**Proceedings of the 17th World Congress**
**The International Federation of Automatic Control**
**Seoul, Korea, July 6-11, 2008**

# Identification of Hammerstein Systems with Set-Valued Observations ⋆

**Yanlong Zhao** *

* *Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China (Tel: 86-10-62651446; e-mail: ylzhao@amss.ac.cn).*

**Abstract:** This work is concerned with identification of Hammerstein systems whose outputs are measured by set-valued sensors. The system consists of a memoryless nonlinearity which is polynomial and possibly non-invertible, followed by a linear subsystem. The parameters of linear and nonlinear parts are unknown but have known orders. Input design, identification algorithms, and their essential properties are presented under the assumptions that the distribution function of the noise is known and the threshold values of set-valued sensors are known. The concept of strongly scaled full rank signals is introduced to capture the essential conditions under which the Hammerstein system can be identified with set-valued observations. Under strongly scaled full rank conditions, a strongly convergent algorithm is constructed. The unbias and efficient properties of the algorithm are investigated.

Keywords: Identification, set-valued observation, Hammerstein system, parameter estimation, sensor threshold, strongly scaled full rank signal.

## 1. INTRODUCTION

Set-valued sensors are commonly employed in practical systems since they are more cost effective than regular sensors. In some applications, they are the only ones available during real-time operations (Wang et al, 2003). More importantly, set-valued observations are the fundamental building blocks for quantized observations that are integrated parts of communication channels. Consequently, understanding system identification under set-valued observations is essential for studying identification of systems involving communication channels.

Set-valued observations supply very limited information on the system outputs, and hence introduce difficulties in system identification. Classical system identification methods, such as least-square algorithms, maximum likelihood methods, etc., assume that the output is measured by a linear sensor and construct estimation algorithms accordingly. However, the information from set-valued observations contains only a finite number of possible values, making it necessary to develop new methodologies and algorithms, and to ensure convergence of estimates.

The first comprehensive treatment on identification with set-valued observations was presented in Wang et al (2003). Based on full rank periodic inputs and empirical distribution, Wang et al (2003) investigated identification errors, time complexity, input design, and impact of disturbances and unmodeled dynamics on identification accuracy and complexity for linear systems that are modeled by impulse responses with binary-valued observations. The work was extended to rational models and unknown noise distributions in Wang et al (2006). Recently, the methodologies have been extended to system identification with quantized observations in Wang & Yin (2007). Most significantly, the optimality of the identification algorithms has been established by showing the Cramér-Rao lower bound is asymptotically achieved (Wang & Yin, 2007). The work on nonlinear systems with set-valued observations started with Wiener systems in Zhao et al (2007). Comparing to Wang et al (2003), the main difficulty of Wiener system identification is how to deal with the nonlinearity. The idea of scaled full rank signals was employed to overcome this difficulty. It was shown that under scaled full rank signals, identification of unknown parameters can be transformed, in an invertible mapping, into a number of simplified core identification problems involving certain intermediate variables, which are solved with the methods of Wang et al (2003).

This paper studies identification of Hammerstein systems whose outputs are measured by set-valued sensors. Hammerstein systems consist of a static nonlinear block followed by a linear dynamic system. They represent typically, but certainly not limited to, linear systems with memoryless nonlinear actuators. Consequently, this paper deals with identification of systems with both nonlinear actuators and nonlinear sensors. In other words, we are in fact dealing with Hammerstein-Wiener systems. When the output of a Hammerstein system must be measured by a set-valued sensor or sent through a communication channel, it can be represented as a Hammerstein system with set-valued observations. Consequently, understanding identification of Hammerstein systems with set-valued observations will be essential for studying both identifica-

⋆ This work was supported by the National Natural Science Foundation of China and the Knowledge Innovation Program of the Chinese Academy of Sciences.

tion of nonlinear systems and impact of communication channels on system models.

Identification of Hammerstein systems, together with Wiener systems, has been studied extensively under linear sensors. Identification methodologies used for Hammerstein/Wiener structures may be loosely classified by iterative algorithms (Hunter & Korenberg, 1986; Korenberg & Hunter, 1998), correlation techniques (Billings, 1980), stochastic recursive algorithms (Chen, 2006; Hu & Chen, 2005), least-squares estimation and singular value decomposition methods (Bai, 1998; Lacy & Bernstein, 2002), Frequency-domain identification methods (Bai, 1998, Ninness & Gibson, 2002) et al. All these approaches require output measurements by linear sensors. In previous literatures on identification of Hammerstein-Wiener systems, the nonlinearities are required to be smooth (Bai, 1998, 2002). However, the switching nonlinearity whose output takes only a finite number of values is fundamentally different from smooth nonlinearities which can provide much more information.

The work in this paper is of essential difference from previous research on Hammerstein system identification, mainly because the output is measured by set-valued sensors, which introduces substantial difficulty in constructing convergent identification algorithms. Technically, this work is more difficult than identification of Wiener systems with set-valued observations, due to much more involved input design. In essence, a periodic full rank input may lose its rank after passing through the input nonlinearity, rendering inapplicable the main ideas of Zhao et al (2007) for identifying Wiener systems.

In this paper, the nonlinearities of Hammerstein systems are polynomial and possibly non-invertible. Polynomial functions are commonly applied as the nonlinearities of Hammerstein systems in current literatures, and polynomial functions can be considered as approximations of some other nonlinear functions. As for polynomial nonlinearities, the concept of strongly scaled full rank input is developed, under which the system parameters can be estimated with the set-valued observations. It is shown that the parameters of linear part are estimated firstly, based on which the nonlinearity is identified.

The remaining part of the paper is organized as follows. The structure of Hammerstein models using set-valued observations is formulated in Section 2. The concepts of strongly scaled full rank signals and their essential properties are introduced in Section 3. Under strongly full rank inputs, algorithms based on each threshold for identifying the variances after the nonlinearity are constructed in Section 4. Then convex combination ones are derived based on all thresholds in Section 5. The identification algorithms for Hammerstein systems are shown to be strongly convergent (in the sense of convergence with probability one). Their efficiency is also investigated. Section 6 develops identification algorithms for the parameters of both linear part and nonlinear part. Examples are presented in Section 7 to illustrate input design, identification algorithms, and convergence results of the methodologies discussed in this paper. Finally, Section 8 provides a brief summary of the findings of this paper.

## 2. PROBLEM FORMULATION

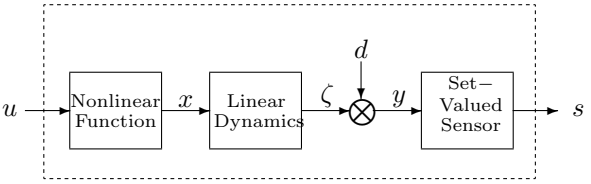


Fig. 1. Hammerstein systems with set-valued observations

Consider the system in Figure 1, in which [1]

$$\begin{cases} y(k) = \sum_{i=0}^{n-1} a_i x(k-i) + d(k), \\ x(k) = b_0 + \sum_{j=1}^{m} b_j u^j(k), \quad b_m = 1, \end{cases} \tag{1}$$

where $u(k)$ is the input, $x(k)$ the intermediate variable, and $d(k)$ the measurement noise. Both $n$ and $m$ are known.

The output $y(k)$ is measured by a sensor of $l$ thresholds $C_1 < \cdots < C_l$. The sensor is represented by

$$s_i(k) = \mathcal{S}(y(k), C_i) = I_{\{y(k)\le C_i\}}, \quad i = 1, \ldots, l,$$

where

$$I_{\{y(k)\in A\}} = \begin{cases} 1, & \text{if } y(k) \in A, \\ 0, & \text{otherwise.} \end{cases}$$

Denote $\theta = [a_0, \ldots, a_{n-1}]^T$, $\phi_0(k) = [1, \ldots, 1]^T$ and $\phi_j(k) = [u^j(k), \ldots, u^j(k-n+1)]^T$, $j = 1, \ldots, m$. Then

$$y(k) = \sum_{j=0}^{m} b_j \phi_j^T(k)\theta + d(k). \tag{2}$$

By using the vector notation, for $k = 1, 2, \ldots$, $Y(k) = [y(2(k-1)(m+1)n+n), \ldots, y(2k(m+1)n+n-1)]^T \in \mathbb{R}^{2n(m+1)}$, $\Phi_j(k) = [\phi_j(2(k-1)(m+1)n+n), \ldots, \phi_j(2k(m+1)n+n-1)]^T \in \mathbb{R}^{2n(m+1)\times n}$, $j = 0, \ldots, m$, $D(k) = [d(2(k-1)(m+1)n+n), \ldots, d(2k(m+1)n+n-1)]^T \in \mathbb{R}^{2n(m+1)}$, $S_i(k) = [s_i(2(k-1)(m+1)n+n), \ldots, s_i(2k(m+1)n+n-1)]^T \in \mathbb{R}^{2n(m+1)}$, $i = 1, \ldots, l$, we can rewrite (2) as

$$Y(l) = \sum_{j=0}^{m} b_j \Phi_j(l)\theta + D(l). \tag{3}$$

The purpose of this paper is to develop identification algorithms of parameters $\theta$ and $\eta = [b_0, \ldots, b_{m-1}]^T$ with the information of the input $u$ and the output of set-valued sensor $s$.

The input signal, that will be used to identify the system, is a $2n(m+1)$-periodic signal $u$ whose one-period values are

$$(v, v, \rho_1 v, \rho_1 v, \ldots, \rho_m v, \rho_m v),$$

where $v = (v_1, \ldots, v_n)$ is to be specified. The scaling factors 1, $\rho_1$, $\cdots$, $\rho_m$ are assumed to be nonzero and

[1] In this paper, the linear subsystem is constraint to be finite impulse response, this is mainly because of simplification. Indeed, the results can be easily extended to infinite impulse response cases by using the techniques in Wang et al (2006).

distinct. Under $2n(m+1)$-periodic inputs, we have $\Phi_j(l) = \Phi_j(1) \triangleq \Phi_j$, $l = 1, 2, \ldots$. Thus, (3) can be written as

$$Y(l) = \sum_{j=0}^{m} b_j \Phi_j \theta + D(l) \triangleq \zeta + D(l). \tag{4}$$

For the above system, the identification algorithm will be divided into two parts: i) to estimate $\zeta$ (which turns to be the estimation of gain systems), ii) to get $\theta$ and $\eta$ with the estimate of $\zeta$.

## 3. STRONGLY SCALED FULL RANK SIGNALS

This section is to introduce a class of input signals, called strongly full rank signals, which will play an important role in subsequent development.

An $n \times n$ circulant matrix (Lancaster & Tismenetsky, 1985)

$$T = \begin{bmatrix} v_n & v_{n-1} & \cdots & v_1 \\ v_1 & v_n & & v_2 \\ \vdots & & \ddots & \vdots \\ v_{n-1} & v_{n-2} & \cdots & v_n \end{bmatrix} \tag{5}$$

is completely determined by its first row $[v_n, \ldots, v_1]$, which will be denoted by $\mathbf{T}([v_n, \ldots, v_1])$.

*Definition 1.* An $n$-periodic signal generated from its one-period values $(v_1, \ldots, v_n)$ is said to be *full rank* if the circulant matrix $\mathbf{T}([v_n, \ldots, v_1])$ is full rank.

*Definition 2.* An $n$-periodic signal generated from its one-period values $(v_1, \ldots, v_n)$ is said to be *strongly m full rank* if the circulant matrices $\mathbf{T}([v_n^i, \ldots, v_1^i])$ are all full rank for $i = 1, \ldots, m$.

*Proposition 1.* An $n$-periodic signal generated from $v = (v_1, \ldots, v_n)$ is strongly $m$ full rank if it is strongly $m+1$ full rank.

*Theorem 1.* An $n$-periodic signal generated from $v = (v_1, \ldots, v_n)$ is strongly $m$ full rank if and only if for $l = 1, 2, \ldots, m$,

$$\gamma_{k,l} = \sum_{j=1}^{n} v_j^l e^{-i\omega_k j}$$

are nonzero at $\omega_k = \frac{2\pi k}{n}$, $k = 1, \ldots, n$.

Due to the page limit, some of the proofs are omitted, interested readers can contact the author for details.

*Remark 1.* Recall that $\mathcal{F}[v^l] = \{\gamma_{1,l}, \ldots, \gamma_{n,l}\}$ $(l = 1, 2, \ldots, m)$ is the frequency samples of the $v^l$, where $\mathcal{F}[\cdot]$ is the discrete Fourier transform. Then, Definition 1 may be equivalently stated as "an $n$-periodic signal $v$ is said to be strongly $m$ full rank if the frequency samples of $v^l$ $(l = 1, 2, \ldots, m)$ do not contain 0."

*Proposition 2.* For $n = 1, 2$, an $n$-periodic signal $u$ generated from $v = (v_1, \ldots, v_n)$ is strongly $m$ full rank if and only if it is full rank.

*Remark 2.* For $n > 2$, the conditions of strongly $m$ full rank may be different from the conditions of full rank.

*Definition 3.* A $2n(m+1)$-periodic signal $u$ is called strongly scaled $m$ full rank if its one-period values are $(v, v, \rho_1 v, \rho_1 v, \ldots, \rho_m v, \rho_m v)$, where $v = (v_1, \ldots, v_n)$ is strongly $m$ full rank; $\rho_j \neq 0$, $\rho_j \neq 1$, $j = 1, \ldots, m$; and $\rho_i \neq \rho_j$, $i \neq j$. Denote $\mathcal{U}_m$ to be the class of such signals.

## 4. ESTIMATE OF $\zeta$ WITH INDIVIDUAL THRESHOLDS

Based on the strongly scaled full rank signal, this section derives the estimation algorithm for $\zeta$ and analyze their unbias and efficient properties. To this end, an estimation algorithm based on the information of individual thresholds is firstly investigated.

*Assumption 1.* The noise $\{d(k)\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with finite variance. The distribution function $F(\cdot)$ of $d(1)$ is known, which is continuously differentiable and has a continuously differentiable inverse $F^{-1}(\cdot)$ and a bounded density $f(\cdot)$ with $f(x) \neq 0$.

*Assumption 2.* The prior information on $\theta = [a_0, \ldots, a_{n-1}]^T$ and $\eta = [b_0, \ldots, b_{m-1}]^T$ is that $\sum_{i=0}^{n-1} a_i \neq 0$, $b_m = 1$, $\eta \neq 0$, $\theta \in \Omega_\theta$ and $\eta \in \Omega_\eta$, where $\Omega_\theta$ and $\Omega_\eta$ are two known compact sets.

The input is a scaled $2n(m+1)-$periodic signal with one period values $(v, v, \rho_1 v, \rho_1 v, \ldots, \rho_m v, \rho_m v)$, where $v = (v_1, \ldots, v_n)$ is strongly $m$ full rank.

By periodicity, $\Phi_j(k) = \Phi_j$ for $j = 0, \ldots, n$, and $\Phi_j$ can be decomposed into $2(m+1)$ submatrices $\Phi_{j,i}$, $i = 1, \ldots, 2(m+1)$, of dimension $n \times n$:

$$\Phi_j = [\Phi_{j,1}^T, \Phi_{j,2}^T, \ldots, \Phi_{j,2(m+1)}^T]^T.$$

And for $k = 1, \ldots, 2(m+1)$, $\Phi_{j,k} = [\phi_j(kn), \phi_j(kn+1), \ldots, \phi_j(kn+n-1)]^T$.

Denote the $n \times n$ circulant matrix $V_0 = T([1, \ldots, 1])$, and $V_j = T([v_n^j, \ldots, v_1^j])$, $j = 1, \ldots, m$. Then, for $j = 0, \ldots, m$, the odd-indexed block matrices satisfy the simple scaling relationship

$$\Phi_{j,1} = V_j, \quad \Phi_{j,3} = \rho_1^j V_j, \quad \ldots, \quad \Phi_{j,2m+1} = \rho_m^j V_j. \tag{6}$$

Denote

$$\tau_j = [\tau_{j,1}, \ldots, \tau_{j,n}]^T = V_j \theta, \quad j = 0, \ldots, m. \tag{7}$$

Then,

$$\Phi_{j,1}\theta = \tau_j, \Phi_{j,3}\theta = \rho_1^j \tau_j, \ldots, \Phi_{j,2m+1}\theta = \rho_m^j \tau_j. \tag{8}$$

Let $\Psi_\theta = [\Phi_0\theta, \Phi_1\theta, \ldots, \Phi_m\theta]$. Then, from (4), we have

$$Y(l) = \Psi_\theta [\eta^T, 1]^T \theta + D(l) = \zeta + D(l). \tag{9}$$

*Remark 3.* In $(v, v, \rho_1 v, \rho_1 v, \ldots, \rho_m v, \rho_m v)$, there are always two identical subsequences $\rho_i v$ $(i = 1, \ldots, m)$ appearing consecutively. The main reason for this input structure is to generate block matrices that satisfy the above scaling relationship (6).

*Remark 4.* We use the following notation for element-wise vector functions. For a scalar function $g(\cdot)$ and a vector $x = [x_1, \ldots, x_l]^T \in \mathbb{R}^l$, the boldface symbol $\mathbf{g}(x)$ represents $\mathbf{g}(x) = [g(x_1), \ldots, g(x_l)]^T \in \mathbb{R}^l$. In addition, if $g(x)$ is invertible, $\mathbf{g}^{-1}(x)$ represents the component-wise inverse $\mathbf{g}^{-1}(x) = [g^{-1}(x_1), \ldots, g^{-1}(x_l)]^T \in \mathbb{R}^l$. Similarly, for $\alpha = [\alpha_1, \ldots, \alpha_l]^T \in \mathbb{R}^l$ and $c = [c_1, \ldots, c_l]^T \in \mathbb{R}^l$, we use the vector notation $\mathbf{I}_{\{\alpha \leq c\}} = [I_{\{\alpha_1 \leq c_1\}}, \ldots, I_{\{\alpha_l \leq c_l\}}]^T$. $\mathbb{1}_\ell$ and $\mathbf{0}_\ell \in \mathbb{R}^\ell$ will denote column vectors with all components being 1 and 0, respectively.

For (9), let

$$\varphi_i(N) = [\varphi_{i,1}(N), \ldots, \varphi_{i,2n(m+1)}(N)]^T = \frac{1}{N}\sum_{k=1}^{N} S_i(k)$$

$$= \frac{1}{N}\sum_{k=1}^{N} \mathbf{I}\{D(k) \le C_i \mathbb{1}_{2n(m+1)} - \Psi_\theta[\eta^T, 1]^T\}, \quad (10)$$

which is the empirical distribution of $D(l)$ at

$$C_i \mathbb{1}_{2n(m+1)} - \zeta = C_i \mathbb{1}_{2n(m+1)} - \Psi_\theta[\eta^T, 1]^T.$$

Then, by the strong law of large numbers,

$$\varphi_i(N) \to p_i = \mathbf{F}(C_i \mathbb{1}_{2n(m+1)} - \Psi_\theta[\eta^T, 1]^T), \text{ w.p.1.} \quad (11)$$

Denote $S_i(N) = [S_{i,1}, \ldots, S_{i,2n(m+1)}]^T$. By Assumption 1, for each $i = 1, \ldots, l$, $S_i(N)$ is i.i.d. Since for $j = 1, \ldots, 2n(m+1)$, $ES_{i,j}(k) = p_{i,j} = \mathbf{F}(C_i - \zeta_j)$ and

$$E(S_{i,j}(k) - p_{i,j})^2 = p_{i,j}(1 - p_{i,j}) := \Delta_{i,j}^2.$$

Then,

$$E\varphi_{i,j}(N) = \frac{1}{N}\sum_{k=1}^{N} ES_{i,j}(k) = p_{i,j},$$

$$E(\varphi_{i,j}(N) - p_{i,j})^2 = \frac{\Delta_{i,j}^2}{N}. \quad (12)$$

Notice that $F$ is a monotonic function by Assumption 1, and $\Omega_\theta$ and $\Omega_\eta$ are bounded by Assumption 2. Then, we can get $z$ such that

$$0 < z < p_{i,j} = F(C_i - \zeta_j) < 1 - z < 1,$$
$$i = 1, \ldots, l, \quad j = 1, \ldots, 2n(m+1).$$

Since $F(\cdot)$ is not invertible at 0 and 1, we modify $\varphi_{i,j}$ to avoid these points. Let

$$\xi_{i,j}(N) = \begin{cases} \varphi_{i,j}(N), & \text{if} \quad z < \varphi_{i,j}(N) < z; \\ z, & \text{if} \quad \varphi_{i,j}(N) < z; \\ 1 - z, & \text{if} \quad \varphi_{i,j}(N) > 1 - z. \end{cases} \quad (13)$$

Since $\varphi_{i,j}(N) \to p_{i,j}$, w.p.1 and $z < p_{i,j} < 1 - z$, we have $\xi_{i,j}(N) \to p_{i,j}$, w.p.1. Denote

$$\xi_i(N) = [\xi_{i,1}(N), \ldots, \xi_{i,2n(m+1)}(N)]^T. \quad (14)$$

By Assumption 1, $F$ has a continuous inverse. Hence, for each $i = 1, \ldots, l$,

$$\begin{aligned} \zeta_i(N) &= [\zeta_{i,1}(N), \ldots, \zeta_{i,2n(m+1)}(N)]^T \\ &:= C_i \mathbb{1}_{2n(m+1)} - \mathbf{F}^{-1}(\xi_i(N)) \\ &\to C_i \mathbb{1}_{2n(m+1)} - \mathbf{F}^{-1}(p_i) = \Psi_\theta[\eta^T, 1]^T \\ &= \zeta = [\zeta_1, \ldots, \zeta_{2n(m+1)}]^T \text{ w.p.1.} \end{aligned} \quad (15)$$

## 5. CONVEX COMBINATION ESTIMATE OF $\zeta$

Since $\zeta_i(N)$ is constructed from each individual threshold $C_i$, this enables us to treat the coefficients of the convex combination as design variables such that the resulting estimate become one with minimal variance. This resulting estimate will be called "convex combination estimate", which is shown to be unbias and asymptotically efficient.

For $j = 1, \ldots, 2n(m+1)$, define $\zeta_{\cdot,j}(N) = [\zeta_{1,j}(N), \ldots, \zeta_{l,j}(N)]^T$ and $c_j(N) = [c_{j,1}(N), \ldots, c_{j,l}(N)]^T$ with $c_{j,1}(N) + \cdots + c_{j,l}(N) = 1$. Construct an estimate of $\zeta_j$ by defining

$$\widehat{\zeta}_j(N) = c_j^T \zeta_{\cdot,j}(N) = \sum_{k=1}^{l} c_{j,k}(N)\zeta_{k,j}(N).$$

Suppose there exists $c_j = [c_{j,1}, \ldots, c_{j,l}]^T$ such that $c_j(N) \to c_j$. Then $c_{j,1} + \cdots + c_{j,l} = 1$, and by (15),

$$\widehat{\zeta}_j(N) \to \zeta_j \sum_{k=1}^{l} c_{j,k} = \zeta_j, \quad \text{w.p.1.}$$

Denote the estimation errors $e_j(N) = \widehat{\zeta}_j(N) - \zeta_j$, $\varepsilon_j(N) = \zeta_{\cdot,j}(N) - \zeta_j \mathbb{1}_l$, and their covariances

$$\sigma_j^2(N) = Ee_j(N)e_j(N)^T, \quad Q_j(N) = E\varepsilon_j(N)\varepsilon_j(N)^T,$$

respectively. Then the covariance of estimation error

$$\begin{aligned} \sigma_j^2(N) &= c_j^T(N) E\varepsilon_j(N)\varepsilon_j(N)^T c_j(N) \\ &= c_j^T(N) Q_j(N) c_j(N). \end{aligned} \quad (16)$$

That is, the variance is a quadratic form with respect to the variable $c_j$. To obtain the convex combination estimate, we choose $c_j$ to

$$\min \sigma_j^2(N), \text{ subject to } \quad c_j^T(N)\mathbb{1}_l = 1. \quad (17)$$

*Theorem 2.* Under Assumptions 1-2, suppose $u \in \mathcal{U}_m$ and $R_j(N) = NQ_j(N) = NE\varepsilon_j(N)\varepsilon_j(N)^T$ ($j = 1, \ldots, 2n(m+1)$) is positive definite. Then, the convex combination estimate can be obtained by choosing

$$c_j^*(N) = \frac{R_j^{-1}(N)\mathbb{1}_l}{\mathbb{1}_l^T R_j^{-1}(N)\mathbb{1}_l}, \quad \widehat{\zeta}_j(N) = \sum_{i=1}^{l} c_{j,i}^* \zeta_{i,j}(N), \quad (18)$$

and the minimal variance satisfies

$$N\sigma_j^{2*}(N) = \frac{1}{\mathbb{1}_l^T R_j^{-1}(N)\mathbb{1}_l}. \quad (19)$$

*Proof.* We solve the constrained optimization problem (17) by using the Lagrange method. Introduce the Hamiltonian

$$\begin{aligned} H(c_j(N), \alpha_j(N)) &= \sigma_j^2(N) + \alpha_j(N)(1 - c_j^T(N)\mathbb{1}_l) \\ &= c_j^T(N) R_j(N) c_j(N)/N + \alpha_j(N)(1 - c_j^T(N)\mathbb{1}_l), \end{aligned}$$

where $\alpha_i(N)$ is a Lagrange multiplier. Differentiating $H(\cdot,\cdot)$ with respect to its arguments and setting $\partial H(c_j, \alpha_j)/\partial c_j = 0$ and $\partial H(c_j, \alpha_j)/\partial \alpha_i = 0$ leads to

$$2R_j(N)c_j(N)/N - \alpha_j(N)\mathbb{1}_l = 0, \quad c_j^T(N)\mathbb{1}_l = 1.$$

Solving the above set of equations yields the stationary point of $H(c_j(N), \alpha_j(N))$

$$\alpha_j^*(N) = \frac{2}{N\mathbb{1}_l^T R_j^{-1}(N)\mathbb{1}_l}, \quad c_j^*(N) = \frac{R_j^{-1}(N)\mathbb{1}_l}{\mathbb{1}_l^T R_j^{-1}(N)\mathbb{1}_l}. \quad (20)$$

It can be verified that the stationary point is indeed a minimum.
Substituting (20) into (16), we obtain (19) as desired. □

### 5.1 Unbias and efficient analysis

From (18), $\widehat{\zeta}_j$ can be regarded as an estimate of $\zeta_j$. In this subsection, the unbias and efficient properties of this estimate will be analyzed.

By Assumption 1, $G(x) = F^{-1}(x)$ is continuous on $(0, 1)$, so $G(x)$ is bounded on the compact set $[z, 1 - z]$. Since

$\zeta_{i,j}(N) = C_i - G(\xi_{i,j}(N)) \to \zeta_{i,j}$ w.p.1, we have $\zeta_{i,j}(N) \to \zeta_{i,j}$ in probability, and furthermore, by Lebesge Dominated Convergence Theorem (Chow, 1978, p. 99), $E\zeta_{i,j}(N) \to \zeta_j$. Hence,

$$E\widehat{\zeta}_j(N) = E\sum_{k=1}^{l} c_{j,k}(N)\zeta_{k,j}(N) \to \zeta_j,$$

which means the estimate of $\zeta_j$ is unbias.

Subsequently, the efficiency of the estimate will be studied. To this end, the properties of $\xi_{i,j}(N)$ in (13) will be firstly introduced.

*Theorem 3.* Suppose $u \in \mathcal{U}_m$. Under Assumptions 1-2, there exist $K_{i,j} \in (0,\infty)$ and $L_{i,j} \in (0,\infty)$, $i = 1,\ldots,l$, $j = 1,\ldots,2n(m+1)$, such that

$$P\{\xi_{i,j}(N) \neq \varphi_{i,j}(N)\} \leq K_{i,j}e^{-L_{i,j}N}. \tag{21}$$

*Proof.* See Appendix A.

*Theorem 4.* Under the conditions of Theorem 3, we have

$$NE(\xi_{i,j}(N) - p_{i,j})^2 \to \Delta_{i,j}^2, \quad N \to \infty. \tag{22}$$

$$NE|(\xi_{i,j}(N) - p_{i,j})|^m \to 0, \quad N \to \infty, \quad m = 3,4,\ldots \tag{23}$$

*Proof.* See Appendix B.

From (19), the covariance of the estimation $\widehat{\zeta}_j(N)$ is decided by $R_j(N)$.

*Theorem 5.* Suppose $u \in \mathcal{U}_m$. If in addition to Assumptions 1-2, the density function $f(x)$ is continuously differentiable, then

$$R_j(N) := NQ_j(N) \to \Lambda_j W_j \Lambda_j := R_j, \quad N \to \infty, \tag{24}$$

where $\varepsilon_j(N) = \zeta_{\cdot,j}(N) - \zeta_j \mathbb{1}_l$, $\Lambda_j = \mathrm{diag}^{-1}\{f(C_j - \zeta_1),\ldots,f(C_j - \zeta_l)\}$ and

$$W_j = \begin{pmatrix} p_{1,j}(1-p_{1,j}) & \cdots & p_{1,j}(1-p_{l,j}) \\ \vdots & \ddots & \vdots \\ p_{1,j}(1-p_{l,j}) & \cdots & p_{l,j}(1-p_{l,j}) \end{pmatrix}. \tag{25}$$

*Proof.* Denote $\varepsilon_{j,i}(N)$ the $i$-th component of $\varepsilon_j(N)$, $G'(x) = dG(x)/dx$ and $G''(x) = dG'(x)/dx$. Then $G'(x) = 1/f(G(x))$ and $G''(x) = -f'(G(x))G'(x)/f^2(G(x))$. Since $f'(x)$ is continuous and by Assumption 1, both $G'(x)$ and $G''(x)$ are continuous, and hence bounded in $[z, 1-z]$. Let $\beta = \sup_{x\in[z,1-z]}\{|G'(x)|\}$ and $\gamma = \sup_{x\in[z,1-z]}\{|G''(x)|\}$. Then, there exists a number $\lambda_{i,j}(N)$ between $p_{i,j}$ and $\xi_{i,j}(N)$ such that

$$\varepsilon_{j,i} = \zeta_{i,j}(N) - \zeta_j = G(\xi_{i,j}(N)) - G(p_{i,j}) = G'(p_{i,j})(\xi_{i,j}(N) - p_{i,j}) + \frac{1}{2}G''(\lambda_{i,j}(N))(\xi_{i,j}(N) - p_{i,j})^2.$$

This implies that for $i,k = 1,\ldots,l$,

$$\begin{aligned} &E\varepsilon_{j,i}(N)\varepsilon_{j,k}(N) = E(\zeta_{i,j}(N) - \zeta_j)(\zeta_{k,j}(N) - \zeta_j) \\ =&G'(p_{i,j})G'(p_{k,j})E(\xi_{i,j}(N) - p_{i,j})(\xi_{k,j}(N) - p_{k,j}) \\ &+EG'(p_{i,j})(\xi_{i,j}(N) - p_{i,j})(\xi_{k,j}(N) - p_{k,j})^2G''(\lambda_{k,j}) \\ &+EG''(\lambda_{i,j})(\xi_{i,j}(N) - p_{i,j})^2(\xi_{k,j}(N) - p_{k,j})G'(p_{k,j}) \\ &+EG''(\lambda_{i,j})(\xi_{i,j}(N) - p_{i,j})^2(\xi_{k,j}(N) - p_{k,j})^2G'(\lambda_{k,j}). \end{aligned} \tag{26}$$

By Holder's inequality and Theorem 4, we have

$$\begin{aligned} &|EG'(p_{i,j})(\xi_{i,j}(N) - p_{i,j})(\xi_{k,j}(N) - p_{k,j})^2G''(\lambda_{k,j})| \\ &\leq \beta\gamma\sqrt{E(\xi_{i,j}(N) - p_{i,j})^2E(\xi_{k,j}(N) - p_{k,j})^4|} \to 0. \end{aligned} \tag{27}$$

Similarly, the last two items in (27) approach to 0.

It can be derived that

$$\begin{aligned} &E(\xi_{i,j}(N) - p_{i,j})(\xi_{k,j}(N) - p_{k,j}) \\ &\quad -E(\varphi_{i,j}(N) - p_{i,j})(\varphi_{k,j}(N) - p_{k,j}) \to 0. \end{aligned} \tag{28}$$

Since $d(k)$, $k = 1,2,\ldots$, are i.i.d.,

$$\begin{aligned} &E(\varphi_{i,j}(N) - p_{i,j})(\varphi_{k,j}(N) - p_{k,j}) \\ &= \frac{1}{N}E\sum_{l_1=1}^{N} I\{d(l_1) \leq p_{i,j}\}I\{d(l_1) \leq p_{k,j}\} - p_{i,j}p_{k,j} \\ &= p_{\min\{i,k\},j} - p_{i,j}p_{k,j} \end{aligned} \tag{29}$$

and

$$G'(p_{i,j}) = 1/f(G(p_{i,j})) = 1/f(C_i - \zeta_j). \tag{30}$$

Therefore, (24) follows from (26)-(30). □

*Proposition 3.* $R_j$ ($j = 1,...,2n(m+1)$) defined by (24) is positive definite, and

$$\mathbb{1}_l^T R_j^{-1} \mathbb{1}_l = \sum_{k=1}^{l+1} \frac{h_{k,j}^2}{\widetilde{p}_{k,j}}, \tag{31}$$

where $\widetilde{p}_{i,j} = F(C_i - \zeta_j) - F(C_{i-1} - \zeta_j)$, $h_{i,j} = f(C_{i-1} - \zeta_j) - f(C_i - \zeta_j)$ for $C_0 = -\infty$, $C_{l+1} = \infty$.

*Lemma 1.* (Wang & Yin, 2007) The Cramér-Rao lower bound for estimating $\zeta_j$ based on observations of $\{s(k)\}$ is

$$\sigma_{CR,j}^2(N) = (N\sum_{j=1}^{l+1} \frac{h_{i,j}^2}{\widetilde{p}_{i,j}})^{-1}.$$

The estimation algorithm of $\zeta_j$ is asymptotically efficient based on the following theorem.

*Theorem 6.* Under the conditions of Theorem 2, for $j = 1,\ldots,2n(m+1)$,

$$\lim_{N\to\infty} N(\sigma_j^{2*}(N) - \sigma_{CR,j}^2(N)) = 0, \quad N \to \infty.$$

*Proof.* This theorem can be proved directly by Theorem 2, Proposition 3 and Lemma 1.

### *5.2 Recursive convex combination estimates*

Since $\sigma_j^2(N) = E\varepsilon_j(N)\varepsilon_j^T(N)$ contains an unknown parameter $\zeta_j$, it cannot be directly computed. As a result, the convex combination estimate $\zeta_j(N)$ in (18) cannot be computed. In this subsection, we will derive computable estimates. The basic idea is to employ a recursive structure in which the unknown $\zeta_j$ is replaced by the current estimate $\tilde{\zeta}_j(N)$. Convergence of the algorithms will be established.

For $i = 1,...,l$ and $j = 1,...,2n(m+1)$, let $\xi_i(0) = \mathbf{0}_{2n(m+1)}$, $\widehat{c}_j(0) = \mathbf{0}_l$, $\widehat{R}_j(0) = \mathbf{0}_{l\times l}$ and $\widehat{\zeta}_j(0) = \mathbf{0}_{2n(m+1)}$. Suppose that at step $N-1$ ($N \geq 1$), $\xi_i(N-1)$, $c_j(N-1)$, and $\widehat{R}_j(N-1)$ have been obtained. Then the estimation algorithms can be constructed as follows.

i) Calculate the sample distribution values

$$\xi_i(N) = \frac{1}{N}S_i(N) + \frac{N-1}{N}\xi_i(N-1).$$

ii) Calculate the data points $\zeta_i(N) = F^{-1}(\xi_i(N))$. Let $\zeta_{\cdot,j}(N) = [\zeta_{1,j}(N),...,\zeta_{l,j}(N)]^T$.

iii) Calculate each covariance estimate $R_j(N)$.
Let $p_{i,j}(N) = F(C_i - \tilde{\zeta}_j(N-1))$ and
$\widehat{\Lambda}_j(N) = \text{diag}^{-1}\{f(p_{1,j}(N)), ..., f(p_{l,j}(N))\}$,
$W_j(N) = \begin{pmatrix} p_{1,j}(N)(1-p_{1,j}(N)) & \cdots & p_{1,j}(N)(1-p_{l,j}(N)) \\ \vdots & \ddots & \vdots \\ p_{1,j}(N)(1-p_{l,j}(N)) & \cdots & p_{l,j}(N)(1-p_{l,j}(N)) \end{pmatrix}$.
Calculate $R_j(N)$ by $\widehat{R}_j(N) = \widehat{\Lambda}_j(N) W_j(N) \widehat{\Lambda}_j(N)$.

iv) If $\widehat{R}_j(N)$ is non-singular, then let

$$\widehat{c}_j(N) = \frac{\widehat{R}_j^{-1}(N)\mathbb{1}}{\mathbb{1}^T \widehat{R}_j^{-1}(N)\mathbb{1}},$$

and $\widehat{\zeta}_j(N) = \widehat{c}_j^T(N)([C_1, ..., C_l]^T - \zeta_{\cdot,j}(N))$. Otherwise, $\widehat{\zeta}_j(N) = \widehat{\zeta}_j(N-1)$.

v) Let $\widehat{\zeta}(N) = [\widehat{\zeta}_1(N), ..., \widehat{\zeta}_{2n(m+1)}(N)]^T$. Go to step 1.

This algorithm depends only on sample paths. At each step, it minimizes estimation variance based on the most recent information on the unknown parameter. In addition, the following asymptotic properties hold.

*Theorem 7.* Under the condition of Theorem 5, for $j = 1, ..., 2n(m+1)$ the above recursive algorithms have the following properties:

$$\widehat{\zeta}_j(N) \to \zeta_j, \quad \text{w.p.1},$$
$$\widehat{R}_j(N) \to R_j, \quad \text{w.p.1},$$
$$NE(\widehat{\zeta}_j(N) - \zeta_j)^2 \to (\mathbb{1}^T R_j^{-1} \mathbb{1})^{-1}, \quad \text{w.p.1} \quad N \to \infty.$$

## 6. ESTIMATION OF SYSTEM PARAMETERS

Identification algorithms of the system parameters will be constructed based on the estimate of $\zeta$. The parameters of linear part are first estimated, based on which the nonlinearity is identified.

### *6.1 Identifiability of the unknown parameters*

*Theorem 8.* Suppose $u \in \mathcal{U}_m$. Then,

$$\Psi_\theta [\eta^T, 1]^T = \zeta$$

has a unique solution $(\theta^*, \eta^*)$.

*Proof.* i) *To obtain $\theta^*$.* By the first component of (15), we have $\zeta = [\zeta_1, \ldots, \zeta_{2n(m+1)}]^T$, and

$$b_0\tau_{0,1} + b_1\tau_{1,1} + \cdots + b_m\tau_{m,1} = \zeta_1.$$

From (8), the $2in+1$ $(i = 1, \ldots, m)$ component of (15) turns to be $b_0\tau_{0,1} + \rho_i b_1 \tau_{1,1} + \cdots + \rho_i^m b_m \tau_{m,1} = \zeta_{2in+1}$, or equivalently,

$$\Re [b_0\tau_{0,1}, \ldots, b_m\tau_{m,1}]^T = [\zeta_1, \zeta_{2n+1}, \ldots, \zeta_{2mn+1}]^T, \quad (32)$$

where $\Re = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \rho_1 & \cdots & \rho_1^m \\ \vdots & \cdots & \cdots & \vdots \\ 1 & \rho_m & \cdots & \rho_m^m \end{pmatrix}$.

Since $\rho_j \neq 0$, $\rho_j \neq 1$, $j = 1, \ldots, m$, and $\rho_i \neq \rho_j$, the determinant of the Vandermonde matrix

$$\det \Re = \prod_{0 \le i < j \le m-1} (\rho_j - \rho_i) \neq 0 \quad \text{with} \quad \rho_0 = 1.$$

Hence, $b_j\tau_{j,1}$, $j = 0, \ldots, m$, can be solved by

$$\Re [b_0\tau_{0,1}, \ldots, b_m\tau_{m,1}]^T = \Re^{-1} [\zeta_1, \zeta_{2n+1}, \ldots, \zeta_{2mn+1}]^T.$$

Similarly, we have

$$\Gamma = \Re^{-1} \Xi, \quad (33)$$

where

$$\Gamma = \begin{pmatrix} b_0\tau_{0,1} & \cdots & b_0\tau_{0,n} \\ \vdots & & \vdots \\ b_m\tau_{m,1} & \cdots & b_m\tau_{m,n} \end{pmatrix}, \ \Xi = \begin{pmatrix} \zeta_1 & \cdots & \zeta_n \\ \vdots & & \vdots \\ \zeta_{2mn+1} & \cdots & \zeta_{(2m+1)n} \end{pmatrix}.$$

Denote $r_i$ as the $i$-th column of $(\Re^{-1})^T$. Then, by $b_m = 1$ we have

$$\tau_m = [\tau_{m,1}, \ldots, \tau_{m,n}]^T = \Xi^T r_m.$$

Notice that $u \in \mathcal{U}(m)$ implies that $V_m$ is full rank. Then, by (7) one can get $\theta^* = V_m^{-1}\tau_m$.

ii) *To obtain $\eta^*$.* By Assumption 2, $\sum_{i=0}^{n-1} a_i \neq 0$, or $V_0\theta \neq \mathbf{0}_n$. For $u \in \mathcal{U}_m$ and $j = 1, \ldots, m$, $V_j = \mathbf{T}([v_n^j, \ldots, v_1^j])$ is full rank by Definition 2, and so $V_j\theta \neq \mathbf{0}_n$. Thus, for each $j = 0, \ldots, m$, $\tau_j = V_j\theta$ has a nonzero component $\tau_{j,i_j^*}$. For any given positive integer $k$ and $j = 1, \ldots, k$, let $\beta_j(k)$ be a $k$-dimensional vector with all components being zero except the $j$-th being 1, that is,

$$\beta_j(k) = [\overbrace{0, \ldots, 0}^{j-1}, 1, \overbrace{0, \ldots, 0}^{k-j}]^T.$$

Then, from (33) we have

$$b_j\tau_{j,i_j^*} = \beta_j^T(m+1)\Re^{-1}\Xi\beta_{i_j^*}(n), \quad j = 0, \ldots, m,$$

which gives $b_j$, $j = 0, \ldots, m$, since $\tau_{j,i_j^*}$ can be calculated from $V_j$ and $\theta^*$ via (7). Thus, $\eta^*$ is obtained. □

### *6.2 Identification algorithms and convergence properties*

The $\zeta(N) = [\zeta_0(N), \ldots, \zeta_{2n(m+1)-1}(N)]^T$ in (14) has $2n(m+1)$ components for a strongly scaled $m$ full rank signal $u \in \mathcal{U}_m$.

Let $V_m = \mathbf{T}([u_n^m, \ldots, u_1^m])$, $[r_1, \ldots, r_m] := (\Re^T)^{-1}$,

$$\Xi(N) = \begin{pmatrix} \zeta_1(N) & \cdots & \zeta_n(N) \\ \vdots & & \vdots \\ \zeta_{2mn+1}(N) & \cdots & \zeta_{(2m+1)n}(N) \end{pmatrix}.$$

Then, we have the following identification algorithm:

i) *Estimate $\theta$.* The estimate of $\theta$ is taken as

$$\theta(N) = V_m^{-1}\Xi^T(N) r_m.$$

ii) *Estimate $\eta$.* Let $b_j(0) = 0$ and $b_j(N) =$

$$\begin{cases} [\zeta_{i_j^*(N)}, \zeta_{2n+i_j^*(N)}, \ldots, \zeta_{2mn+i_j^*(N)}] r_{i_j^*(N)} / \tau_{j,i_j^*}(N), & \text{if } \tau_{j,i_j^*(N)} \neq 0, \\ b_j(N-1), & \text{if } \tau_{j,i_j^*(N)} = 0, \end{cases}$$

where, for $j = 0, 1, \ldots, m-1$,

$$i_j^*(N) = \min\{\text{argmax}_{1 \le i \le n} |\tau_{j,i}(N)|\}, \quad (34)$$

$r_{i_j^*(N)}$ is the $i_j^*(N)$-th column of $(\Re^T)^{-1}$, and $\tau_{j,i_j^*}(N)$ is the $i_j^*(N)$-th component of $\tau_j(N) = V_j\theta(N)$. Then, the estimate of $\eta$ is taken as
$\eta(N) = [b_0(N), \ldots, b_{m-1}(N)]^T$.

*Theorem 9.* Suppose $u \in \mathcal{U}_m$. Then, under Assumptions 1-2,

$$\theta(N) \to \theta, \ \eta(N) \to \eta \ \text{ w.p.1 as } \ N \to \infty.$$

## 7. ILLUSTRATIVE EXAMPLES

This section gives two examples to illustrate the convergence of the estimate algorithms developed in this paper.

*Example 1.* Consider a gain system $y(k) = a + d(k)$. Here the actual value of the unknown $a$ is 5. The disturbance is a sequence of i.i.d. Gaussian variables with zero mean and standard deviation $\sigma = 5$. The sensor has 3 switching thresholds $C_1 = 2$, $C_2 = 6$, and $C_3 = 10$. Then, the recursive algorithm in Section 5 is used to generate convex combination estimates.

For comparison, estimates derived by using each switching threshold individually are also calculated. Figure 2 compares convex combination estimates to those using individual switching thresholds. It is shown that the estimate with 3 thresholds converges faster than the ones with individual thresholds. The weights of the estimates of each threshold are shown in Figure 3, which illustrates that the weights are not sure to be positive.

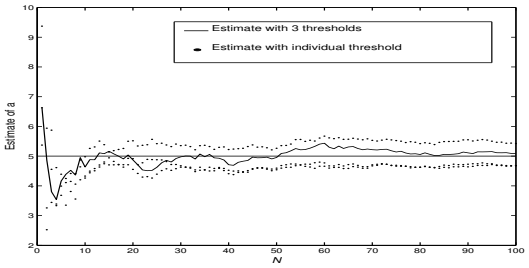


Fig. 2. Identification with set-valued output observations

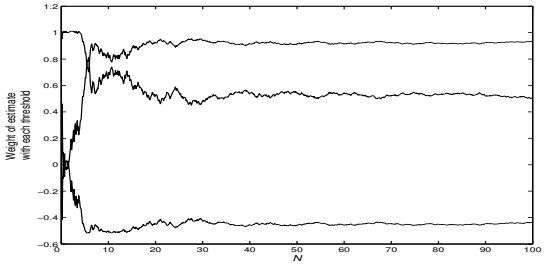


Fig. 3. Weights of estimates with each threshold

*Example 2.* For some prior information, the algorithms in Subsection 6.2 can be simplified. For example, the estimate algorithms of $\eta$ can be simplified, when the prior information on $\theta$ is known to be positive and the periodic input $u$ is positive.

Consider

$$\begin{cases} y(k) = a_0 x(k) + a_1 x(k-1) + d(k), \\ x(k) = b_0 + b_1 u(k) + u^2(k), \end{cases}$$

where the noise $\{d(k)\}$ is a sequence of i.i.d. Gaussian variables with $Ed_1 = 2$, $\sigma_d^2 = 4$. The output is measured by a binary-valued sensor with threshold $C = 13$. The linear subsystem has order $n = 2$, and the nonlinear function has order $m = 2$. The prior information on $a_i$, $i = 0, 1$, is that $a_i \in [0.5, 5]$. Suppose the true values of unknown parameters are $\theta = [a_0, a_1]^T = [1.17, 0.95]^T$ and $\eta = [b_0, b_1]^T = [3, 1.3]^T$.

The input is 12-periodic with one period $(v, v, \rho_1 v, \rho_1 v, \rho_2 v, \rho_2 v)$, where $v = [1.2, 0.85]$, $\rho_1 = 0.65$ and $\rho_2 = 1.25$. Define the block variables $X(l)$, $Y(l)$, $\Phi_j(l)$, $D(l)$ and $S(l)$, in the case of an 12-periodic input. Using (14), we can construct the algorithms in Subsection 6.2 to identify $\theta$.

Notice that $a_i \in [0.5, 5]$, $i = 1, 2$, and $u$ is positive. Then, $\tau_{j,1}$, the first component of $V_j\theta$, is $\tau_{j,1} = v_2^2 a_0 + v_1^2 a_1 \geq 0.5(v_2^2 + v_1^2) \neq 0$, where the last inequality is derived from the fact that $v$ is strongly 2 full rank. So, it is not necessary to calculate $i_j^*(N)$ in (34), which aims to find the nonzero component of $\tau_j$. And $\eta$ can be estimated as following: $\eta(0) = \mathbf{0}$ and

$$\eta(N) = \begin{cases} \Lambda(N)\Re^c[\zeta_1(N), \zeta_{2n+1}(N), \ldots, \zeta_{2mn+1}(N)]^T, \\ \qquad\qquad \text{if } \prod_{j=0}^{m-1} \tau_{j,1}(N) \neq 0, \\ \eta(N-1), \qquad \text{if } \prod_{j=0}^{m-1} \tau_{j,1}(N) = 0, \end{cases}$$

where $\Lambda(N) = \text{diag}^{-1}(\tau_{0,1}(N), \ldots, \tau_{m-1,1}(N))$, $\Re^c$ is a $m \times (m+1)$ matrix containing from 1st to $m-1$-th rows of $\Re^{-1}$, and $\tau_{j,1}(N)$ is the 1st component of $\tau_j(N) = V_j\theta(N)$.

The estimate errors of $\theta$ and $\eta$ are shown in Figure 4, where the errors are measured by the Euclidean norm. Both parameter estimates of the linear and nonlinear subsystems converge to their true values.

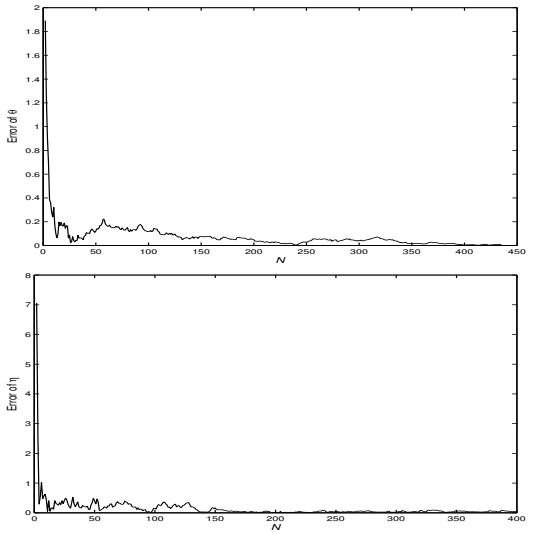


Fig. 4. Identification errors of $\theta$ and $\eta$

## 8. CONCLUDING REMARKS

In this paper, identification of Hammerstein systems with set-valued output observations is studied. Under assumptions of known disturbance distribution functions and strongly scaled full rank inputs, identification algorithms, convergence properties, and identification efficiency are derived.

There are many potential extensions of the results in this paper. For example, when the sensor threshold value and/or the noise distribution function are unknown, combined identification of systems, distribution functions and sensor thresholds is of practical importance.

## REFERENCES

E.W. Bai. An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear system. *Automatica*, volume 34(3), pages 333–338, 1998.

E.W. Bai. A blind approach to the Hammerstein-Wiener model identification. *Automatica*, volume 38(6), pages 967-979, 2002.

P. Billingsley. *Convergence of Probability Measures*, 2nd Edition, J. Wiley, New York, 1999.

S. Billings. Identification of Nonlinear Systems- A survey. *Proc. of IEE*, volume 127(6), pages 272-285, 1980.

P. Celka, N.J. Bershad, and J.M. Vesin. Stochastic gradient identification of polynomial Wiener systems: Analysis and application. *IEEE Trans. Signal Process.*, volume 49(2), pages 301-313, 2001.

H.F. Chen. Recursive identification for Wiener model with discontinuous piece-wise linear function" *IEEE Trans. Autom. Control*, volume 51(3), pages 390-400, 2006.

Y.S. Chow, *Probability Theory*, Springer Verlag, New York, 1978.

I.W. Hunter, and M.J. Korenberg. The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Bio Cybernetics*, volume 55(3), pages 135-144, 1986.

M.J. Korenberg, and I.W. Hunter. Two methods for identifying Wiener cascades having noninvertible static nonlinearities. *Annals of Biomedical Engineering*, volume 27(6), pages 793-804, 1998.

H.J. Kushner and G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd Ed., Springer-Verlag, New York, 2003.

S.L. Lacy, and D.S. Bernstein. Identification of FIR Wiener systems with unknown, noninvertible polynomial nonlinearities. *Proc. Amer. Contr. Conf.*, pages 893-899, 2002.

P. Lancaster, and M. Tismenetsky, *The Theory of Matrices*, 2nd Ed., Academic Press, 1985.

B. Ninness, and S. Gibson. Quantifying the accuracy of Hammerstein model estimation. *Automatica*, volume 38(12), pages 2037-2051, 2002.

R.J. Serfling, *Approximation Theorems of Mathematical Statistics*, J. Wiley, New York, 1980.

L.Y. Wang and G. Yin. Asymptotically efficient parameter estimation using quantized output observations. *Automatica*, volume 43(7), pages 1178-1191, 2007.

L.Y. Wang, G. Yin, and H. Wang. Identification of Wiener models with anesthesia applicaitons. *Int. J. Pure & Appl. Sci*, volume 1(1), pages 35-61, 2004.

L.Y. Wang, G. Yin, and J.F. Zhang. Joint identification of plant rational models and noise distribution functions using binary-valued observations. *Automatica*, volume 42(4), pages 535-547, 2006.

L.Y. Wang, J.F. Zhang, and G. Yin. System identification using binary sensors. *IEEE Trans. Automat. Control*, volume 48(11), pages 1892-1907, 2003.

T. Wigren. Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model. *IEEE Trans. Automatic Control*, volume 39(11), pages 2191-2206, 1994.

Y.L. Zhao, L.Y. Wang, G. Yin, and J.F. Zhang. Identification of Wiener Systems with Binary-Valued Output Observations. *Automatica*, volume 43(10), October 2007.

## Appendix A. PROOF OF THEOREM 3

*Proof.* Denote $X_{i,j} = (S_{i,j}(1) - p_{i,j})/\Delta_{i,j}$. Note that $EX_{i,j} = 0$ and $EX_{i,j}^2 = 1$. By the i.i.d. assumption, taking a Taylor expansion of $M_{i,j}(h,N) = [E\exp(hX_{i,j}/\sqrt{N})]^N$, the moment generating function of $\sqrt{N}(\varphi_{i,j}(N) - p_{i,j})/\Delta_{i,j}$, we obtain

$$\begin{aligned} M_{i,j}(h,N) &= \Big[E[1 + \frac{hX_{i,j}}{\sqrt{N}} + \frac{h^2X_{i,j}^2}{2N} + O(N^{-3/2})]\Big]^N \\ &= [1 + \frac{h^2}{2N} + O(N^{-(3/2)})]^N. \end{aligned}$$

Consequently, for any $t \in \mathbb{R}$,

$$\inf_h e^{-ht} M_{i,j}(h,N) =\leq Ke^{-\frac{t^2}{2}}, \tag{A.1}$$

where $K > 0$ is a positive constant.

By means of the Chernoff bound (Serfling, 1980, p. 326), for any $t \in (-\infty, p_{i,j}]$,

$$\begin{aligned} P\{\varphi_{i,j}(N) \leq t\} &= P\{\textstyle\sum_{k=1}^N (S_{i,j}(k) - p_{i,j}) \leq N\frac{(t-p_{i,j})}{\Delta_{i,j}}\} \\ &\leq \{\inf_h [e^{-\frac{h(t-p_{i,j})}{\Delta_{i,j}}} M_{i,j}(h,N)]\}^N \end{aligned} \tag{A.2}$$

and for any $p_{i,j} \leq t < \infty$,

$$P\{\varphi_{i,j}(N) \geq t\} \leq \{\inf_h [e^{-\frac{h(t-p_{i,j})}{\Delta_{i,j}}} M_{i,j}(h,N)]\}^N. \tag{A.3}$$

Considering

$$P\{\xi_{i,j}(N) \neq \varphi_{i,j}(N)\} = P(\varphi_N \leq z) + P(\varphi_N \geq 1 - z)$$

and (A.1)-(A.3), we have (21) is true. □

## Appendix B. PROOF OF THEOREM 4

*Proof.* i) By Theorem 3, there exist $K_{i,j} \in (0,\infty)$ and $L_{i,j} \in (0,\infty)$ such that

$$EN(\xi_{i,j}(N) - \varphi_{i,j}(N))^2 \leq NzP\{\xi_{i,j}(N) \neq \varphi_{i,j}(N)\} \to 0.$$

This together with

$$\begin{aligned} &EN(\varphi_{i,j}(N) - p_{i,j})(\xi_{i,j}(N) - \varphi_{i,j}(N)) \\ \leq &\sqrt{EN(\varphi_{i,j}(N) - p_{i,j})^2 EN(\xi_{i,j}(N) - \varphi_{i,j}(N))^2} \end{aligned}$$

implies that

$$\begin{aligned} &EN(\xi_{i,j}(N) - p_{i,j})^2 - EN(\varphi_{i,j}(N) - p_{i,j})^2 \\ =&2EN(\varphi_{i,j}(N) - p_{i,j})(\xi_{i,j}(N) - \varphi_{i,j}(N)) \\ &+EN(\xi_{i,j}(N) - \varphi_{i,j}(N))^2 \to 0. \end{aligned} \tag{B.1}$$

Thus, by (12), we get (22).

ii) Similarly, for $m = 3, 4, \ldots$, one can get

$$NE|(\xi_{i,j}(N) - p_{i,j})|^m - NE|(\varphi_{i,j}(N) - p_{i,j})|^m \to 0.$$

By Holder's inequality

$$\begin{aligned} &NE|\varphi_{i,j}(N) - p_{i,j}|^m \\ &\leq \Delta_{i,j}\sqrt{NE(\varphi_{i,j}(N) - p_{i,j})^{2(m-1)}}. \end{aligned} \tag{B.2}$$

Notice that for each $i, j$, $S_{i,j}(k)$ is i.i.d. Then, we have

$$\begin{aligned} &NE(\varphi_{i,j}(N) - p_{i,j})^{2(m-1)} \\ =&NE[\frac{1}{N}\sum_{k=1}^N (S_{i,j}(k) - p_{i,j})]^{2(m-1)} \leq N^{-2(m-2)}, \end{aligned}$$

which together with (B.2) results in

$$NE|\varphi_{i,j}(N) - p_{i,j}|^m \leq \Delta_{i,j}N^{-(m-2)} \to 0.$$

Hence, (23) is obtained. □