other hand, input/output data from the system contain information about the noise distribution. By viewing unknown distributions and system parameters jointly as uncertainties, this section investigates the joint identification of them. To avoid undue complexity, we consider the case that $m=1$ (the output sensor has only one threshold C) and the noise distribution function $F(\cdot)$ contains $\bar{n}$ unknown parameters $\rho_1,\ldots,\rho_{\bar{n}}$, i.e., $F(z)=F(z;\rho)$ for any $z\in\mathbb{R}$ by denoting $\rho=[\rho_1,\ldots,\rho_{\bar{n}}]'\in\mathbb{R}^{\bar{n}}$.

7.1. Identifiability

Since $\pi_{l_0+1},\ldots,\pi_l$ are sparse in the regressor sequence $\{\phi_k',k\ge 1\}$ (i.e., $\beta_j=0$ for $j\in L_0^-$), for simplicity they will not be used to design algorithms in this section. From $\{s_k\}$, $F(C-w^j;\rho)$ can be estimated by the empirical-measure-based method for $j\in L_0$. If $F(C-w^j;\rho)$ are precisely known, then there exists a vector $\alpha=[\alpha_1,\ldots,\alpha_{l_0}]'\in\mathbb{R}^{l_0}$ such that

$$F(w^j;\rho)=\alpha_j,\quad j=1,\ldots,l_0.\tag{36}$$

The equations above are usually unsolvable because there are $l_0+\bar{n}$ unknowns ($w_1,\ldots,w_{l_0}$ and ρ) but only l_0 equations. Thus, at least another $\bar{n}$ independent equations are needed, This is achieved by special input design. Assume that the input can make $w^1, \ldots, w^{l_0}$ satisfy the following T ($\geq \bar{n}$) equations

$$\mathcal{F}_i(w^1, \ldots, w^{l_0}) = 0, \quad i = 1, \ldots, T, \tag{37}$$

where $\mathcal{F}_i(\cdot)$ is a function from $\mathbb{R}^{l_0}$ to $\mathbb{R}$, $i = 1, \ldots, T$.

If we can obtain $w^1, \ldots, w^{l_0}$ and ρ from (36) and (37), then θ will be given by $w^1, \ldots, w^{l_0}$, i.e.,

$$\theta = \left(\Psi'\Psi\right)^{-1} \Psi W^0, \quad \text{with } W^0 = (w^1, \ldots, w^{l_0})' \tag{38}$$

if Ψ is full column rank.

Remark 4. A simple way for realizing (37) is the scaling factor method proposed in Wang et al. (2010).

Assumption 3. There exists a convex compact set $\Xi \subseteq \mathbb{R}^{l_0}$ such that $\alpha \in \Xi$ and for any $\varsigma = [\varsigma_1, \ldots, \varsigma_{l_0}]' \in \Xi$ the equations $F\left(x_j; \rho\right) = \varsigma_j, j = 1, \ldots, l_0, \mathcal{F}_i\left(x_1, \ldots, x_{l_0}\right) = 0, i = 1, \ldots, T$ have a unique solution $[x_1, \ldots, x_{l_0}, \rho']' \in \mathbb{R}^{l_0+\bar{n}}$, also denoted by $\tau(\varsigma)$ that is continuous in Ξ with respect to some vector norm.

7.2. Identification algorithms

Under Assumption 3, by (36) and (38)

$$\tau(\alpha) = [w^1, \ldots, w^{l_0}, \rho']' = (W^0, \rho')'. \tag{39}$$

This, together with the continuity of $\tau(\cdot)$, implies that one can derive an estimate of α from the estimates of W^0 and ρ. A detailed process is given by the following algorithm.

Identification algorithm:

(1) At N, for $j \in L_0$, the system outputs under π_j are $y_k^j = w^j + d_k^j$ and the corresponding binary-valued observations are $s_k^j = I_{\{y_k^j \leq C\}}$. Define $\bar{S}_N^j = \begin{cases} \frac{1}{N_j}\sum_{i=1}^{N_j} s_i^j, & N_j \neq 0; \\ \frac{1}{2}, & N_j = 0, \end{cases}$ and let $\widehat{\alpha}_N^0 = [\bar{S}_N^1, \ldots, \bar{S}_N^{l_0}]'$ and $\widehat{\alpha}_N = \Pi_\Xi(\widehat{\alpha}_N^0)$ where $\Pi_\Xi(\cdot)$ is a projection operator given by $\Pi_\Xi(z) = \text{argmin}_{\omega\in\Xi}\|z-\omega\|$ for any $z \in \mathbb{R}^{l_0}$, and $\|\cdot\|$ is the vector norm in Assumption 3.

(2) By Assumption 3, let $\tau_N = \tau(\widehat{\alpha}_N)$. Using $\tau_{N,i}$ to represent the ith component of τ_N for $i = 1, \ldots, l_0 + \bar{n}$, the estimates of W^0 and ρ are $\widehat{W}_N^0 = [\tau_{N,1}, \ldots, \tau_{N,l_0}]'$, $\widehat{\rho}_N = [\tau_{N,l_0+1}, \ldots, \tau_{N,l_0+m}]'$.

(3) Suppose that $(\Phi_N^0)'\bar{\Lambda}\Phi_N^0$ is full rank, the estimate of θ is

$$\widehat{\theta}_N = \left(\frac{1}{N}(\Phi_N^0)'\bar{\Lambda}\Phi_N^0\right)^{-1} \frac{1}{N}(\Phi_N^0)'\bar{\Lambda}\text{diag}[\sqrt{N_1}, \ldots, \sqrt{N_{l_0}}]\widehat{W}_N^0 \tag{40}$$

where $\Phi_N^0 = [\sqrt{N_1}\pi_1, \ldots, \sqrt{N_{l_0}}\pi_{l_0}]'$ and $\bar{\Lambda} = \text{diag}[\bar{\lambda}_1, \ldots, \bar{\lambda}_{l_0}] > 0$ is a weighting matrix.

7.3. Convergence

Theorem 11. *For system (1) with binary observations, if $\{d_k\}$ is i.i.d and Assumptions 2 and 3 hold, then τ_N converges strongly to the real value, $\tau_N \to \tau(\alpha)$ w.p.1 as $N \to \infty$.*

Proof. By the law of large numbers and the definition of $\bar{S}_N^j$, we have $\bar{S}_N^j \to F\left(C - w^j; \rho\right)$ for $j \in L_0$. Thus, it follows that $\widehat{\alpha}_N \to \alpha$ w.p.1 as $N \to \infty$. Since $\tau(\varsigma)$ is continuous in Ξ by Assumption 3, it is known that $\tau_N \to \tau(\alpha)$ by (39). This completes the proof. □

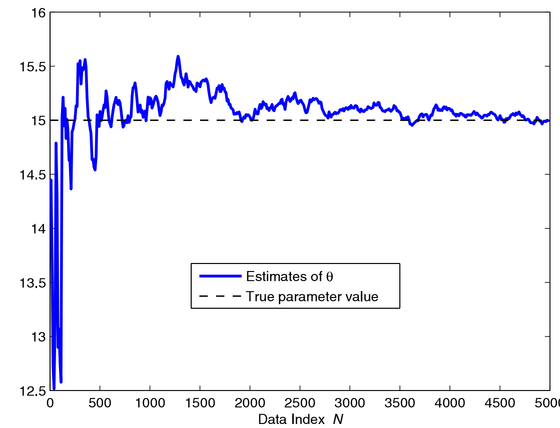


Fig. 2. Convergence of $\widehat{\theta}_N$ from (12).

Theorem 12. *Under the conditions of Theorem 11, if u is persistently exciting, then $\widehat{\theta}_N$ from (40) converges strongly to the true value, $\widehat{\theta}_N \to \theta$ w.p.1 as $N \to \infty$.*

Proof. By virtue of Theorem 11, we know that $\tau_{N,j} \to w^j$ for $j \in L_0$. Similar to (14) and (15), one can obtain $\frac{1}{N}(\Phi_N^0)'\bar{\Lambda}\text{diag}[\sqrt{N_1}, \ldots, \sqrt{N_{l_0}}]\widehat{W}_N^0 \to \Psi' H_1 \Psi\theta$ and $\frac{1}{N}(\Phi_N^0)'\bar{\Lambda}\Phi_N^0 \to \Psi' H_1 \Psi$, as $N \to \infty$. This together with $\Psi' H_1 \Psi > 0$ proves the theorem. □

8. Numerical examples

Example 1. Consider a gain system $y_k = u_k\theta + d_k$, where the true value $\theta = 15$ and $\{d_k\}$ is a sequence of i.i.d. normal random variables with zero mean and standard deviation $\sigma = 25$. The output is measured by a sensor that has three thresholds $C_1 = 32$, $C_2 = 53$, $C_3 = 60$. The input is quantized and takes values from $\mathcal{U} = \{\pi_1, \pi_2, \pi_3, \pi_4\} = \{1, 2, 3.1, 5\}$. Since $\theta \in \mathbb{R}$, we have $\mathcal{P} = \mathcal{U}$. At N, assume that $N_1 = N - N_2 - N_3 - N_4$, $N_2 = \lceil 0.6(N - N_3 - N_4)\rceil$, $N_3 = \min\{110, |\lceil \log N \rceil|\}$, $N_4 = \lceil \sqrt{N} \rceil$, where $\lceil z \rceil$ denotes the smallest integer greater than or equal to $z \in \mathbb{R}$. Thus, $\pi_1 = 1$ and $\pi_2 = 2$ are persistent, $P_u = \{1, 2\}$ and $\Psi = [1, 2]'$.

In (8)–(10), we use the algorithm (20) proposed in Wang and Yin (2007) to simulate the optimal QCCE, and $\widehat{\theta}_N$ is computed using (12) with $\lambda = \text{diag}[4, 2, 1, 1]$. The convergence is shown by Fig. 2.

From Theorem 9, one can get

$$[\zeta_{i,j}]_{4\times 4} = \begin{bmatrix} 0.7517 & 0.5319 & 0.2810 & 0.0427 \\ 0.1840 & 0.2893 & 0.3216 & 0.1467 \\ 0.0284 & 0.0637 & 0.1028 & 0.0849 \\ 0.0359 & 0.1151 & 0.2946 & 0.7257 \end{bmatrix}$$

and

$$[\eta_{i,j}]_{4\times 4} = \begin{bmatrix} -0.0127 & -0.0159 & -0.0135 & -0.0036 \\ 0.0077 & 0.0054 & -0.0019 & -0.0072 \\ 0.0018 & 0.0027 & 0.0016 & -0.0025 \\ 0.0032 & 0.0078 & 0.0138 & 0.0133 \end{bmatrix},$$

which together with the definition of N_i, $i = 1, \ldots, 4$, leads to the CR lower bound $\Sigma_{CR}(N)$. Fig. 3 gives the sample variances $\Sigma(N)$ with $\Lambda = \text{diag}[4, 2, 1, 1]$ and $\Lambda = \Lambda^* = \text{diag}[0.0009, 0.0012, 1, 1]$, and the theoretical CR bound. It can be seen that the curve with $\Lambda = \text{diag}[4, 2, 1, 1]$ is higher than the one with $\Lambda = \Lambda^*$, which illustrates the impact of Λ on the convergence rate. Especially, the curve with $\Lambda = \Lambda^*$ converges to the CR lower bound, which indicates the asymptotic efficiency.

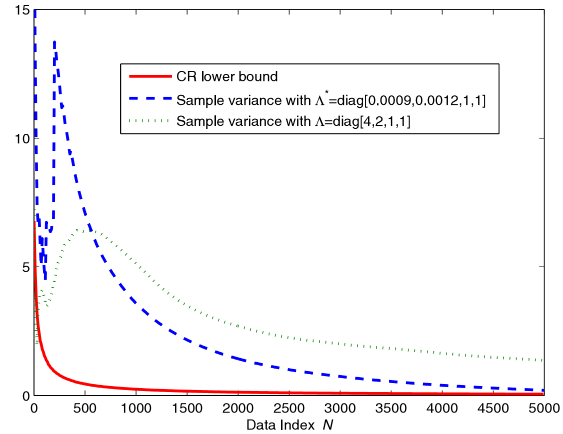


Fig. 3. Asymptotic efficiency of $\widehat{\theta}_N$ from (12).

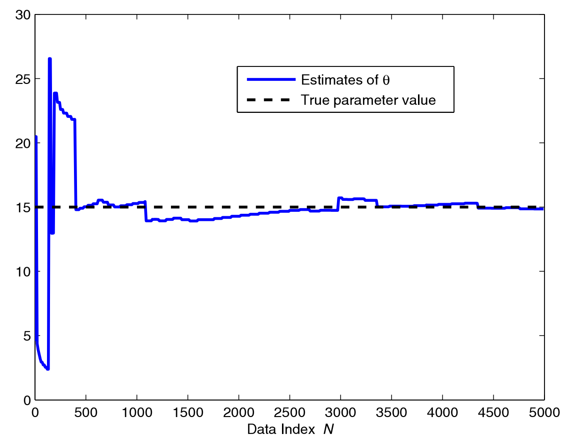


Fig. 4. Convergence of algorithm (27)–(29).

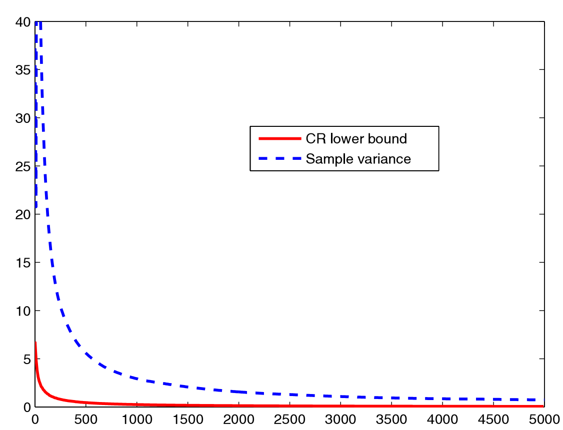


Fig. 5. Convergence rate of algorithm (27)–(29).

Using $\widehat{\theta}_N$ to substitute for θ in Λ^*, we simulate the adaptive asymptotically efficient algorithm (27)–(29), whose convergence and convergence rate are shown by Figs. 4 and 5.

Example 2. If $\theta \in \mathbb{R}$, then $\pi_j'\pi_j \in \mathbb{R}$. By (35), $Q(\beta_1, \ldots, \beta_l) = \sum_{j=1}^{l} \beta_j \sum_{i=1}^{m+1} \frac{\eta_{i,j}^2}{\zeta_{i,j}} \pi_j'\pi_j$. Assume that $\sum_{i=1}^{m+1} \frac{\eta_{i,j_0}^2}{\zeta_{i,j_0}} \pi_{j_0}'\pi_{j_0} = \max_{1\le j\le l} \sum_{i=1}^{m+1} \frac{\eta_{i,j}^2}{\zeta_{i,j}} \pi_j'\pi_j$. Then, $\beta_j = \begin{cases} 1, & j = j_0 \\ 0, & j \neq j_0 \end{cases}$ gives a solution of (33). In addition, the uniqueness of the solution of (33) is determined by the uniqueness of j_0.

Example 3. Consider problem (34). The input is quantized and takes values from $\mathcal{U} = \{-1, 1\}$, which implies that $\mathcal{P} = \{\pi_1, \pi_2, \pi_3, \pi_4\} = \{[-1, -1], [-1, 1], [1, 1], [1, -1]\}$. Observe that

$$\bar{\Psi}'\mathrm{diag}[\beta_1, \ldots, \beta_l]\bar{\Psi} = \begin{bmatrix} \sum_{j=1}^{4} \beta_j & \sum_{j=1}^{4} (-1)^{j+1}\beta_j \\ \sum_{j=1}^{4} (-1)^{j+1}\beta_j & \sum_{j=1}^{4} \beta_j \end{bmatrix}.$$

From this expression, one can derive $\overline{Q}(\beta_1, \beta_2, \beta_3, \beta_4) = 2\min\{\beta_1 + \beta_3, \beta_2 + \beta_4\} = \sum_{j=1}^{4} \beta_j - \left|\sum_{j=1}^{4} (-1)^{j+1}\beta_j\right|$. Since $\sum_{j=1}^{4} \beta_j = 1$, $\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*$ is a solution of (34) if and only if $\beta_1^* + \beta_3^* = \beta_2^* + \beta_4^* = 1/2$ with $0 \le \beta_j^* \le 1$ for $j = 1, \ldots, 4$, and $\max \overline{Q}(\beta_1, \beta_2, \beta_3, \beta_4) = 1$. Thus, $\beta_1^* = \beta_2^* = 1/2$ and $\beta_3^* = \beta_4^* = 0$ is a solution, which indicates that a 2-periodic signal with its one-period $\phi_{n+1} = [-1, -1]'$ and $\phi_{n+2} = [-1, 1]'$ is an optimal input sequence.

9. Concluding remarks

This paper resolves a critical standing issue in quantized system identification: When the input is periodic and full rank, algorithms and their key convergence properties are available, including strong and mean-square convergence, strong and mean-square convergence rates, asymptotic normality, and asymptotic efficiency. However, at present it is unclear what constructive algorithms will retain all these properties when the input is not periodic. Under quantized output observations and general quantized inputs, this paper introduces identification algorithms and input excitation conditions under which parameter estimates attain all these properties. The results and methods developed in this paper can be potentially extended to different systems, noise characterizations, and uncertainties.

References

Agüero, J.C., Goodwin, G.C., & Yuz, J.I. (2007). System identification using quantized data. In *Proceedings of 46th IEEE conference on decision and control* (pp. 4263–4268).

Akyildiz, I., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: A survey. *Computer Networks*, *38*, 393–422.

Casini, M., Garulli, A., & Vicino, A. (2011). Input design in worst-case system identification using binary sensors. *IEEE Transactions on Automatic Control*, *56*(5), 1186–1191.

Casini, M., Garulli, A., & Vicino, A. (2012). Input design in worst-case system identification with quantized measurements. *Automatica*, *48*(12), 2997–3007.

Cerone, V., Piga, D., & Regruto, D. (2013). Fixed-order FIR approximation of linear systems from quantized input and output data. *Systems & Control Letters*, *62*(12), 1136–1142.

Chen, T.S., Zhao, Y.L., & Ljung, L. (2012). Impulse response estimation with quantized measurements: a regularized FIR model approach. In *Proceedings of the 16th IFAC symposium on system identification* (pp. 113–118).

Chiuso, A. (2008). A note on estimation using quantized data. In *Proceedings of the 17th IFAC world congress* (pp. 12486–12491).

Colinet, E., & Juillard, J. (2010). A weighted least-squares approach to parameter estimation problems based on binary measurements. *IEEE Transactions on Automatic Control*, *55*(1), 148–152.

Godoy, B., Goodwin, G. C., Agüero, J., Marelli, D., & Wigren, T. (2011). On identification of FIR systems having quantized output data. *Automatica*, *47*(9), 1905–1915.

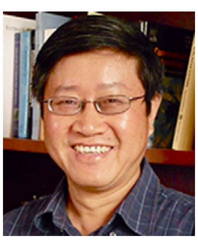
Guo, J., & Zhao, Y. L. (2013). Recursive projection algorithm on FIR system identification with binary-valued observations. *Automatica*, *49*, 3396–3401.

Guo, J., & Zhao, Y. L. (2014). Identification of the gain system with quantized observations and bounded persistent excitations. *Science in China Series F: Information Sciences*, *57*, 012205:1–012205:15.

Guo, L. (1993). *Time-varying stochastic systems—stability, estimation and control.* Changchun, PR China: Jilin Science and Technology Press.

Gustafsson, F., & Karlsson, R. (2009). Statistical results for system identification based on quantized observations. *Automatica*, *45*(12), 2794–2801.
He, Q., Wang, L. Y., & Yin, G. (2013). *Regular and quantized observations: Applications of large deviations principles*. New York, Heidelberg, Dordrecht, London: Springer.
Li, Q., Zhang, B. H., Cui, L. G., Fan, Z., & Athanasios, V. V. (2014). Epidemics on small worlds of tree-based wireless sensor networks. *Journal of Systems Science and Complexity*, *27*(6), 1095–1120.
Ljung, L. (1987). *System identification: Theory for the user*. Englewood Cliffs, NJ: Prentice-Hall.
Marelli, D., You, K. Y., & Fu, M. Y. (2013). Identification of ARMA models using intermittent and quantized output observations. *Automatica*, *49*, 360–369.
Mei, H. W., Yin, G., & Wang, L. Y. (2014). Almost sure convergence rates for system identification using binary, quantized, and regular sensors. *Automatica*, *50*(8), 2120–2127.
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., & Zhao, Y. L. (2010). *System identification with quantized observations*. Boston, MA: Birkhäuser.
Wang, L. Y., Zhang, J. F., & Yin, G. (2003). System identification using binary sensors. *IEEE Transactions on Automatic Control*, *48*, 1892–1907.
You, K. Y., Xie, L. H., Sun, S. L, & Xiao, W. D. (2011). Quantized filtering of linear stochastic system. *Transactions of the Institute of Measurement and Control*, *33*(6), 683–698.
Zhao, Y. L., Wang, L. Y., Yin, G., & Zhang, J. F. (2007). Identification of Wiener systems with binary-valued output observations. *Automatica*, *43*(10), 1752–1765.

Jin Guo received the B.S. degree in mathematics from Shandong University, China, in 2008, and Ph.D. degree in system modeling and control theory from Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2013. Now, he is a postdoctoral researcher in the School of Automation and Electrical Engineering, University of Science and Technology Beijing. His current research interests are identification and control of set-valued output systems and systems biology.



Le Yi Wang received the Ph.D. degree in electrical engineering from McGill University, Montreal, Canada, in 1990. Since 1990, he has been with Wayne State University, Detroit, Michigan, where he is currently a professor in the Department of Electrical and Computer Engineering. His research interests are in the areas of complexity and information, system identification, robust control, H-infinity optimization, time-varying systems, adaptive systems, hybrid and nonlinear systems, information processing and learning, as well as medical, automotive, communications, power systems, and computer applications of control methodologies. He was a keynote speaker in several international conferences. He was an Associate Editor of the IEEE Transactions on Automatic Control and several other journals, and currently is an Associate Editor of the Journal of Control Theory and Applications. He is a Fellow of IEEE.

George Yin received the B.S. degree in mathematics from the University of Delaware in 1983, M.S. degree in Electrical Engineering, and Ph.D. in Applied Mathematics from Brown University in 1987. He joined Wayne State University in 1987, and became a professor in 1996. His research interests include stochastic systems and applications. He severed on the IFAC Technical Committee on Modeling, Identification and Signal Processing, and many conference program committees; he was Co-Chair of SIAM Conference on Control & Its Application, 2011, and Co-Chair of a couple AMS-IMS-SIAM Summer Research Conferences; he also chaired a number SIAM related prize selection committees. He is Chair SIAM Activity Group on Control and Systems Theory, and is serving as the Board of Directors of American Automatic Control Council. He is an associate editor of SIAM Journal on Control and Optimization, and is on the editorial board of a number of other journals. He was an Associate Editor of Automatica (2005–2011) and IEEE Transactions on Automatic Control (1994–1998). He is a Fellow of IEEE and a fellow of SIAM.

Yanlong Zhao received the B.S. degree in mathematics from Shandong University, Jinan, 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 AMSS, CAS, where he is currently an associate professor. His current research interests include identification and control of quantized systems. He has been an Associate Editor of Asian Journal of Control since April 2012, and Journal of Systems and Science and Complexity since June 2014. He is now serving as a Vice-Chair of IEEE CSS Beijing Chapter, Vice-General Secretary of Chinese Association of Automation, and member of the International Federation of Automatic Control (IFAC) Technical Committee on Modeling.

Ji-Feng Zhang received the B.S. degree in mathematics from Shandong University, China, in 1985, and Ph.D. degree in control theory from the Institute of Systems Science (ISS), Chinese Academy of Sciences, China, in 1991. Since 1985, he has been with the ISS, and now is the Director of the ISS. His current research interests include system modeling and identification, adaptive control, stochastic systems, and multi-agent systems. He is/was Editor-in-Chief or Deputy Editor-in-Chief of several journals such as "Journal of Systems Science and Mathematical Sciences" and "Science China: Information Sciences", and Associate Editor of more than 10 journals including "IEEE Transactions on Automatic Control" and "SIAM Journal on Control and Optimization". He received the Second Prize of the State Natural Science Award of China in 2010, the Distinguished Young Scholar Fund from National Natural Science Foundation of China in 1997. He was General Co-Chair, Program Chair of many control conferences. He is now serving as a Vice-President of the Systems Engineering Society of China, Vice-President of Chinese Association of Automation, and Vice-Chair of the IFAC Technical Board. He is a Fellow of IEEE.