Preprints of the
9th International Symposium on Advanced Control of Chemical Processes
The International Federation of Automatic Control
June 7-10, 2015, Whistler, British Columbia, Canada

Multi-innovation parameter estimation for Hammerstein MIMO output-error systems based on the key-term separation ⋆

Qianyan Shen, Feng Ding

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, PR China (shenqianyan1990@163.com; fding@jiangnan.edu.cn)

Abstract: This paper uses the key-term separation principle and develops a multi-innovation stochastic gradient algorithm for Hammerstein MIMO output error systems. The basic idea is to decompose a Hammerstein MIMO system into two subsystems and to identify the parameters of each subsystem interactively. The parameter estimation accuracy can be improved with the innovation length increasing. The simulation example verifies the effectiveness of the proposed algorithm.

Keywords: Gradient search; Multi-innovation identification; Hammerstein system; Hierarchical identification; Key-term separation

1. INTRODUCTION

Parameter identification of nonlinear systems, whose block structure models can be described as Hammerstein models (Rebillat, Hajrya & Mechbal, 2014; Wang, Zhao, Xu & Zhao, 2014; Shen & Ding, 2014), Wiener models (Hu, Liu, Zhou & Yang, 2014; Lindstena, Schöna & Jordan, 2013; Ikhouane & Giri, 2014), Hammerstein-Wiener models (Yu, Mao, Jia & Yuan, 2014; Averin, 2003; Wills, Schon, Ljung & Ninness, 2013) and Wiener-Hammerstein models (Vanbeylen,2014; Schoukens, Pintelon & Rolain, 2014; Haryanto & Hong, 2013), has received much attention in recent years. For example, Sun and Liu (2013) proposed a particle swarm optimization aided maximum likelihood identification algorithm for Hammerstein systems. Hong, Mitchell and Chen (2013) used a De Boor algorithm combining with Gauss-Newton algorithm for the parameter estimation of Wiener systems. He, Xu and Masri (2012) derived a weighted adaptive iterative least-squares algorithm for nonlinear systems.

Many identification methods and identification techniques have been developed and have wide applications in optimization (Alevizos, 2013), chemical process modeling (Jiang, Yan & Liu, 2013; Zhang & Liu, 2013), signal filtering (Scarpiniti, Comminiello & Parisi, 2013; Li & Shi, 2012; Wang & Tang, 2014), control theory (Shi & Yu, 2011; Xu, 2014; Xu, Chen, & Xiong, 2015), parameter identification (Zhao & Qin, 2014; Vicario, Phan, Betti & Longman, 2014; Ding & Lin, 2014), and so on. Recently, Carini, Sicuranza and Mathews (2013) discussed NLMS and RLS adaptive algorithms for identifying linear-in-the-parameters nonlinear systems using periodic input sequences. Mileounis and Kalouptsidis (2013) studied the blind identification algorithms and order determination of sparse systems. Kim (2014) proposed a sensitivity based system identification algorithm for extracting the natural frequencies and the damping ratio from ambient cable vibration data.

Much attention has been paid to output error type systems, including output-error (OE) systems (Piga & Toth, 2014; Söderström, Hong, Schoukens & Pintelon, 2010), output-error autoregressive systems (Chen, Zhang & Ding, 2014), output-error moving average (OEMA) systems (Wang, 2011; Hu, Liu, & Zhou, 2014; Ding, Fan & Lin, 2013) and Box-Jenkins systems (Pintelona, Schoukensa & Guillaumeb, 2007; Liu & Wang, 2010). Zhang and Cui (2011) presented a bias compensation recursive least squares identification algorithm for output error systems with colored noises. Ding, Liu and Liu (2010) derived a gradient based iterative algorithm and a least squares based iterative algorithm for OE and OEMA systems.

This paper studies the multi-innovation parameter estimation algorithm for Hammerstein MIMO output-error systems by the key-term separation principle and the hierarchical identification principle. One common method for the identification of Hammerstein systems is the over-parameterization method, which is to directly transform the Hammerstein system into a pseudo-linear model, but the unknown parameter space contains the products of the unknown parameters. However, the over-parameterization mehod is hardly applied to identify Hammerstein MIMO systems for that its drawback is that the dimension of the resulting unknown parameter vector is large, and the identification algorithms have heavy computation. Using the key-term separation principle, the proposed algorithm can reduce redundant parameters to enhance the computing efficiency. Moreover, using the hierarchical identification principle, we separate a large scale system into two subsystems and decrease the dimensions of the information

⋆ This work was supported by the National Natural Science Foundation of China and the PAPD of Jiangsu Higher Education Institutions.

matrices for the purpose of reducing the computational burden further.

The rest of this paper is organized as follows. Section 2 introduces the identification problems for Hammerstein MIMO OE systems by the key-term separation principle. Section 3 discuss the multi-innovation stochastic gradient algorithm. Section 4 gives an illustrative example to show the effectiveness of the proposed algorithms. Finally, we offer some concluding remarks in Section 5.

2. THE DESCRIPTION OF HAMMERSTEIN MIMO OE SYSTEMS

Consider the following Hammerstein MIMO output-error system,

$$\boldsymbol{y}(t)=\boldsymbol{x}(t)+\boldsymbol{v}(t), \tag{1}$$

$$\boldsymbol{x}(t)=\boldsymbol{A}^{-1}(z)\boldsymbol{B}(z)\bar{\boldsymbol{u}}(t), \tag{2}$$

where $\boldsymbol{y}(t)\in\mathbb{R}^m$ is the output vector, $\bar{\boldsymbol{u}}(t):=[\bar{u}_1(t),\cdots,\bar{u}_m(t)]\in\mathbb{R}^m$ is the output vector of the nonlinear block, $\boldsymbol{v}(t)\in\mathbb{R}^m$ is an additive noise vector with zero mean, $\bar{u}_i(t)$ is a linear combination with unknown coefficients c_{ij} of a known basis $\boldsymbol{f}(u_i(t)):=[f_1(u_i(t)),f_2(u_i(t)),\cdots,f_{n_c}(u_i(t))]^{\mathrm{T}}\in\mathbb{R}^{n_c}$, i.e.,

$$\bar{u}_i(t)=c_{i1}f_1(u_i(t))+c_{i2}f_2(u_i(t))+\cdots+c_{in_c}f_{n_c}(u_i(t)),$$

$u_i(t)\in\mathbb{R}$ is the ith element of system input vector $\boldsymbol{u}(t)\in\mathbb{R}^m$, $\boldsymbol{A}(z)$ and $\boldsymbol{B}(z)$ are polynomial matrices in the unit backward shift operator $[z^{-1}y(t)=y(t-1)]$:

$$\begin{aligned}\boldsymbol{A}(z)&:=\boldsymbol{I}+\boldsymbol{A}_1z^{-1}+\boldsymbol{A}_2z^{-2}+\cdots+\boldsymbol{A}_{n_a}z^{-n_a},\\ \boldsymbol{A}_l&=[a_{ij}^l]\in\mathbb{R}^{m\times m},\\ \boldsymbol{B}(z)&:=\boldsymbol{B}_0+\boldsymbol{B}_1z^{-1}+\boldsymbol{B}_2z^{-2}+\cdots+\boldsymbol{B}_{n_b}z^{-n_b},\\ \boldsymbol{B}_l&=[b_{ij}^l]\in\mathbb{R}^{m\times m}.\end{aligned}$$

The symbol $\boldsymbol{I}$ stands for an identity matrix of appropriate sizes. Assume that the orders n_a, n_b and n_c are known and $\boldsymbol{y}(t)=0$, $\boldsymbol{u}(t)=0$ and $\boldsymbol{v}(t)=0$ for $t\leqslant 0$. Rewrite Equation (2) as

$$\begin{aligned}\boldsymbol{x}(t)&=[\boldsymbol{I}-\boldsymbol{A}(z)]\boldsymbol{x}(t)+\boldsymbol{B}(z)\bar{\boldsymbol{u}}(t)\\ &=[\boldsymbol{I}-\boldsymbol{A}(z)]\boldsymbol{x}(t)+\boldsymbol{B}_0\bar{\boldsymbol{u}}(t)+\cdots+\boldsymbol{B}_{n_b}\bar{\boldsymbol{u}}(t-n_b),\end{aligned}$$

from which we get more details:

$$\boldsymbol{B}_l\bar{\boldsymbol{u}}(t-l)=\left[\sum_{j=1}^m b_{1j}^l\bar{u}_j(t-l),\sum_{j=1}^m b_{2j}^l\bar{u}_j(t-l),\cdots,\sum_{j=1}^m b_{mj}^l\bar{u}_j(t-l)\right]^{\mathrm{T}}, \tag{3}$$

$$\sum_{j=1}^m b_{ij}^l\bar{u}_j(t-l)=\sum_{j=1}^m b_{ij}^l[c_{j1},c_{j2},\cdots,c_{jn_c}]\boldsymbol{f}(u_j(t-l)),\quad l=0,1,\cdots,n_b. \tag{4}$$

According to (3)–(4), it is clear that b_{ij}^l and c_{ij} cannot be identified separately. Therefore, to obtain unique parameter estimates, we take $\bar{\boldsymbol{u}}(t)$ as a key term and fix $\boldsymbol{B}_0=\boldsymbol{I}$ and obtain the identification model:

$$\begin{aligned}\boldsymbol{x}(t)&=[\boldsymbol{I}-\boldsymbol{A}(z)]\boldsymbol{x}(t)+\bar{\boldsymbol{u}}(t)+[\boldsymbol{B}(z)-\boldsymbol{I}]\bar{\boldsymbol{u}}(t)\\ &=\boldsymbol{\vartheta}^{\mathrm{T}}\boldsymbol{\phi}(t)+\bar{\boldsymbol{u}}(t),\\ \boldsymbol{y}(t)&=\boldsymbol{x}(t)+\boldsymbol{v}(t)=\boldsymbol{\vartheta}^{\mathrm{T}}\boldsymbol{\phi}(t)+\bar{\boldsymbol{u}}(t)+\boldsymbol{v}(t),\end{aligned} \tag{5}$$

where

$$\begin{aligned}\boldsymbol{\vartheta}^{\mathrm{T}}&:=[\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_{n_a},\boldsymbol{B}_1,\boldsymbol{B}_2,\cdots,\boldsymbol{B}_{n_b}]\in\mathbb{R}^{m\times(mn)},\\ n&:=n_a+n_b,\\ \boldsymbol{\phi}(t)&:=[-\boldsymbol{x}^{\mathrm{T}}(t-1),-\boldsymbol{x}^{\mathrm{T}}(t-2),\cdots,-\boldsymbol{x}^{\mathrm{T}}(t-n_a),\\ &\qquad\bar{\boldsymbol{u}}^{\mathrm{T}}(t-1),\bar{\boldsymbol{u}}^{\mathrm{T}}(t-2),\cdots,\bar{\boldsymbol{u}}^{\mathrm{T}}(t-n_b)]^{\mathrm{T}}\in\mathbb{R}^{mn}.\end{aligned}$$

3. THE MULTI-INNOVATION STOCHASTIC GRADIENT ALGORITHM

Define the parameter vector $\boldsymbol{\theta}$, the information matrix $\boldsymbol{\Psi}(t)$, and the fictitious output vectors $\boldsymbol{\xi}_1(t)$ and $\boldsymbol{\xi}_2(t)$ as

$$\begin{aligned}\boldsymbol{\theta}&:=[c_{11},c_{12},\cdots,c_{1n_c},c_{21},c_{22},\cdots,c_{2n_c},\cdots,\\ &\qquad c_{m1},c_{m2},\cdots,c_{mn_c}]^{\mathrm{T}}\in\mathbb{R}^{mn_c},\\ \boldsymbol{\Psi}(t)&:=\mathrm{blockdiag}[\boldsymbol{f}^{\mathrm{T}}(u_1(t)),\cdots,\boldsymbol{f}^{\mathrm{T}}(u_m(t))]\in\mathbb{R}^{m\times(mn_c)},\\ \boldsymbol{\xi}_1(t)&:=\boldsymbol{y}(t)-\bar{\boldsymbol{u}}(t)=\boldsymbol{y}(t)-\boldsymbol{\Psi}(t)\boldsymbol{\theta},\\ \boldsymbol{\xi}_2(t)&:=\boldsymbol{y}(t)-\boldsymbol{\vartheta}^{\mathrm{T}}\boldsymbol{\phi}(t).\end{aligned}$$

From (5), we can obtain two subsystems:

$$\mathrm{S}_1:\ \boldsymbol{\xi}_1(t)=\boldsymbol{\vartheta}^{\mathrm{T}}\boldsymbol{\phi}(t)+\boldsymbol{v}(t), \tag{6}$$

$$\mathrm{S}_2:\ \boldsymbol{\xi}_2(t)=\boldsymbol{\Psi}(t)\boldsymbol{\theta}+\boldsymbol{v}(t). \tag{7}$$

Notice that the associate parameters between two subsystems are c_{ij}. Define the parameter estimation matrix $\hat{\boldsymbol{\vartheta}}(t)$ of $\boldsymbol{\vartheta}$, and the parameter estimation vector $\hat{\boldsymbol{\theta}}(t)$ of $\boldsymbol{\theta}$ at current time t as

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}^{\mathrm{T}}(t)&:=[\hat{\boldsymbol{A}}_1(t),\hat{\boldsymbol{A}}_2(t),\cdots,\hat{\boldsymbol{A}}_{n_a}(t),\\ &\qquad\hat{\boldsymbol{B}}_1(t),\hat{\boldsymbol{B}}_2(t),\cdots,\hat{\boldsymbol{B}}_{n_b}(t)]\in\mathbb{R}^{m\times(mn)},\\ \hat{\boldsymbol{\theta}}(t)&:=[\hat{c}_{11}(t),\cdots,\hat{c}_{1n_c}(t),\hat{c}_{21}(t),\cdots,\hat{c}_{2n_c}(t),\cdots,\\ &\qquad\hat{c}_{m1}(t),\cdots,\hat{c}_{mn_c}(t)]^{\mathrm{T}}\in\mathbb{R}^{mn_c}.\end{aligned}$$

In order to clearly present the multi-innovation stochastic gradient (MISG) algorithm in this paper, it is necessary to introduce the stochastic gradient (SG) identification algorithm first. Using the gradient search, we can obtain the SG identification algorithm for estimating the parameter matrix $\boldsymbol{\vartheta}$ and the parameter vector $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\vartheta}}(t)=\hat{\boldsymbol{\vartheta}}(t-1)+\frac{\boldsymbol{\phi}(t)}{r_1(t)}\boldsymbol{e}_1^{\mathrm{T}}(t), \tag{8}$$

$$\boldsymbol{e}_1(t)=\boldsymbol{\xi}_1(t)-\hat{\boldsymbol{\vartheta}}^{\mathrm{T}}(t-1)\boldsymbol{\phi}(t), \tag{9}$$

$$r_1(t)=r_1(t-1)+\|\boldsymbol{\phi}(t)\|^2,\ r_1(0)=1, \tag{10}$$

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\frac{\boldsymbol{\Psi}^{\mathrm{T}}(t)}{r_2(t)}\boldsymbol{e}_2(t), \tag{11}$$

$$\boldsymbol{e}_2(t)=\boldsymbol{\xi}_2(t)-\boldsymbol{\Psi}(t)\hat{\boldsymbol{\theta}}(t-1), \tag{12}$$

$$r_2(t)=r_2(t-1)+\|\boldsymbol{\Psi}(t)\|^2,\ r_2(0)=1. \tag{13}$$

Here, the norm of a matrix $\boldsymbol{X}$ is defined by $\|\boldsymbol{X}\|^2:=\mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$, $\boldsymbol{e}_1(t)\in\mathbb{R}^m$ and $\boldsymbol{e}_2(t)\in\mathbb{R}^m$ represent innovation vectors and each element of $\boldsymbol{e}_1(t)$ and $\boldsymbol{e}_2(t)$ is a scalar innovation at the current time.

To enhance the convergence rate of the SG algorithm, we derive an MISG algorithm with high computational efficiency and high accuracy for Subsystems (6) and (7) by expanding the single innovation vectors $\boldsymbol{e}_1(t)$ and $\boldsymbol{e}_2(t)$ to an innovation matrix $\boldsymbol{E}_1(p,t) \in \mathbb{R}^{p\times m}$ and an innovation vector $\boldsymbol{E}_2(p,t) \in \mathbb{R}^{mp}$ as

$$\boldsymbol{E}_1(p,t) := \begin{bmatrix} \boldsymbol{\xi}_1^{\mathrm{T}}(t) - \boldsymbol{\phi}^{\mathrm{T}}(t)\hat{\boldsymbol{\vartheta}}(t-1) \\ \boldsymbol{\xi}_1^{\mathrm{T}}(t-1) - \boldsymbol{\phi}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\vartheta}}(t-1) \\ \vdots \\ \boldsymbol{\xi}_1^{\mathrm{T}}(t-p+1) - \boldsymbol{\phi}^{\mathrm{T}}(t-p+1)\hat{\boldsymbol{\vartheta}}(t-1) \end{bmatrix}, \tag{14}$$

$$\boldsymbol{E}_2(p,t) := \begin{bmatrix} \boldsymbol{\xi}_2(t) - \boldsymbol{\Psi}(t)\hat{\boldsymbol{\theta}}(t-1) \\ \boldsymbol{\xi}_2(t-1) - \boldsymbol{\Psi}(t-1)\hat{\boldsymbol{\theta}}(t-1) \\ \vdots \\ \boldsymbol{\xi}_2(t-p+1) - \boldsymbol{\Psi}(t-p+1)\hat{\boldsymbol{\theta}}(t-1) \end{bmatrix}, \tag{15}$$

where p represents the innovation length.

Define the information matrices $\boldsymbol{\Phi}(p,t)$ and $\boldsymbol{\Omega}(p,t)$ and the stacking output matrix/vector $\boldsymbol{Y}_1(p,t)/\boldsymbol{Y}_2(p,t)$ as

$$\begin{aligned}
\boldsymbol{\Phi}(p,t) &:= [\boldsymbol{\phi}(t), \boldsymbol{\phi}(t-1), \cdots, \boldsymbol{\phi}(t-p+1)] \in \mathbb{R}^{mn\times p}, \\
\boldsymbol{\Omega}(p,t) &:= [\boldsymbol{\Psi}^{\mathrm{T}}(t), \cdots, \boldsymbol{\Psi}^{\mathrm{T}}(t-p+1)] \in \mathbb{R}^{(mn_c)\times(mp)}, \\
\boldsymbol{Y}_1(p,t) &:= [\boldsymbol{\xi}_1(t), \boldsymbol{\xi}_1(t-1), \cdots, \boldsymbol{\xi}_1(t-p+1)]^{\mathrm{T}} \in \mathbb{R}^{p\times m}, \\
\boldsymbol{Y}_2(p,t) &:= [\boldsymbol{\xi}_2^{\mathrm{T}}(t), \boldsymbol{\xi}_2^{\mathrm{T}}(t-1), \cdots, \boldsymbol{\xi}_2^{\mathrm{T}}(t-p+1)]^{\mathrm{T}} \in \mathbb{R}^{mp}.
\end{aligned}$$

From (14)–(15), we have

$$\begin{aligned}
\boldsymbol{E}_1(p,t) &= \boldsymbol{Y}_1(p,t) - \boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1), \\
\boldsymbol{E}_2(p,t) &= \boldsymbol{Y}_2(p,t) - \boldsymbol{\Omega}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1).
\end{aligned}$$

The MISG algorithm with the innovation length p can be expressed as

$$\begin{aligned}
\hat{\boldsymbol{\vartheta}}(t) &= \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\Phi}(p,t)}{r_1(t)}\boldsymbol{E}_1(p,t) \\
&= \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\Phi}(p,t)}{r_1(t)} \\
&\quad \times[\boldsymbol{Y}_1(p,t) - \boldsymbol{\Phi}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1)], \qquad (16) \\
r_1(t) &= r_1(t-1) + \|\boldsymbol{\phi}(t)\|^2,\ r_1(0) = 1, \qquad (17) \\
\hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\Omega}(p,t)}{r_2(t)}\boldsymbol{E}_2(p,t) \\
&= \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\Omega}(p,t)}{r_2(t)} \\
&\quad \times[\boldsymbol{Y}_2(p,t) - \boldsymbol{\Omega}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)], \qquad (18) \\
r_2(t) &= r_2(t-1) + \|\boldsymbol{\Psi}(t)\|^2,\ r_2(0) = 1. \qquad (19)
\end{aligned}$$

Because $\boldsymbol{\Phi}(p,t)$ contains the unknown vectors $\boldsymbol{\phi}(t)$, $\boldsymbol{Y}_1(p,t)$ contains the unknown parameter vector $\boldsymbol{\theta}$ and $\boldsymbol{Y}_2(p,t)$ contains the unknown parameter matrix $\boldsymbol{\vartheta}$, so the algorithm in (16)–(19) cannot be implemented. The common solution is to replace the unmeasurable inner vectors $\boldsymbol{x}(t-i)$ and $\bar{\boldsymbol{u}}(t-i)$ with their estimates $\hat{\boldsymbol{x}}(t-i)$ and $\hat{\bar{\boldsymbol{u}}}(t-i)$, and define

$$\begin{aligned}
\hat{\boldsymbol{\Phi}}(p,t) &:= [\hat{\boldsymbol{\phi}}(t), \hat{\boldsymbol{\phi}}(t-1), \cdots, \hat{\boldsymbol{\phi}}(t-p+1)] \in \mathbb{R}^{(mn)\times p}, \\
\hat{\boldsymbol{\xi}}_1(t) &:= \boldsymbol{y}(t) - \boldsymbol{\Psi}(t)\hat{\boldsymbol{\theta}}(t-1) \in \mathbb{R}^m, \\
\hat{\boldsymbol{\xi}}_2(t) &:= \boldsymbol{y}(t) - \hat{\boldsymbol{\vartheta}}^{\mathrm{T}}(t-1)\hat{\boldsymbol{\phi}}(t) \in \mathbb{R}^m, \\
\hat{\boldsymbol{Y}}_1(p,t) &:= [\hat{\boldsymbol{\xi}}_1(t), \hat{\boldsymbol{\xi}}_1(t-1), \cdots, \hat{\boldsymbol{\xi}}_1(t-p+1)]^{\mathrm{T}} \in \mathbb{R}^{p\times m}, \\
\hat{\boldsymbol{Y}}_2(p,t) &:= [\hat{\boldsymbol{\xi}}_2^{\mathrm{T}}(t), \hat{\boldsymbol{\xi}}_2^{\mathrm{T}}(t-1), \cdots, \hat{\boldsymbol{\xi}}_2^{\mathrm{T}}(t-p+1)]^{\mathrm{T}} \in \mathbb{R}^{mp}.
\end{aligned}$$

In order to improve the convergence rate of the MISG algorithm, we introduce a forgetting factor λ, so Equations (17) and (19) can be rewritten,

$$\begin{aligned}
r_1(t) &= \lambda r_1(t-1) + \|\hat{\boldsymbol{\phi}}(t)\|^2,\ r_1(0) = 1, \\
r_2(t) &= \lambda r_2(t-1) + \|\boldsymbol{\Psi}(t)\|^2,\ r_2(0) = 1.
\end{aligned}$$

Replacing $\boldsymbol{Y}_1(p,t)$ and $\boldsymbol{\Phi}(p,t)$ in (16) with their estimates $\hat{\boldsymbol{Y}}_1(p,t)$ and $\hat{\boldsymbol{\Phi}}(p,t)$, $\boldsymbol{\phi}(t)$ in (17) with $\hat{\boldsymbol{\phi}}(t)$, $\boldsymbol{Y}_2(p,t)$ in (18) with $\hat{\boldsymbol{Y}}_2(p,t)$, we can obtain the MISG algorithm for computing $\hat{\boldsymbol{\vartheta}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$:

$$\begin{aligned}
\hat{\boldsymbol{\vartheta}}(t) &= \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}(p,t)}{r_1(t)} \\
&\quad \times[\hat{\boldsymbol{Y}}_1(p,t) - \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1)], \qquad (20) \\
r_1(t) &= \lambda r_1(t-1) + \|\hat{\boldsymbol{\phi}}(t)\|^2,\ r_1(0) = 1, \qquad (21) \\
\hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\Omega}(p,t)}{r_2(t)} \\
&\quad \times[\hat{\boldsymbol{Y}}_2(p,t) - \boldsymbol{\Omega}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\theta}}(t-1)], \qquad (22) \\
r_2(t) &= \lambda r_2(t-1) + \|\boldsymbol{\Psi}(t)\|^2,\ r_2(0) = 1, \qquad (23) \\
\hat{\boldsymbol{\phi}}(t) &= [-\hat{\boldsymbol{x}}^{\mathrm{T}}(t-1), -\hat{\boldsymbol{x}}^{\mathrm{T}}(t-2), \cdots, -\hat{\boldsymbol{x}}^{\mathrm{T}}(t-n_a), \\
&\quad \hat{\bar{\boldsymbol{u}}}^{\mathrm{T}}(t-1), \hat{\bar{\boldsymbol{u}}}^{\mathrm{T}}(t-2), \cdots, \hat{\bar{\boldsymbol{u}}}^{\mathrm{T}}(t-n_b)]^{\mathrm{T}}, \qquad (24) \\
\hat{\boldsymbol{\Phi}}(p,t) &= [\hat{\boldsymbol{\phi}}(t), \hat{\boldsymbol{\phi}}(t-1), \cdots, \hat{\boldsymbol{\phi}}(t-p+1)], \qquad (25) \\
\boldsymbol{\Psi}(t) &= \mathrm{blockdiag}[\boldsymbol{f}^{\mathrm{T}}(u_1(t)), \cdots, \boldsymbol{f}^{\mathrm{T}}(u_m(t))], \qquad (26) \\
\boldsymbol{\Omega}(p,t) &= [\boldsymbol{\Psi}^{\mathrm{T}}(t), \boldsymbol{\Psi}^{\mathrm{T}}(t-1), \cdots, \boldsymbol{\Psi}^{\mathrm{T}}(t-p+1)], \qquad (27) \\
\hat{\boldsymbol{\xi}}_1(t) &= \boldsymbol{y}(t) - \boldsymbol{\Psi}(t)\hat{\boldsymbol{\theta}}(t-1), \qquad (28) \\
\hat{\boldsymbol{\xi}}_2(t) &= \boldsymbol{y}(t) - \hat{\boldsymbol{\vartheta}}^{\mathrm{T}}(t)\hat{\boldsymbol{\phi}}(t), \qquad (29) \\
\hat{\boldsymbol{Y}}_1(p,t) &= [\hat{\boldsymbol{\xi}}_1(t), \hat{\boldsymbol{\xi}}_1(t-1), \cdots, \hat{\boldsymbol{\xi}}_1(t-p+1)]^{\mathrm{T}}, \qquad (30) \\
\hat{\boldsymbol{Y}}_2(p,t) &= [\hat{\boldsymbol{\xi}}_2^{\mathrm{T}}(t), \hat{\boldsymbol{\xi}}_2^{\mathrm{T}}(t-1), \cdots, \hat{\boldsymbol{\xi}}_2^{\mathrm{T}}(t-p+1)]^{\mathrm{T}}, \qquad (31) \\
\hat{\bar{\boldsymbol{u}}}(t) &= \boldsymbol{\Psi}(t)\hat{\boldsymbol{\theta}}(t), \qquad (32) \\
\hat{\boldsymbol{x}}(t) &= \hat{\boldsymbol{\vartheta}}^{\mathrm{T}}(t)\hat{\boldsymbol{\phi}}(t) + \hat{\bar{\boldsymbol{u}}}(t). \qquad (33)
\end{aligned}$$

The procedures of computing $\hat{\boldsymbol{\vartheta}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ in the MISG algorithm are listed in the following.

1) Let $t = 1$, set the initial values $\lambda = 0.99$, $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_{m\times(mn)}/p_0$, $\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_{mn_c}/p_0$, $\hat{\boldsymbol{x}}(i) = \mathbf{0}$ and $\hat{\bar{\boldsymbol{u}}}(i) = \mathbf{0}$ for $i \leqslant 0$ (p_0 is a large number, e.g., $p_0 = 10^6$).
2) Collect the input-output data $u_i(t)$ and $y_i(t)$, form $\hat{\boldsymbol{\phi}}(t)$, $\hat{\boldsymbol{\Phi}}(p,t)$, $\boldsymbol{\Psi}(t)$ and $\boldsymbol{\Omega}(p,t)$ using (24)–(27), respectively.
3) Compute $\hat{\boldsymbol{\xi}}_1(t)$ using (28) and form $\hat{\boldsymbol{Y}}_1(p,t)$ using (30), compute $r_1(t)$ using (21) and update $\hat{\boldsymbol{\vartheta}}(t)$ using (20).
4) Compute $\hat{\boldsymbol{\xi}}_2(t)$ using (29) and form $\hat{\boldsymbol{Y}}_2(p,t)$ using (31), compute $r_2(t)$ using (23) and update $\hat{\boldsymbol{\theta}}(t)$ using (22).

5) Compute $\hat{\boldsymbol{u}}(t)$ using (32) and $\hat{\boldsymbol{x}}(t)$ using (33).
6) Increase t by 1 and go to Step 2.

4. EXAMPLE

Consider the following Hammerstein MIMO OE system,

$$\boldsymbol{y}(t)=\boldsymbol{x}(t)+\boldsymbol{v}(t),$$
$$\boldsymbol{x}(t)=\boldsymbol{A}^{-1}(z)\boldsymbol{B}(z)\bar{\boldsymbol{u}}(t),$$
$$\boldsymbol{y}(t)=\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}, \quad \boldsymbol{u}(t)=\begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}, \quad \boldsymbol{v}(t)=\begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix},$$
$$\boldsymbol{A}(z)=\boldsymbol{I}_2+\boldsymbol{A}_1 z^{-1}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}+\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} z^{-1}$$
$$=\begin{bmatrix} 1+0.39z^{-1} & 0.10z^{-1} \\ 0.33z^{-1} & 1-0.76z^{-1} \end{bmatrix},$$
$$\boldsymbol{B}(z)=\boldsymbol{I}_2+\boldsymbol{B}_1 z^{-1}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}+\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} z^{-1}$$
$$=\begin{bmatrix} 1+1.18z^{-1} & 2.39z^{-1} \\ 0.79z^{-1} & 1+1.53z^{-1} \end{bmatrix},$$
$$\bar{\boldsymbol{u}}(t)=\begin{bmatrix} \bar{u}_1(t) \\ \bar{u}_2(t) \end{bmatrix}=\begin{bmatrix} c_{11}f_1(u_1(t))+c_{12}f_2(u_1(t)) \\ c_{21}f_1(u_2(t))+c_{22}f_2(u_2(t)) \end{bmatrix}$$
$$=\begin{bmatrix} -2.25f_1(u_1(t))+1.24f_2(u_1(t)) \\ 0.76f_1(u_2(t))+1.30f_2(u_2(t)) \end{bmatrix},$$

where $f_1(u_i(t))=u_i^2(t)$ and $f_2(u_i(t))=u_i(t)$, the parameter matrix $\boldsymbol{\vartheta}$ and the parameter vector $\boldsymbol{\theta}$ are

$$\boldsymbol{\vartheta}^{\mathrm{T}}=[\boldsymbol{A}_1,\boldsymbol{B}_1]=\begin{bmatrix} 0.39 & 0.10 & 1.18 & 2.39 \\ 0.33 & \text{-}0.76 & 0.79 & 1.53 \end{bmatrix},$$
$$\boldsymbol{\theta}=[c_{11},c_{12},c_{21},c_{22}]^{\mathrm{T}}=[-2.25,1.24,0.76,1.30]^{\mathrm{T}}.$$

The inputs $\{u_1(t)\}$ and $\{u_2(t)\}$ are taken as two persistent excitation sequences with zero mean and unit variances, $\{v_1(t)\}$ and $\{v_2(t)\}$ are taken as two white noise sequences with zero mean and variances $\sigma_1^2=0.10^2$ for $v_1(t)$ and $\sigma_2^2=0.10^2$ for $v_2(t)$. Applying the MISG algorithm in (20)–(33) to estimate parameters of this example system, the parameter estimates and the estimation errors δ are shown in Tables 1–3 and Figure 1, where

$$\delta:=\sqrt{\frac{\|\hat{\boldsymbol{\vartheta}}(t)-\boldsymbol{\vartheta}\|^2+\|\hat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}\|^2}{\|\boldsymbol{\vartheta}\|^2+\|\boldsymbol{\theta}\|^2}}\times 100\%.$$

From Tables 1–3 and Figure 1, it is clear that the parameter estimation errors given by the MISG algorithm become smaller with t increasing. The MISG algorithm can give more accurate parameter estimates than the SG algorithm, and the MISG estimation errors become smaller with the increasing of the innovation length p. The proposed algorithm can reduce computational load by decomposing a multi-variable system into two subsystems and has high computational efficiencies in contrast to the over-parameterization least squares methods by the key-term separation principle.

To be more exactly, using the over-parameterization method for the second-order Hammerstein MIMO output error system ($m=2$), we obtain complex identification model:

$$y_i(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}_i+\boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\vartheta}_i+v_i(t),\ i=1,2$$

$$\boldsymbol{\varphi}(t):=[-\boldsymbol{x}^{\mathrm{T}}(t-1),\cdots,-\boldsymbol{x}^{\mathrm{T}}(t-n_a)]^{\mathrm{T}}\in\mathbb{R}^{2n_a},$$
$$\boldsymbol{\theta}_i:=[\boldsymbol{a}_i^{1\mathrm{T}},\boldsymbol{a}_i^{2\mathrm{T}},\cdots,\boldsymbol{a}_i^{n_a\mathrm{T}}]^{\mathrm{T}}\in\mathbb{R}^{2n_a},$$
$$\boldsymbol{a}_i^j:=[a_{i1}^j,a_{i2}^j]^{\mathrm{T}},\ j=1,2,\cdots,n_a$$
$$\boldsymbol{\phi}(t):=[\boldsymbol{f}^{\mathrm{T}}(u_1(t)),\boldsymbol{f}^{\mathrm{T}}(u_2(t)),\cdots,\boldsymbol{f}^{\mathrm{T}}(u_1(t-n_b)),$$
$$\boldsymbol{f}^{\mathrm{T}}(u_2(t-n_b))]^{\mathrm{T}}\in\mathbb{R}^{2n_bn_c},$$
$$\boldsymbol{f}(u_i(t-k)):=[f_1(u_i(t-k)),f_2(u_i(t-k)),\cdots,$$
$$f_{n_c}(u_i(t-k))]^{\mathrm{T}}\in\mathbb{R}^{n_c},\ k=1,2,\cdots,n_b$$
$$\boldsymbol{\vartheta}_i:=[b_{i1}^0\boldsymbol{c}_1^{\mathrm{T}},b_{i2}^0\boldsymbol{c}_2^{\mathrm{T}},\cdots,b_{i1}^{n_b}\boldsymbol{c}_1^{\mathrm{T}},b_{i2}^{n_b}\boldsymbol{c}_2^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{R}^{2n_bn_c},$$
$$\boldsymbol{c}_i:=[c_{i1},c_{i2},\cdots,c_{in_c}]^{\mathrm{T}}\in\mathbb{R}^{n_c}.$$

Notice that there are $4n_a+4n_bn_c$ parameters to be identified, which is greater than $4(n_a+n_b)+2n_c$ parameters in the actual system for $n_b\geqslant 1$ and $n_c\geqslant 2$.

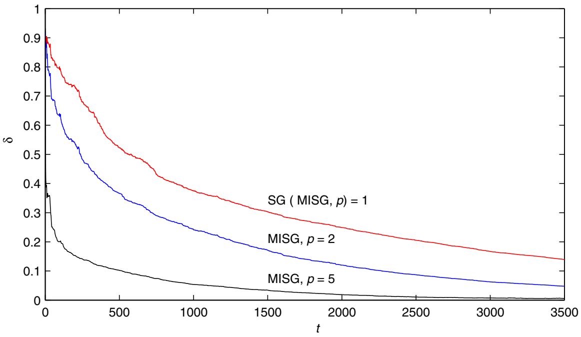

Fig. 1. The H-MISG estimation errors δ versus t

5. CONCLUSIONS

This paper discusses identification problems for a Hammerstein MIMO output error system. The identification model is formulated by using the key term separation principle, and a multi-innovation stochastic gradient algorithm is proposed to estimate the parameters of the system. Using the key term separation principle we can obtain unique parameter estimates directly, which means reduce the redundant parameters to improve the computational efficiency. The simulation results verify the proposed method.

REFERENCES

Alevizos P.D. (2013). An optimal algorithm for computing the non-trivial circuits of a union of iso-oriented rectangles. *Information Processing Letters*, 113(3), 55-59.

Averin G.R. (2003). The Hammerstein-Wiener model for identification of stochastic systems. *Automation and Remote Control*, 64(9), 1418-1413.

Carini A., Sicuranza G.L., and Mathews V.J. (2013). Efficient adaptive identification of linear-in-the-parameters nonlinear filters using periodic input sequences. *Signal Processing*, 93(5), 1210-1220.

Chen H.B., Zhang W.G., and Ding F. (2014). Data filtering based least squares iterative algorithm for parameter identification of output error autoregressive systems. *Information Processing Letters*, 114(10), 573-578.

Ding J., Fan C.X., and Lin J.X. (2013). Auxiliary model based parameter estimation for dual-rate output error

Table 1. The SG estimates and errors

t	100	200	500	1000	2000	3000
$a_{11} = 0.39$	0.55845	0.59989	0.33743	0.20952	0.26741	0.35306
$a_{12} = 0.10$	1.13342	0.85411	0.55758	0.21929	0.12845	0.12664
$b_{11} = 1.18$	-0.15865	-0.13555	0.09474	0.40273	0.72637	0.96039
$b_{12} = 2.39$	0.63605	0.72725	0.94059	1.23714	1.56794	1.79092
$a_{21} = 0.33$	0.13492	0.14114	0.06319	-0.00685	0.07708	0.16783
$a_{22} = -0.76$	-0.60019	-0.64795	-0.72329	-0.69845	-0.65904	-0.64284
$b_{21} = 0.79$	-0.15689	-0.14304	-0.03109	0.13895	0.31684	0.47425
$b_{22} = 1.53$	1.41326	1.44280	1.50654	1.56976	1.60099	1.59480
$c_{11} = -2.25$	0.00349	-0.17849	-1.24422	-1.77991	-2.19590	-2.24719
$c_{12} = 1.24$	1.03730	0.99041	1.10142	1.31417	1.37330	1.25128
$c_{21} = 0.76$	0.51654	0.50175	0.44244	0.61155	0.72286	0.75612
$c_{22} = 1.30$	0.71219	0.88182	1.13677	1.24468	1.35963	1.30310
$\delta(\%)$	79.36473	73.29598	52.22400	37.44584	24.89365	16.72228

Table 2. The MISG estimates and errors with $p = 2$

t	100	200	500	1000	2000	3000
$a_{11} = 0.39$	0.07116	0.04501	0.02790	0.22471	0.41486	0.41686
$a_{12} = 0.10$	0.51165	0.34295	0.15830	0.12580	0.09997	0.08520
$b_{11} = 1.18$	0.04675	0.13262	0.40288	0.70401	1.02286	1.17091
$b_{12} = 2.39$	0.71269	0.88721	1.23316	1.60074	1.94323	2.13110
$a_{21} = 0.33$	0.07297	0.03570	-0.00555	0.09765	0.24674	0.29118
$a_{22} = -0.76$	-0.51441	-0.68416	-0.73023	-0.70783	-0.72255	-0.73852
$b_{21} = 0.79$	0.01315	0.08014	0.23767	0.39085	0.58292	0.70650
$b_{22} = 1.53$	1.08999	1.17613	1.29560	1.40032	1.47067	1.50218
$c_{11} = -2.25$	-0.76437	-1.06362	-1.96017	-2.15563	-2.25214	-2.26335
$c_{12} = 1.24$	1.40896	1.21716	1.34583	1.40573	1.30403	1.23590
$c_{21} = 0.76$	0.85135	0.78706	0.69434	0.77797	0.73501	0.74712
$c_{22} = 1.30$	0.65174	1.02584	1.28701	1.29156	1.32601	1.29681
$\delta(\%)$	63.41373	53.85604	36.59621	24.07816	12.00300	6.28417

Table 3. The MISG estimates and errors with $p = 5$

t	100	200	500	1000	2000	3000
$a_{11} = 0.39$	0.50098	0.48646	0.46849	0.43182	0.41151	0.39511
$a_{12} = 0.10$	0.06477	0.05526	0.05654	0.07234	0.08143	0.09439
$b_{11} = 1.18$	1.36225	1.33118	1.27629	1.25199	1.20337	1.18657
$b_{12} = 2.39$	1.81148	1.92662	2.06705	2.20690	2.32397	2.36280
$a_{21} = 0.33$	0.08178	0.13757	0.22748	0.28012	0.32285	0.32739
$a_{22} = -0.76$	-0.72378	-0.75076	-0.73493	-0.75379	-0.76104	-0.76267
$b_{21} = 0.79$	0.40431	0.46801	0.58577	0.70866	0.77319	0.78549
$b_{22} = 1.53$	1.21121	1.29345	1.37162	1.45683	1.50863	1.52360
$c_{11} = -2.25$	-2.01624	-2.16081	-2.26233	-2.22271	-2.23369	-2.25991
$c_{12} = 1.24$	1.25857	1.20812	1.30204	1.23358	1.23841	1.25021
$c_{21} = 0.76$	0.77929	0.84564	0.74709	0.74583	0.74837	0.75990
$c_{22} = 1.30$	1.08487	1.34324	1.35360	1.32646	1.28979	1.30319
$\delta(\%)$	20.06273	15.31056	10.21841	5.38646	1.88270	0.75524

systems with colored noise. *Applied Mathematical Modelling*, 37(6), 4051-4058.

Ding J., and Lin J.X. (2014). Modified subspace identification for periodically non-uniformly sampled systems by using the lifting technique. *Circuits Systems and Signal Processing*, 33(5), 1439-1449.

Ding F., Liu X.P., and Liu G. (2010). Gradient based and least-squares based iterative identification methods for OE and OEMA systems. *Digital Signal Processing*, 20(3), 664-677.

Haryanto A., and Hong K.S. (2013). Maximum likelihood identification of Wiener-Hammerstein models. *Mechanical Systems and Signal Processing*, 41(1-2), 54-70.

He J., Xu B., and Masri S.F. (2012). Restoring force and dynamic loadings identification for a nonlinear chain-like structure with partially unknown excitations. *Nonlinear Dynamics*, 69(1-2), 231-245.

Hong X., Mitchell R.J., and Chen S. (2013). System identification of Wiener systems with B-spline functions using De Boor recursion. *International Journal of Systems Science*, 44(9), 1666-1674.

Hu Y.B., Liu B.L., and Zhou Q. (2014). A multi-innovation generalized extended stochastic gradient algorithm for output nonlinear autoregressive moving average systems. *Applied Mathematics and Computation*, 247, 218-224.

Hu Y.B., Liu B.L., Zhou Q., and Yang C. (2014). Recursive extended least squares parameter estimation for Wiener nonlinear systems with moving average noises. *Circuits, Systems and Signal Processing*, 33(2), 655-664.

Ikhouane F., and Giri F. (2014). A unified approach for the parametric identification of SISO/MIMO Wiener and Hammerstein systems. *Journal of The Franklin Institute*, 351(3), 1717-1727.

Jiang H.Q., Yan Z.B., and Liu X.G. (2013). Melt index prediction using optimized least squares support vector machines based on hybrid particle swarm optimization algorithm. *Neurocomputing*, 119, 469-477.

Kim B.H. (2014). Extracting modal parameters of a cable on shaky motion pictures. *Mechanical Systems and Signal Processing*, 49(1-2), 3-12.

Li H.P., and Shi Y. (2012). Robust H-infinity filtering for nonlinear stochastic systems with uncertainties and Markov delays. *Automatica*, 48(1), 159-166.

Lindstena F., Schöna T.B., and Jordan M.I. (2013). Bayesian semiparametric Wiener system identification. *Automatica*, 49(7), 2053-2063.

Liu Y.J., and Wang D.Q. (2010). Least squares based iterative algorithms for identifying Box-Jenkins models with finite measurement data. *Digital Signal Processing*, 20(5), 1458-1467.

Mileounis G., and Kalouptsidis N. (2013). A sparsity driven approach to cumulant based identification and order determination. *Signal Processing*, 93(7), 1892-1899.

Piga D., and Toth R. (2014). A bias-corrected estimator for nonlinear systems with output-error type model structures. *Automatica*, 50(9), 2373-2380.

Pintelona R., Schoukensa J., and Guillaumeb P. (2007). Box-Jenkins identification revisited–Part III: Multivariable systems. *Automatica*, 43(5), 868-875.

Rebillat M., Hajrya R., and Mechbal N. (2014). Nonlinear structural damage detection based on cascade of Hammerstein models. *Mechanical Systems and Signal Processing*, 48(1-2), 247-259.

Scarpiniti M., Comminiello D., Parisi R. and Uncini A. (2013). Nonlinear spline adaptive filtering. *Signal Processing*, 93(4), 772-783.

Schoukens M., Pintelon R., and Rolain Y. (2014). Identification of Wiener-Hammerstein systems by a nonparametric separation of the best linear approximation. *Automatica*, 50(2), 628-634.

Shen Q.Y., and Ding F. (2014). Iterative estimation methods for Hammerstein controlled autoregressive moving average systems based on the key-term separation principle. *Nonlinear Dynamics*, 75(4), 709-716.

Shi Y., and Yu B. (2011). Robust mixed H_2/H_infinity control of networked control systems with random time delays in both forward and backward communication links. *Automatica*, 47(4), 754-760.

Söderström T., Hong M., Schoukens J., and Pintelon R. (2010). Accuracy analysis of time domain maximum likelihood method and sample maximum likelihood method for errors-in-variables and output error identification. *Automatica*, 46(4), 721-727.

Sun J.L., and Liu X.G. (2013). A novel APSO-aided maximum likelihood identification method for Hammerstein systems. *Nonlinear Dynamics*, 73(1-2), 449-462.

Vanbeylen L. (2014). A fractional approach to identify Wiener-Hammerstein systems. *Automatica*, 50(3), 903-909.

Vicario F., Phan M.Q., Betti R., and Longman R.W. (2014). Linear state representations for identification of bilinear discrete-time models by interaction matrices. *Nonlinear Dynamics*, 77(4), 1561-1576.

Wang D.Q. (2011). Least squares-based recursive and iterative estimation for output error moving average systems using data filtering. *IET Control Theory and Application*, 5(14), 1648-1657.

Wang C., and Tang T. (2014). Several gradient-based iterative estimation algorithms for a class of nonlinear systems using the filtering technique. *Nonlinear Dynamics*, 77(3), 769-780.

Wang H.K., Zhao J., Xu Z.H., and Shao Z.J. (2014). Model predictive control for Hammerstein systems with unknown input nonlinearities. *Industrial & Engineering Chemistry Research*, 53(18), 7714-7722.

Wills A., Schon T.B., Ljung L., and Ninness B. (2013). Identification of Hammerstein-Wiener models. *Automatica*, 49(1), 70-81.

Xu L. (2014). A proportional differential control method for a time-delay system using the Taylor expansion approximation. *Applied Mathematics and Computation*, 236, 391-399.

Xu L., Chen L., Xiong W.L., and Yuan P. (2015). Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. *Nonlinear Dynamics*, 79(3), 2155-2163.

Yu F., Mao Z.Z., Jia M.X., and Yuan P. (2014). Recursive parameter identification of Hammerstein-Wiener systems with measurement noise. *Signal Processing*, (105), 137-147.

Zhang M.M., and Liu X.G. (2013). Melt index prediction by fuzzy functions and weighted least squares support vector machines. *Chemical Engineering & Technology*, 36(9), 1577-1584.

Zhang Y., and Cui G.M. (2011). Bias compensation methods for stochastic systems with colored noise. *Applied Mathematical Modelling*, 35(4), 1709-1716.

Zhao Y., and Qin S.J. (2014). Subspace identification with non-steady Kalman filter parameterization. *Journal of Process Control*, 24(9), 1337-1345.

