Automatica 78 (2017) 280–286

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Identification of Wiener systems with quantized inputs and binary-valued output observations☆

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

[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

[e] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, PR China

ARTICLE INFO

Article history:
Received 27 March 2016
Received in revised form
2 August 2016
Accepted 17 November 2016
Available online 25 January 2017

Keywords:
Identification
Wiener system
Quantized input
Binary-valued observation
Asymptotic efficiency

ABSTRACT

This paper investigates identification of Wiener systems with quantized inputs and binary-valued output observations. By parameterizing the static nonlinear function and incorporating both linear and nonlinear parts, we begin by investigating system identifiability under the input and output constraints. Then a three-step algorithm is proposed to estimate the unknown parameters by using the empirical measure, input persistent patterns, and information on noise statistics. Convergence properties of the algorithm, including strong convergence and mean-square convergence rate, are established. Furthermore, by selecting a suitable transformation matrix, the asymptotic efficiency of the algorithm is proved in terms of the Cramér–Rao lower bound. Finally, numerical simulations are presented to illustrate the main results of this paper.

1. Introduction

Wiener systems are often used to describe nonlinear systems in practice. Such systems are typically comprised of two blocks: a linear dynamic system followed by a nonlinear static function. Practical Wiener systems are exemplified by distillation columns (Zhu, 1999), pH control processes (Kalafatis, Arifin, Wang, & Cluett, 1995), and biological systems (Hunter & Korenberg, 1986). Theoretically, some nonlinear systems, which are not of a Wiener structure, may be represented or approximated by a multivariate Wiener model (Boyd & Chua, 1985). Consequently, its study carries profound theoretical and practical significance.

Identification of Wiener systems has drawn great attention and experienced substantial advancement. Fundamental progress has been achieved in methodology development, identification algorithms, essential convergence properties, and applications (Chen & Zhao, 2014; Giri & Bai, 2010; Greblicki, 1997; Hagenblad, Ljung, & Wills, 2008; Wang & Ding, 2011; Wills, Schon, Ljung, & Ninness, 2011; Zhu, 1999). Zhu (1999) extended an identification method for multi-input single-output Wiener models and applied it to identify two distillation columns. Hagenblad et al. (2008) employed the Maximum Likelihood (ML) method to identify Wiener systems, and discussed efficient implementation issues for Wiener systems under disturbances. Wang and Ding (2011) derived an LS-type gradient-based iterative identification algorithm for Wiener systems. Chen and Zhao (2014) used stochastic approximation algorithms with expanding truncation to identify Wiener systems. Greblicki (1997) introduced a nonparametric approach to Wiener system identification. Wills et al. (2011) developed a new ML-based algorithm for identifying Hammerstein–Wiener models. Giri and Bai (2010) summarized progress on identification methods of block-oriented nonlinear systems.

Along with the rapid advancement of sensor and communication technologies (Shen, Tan, Wang, Wang, & Lee, 2015; Xie & Wang, 2014), system identification under binary-valued/quantized observations has also drawn a lot of attention during the past decade (Casini, Garulli, & Vicino, 2012; Godoy, Goodwin, Agüero, Marelli, & Wigren, 2011; Guo & Zhao, 2013; Wang, Yin, Zhang, & Zhao, 2010; Wang, Zhang, & Yin, 2003; Wigren, 1998).

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Wei Xing Zheng 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).

http://dx.doi.org/10.1016/j.automatica.2016.12.034
0005-1098/

Identification of Wiener systems under binary-valued/quantized observations becomes naturally an interesting problem. Zhao, Wang, Yin, and Zhang (2007) presented the first algorithm to this problem. Under scaled full-rank periodic inputs and binary-valued observations, Zhao et al. (2007) showed that the identification of Wiener systems could be decomposed into a finite number of core identification problems. The concept of joint identifiability of the core problem was introduced to capture the essential conditions under which a Wiener system could be identified with binary-valued observations. A strongly convergent algorithm was constructed and proved to be asymptotically efficient for the core identification problems, achieving asymptotic optimality in its convergence rate. The idea and technique developed in Zhao et al. (2007) has also been successfully applied to identification of Hammerstein systems with quantized observations (Zhao, Wang, Yin, & Zhang, 2010).

However, commonly encountered inputs are not necessarily periodic. Input signals often cannot be arbitrarily selected to be periodic (Kang, Zhai, Liu, & Zhao, 2015; Ljung, 1987), and in adaptive control the control input is adjusted in real time and is usually non-periodic (Guo, 1993; He, Zhang, & Ge, 2014). Under both quantized inputs and quantized output observations, Guo, Wang, Yin, Zhao, and Zhang (2015) offered a constructive method to identify finite impulse response (FIR) systems, in which regressor sequences were classified into distinct pattern sets according to their values. It was shown that input–output data could be grouped, without losing any information, on the basis of both quantized output observations and input regressor patterns and used to derive an asymptotically efficient algorithm. This paper extends this idea to identify Wiener systems under quantized inputs and binary-valued output observations.

Different from the identification algorithms for linear systems in Guo et al. (2015), identification of Wiener systems is more complex, mainly because the internal variables between the linear and nonlinear subsystems are unmeasured, making it hard to identify the subsystems individually. In this paper, for identifiable Wiener systems, a three-step identification algorithm is proposed. The first step aims to estimate the output of the nonlinear function by using empirical measures and organize its inputs a finite number of possible values defined as the products of basic persistent patterns and parameters of the linear dynamics. Then the second step estimates the parameters of the nonlinear function and its input values jointly. Finally, the third step estimates the parameters of the linear dynamics. Under some typical assumptions on system order, input persistent excitation, and noise distribution functions, the algorithm is shown to be strongly convergent and asymptotically efficient in terms of the Cramér–Rao (CR) lower bound.

The rest of the paper is organized into the following sections. Section 2 formulates the Wiener systems identification problem with quantized inputs and binary-valued observations. System identifiability under input and output quantization is discussed in Section 3. A three-step identification algorithm is introduced in Section 4 based on empirical measures, persistent patterns, relations between the linear and nonlinear subsystems. Section 5 establishes convergence properties of the algorithm, including strong convergence, mean-square convergence rate, and asymptotic efficiency. A numerical case study is presented in Section 6 to demonstrate effectiveness of the algorithm and the convergence properties. Finally, findings of the paper are summarized in Section 7, together with remarks on some open issues.

2. Problem formulation

Consider a single-input-single-output discrete-time Wiener system described by

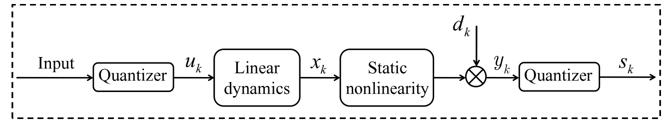

Fig. 1. System configuration.

$$\begin{cases} x_k = \sum_{i=1}^{n} a_i u_{k-i+1} \\ y_k = H(x_k, \eta) + d_k \end{cases} \tag{1}$$

where u_k, x_k and d_k are the input, the intermediate variable, and the system noise, respectively. $H(\cdot, \eta) : \mathcal{D}_H \to \mathbb{R}$ is a parameterized static nonlinear function with domain $\mathcal{D}_H \subseteq \mathbb{R}$ and vector-valued parameter $\eta \in \Omega_\eta \subseteq \mathbb{R}^m$. Both n and m are known. By defining the regressor $\phi_k = [u_k, \ldots, u_{k-n+1}]'$ and $\theta = [a_1, \ldots, a_n]'$, the linear dynamics can be expressed compactly as $x_k = \phi_k'\theta$. Here z' denotes the transpose of $z \in \mathbb{R}^{\iota_1 \times \iota_2}$ for a vector or matrix.

The system structure is shown in Fig. 1, in which the input u_k is quantized and takes a finite number of possible values, $u_k \in \mathcal{U} = \{\mu_1, \ldots, \mu_r\}$. The output y_k is measured by a binary sensor with a finite threshold $C \in \mathbb{R}$, which can be represented by an indicator function

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

Based on $\{u_k\}$ and $\{s_k\}$, this paper will first discuss the issue of identifiability, then design an algorithm to identify θ and η for identifiable systems, and finally establish key convergence properties of the algorithm.

Assumption 2.1. Suppose that $\{d_k\}$ 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 the inverse function denoted by $F^{-1}(\cdot)$ is twice continuously differentiable. The moment generating function of d_1 exists.

Remark 2.1. In this paper, the output quantizer is binary-valued with the threshold C. For multi-threshold quantizers, the reader is referred to Wang et al. (2010) in which a quasi-convex combination technique was introduced to combine information from different thresholds and to achieve asymptotic efficiency. For more general quantizers, Wigren (1998) introduced a stochastic gradient-based adaptive filtering algorithm. Its analysis method with an associated differential equation may be useful for other types of systems.

3. System identifiability

System identification addresses the fundamental issue: Under what conditions, the parameters of a Wiener system can be uniquely determined from its noise-free input–output observations? For identifiable systems, algorithms can then be developed to estimate system parameters under noisy observations.

Suppose that $u = \{u_k, k = 1, 2, \ldots\}$ is an arbitrary input sequence taking quantized values in $\mathcal{U} = \{\mu_1, \ldots, \mu_r\}$. The input u generates a regressor sequence $\{\phi_k'\}$ that takes values in $l = r^n$ possible (row vector) patterns denoted by $\mathcal{P} = \{\pi_1, \ldots, \pi_l\}$. Pattern examples include $\pi_1 = [\mu_1, \ldots, \mu_1, \mu_1]$, $\pi_2 = [\mu_1, \ldots, \mu_1, \mu_2]$, etc.

For a given input sequence u and its corresponding regressor sequence $\{\phi_{n+1}', \ldots, \phi_{n+N}'\}$, denote ($N$-dependent) $N_j = \sum_{i=1}^{N} I_{\{\phi_{n+i}'=\pi_j\}}$, $j \in L = \{1, \ldots, l\}$. That is, $\{\phi_{n+1}', \ldots, \phi_{n+N}'\}$ contains N_j copies of the pattern π_j.

Assumption 3.1. The input sequence u is deterministic. There exists $\beta_j \geq 0$ such that $\lim_{N\to\infty} N_j/N = \beta_j$. If $\beta_j > 0$, then the pattern π_j is said to be *persistent* in u. Without loss of generality, suppose that for the input u under study, $\beta_j \neq 0$ for $j \in L_0 = \{1, \dots, l_0\}$ and $\beta_j = 0$ for $j \in L_0^- = \{l_0+1, \dots, l\}$.

It will become clear that convergence properties depend only on persistent patterns. As a result, the non-persistent patterns $\pi_{l_0+1}, \dots, \pi_l \in L_0^-$ will not be used in designing algorithms.

Denote $w_j = \pi_j\theta$. Then, the input–output mapping of the nonlinear function becomes

$$\begin{cases} H(w_1, \eta) & = & \overline{\alpha}_1, \\ \quad\vdots & & \\ H(w_{l_0}, \eta) & = & \overline{\alpha}_{l_0}, \end{cases} \tag{3}$$

where the output $\overline{\alpha} = [\overline{\alpha}_1, \dots, \overline{\alpha}_{l_0}]' \in \mathbb{R}^{l_0}$ can be derived from observed data and are considered to be known in study of identifiability.

Let

$$\Psi = \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_{l_0} \end{bmatrix}, \tag{4}$$

and assume that Ψ has full column rank (then $l_0 \geq n$), which is a basic persistent excitation condition in identification of FIR systems with quantized inputs and outputs (Guo et al., 2015). Hence, one can always select n patterns from $\pi_1, \dots, \pi_{l_0}$ such that their transposes constitute a basis of $\mathbb{R}^n$. Without loss of generality, let these n patterns be $\pi_1, \dots, \pi_n$, which will be called *the set of basic persistent patterns*.

Consequently, all other persistent patterns can be represented by the basis, $\pi_j' = \sum_{i=1}^n \gamma_{j,i}\pi_i'$, $j = n+1, \dots, l_0$, which implies that

$$w_j = \sum_{i=1}^n \gamma_{j,i} w_i, \quad j = n+1, \dots, l_0.$$

Substituting these into (3), we have

$$\mathbb{H}(w_1, \dots, w_n, \eta) = \overline{\alpha}, \tag{5}$$

with $\mathbb{H}(w_1, \dots, w_n, \eta) = [H(\sum_{i=1}^n \gamma_{1,i}w_i, \eta), \dots, H(\sum_{i=1}^n \gamma_{l_0,i}w_i, \eta)]'$.

Equality (5) contains l_0 equations and $n+m$ unknowns. If the solution to (5) exists and is unique, then η and $W \triangleq [w_1, \dots, w_n]'$ can be obtained. Furthermore, θ can be derived from W since $W = \bar{\Psi}\theta$ and $\bar{\Psi}$ is full rank with $\bar{\Psi} \triangleq [\pi_1', \dots, \pi_n']'$. In this sense, the system (1) is identifiable. To ensure this identifiability in the algorithm design, we give an assumption as follows.

Assumption 3.2. There exists a compact set $\Xi \subseteq \mathbb{R}^{l_0}$ that contains the true output $\overline{\alpha} \in \Xi$ such that for any $\xi = [\xi_1, \dots, \xi_{l_0}]' \in \Xi$, the equations

$$\mathbb{H}(x_1, \dots, x_n, \eta) = \xi \tag{6}$$

have a unique solution $[x_1, \dots, x_n, \eta']' \in \mathbb{R}^{n+m}$, denoted by $\varrho(\xi)$. Moreover, $\varrho(\xi)$ is bounded and continuous in Ξ.

For a given full (row) rank matrix $\Gamma \in \mathbb{R}^{(n+m)\times l_0}$, denote its range from Ξ as $\Xi^0 = \{\zeta : \zeta = \Gamma\xi, \xi \in \Xi\} \subseteq \mathbb{R}^{n+m}$. Let

$$\mathbb{G}(x_1, \dots, x_n, \eta) = \Gamma\mathbb{H}(x_1, \dots, x_n, \eta). \tag{7}$$

Then, for any $\zeta \in \Xi^0$, there exist $\xi \in \Xi$ and $\xi^0 \in \mathbb{R}^{l_0}$ such that $\zeta = \Gamma\xi$ and $\xi = \Gamma^+\zeta + (I - \Gamma^+\Gamma)\xi^0$, where $^+$ represents the Moore–Penrose inverse and I is the identity matrix of suitable dimension. Under Assumption 3.2, $\mathbb{H}(\varrho(\xi)) = \xi$. Hence $\mathbb{G}(\varrho(\xi)) = \Gamma\mathbb{H}(\varrho(\xi)) = \Gamma\xi = \zeta$, which implies that

$$\mathbb{G}\left(\varrho(\Gamma^+\zeta + (I - \Gamma^+\Gamma)\xi^0)\right) = \zeta.$$

Therefore, $\varrho(\Gamma^+\zeta + (I - \Gamma^+\Gamma)\xi^0)$ is a solution of the equation $\mathbb{G}(x_1, \dots, x_n, \eta) = \zeta$. We denote it as $\tau(\zeta) = [\tau_1(\zeta), \dots, \tau_{n+m}(\zeta)]'$, which means that $\tau_i(\zeta) = x_i$ for $i = 1, \dots, n$ and $[\tau_{n+1}(\zeta), \dots, \tau_{n+m}(\zeta)]' = \eta$. By Assumption 3.2, $\alpha = \Gamma\overline{\alpha} \in \Xi^0$ and $\tau(\zeta)$ is bounded and continuous in Ξ^0.

In fact, (7) defines a linear transformation on $\mathbb{H}(x_1, \dots, x_n, \eta)$ by the left multiplication of a full-rank matrix. This transformation has no effect on the existence of solutions to Eq. (6). In the subsequent algorithm design, Γ will be used to improve convergence properties of the algorithm in Section 5.

Remark 3.1. If specific parametric models are considered on the nonlinear function, certain normalization is often needed to ensure that the parameters are independent. For example, consider the nonlinear model $y_k = b_0 + b_1x_k + b_2x_k^2$ with unknown parameters $[b_0, b_1, b_2]$; and the linear dynamic system $x_k = a_1u_{k-1}$. They result in the combined system $y_k = b_0 + b_1a_1u_{k-1} + b_2a_1^2u_{k-1}^2$, where there are only three independent parameters, but four unknowns. Without normalization on the scaling factor, the parameters cannot be uniquely determined from any input–output sequence. As a remedy, one may impose $b_2 = 1$. In this paper, we assume that such normalization has already been included in the model parameterization.

4. Identification algorithm

Under Assumption 3.2, by (5) it is known that

$$\tau(\alpha) = [w_1, \dots, w_n, \eta']' = \begin{bmatrix} W \\ \eta \end{bmatrix} := \vartheta. \tag{8}$$

Since this mapping $\tau(\cdot)$ is known and continuous, one may estimate α first, and then derive estimates for W and η via (8). With this in mind, an identification algorithm is constructed as follows, which is divided into three steps.

Identification Algorithm:

(1) (**Estimate** α). At N (after $N+n$ observations), for $j \in L_0$, by (1) the system outputs under π_j can be described by

$$y_k^j = H(\pi_j\theta, \eta) + d_k^j = H(w_j, \eta) + d_k^j$$

and the corresponding binary-valued observations are denoted by $s_k^j = I_{\{y_k^j \leq C\}}$. Define

$$\overline{S}_N^j = \begin{cases} \dfrac{1}{N_j}\displaystyle\sum_{i=1}^{N_j} s_i^j, & N_j \neq 0; \\ \dfrac{1}{2}, & N_j = 0, \end{cases} \tag{9}$$

and let

$$\begin{aligned} \widehat{\overline{\alpha}}_N &= \left[C - F^{-1}(\overline{S}_N^1), \dots, C - F^{-1}(\overline{S}_N^{l_0})\right]', \\ \widehat{\alpha}_N &= \Gamma\Pi_\Xi\left(\widehat{\overline{\alpha}}_N\right), \end{aligned} \tag{10}$$

where $\Pi_\Xi(z)$ is a projection from z to Ξ and can be any one in $\{\xi \in \Xi : \|\xi - z\| = \min_{\nu\in\Xi}\|\nu - z\|\}$ for $z \in \mathbb{R}^{l_0}$, and $\|\cdot\|$ is a vector norm.

(2) (**Estimate** ϑ). Under Assumption 3.2, an estimate of ϑ denoted by $\widehat{\vartheta}_N$ can be derived by

$$\widehat{\vartheta}_N = \tau(\widehat{\alpha}_N). \tag{11}$$

By using component-wise extraction, an estimate of W denoted by $\widehat{W}_N$ and an estimate of η denoted by $\widehat{\eta}_N$ can be expressed as

$$\widehat{W}_N = [\tau_1(\widehat{\alpha}_N), \ldots, \tau_n(\widehat{\alpha}_N)]', \tag{12}$$

$$\widehat{\eta}_N = [\tau_{n+1}(\widehat{\alpha}_N), \ldots, \tau_{n+m}(\widehat{\alpha}_N)]'. \tag{13}$$

(3) (**Estimate** θ). Considering that $\widehat{W}_N = W + e_N = \bar{\Psi}\theta + e_N$ with the estimation error $e_N = \widehat{W}_N - \bar{\Psi}\theta$, an estimate of θ can be constructed as

$$\widehat{\theta}_N = \bar{\Psi}^{-1}\widehat{W}_N. \tag{14}$$

5. Convergence properties

This section establishes key convergence properties of the identification algorithm, including strong convergence, mean-square convergence rate, and asymptotic efficiency.

5.1. Strong convergence

Theorem 5.1. *Consider system (1) with binary-valued observations (2). If Ψ given by (4) is full column rank, and Assumptions 2.1, 3.1, and 3.2 hold, then $\widehat{\vartheta}_N$ from (11) converges strongly to the true ϑ,*

$$\widehat{\vartheta}_N \to \vartheta \ \text{w.p.1} \ \text{as} \ N \to \infty.$$

Proof. By the strong law of large numbers and (9), we have $\bar{S}_N^j \to F\left(C - H(w_j, \eta)\right)$, w.p.1. as $N \to \infty$, which together with Assumption 2.1 implies that

$$C - F^{-1}(\bar{S}_N^j) \to H(w_j, \eta) \ \text{w.p.1.} \ \text{as} \ N \to \infty, \quad j \in L_0.$$

From (9) and (10), it follows that $\widehat{\alpha}_N \to \alpha$ w.p.1 as $N \to \infty$. Since $\tau(\zeta)$ is continuous in Ξ^0, $\widehat{\vartheta}_N = \tau(\widehat{\alpha}_N) \to \tau(\alpha) = \vartheta$ by (8). This completes the proof. □

Theorem 5.2. *Under the conditions of Theorem 5.1, $\widehat{\theta}_N$ from (14) converges strongly to the true θ,*

$$\widehat{\theta}_N \to \theta \ \text{w.p.1} \ \text{as} \ N \to \infty.$$

Proof. By virtue of Theorem 5.1, we have $\widehat{W}_N \to W = \bar{\Psi}\theta$, w.p.1. as $N \to \infty$, which together with (14) yields that

$$\widehat{\theta}_N = \bar{\Psi}^{-1}\widehat{W}_N \to \bar{\Psi}^{-1}\bar{\Psi}\theta = \theta, \ w.p.1 \text{ as } N \to \infty. \quad \square$$

5.2. Asymptotic efficiency

For convenience, denote $F^d(x) = dF(x)/dx$, $F_{CH}(x) = F(C - H(x, \eta))$, $F_{CH}^d(x) = F^d(C - H(x, \eta))$,

$$\Lambda = \operatorname{diag}\left[\frac{\sqrt{F_{CH}(w_j)(1 - F_{CH}(w_j))}}{F_{CH}^d(w_j)}\right]_{j=1,\ldots,l_0}, \tag{15}$$

$\tau_j^d(\zeta) = \frac{\partial \tau_j(\zeta)}{\partial \zeta} = \left[\frac{\partial \tau_j(\zeta)}{\partial \zeta_1}, \ldots, \frac{\partial \tau_j(\zeta)}{\partial \zeta_{n+m}}\right]'$, and $J = [\tau_1^d(\alpha), \ldots, \tau_{n+m}^d(\alpha)]$, also define

$$B = \frac{\partial \mathbb{G}}{\partial \vartheta}, \qquad D = \frac{\partial \mathbb{H}}{\partial \vartheta}, \tag{16}$$

where $\operatorname{diag}[z_j]_{j=1,\ldots,\iota} = \operatorname{diag}[z_1, \ldots, z_\iota]$ is a diagonal matrix. Let $\Sigma(N; \vartheta)$ represent the covariance matrix of the estimation error of $\widehat{\vartheta}_N$, i.e.,

$$\Sigma(N; \vartheta) = E(\widehat{\vartheta}_N - \vartheta)(\widehat{\vartheta}_N - \vartheta)', \quad N = 1, 2, \ldots,$$

where $E(\cdot)$ is the expectation.

Lemma 5.1. *If $\mathbb{H}(x_1, \ldots, x_n, \eta)$ is differentiable at ϑ, then the Cramér–Rao lower bound for estimating ϑ based on observations of $\{s_k, 1 \le k \le N\}$ is*

$$\Sigma_{CR}(N; \vartheta) = \left(\sum_{j=1}^{l} \frac{N_j (F_{CH}^d(w_j))^2 \frac{\partial H(w_j, \eta)}{\partial \vartheta} \left(\frac{\partial H(w_j, \eta)}{\partial \vartheta}\right)'}{F_{CH}(w_j)(1 - F_{CH}(w_j))}\right)^{-1}. \tag{17}$$

Proof. Let z_k be some possible sample value of s_k. Since $\{d_k\}$ is i.i.d, the likelihood function of $s_1, \ldots, s_N$ taking values $z_1, \ldots, z_N$ conditioned on ϑ is

$$\begin{aligned}
&\ell(z_1, \ldots, z_N; \vartheta) \\
&\quad = \Pr\{s_1 = z_1, \ldots, s_N = z_N; \vartheta\} \\
&\quad = \prod_{k=1}^{N} [F(C - H(x_k, \eta))]^{z_k} [1 - F(C - H(x_k, \eta))]^{1 - z_k} \\
&\quad = \prod_{k=1}^{N} [F_{CH}(x_k)]^{z_k} [1 - F_{CH}(x_k)]^{1 - z_k}.
\end{aligned}$$

Replace the particular realizations z_k by their corresponding random variables s_k, and denote the resulting quantity by $\ell = \ell(s_1, \ldots, s_N; \vartheta)$. Set $M_N^j = \{k : \phi_k' = \pi_j, 1 \le k \le N\}$ and $\chi_j = \frac{1}{N_j} \sum_{k \in M_N^j} s_k$. It is apparent that $E\chi_j = F_{CH}(w_j)$. Then, we have

$$\begin{aligned}
\ell &= \prod_{j=1}^{l} \prod_{k \in M_N^j} [F_{CH}(w_j)]^{s_k} \left[1 - F_{CH}(w_j)\right]^{1 - s_k} \\
&= \prod_{j=1}^{l} \left[F_{CH}(w_j)\right]^{N_j \chi_j} \left[1 - F_{CH}(w_j)\right]^{N_j - N_j \chi_j},
\end{aligned}$$

which leads to $\log \ell = \sum_{j=1}^{l} \left\{N_j \chi_j \log F_{CH}(w_j) + (N_j - N_j \chi_j) \log[1 - F_{CH}(w_j)]\right\}$ and

$$\begin{aligned}
\frac{\partial \log \ell}{\partial \vartheta} = \sum_{j=1}^{l} \Bigg\{ &- \frac{N_j \chi_j F_{CH}^d(w_j)}{F_{CH}(w_j)} \frac{\partial H(w_j, \eta)}{\partial \vartheta} \\
&+ \frac{(N_j - N_j \chi_j) F_{CH}^d(w_j)}{1 - F_{CH}(w_j)} \frac{\partial H(w_j, \eta)}{\partial \vartheta} \Bigg\}.
\end{aligned}$$

Consequently, it can be verified that

$$E \frac{\partial^2 \log \ell}{\partial \vartheta^2} = - \sum_{j=1}^{l} \frac{N_j (F_{CH}^d(w_j))^2 \frac{\partial H(w_j, \eta)}{\partial \vartheta} \left(\frac{\partial H(w_j, \eta)}{\partial \vartheta}\right)'}{F_{CH}(w_j)(1 - F_{CH}(w_j))},$$

and (17) follows. □

Remark 5.1. The assumption that the input u is deterministic simplifies analysis. If the input u is stochastic, the regressor sequence $\{\phi_k\}$ will be dependent, unless the system is just a gain. As a result, $\{s_k\}$ becomes dependent even if $\{u_k\}$ is i.i.d. This makes it difficult to derive the joint distribution of $\{s_k\}$ or obtain the CR lower bound for estimating ϑ.

Theorem 5.3. *Under the conditions of Theorem 5.1 and Lemma 5.1, if $\tau(\zeta)$ is differentiable at α and D in (16) has full row rank, then the estimate $\widehat{\vartheta}_N$ from (11) has the mean-square convergence rate*

$$N\Sigma(N; \vartheta) \to (\Gamma D')^{-1} \Gamma \Upsilon^2 \Lambda^2 \Gamma' (D\Gamma')^{-1} \quad \text{as } N \to \infty,$$

where $\Upsilon = \operatorname{diag}[1/\sqrt{\beta_1}, \ldots, 1/\sqrt{\beta_{l_0}}]$ and Λ is given by (15).

Proof. By the mean value theorem, from (8) and (11) there exist $\widetilde{\alpha}_N^1, \ldots, \widetilde{\alpha}_N^{n+m}$ on the line segment $\widehat{\alpha}_N$ and α such that

$$\widehat{\vartheta}_N - \vartheta = \left[\tau_1^d(\widetilde{\alpha}_N^1), \ldots, \tau_{n+m}^d(\widetilde{\alpha}_N^{n+m})\right]' (\widehat{\alpha}_N - \alpha). \tag{18}$$

By Wang et al. (2010), it can be concluded that $\sqrt{N_j}(C - F^{-1}(\bar{S}_N^j) - H(w_j, \eta)) \xrightarrow{d} \mathcal{N}\left(0, \frac{F_{CH}(w_j)(1-F_{CH}(w_j))}{(F_{CH}^d(w_j))^2}\right)$, which implies that

$$\Upsilon_N \begin{bmatrix} C - F^{-1}(\bar{S}_N^1) - H(w_1, \eta) \\ \vdots \\ C - F^{-1}(\bar{S}_N^{l_0}) - H(w_{l_0}, \eta) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Lambda^2\right)$$

as $N \to \infty$, where $\Upsilon_N = \operatorname{diag}[\sqrt{N_1}, \ldots, \sqrt{N_{l_0}}]$ and $\xrightarrow{d}$ denotes convergence in distribution. In the light of Assumption 3.1, we have $\Upsilon_N/\sqrt{N} \to \Upsilon^{-1}$ as $N \to \infty$. By (10), one can get

$$\sqrt{N}\,(\widehat{\alpha}_N - \alpha) \xrightarrow{d} \mathcal{N}\left(0, \Gamma\Upsilon^2\Lambda^2\Gamma'\right) \quad \text{as } N \to \infty. \tag{19}$$

Note that $\widetilde{\alpha}_N^j \to \alpha$ w.p.1. as $N \to \infty$, by (7) and (18)–(19) it can be seen that

$$\begin{aligned} &N\Sigma(N; \vartheta) \\ &\quad \to [\tau_1^d(\alpha), \ldots, \tau_{n+m}^d(\alpha)]'\Gamma\Upsilon^2\Lambda^2\Gamma'[\tau_1^d(\alpha), \ldots, \tau_{n+m}^d(\alpha)] \\ &\quad = J'\Gamma\Upsilon^2\Lambda^2\Gamma' J, \quad \text{as } N \to \infty. \end{aligned} \tag{20}$$

In view of (7), one can have $B = D\Gamma'$. Since both D and Γ have full row ranks, B is full rank. Noticing that $\mathbb{G}(\tau(\alpha)) = \alpha$ and $\tau(\alpha) = \vartheta$, we have $BJ = JB = I$ by Theorem 12 in Magnus and Neudecker (2007). Therefore, $J = (D\Gamma')^{-1}$ and $J' = ((D\Gamma')^{-1})' = (\Gamma D')^{-1}$, which together with (20) implies that

$$N\Sigma(N; \vartheta) \to (\Gamma D')^{-1}\Gamma\Upsilon^2\Lambda^2\Gamma'(D\Gamma')^{-1} \quad \text{as } N \to \infty,$$

as claimed. □

Theorem 5.4. *Under the conditions of Theorem 5.3, if one selects*

$$\Gamma = \Gamma^* = \left(D\Upsilon^{-2}\Lambda^{-2}D'\right)^{-1} D\Upsilon^{-2}\Lambda^{-2}, \tag{21}$$

then the estimate $\widehat{\vartheta}_N$ from (11) is asymptotically efficient in the sense that

$$N\Sigma(N; \vartheta) - N\Sigma_{CR}(N; \vartheta) \to 0 \quad \text{as } N \to \infty.$$

Proof. Since $\Upsilon, \Lambda > 0$ and D is full row rank, $D\Upsilon^{-2}\Lambda^{-2}D' > 0$. Under the hypothesis, $\Gamma^*D' = \left(D\Upsilon^{-2}\Lambda^{-2}D'\right)^{-1} D\Upsilon^{-2}\Lambda^{-2}D' = I$ by (21). Consequently, we have

$$\begin{aligned} &(\Gamma^*D')^{-1}\Gamma^*\Upsilon^2\Lambda^2(\Gamma^*)'(D(\Gamma^*)')^{-1} \\ &\quad = \Gamma^*\Upsilon^2\Lambda^2(\Gamma^*)' \\ &\quad = \left(D\Upsilon^{-2}\Lambda^{-2}D'\right)^{-1} D\Upsilon^{-2}\Lambda^{-2}D'((D\Upsilon^{-2}\Lambda^{-2}D')^{-1})' \\ &\quad = \left(D\Upsilon^{-2}\Lambda^{-2}D'\right)^{-1}. \end{aligned} \tag{22}$$

By Lemma 5.1, it can be seen that

$$\begin{aligned} N\Sigma_{CR}(N; \vartheta) \to &\left(\left[\frac{\partial H(w_1, \eta)}{\partial \vartheta}, \ldots, \frac{\partial H(w_{l_0}, \eta)}{\partial \vartheta}\right]\Upsilon^{-2}\Lambda^{-2}\right. \\ &\quad \left.\times \left[\frac{\partial H(w_1, \eta)}{\partial \vartheta}, \ldots, \frac{\partial H(w_{l_0}, \eta)}{\partial \vartheta}\right]'\right)^{-1} \\ &= \left(D\Upsilon^{-2}\Lambda^{-2}D'\right)^{-1} \quad \text{as } N \to \infty. \end{aligned}$$

This and (22) prove the theorem by virtue of Theorem 5.3. □

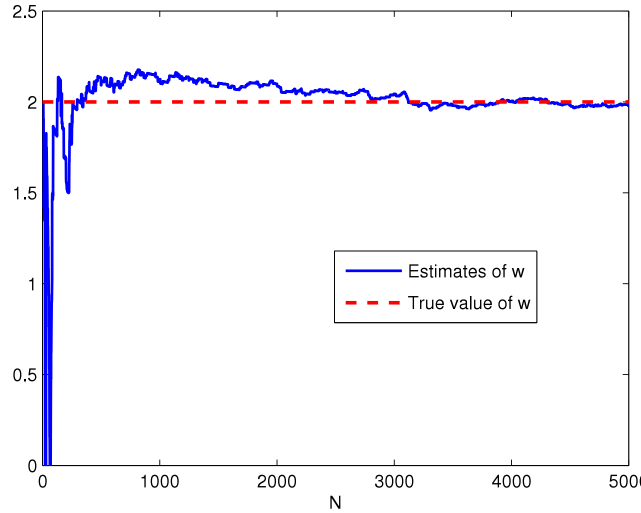

Fig. 2. Convergence of $\widehat{W}_N$ from (12).

6. Simulation example

Consider a Wiener system, in which the linear dynamics is a gain system and the output nonlinearity is an exponential function $H(x, \eta) = 2^x + \eta$,

$$\begin{cases} x_k = u_k\theta, \\ y_k = H(x_k, \eta) + d_k = 2^{x_k} + \eta + d_k, \end{cases}$$

where the true values are $\theta = 20$, $\eta = 30$ and $\{d_k\}$ is a sequence of i.i.d. normal random variables with zero mean and standard deviation $\sigma = 5$. The output y_k is measured by a sensor with threshold $C = 39$, and hence $s_k = I_{\{y_k \le 39\}}$. The input u_k is quantized and takes values from $\mathcal{U} = \{\mu_1, \mu_2, \mu_3, \mu_4\} = \{0.1, 0.2, 3, 5\}$. Since $\theta \in \mathbb{R}$, we have $\mathcal{P} = \mathcal{U}$. Suppose that at step N, the input sequence generates patterns with the following frequencies

$$\begin{aligned} &N_1 = N - N_2 - N_3 - N_4, \quad N_2 = \lceil 0.6(N - N_3 - N_4) \rceil, \\ &N_3 = \min\{110, |\lceil \log N \rceil|\}, \quad N_4 = \lceil \sqrt{N} \rceil, \end{aligned}$$

where $\lceil z \rceil$ denotes the smallest integer greater than or equal to $z \in \mathbb{R}$. As a result, $\beta_1 = \lim_{N\to\infty} N_1/N = 0.4$, $\beta_2 = 0.6$, $\beta_3 = \beta_4 = 0$; and $\Psi = [1, 2]'$, $[w_1, w_2]' = [2, 4]'$, $\bar{\alpha} = [\bar{\alpha}_1, \bar{\alpha}_2]' = [34, 46]'$ by (3) and (4). Let π_1 be the basic persistent pattern. Then $\pi_2 = 2\pi_1$, $w_2 = 2w_1$ and $W = w_1 = 2$.

Let $\Gamma = I_{2\times 2}$ and $\Xi^0 = \Xi = \{[z_1, z_2]' \in \mathbb{R}^2 : 28 \le z_1 \le 40, 40 \le z_2 \le 52\}$. It follows that $\alpha = \bar{\alpha} \in \Xi$. For any $\xi = [\xi_1, \xi_2]' \in \Xi$, it can be derived that the following equations

$$\mathbb{H}(x_1, \eta) = \begin{bmatrix} H(x_1, \eta) \\ H(2x_1, \eta) \end{bmatrix} = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}$$

have a unique solution $\tau_1(\xi) = x_1 = \log_2\left(\frac{1+\sqrt{1+4(\xi_2-\xi_1)}}{2}\right)$ and $\tau_2(\xi) = \eta = \frac{2\xi_1 - 1 - \sqrt{1+4(\xi_2-\xi_1)}}{2}$, indicating that Assumption 3.2 holds.

Using (12) and (13) to compute $\widehat{W}_N$ and $\widehat{\eta}_N$, the convergence is shown by Figs. 2 and 3. Furthermore, Fig. 4 demonstrates the convergence of $\widehat{\theta}_N$ given by (14).

By $H(x, \eta) = 2^x + \eta$, one can get

$$D = \left[\frac{\partial H(w_1, \eta)}{\partial \vartheta}, \frac{\partial H(w_2, \eta)}{\partial \vartheta}\right] = \begin{bmatrix} 4\ln 2 & 32\ln 2 \\ 1 & 1 \end{bmatrix},$$

which illustrates that D is full rank. Since $l_0 = n + m = 2$, $\Gamma^* = \left(D\Upsilon^{-2}\Lambda^{-2}D'\right)^{-1} D\Upsilon^{-2}\Lambda^{-2} = (D')^{-1}$. With $\Gamma = \Gamma^* = \frac{1}{28\ln 2}\begin{bmatrix} -1 & 1 \\ 32\ln 2 & -4\ln 2 \end{bmatrix}$, Theorem 5.4 is true and hence $\widehat{\vartheta}_N$ is asymptotically efficient, which is shown by Fig. 5.

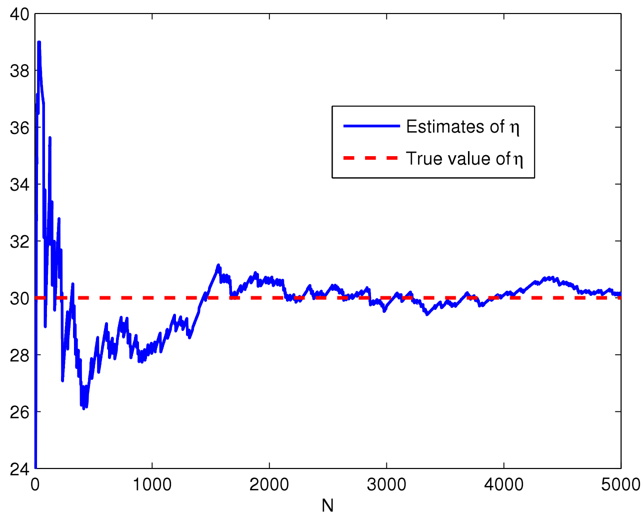

Fig. 3. Convergence of $\widehat{\eta}_N$ from (13).

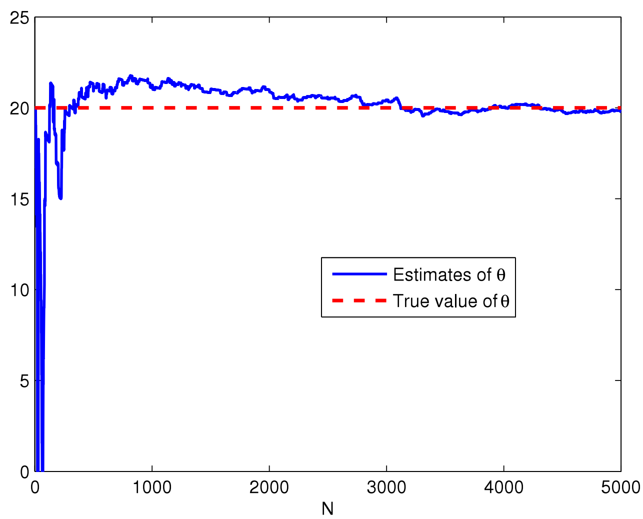

Fig. 4. Convergence of $\widehat{\theta}_N$ from (14).

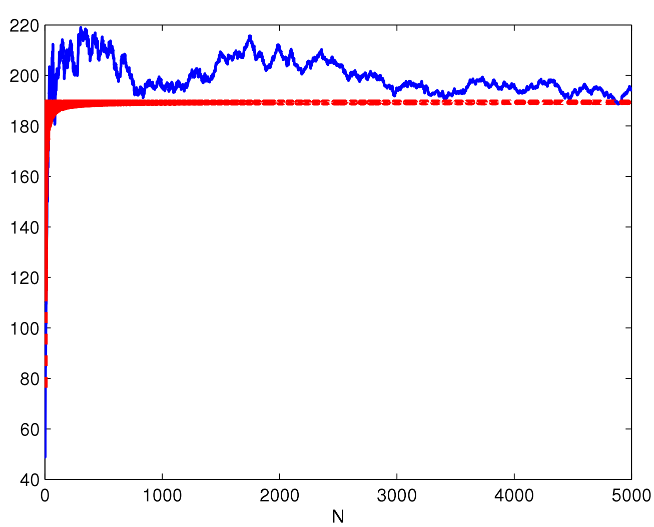

Fig. 5. Asymptotic efficiency of $\widehat{\vartheta}_N$ from (11): The red dash line is the Frobenius norm of $N\Sigma_{CR}(N;\vartheta)$ and the blue solid line comes from the average of 100 trajectories of the Frobenius norm of $N(\widehat{\vartheta}_N-\vartheta)(\widehat{\vartheta}_N-\vartheta)'$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

7. Concluding remarks

This paper studies identification of Wiener systems under quantized inputs and binary-valued output observations. After establishing identifiability conditions, a three-step algorithm is introduced to estimate unknown parameters. The algorithm is shown to be strongly convergent and asymptotically efficient in terms of the CR lower bound. The results of this paper can be extended to more general cases of multi-threshold quantized observations. The method can potentially be extended to identify Hammerstein systems and other nonlinear systems under quantized inputs and quantized observations.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (61403027), in part by the Fundamental Research Funds for the Central Universities (FRF-TP-15-071A1), in part by the Open Research Project under Grant 20160105 from SKLMCCS, in part by the U.S. Army Research Office under Grant W911NF-15-1-0218, and in part by the National Key Basic Research Program of China (973 Program) under Grant 2014CB845301.

References

Boyd, S., & Chua, L. O. (1985). Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Transactions on Circuits and Systems*, *32*, 1150–1161.
Casini, M., Garulli, A., & Vicino, A. (2012). Input design in worst-case system identification with quantized measurements. *Automatica*, *48*(12), 2997–3007.
Chen, H., & Zhao, W. (2014). *Recursive identification and parameter estimation*. CRC Press.
Giri, F., & Bai, E. W. (2010). *Block-oriented nonlinear system identification*. Springer.
Godoy, B., Goodwin, G., Agüero, J., Marelli, D., & Wigren, T. (2011). On identification of FIR systems having quantized output data. *Automatica*, *47*(9), 1905–1915.
Greblicki, W. (1997). Nonparametric approach to Wiener system identification. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, *44*(6), 538–545.
Guo, L. (1993). *Time-varying stochastic systems—stability, estimation and control*. Changchun, PR China: Jilin Science and Technology Press.
Guo, J., Wang, L. Y., Yin, G., Zhao, Y., & 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. (2013). Recursive projection algorithm on FIR system identification with binary-valued observations. *Automatica*, *49*, 3396–3401.
Hagenblad, A., Ljung, L., & Wills, A. (2008). Maximum likelihood identification of Wiener models. *Automatica*, *44*(11), 2697–2705.
He, W., Zhang, S., & Ge, S. S. (2014). Adaptive control of a flexible crane system with the boundary output constraint. *IEEE Transactions on Industrial Electronics*, *61*(8), 4126–4133.
Hunter, I. W., & Korenberg, M. J. (1986). The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics*, *55*, 135–144.
Kalafatis, A., Arifin, N., Wang, L., & Cluett, W. R. (1995). A new approach to the identification of pH processes based on the Wiener model. *Chemical Engineering Science*, *50*(23), 3693–3701.
Kang, Y., Zhai, D. H., Liu, G. P., & Zhao, Y. B. (2015). On input-to-state stability of switched stochastic nonlinear systems under extended asynchronous switching. *IEEE Transactions on Cybernetics*, *46*(5), 1092–1105.
Ljung, L. (1987). *System identification: theory for the user*. Englewood Cliffs, NJ: Prentice-Hall.
Magnus, J. R., & Neudecker, H. (2007). *Matrix differential calculus with applications in statistics and econometrics* (3rd ed.). John Wiley & Sons.
Shen, J., Tan, H., Wang, J., Wang, J. W., & Lee, S. (2015). A novel routing protocol providing good transmission reliability in underwater sensor networks. *Journal of Internet Technology*, *16*(1), 171–178.
Wang, D., & Ding, F. (2011). Least squares based and gradient based iterative identification for Wiener nonlinear systems. *Signal Processing*, *91*(5), 1182–1189.
Wang, L. Y., Yin, G., Zhang, J. F., & Zhao, Y. (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.
Wigren, T. (1998). Adaptive filtering using quantized output measurements. *IEEE Transactions on Signal Processing*, *46*(12), 3423–3426.
Wills, A., Schon, T., Ljung, L., & Ninness, B. (2011). Identification of Hammerstein-Wiener models. *Automatica*, *1*(1), 1–14.
Xie, S., & Wang, Y. (2014). Construction of tree network with limited delivery latency in homogeneous wireless sensor networks. *Wireless Personal Communications*, *78*(1), 231–246.
Zhao, Y., Wang, L. Y., Yin, G., & Zhang, J. F. (2007). Identification of Wiener systems with binary-valued output observations. *Automatica*, *43*, 1752–1765.

Zhao, Y., Wang, L. Y., Yin, G., & Zhang, J. F. (2010). Identification of Hammerstein systems with quantized observations. *SIAM Journal on Control and Optimization*, *48*(7), 4352–4376.

Zhu, Y. (1999). Distillation column identification for control using Wiener model. In *Proceding of 1999 American control conference*, Hyatt Regency San Diego, California, USA.

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. He is currently an associate professor in the School of Automation and Electrical Engineering, University of Science and Technology Beijing. His 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 serves on the IFAC Technical Committee on Modeling, Identification and Signal Processing. He was an Associate Editor of the IEEE Transactions on Automatic Control and several other journals. He was a Visiting Faculty at University of Michigan in 1996 and Visiting Faculty Fellow at University of Western Sydney in 2009 and 2013. He is a member of the Core International Expert Group at Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and an International Expert Adviser at Beijing Jiao Tong University. 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 served as a member of the program committee for many IEEE Conference on Control and Decision, severed on the IFAC Technical Committee on Modeling, Identification and Signal Processing. He was Co-Chair of SIAM Conference on Control & Its Application in 2011, and Co-Chair of two AMS-IMS-SIAM Summer Research Conferences. He also chaired a number SIAM prize committees, was Chair of SIAM Activity Group on Control and Systems Theory, and served on the Board of Directors of American Automatic Control Council. He is an associate editor of SIAM Journal on Control and Optimization, and 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 IFAC, 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 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, and now is the Director of ISS. His current research interests include system modeling, adaptive control, stochastic systems, and multi-agent systems.

He is a Fellow of IFAC, a Fellow of IEEE, and an Academician of the International Academy for Systems and Cybernetic Sciences. He is the Editor-in-Chief of "Journal of Systems Science and Mathematical Sciences", the Founding Editor-in-Chief of "All About Systems and Control", Deputy Editor-in-Chief of "Science China: Information Sciences" and "Systems Engineering—Theory and Practice". He was associate editor of a number of other journals, including the "SIAM Journal on Control and Optimization" and "IEEE Transactions on Automatic Control".

He is a Vice-Chair of the IFAC Technical Board, Vice President of the Systems Engineering Society of China, Vice President of the Chinese Association of Automation, and was an appointed member of the Board of Governors of the IEEE Control Systems Society. He twice received the Second Prize of the State Natural Science Award of China in 2010 and 2015, respectively, the Excellent Chinese Doctoral Dissertation Supervisor in 2009, and the Distinguished Young Scholar Fund from National Natural Science Foundation of China in 1997. He has served as General Co-Chair, IPC Chair/Co-Chair/Vice-Chair for many control conferences, including the 33rd Chinese Control Conference, the 17th IFAC Symposium on System Identification, and the 20th IFAC World Congress.