CrossMark

Recursive parameter identification of the dynamical models for bilinear state space systems

Xiao Zhang · Feng Ding · Fuad E. Alsaadi · Tasawar Hayat

Abstract This paper investigates the recursive parameter and state estimation algorithms for a special class of nonlinear systems (i.e., bilinear state space systems). A state observer-based stochastic gradient (O-SG) algorithm is presented for the bilinear state space systems by using the gradient search. In order to improve the parameter estimation accuracy and the convergence rate of the O-SG algorithm, a state observer-based multi-innovation stochastic gradient algorithm and a state observer-based recursive least squares identification algorithm are derived by means of the multi-innovation theory. Finally, a numerical example is provided to demonstrate the effectiveness of the proposed algorithms.

Keywords Dynamical system · Parameter estimation · State estimation · Multi-innovation theory · State space model · Bilinear system

X. Zhang · F. Ding (✉) · F. E. Alsaadi · T. Hayat
Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, People's Republic of China
e-mail: fding@jiangnan.edu.cn

X. Zhang
e-mail: xzhang13@126.com

F. E. Alsaadi
e-mail: fuad_alsaadi@yahoo.com

T. Hayat
e-mail: fmgpak@gmail.com

F. Ding
College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266042, People's Republic of China

F. Ding · F. E. Alsaadi · T. Hayat
Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

T. Hayat
Department of Mathematics, Quaid-I-Azam University, Islamabad 44000, Pakistan

1 Introduction

System identification is the methodology of establishing mathematical models [1–3] and has wide applications in many areas such as linear system modeling [4,5] and nonlinear system modeling [6–9]. The identification of linear systems has reached a high level of maturity [10], and nonlinear systems generally exist in industry areas, so their identification has received extensive attention. Bilinear systems are considered as a special class of nonlinear systems and are linear to the state and the control input, respectively, but not to them jointly. The bilinear system is a simple nonlinear extension of a linear system. So it is necessary to introduce some work about the identification of nonlinear systems such as Hammerstein systems. Many nonlinear parameter estimation methods [11–13] have been developed, such as the subspace methods [14,15], the hierarchical methods [16,17] and the key term separation methods [18]. For example, under the assumption of a white unobserved Gaus-

sian input signal, errorless output observations and an invertible nonlinearity, Vanbeylen et al. [19] proposed a maximum likelihood estimator for Hammerstein systems with output measurements. Ase and Katayama [20] presented a subspace-based method to identify the Wiener–Hammerstein benchmark model by using the orthogonal projection subspace method and the separable least squares method.

The parameter identification [21–23] and the design of state observers [24–26] for bilinear systems have been carried out throughout the years both for continuous-time bilinear systems [27,28] and for discrete-time bilinear systems [29,30]. In the literature, Jan et al. [31] utilized the block pulse functions for the bilinear system identification in order to reduce the computation time. Dai et al. [32] proposed a robust recursive least squares method for bilinear systems identification. dos Santos et al. [33] presented a subspace state space identification algorithm for multi-input multi-output bilinear systems driven by white noise inputs by utilizing the Kalman filter idea.

State space systems can describe not only the system input and output characteristics but also the system internal structure characteristics, and play an important role in dynamical system state estimation [34,35] and parameter estimation [36]. Many identification methods have been proposed for linear state space systems, but the identification of nonlinear state space systems is still difficult and has not been fully investigated. Schön et al. [37] derived an expectation maximization algorithm for the parameter estimation of nonlinear state space systems using a particle smoother. Marconato et al. [38] presented an identification method for nonlinear state space models on a benchmark problem based on the classical identification techniques and regression methods.

The least squares methods contain the recursive least squares algorithms [39–41] and the least squares-based iterative algorithms [42,43]. In the literature, Arablouei et al. [44] studied an unbiased recursive least squares algorithm for errors-in-variables systems utilizing the dichotomous coordinate-descent iterations. Xu et al. [45] proposed the recursive least squares and multi-innovation stochastic gradient parameter estimation methods for signal modeling. Wan et al. [46] presented a novel method for the T-wave alternans assessment based on the least squares curve fitting technique. This paper addresses the identification issue for the bilinear state space systems with unmeasurable state variables.

The basic idea is to transform a bilinear state space system into its observer canonical form and then to derive the recursive parameter estimation algorithms based on the multi-innovation theory and design the state observer to estimate the unknown states. The main contributions of this paper lie in the following.

- By using the gradient search and the state observer, we overcome the difficulty that there exists both unmeasurable states and unknown parameters in the identification models, and propose the state observer-based gradient identification algorithms for bilinear state space systems.
- By utilizing the current innovations and past innovations, this paper presents a state observer-based multi-innovation stochastic gradient (O-MISG) algorithm for bilinear state space systems to improve the parameter estimation accuracy and convergence rates.

To close this section, we give an outline of this paper. Section 2 derives the observer canonical state space model for bilinear systems. Section 3 introduces the identification model for the bilinear state space models. A state estimation-based recursive parameter identification algorithm is presented in Sect. 4. Section 5 provides an illustrative example for the results in this paper. Finally, some concluding remarks are given in Sect. 6.

2 The observer canonical state space model for bilinear systems

First of all, let us introduce some notation. "$A =: X$" or "$X := A$" stands for "A is defined as X"; the symbol I (I_n) represents an identity matrix of appropriate size ($n \times n$); z denotes a unit forward shift operator like $zx(t) = x(t + 1)$ and $z^{-1}x(t) = x(t - 1)$; the superscript T symbolizes the vector/matrix transpose; $\hat{\theta}(t)$ denotes the estimate of θ at time t.

Consider the bilinear state space system described by

$$\bar{x}(t + 1) = \bar{A}\bar{x}(t) + \bar{B}\bar{x}(t)u(t) + \bar{f}u(t), \quad (1)$$

$$y(t) = \bar{h}\bar{x}(t) + v(t), \quad (2)$$

where $\bar{x}(t) := [\bar{x}_1(t), \bar{x}_2(t), \ldots, \bar{x}_n(t)]^T \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}$ and $y(t) \in \mathbb{R}$ are the system input and output variables, $v(t) \in \mathbb{R}$ is a random noise with zero mean and variance σ^2, and $\bar{A} \in \mathbb{R}^{n \times n}$,

$\bar{B} \in \mathbb{R}^{n \times n}$, $\bar{f} \in \mathbb{R}^n$ and $\bar{h} \in \mathbb{R}^{1 \times n}$ are the parameter matrices/vectors of the system.

Suppose that the system in (1) and (2) is controllable and observable. The bilinear system can be transformed into an observer canonical form under the non-singular linear transformation such as $\bar{x}(t) = T_o x(t)$, where $T_o \in \mathbb{R}^{n \times n}$ stands for the non-singular matrix. The transformation is as follows.

$$
\begin{aligned}
x(t+1) &= T_o^{-1} \bar{x}(t+1) = T_o^{-1}[\bar{A}\bar{x}(t) \\
&\quad + \bar{B}\bar{x}(t)u(t) + \bar{f}u(t)] \\
&= T_o^{-1}\bar{A}\bar{x}(t) + T_o^{-1}\bar{B}\bar{x}(t)u(t) + T_o^{-1}\bar{f}u(t) \\
&= T_o^{-1}\bar{A}T_o x(t) + T_o^{-1}\bar{B}T_o x(t)u(t) + T_o^{-1}\bar{f}u(t) \\
&= Ax(t) + Bx(t)u(t) + fu(t), \qquad (3)
\end{aligned}
$$

$$
\begin{aligned}
y(t) &= \bar{h}\bar{x}(t) + v(t) \\
&= hx(t) + v(t), \qquad (4)
\end{aligned}
$$

where

$$
A := T_o^{-1}\bar{A}T_o = \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -a_{n-1} & 0 & \cdots & 0 & 1 \\ -a_n & 0 & \cdots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{n \times n},
$$

$$
B := T_o^{-1}\bar{B}T_o = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad b_i \in \mathbb{R}^{1 \times n},
$$

$$
\begin{aligned}
f := T_o^{-1}\bar{f} &= [f_1, f_2, \ldots, f_n]^T \in \mathbb{R}^n, \quad h := \bar{h}T_o \\
&= [1, 0, \ldots, 0] \in \mathbb{R}^{1 \times n}. \qquad (5)
\end{aligned}
$$

The transformation matrix is given by

$$
T_o := [\bar{A}^{n-1}l_n, \bar{A}^{n-2}l_n, \ldots, l_n] \in \mathbb{R}^{n \times n},
$$

where l_n is the nth column of T_{ob}.

$$
T_{ob} := \begin{bmatrix} \bar{h} \\ \bar{h}\bar{A} \\ \vdots \\ \bar{h}\bar{A}^{n-1} \end{bmatrix}^{-1} \in \mathbb{R}^{n \times n},
$$

where T_{ob} is another non-singular matrix which transform the general system into its observability canonical form.

Proof As $T = [*, l_n]$, we have

$$
T_{ob}^{-1}T_{ob} := \begin{bmatrix} \bar{h} \\ \bar{h}\bar{A} \\ \vdots \\ \bar{h}\bar{A}^{n-1} \end{bmatrix}[*, l_n] = I_n,
$$

and the nth column can be written as

$$
\begin{cases}
\bar{h}l_n = 0, \\
\bar{h}\bar{A}l_n = 0, \\
\quad\vdots \\
\bar{h}\bar{A}^{n-2}l_n = 0, \\
\bar{h}\bar{A}^{n-1}l_n = 1.
\end{cases}
$$

Hence, we have

$$
\begin{aligned}
h &= \bar{h}T_o = \bar{h}[\bar{A}^{n-1}l_n, \bar{A}^{n-2}l_n, \ldots, l_n] \\
&= [\bar{h}\bar{A}^{n-1}l_n, \bar{h}\bar{A}^{n-2}l_n, \ldots, \bar{h}l_n] \\
&= [1, 0, 0, \ldots, 0],
\end{aligned}
$$

$$
\begin{aligned}
T_o^{-1}T_o &= T_o^{-1}[\bar{A}^{n-1}l_n, \bar{A}^{n-2}l_n, \ldots, l_n] \\
&= [T_o^{-1}\bar{A}^{n-1}l_n, T_o^{-1}\bar{A}^{n-2}l_n, \ldots, T_o^{-1}l_n] = I_n,
\end{aligned}
$$

or

$$
T_o^{-1}\bar{A}^{n-1}l_n = [1, 0, 0, 0, \ldots, 0]^T = e_1,
$$
$$
T_o^{-1}\bar{A}^{n-2}l_n = [0, 1, 0, 0, \ldots, 0]^T = e_2,
$$
$$
T_o^{-1}\bar{A}^{n-3}l_n = [0, 0, 1, 0, \ldots, 0]^T = e_3,
$$
$$
\vdots
$$
$$
T_o^{-1}l_n = [0, 0, 0, 0, \ldots, 1]^T = e_n. \qquad (6)
$$

Therefore, we obtain

$$
\begin{aligned}
A &= T_o^{-1}\bar{A}T_o \\
&= T_o^{-1}\bar{A}[\bar{A}^{n-1}l_n, \bar{A}^{n-2}l_n, \ldots, l_n] \\
&= [T_o^{-1}\bar{A}^n l_n, T_o^{-1}\bar{A}^{n-1}l_n, T_o^{-1}\bar{A}^{n-2}l_n, \ldots, T_o^{-1}\bar{A}l_n] \\
&= [T_o^{-1}\bar{A}^n l_n, e_1, e_2, \ldots, e_{n-1}]. \qquad (7)
\end{aligned}
$$

Suppose that the characteristic equation of $\bar{A}$ can be represented as

$$
\det[sI_n - \bar{A}] = s^n + a_1 s^{n-1} + a_2 s^{n-2} + \cdots + a_n = 0.
$$

According to the Cayley–Hamilton theorem, we have

$$
\bar{A}^n + a_1\bar{A}^{n-1} + a_2\bar{A}^{n-2} + \cdots + a_n I_n = 0,
$$

or

$$
\bar{A}^n = -a_1\bar{A}^{n-1} - a_2\bar{A}^{n-2} - \cdots - a_n I_n. \qquad (8)
$$

Pre- and post-multiplying (8) by T^{-1} and l_n gives

$$
\begin{aligned}
T_o^{-1}\bar{A}^n l_n &= -a_1 T_o^{-1}\bar{A}^{n-1}l_n - a_2 T_o^{-1}\bar{A}^{n-2}l_n \\
&\quad - \cdots - a_n T_o^{-1}l_n \\
&= -a_1 e_1 - a_2 e_2 - \cdots - a_n e_n \\
&= [-a_1, -a_2, \ldots, -a_n]^T. \qquad (9)
\end{aligned}
$$

By substituting the above equation into (7), the proof is finished. $\qquad \square$

The model in (1) and (2) contains $2n^2 + 2n$ parameters, but the observer canonical form in (3) and (4) contains only $n^2 + 2n$ parameters. The decrease in the parameters to be identified leads to the reduction in the identification computation because both the initial parameters and the transformed parameters describe the relationship between input and output of the same bilinear state space system.

It is pointed out that this paper does not consider the process noise in the state equation, because if adding the process noise, the problem is more complex and more difficult, more challenging and more interesting, the related work can be found in [15,47] without considering the process noise. In the future work, we will investigate the case with process noise.

3 The identification model for the bilinear state space system

From (3) to (5), we have the following n equations:

$$\begin{cases} x_1(t+1) &= -a_1 x_1(t) + x_2(t) + \boldsymbol{b}_1 \boldsymbol{x}(t)u(t) + f_1 u(t), \\ x_2(t+1) &= -a_2 x_1(t) + x_3(t) + \boldsymbol{b}_2 \boldsymbol{x}(t)u(t) + f_2 u(t), \\ &\vdots \\ x_{n-1}(t+1) &= -a_{n-1} x_1(t) + x_n(t) \\ & \quad + \boldsymbol{b}_{n-1}\boldsymbol{x}(t)u(t) + f_{n-1}u(t), \\ x_n(t+1) &= -a_n x_1(t) + \boldsymbol{b}_n \boldsymbol{x}(t)u(t) + f_n u(t), \end{cases}$$

which can be simplified as

$$\begin{aligned} x_i(t+1) &= -a_i x_1(t) + x_{i+1}(t) + \boldsymbol{b}_i \boldsymbol{x}(t)u(t) \\ & \quad + f_i u(t), \quad i = 1, 2, \ldots, n-1, \end{aligned} \quad (10)$$

$$x_n(t+1) = -a_n x_1(t) + \boldsymbol{b}_n \boldsymbol{x}(t)u(t) + f_n u(t). \quad (11)$$

Multiplying both sides of (10) by z^{-i} gives

$$\begin{aligned} x_i(t-i+1) &= -a_i x_1(t-i) + x_{i+1}(t-i) \\ & \quad + \boldsymbol{b}_i \boldsymbol{x}(t-i)u(t-i) + f_i u(t-i). \end{aligned} \quad (12)$$

Summing (12) for i from $i = 1$ to $i = (n-1)$, we have

$$\begin{aligned} x_1(t) &= -\sum_{i=1}^{n-1} a_i x_1(t-i) + x_n(t-n+1) \\ & \quad + \sum_{i=1}^{n-1} \boldsymbol{b}_i \boldsymbol{x}(t-i)u(t-i) + \sum_{i=1}^{n-1} f_i u(t-i). \end{aligned} \quad (13)$$

Then, multiplying both sides of (11) by z^{-n} gives

$$\begin{aligned} x_n(t-n+1) &= -a_n x_1(t-n) + \boldsymbol{b}_n \boldsymbol{x}(t-n)u(t-n) \\ & \quad + f_n u(t-n). \end{aligned} \quad (14)$$

Substituting (14) into (13), we obtain

$$\begin{aligned} x_1(t) &= -\sum_{i=1}^{n} a_i x_1(t-i) + \sum_{i=1}^{n} \boldsymbol{b}_i \boldsymbol{x}(t-i)u(t-i) \\ & \quad + \sum_{i=1}^{n} f_i u(t-i) \\ &= [x_1(t-1), x_1(t-2), \ldots, x_1(t-n)] \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_n \end{bmatrix} \\ & \quad + [u(t-1), u(t-2), \ldots, u(t-n)] \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \\ & \quad + [\boldsymbol{x}^{\mathrm{T}}(t-1)u(t-1), \boldsymbol{x}^{\mathrm{T}}(t-2)u(t-2), \ldots, \\ & \quad \boldsymbol{x}^{\mathrm{T}}(t-n)u(t-n)] \begin{bmatrix} \boldsymbol{b}_1^{\mathrm{T}} \\ \boldsymbol{b}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{b}_n^{\mathrm{T}} \end{bmatrix}. \end{aligned} \quad (15)$$

From (4), we have

$$y(t) = x_1(t) + v(t). \quad (16)$$

Define the information vector $\boldsymbol{\varphi}(t)$ and the parameter vector $\boldsymbol{\theta}$ as

$$\begin{aligned} \boldsymbol{\varphi}(t) &:= [\boldsymbol{\varphi}_x^{\mathrm{T}}(t), \boldsymbol{\varphi}_{xu}^{\mathrm{T}}(t), \boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}} \in \mathbb{R}^{n^2+2n}, \\ \boldsymbol{\varphi}_x(t) &:= [-x_1(t-1), -x_1(t-2), \ldots, -x_1(t-n)]^{\mathrm{T}} \in \mathbb{R}^n, \\ \boldsymbol{\varphi}_{xu}(t) &:= [\boldsymbol{x}^{\mathrm{T}}(t-1)u(t-1), \boldsymbol{x}^{\mathrm{T}}(t-2)u(t \\ & \quad -2), \ldots, \boldsymbol{x}^{\mathrm{T}}(t-n)u(t-n)]^{\mathrm{T}} \in \mathbb{R}^{n^2}, \\ \boldsymbol{\varphi}_u(t) &:= [u(t-1), u(t-2), \ldots, u(t-n)]^{\mathrm{T}} \in \mathbb{R}^n, \\ \boldsymbol{\theta} &:= [\boldsymbol{a}^{\mathrm{T}}, \boldsymbol{b}^{\mathrm{T}}, \boldsymbol{f}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n^2+2n}, \\ \boldsymbol{a} &:= [a_1, a_2, \ldots, a_n]^{\mathrm{T}} \in \mathbb{R}^n, \\ \boldsymbol{b} &:= [\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n]^{\mathrm{T}} \in \mathbb{R}^{n^2}, \\ \boldsymbol{f} &:= [f_1, f_2, \ldots, f_n]^{\mathrm{T}} \in \mathbb{R}^n. \end{aligned}$$

Substituting (15) into (16), we obtain the identification model of the bilinear state space system in (3) and (4):

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta} + v(t). \quad (17)$$

4 The recursive state and parameter estimation algorithm

The recursive state and parameter estimation algorithm is composed of the state estimation and parameter identification.

4.1 The state estimation algorithm

Although the open-loop observer [26], closed-loop observer [48] and the Kalman filter [35] all can estimate the system states, in the state estimate-based parameter identification methods in this paper, we use the open-loop observer, which can also achieve good performance. From the simulation results, we can see that the estimated states are very close to the actual states of the system (see Figs. 3, 4, 7, 8), which show that the open-loop state observer we propose is valid. Compared with the closed-loop state observer, the open-loop state observer is simpler in structure and has less computation cost. So when the open-loop state observer can satisfy the estimation accuracy, we will give priority to the design of the open-loop state observer instead of the closed-loop state observer to reduce the computation cost. Of course, we may use the closed-loop observer like in Refs. [25] and [48]. Then we carry on the design of the open-loop observer.

If the parameter matrices/vector $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times n}$ and $f \in \mathbb{R}^n$ are known, we can apply the following state observer to generate the estimate $\hat{x}(t)$ of the unknown state vector $x(t)$:

$$\hat{x}(t+1) = A\hat{x}(t) + B\hat{x}(t)u(t) + fu(t). \tag{18}$$

When the parameter matrices/vector $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times n}$ and $f \in \mathbb{R}^n$ are unknown, we replace them with their estimated parameter matrices/vector $\hat{A}(t)$, $\hat{B}(t)$ and $\hat{f}(t)$ in the following parameter estimation algorithms to compute the estimate $\hat{x}(t)$ of the state vector $x(t)$ and obtain

$$\hat{x}(t+1) = \hat{A}(t)\hat{x}(t) + \hat{B}(t)\hat{x}(t)u(t) + \hat{f}(t)u(t). \tag{19}$$

In (18) and (19), the initial state $\hat{x}(1)$ is usually defined as a real vector with small entries, e.g., $\hat{x}(1) = \mathbf{1}_n/p_0$, where $\mathbf{1}_n := [1, 1, \ldots, 1]^\mathsf{T} \in \mathbb{R}^n$, and p_0 is a large number, e.g., $p_0 = 10^6 \gg 1$.

4.2 The state observer-based multi-innovation stochastic gradient algorithm

Based on the identification model in (17), we propose the stochastic gradient (SG) algorithm as

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\varphi(t)}{r(t)}e(t), \tag{20}$$

$$e(t) = y(t) - \varphi^\mathsf{T}(t)\hat{\theta}(t-1), \tag{21}$$

$$r(t) = r(t-1) + \|\varphi(t)\|^2, \quad r(0) = 1, \tag{22}$$

where the norm of a matrix X is defined as $\|X\|^2 := \mathrm{tr}[XX^\mathsf{T}]$, $1/r(t)$ represents the convergence factor or the step size, and $e(t) := y(t) - \varphi^\mathsf{T}(t)\hat{\theta}(t-1)$ is defined as the scalar innovation. As is known to us, the SG algorithm has slow convergence rates. In order to overcome this weakness, we take advantage of the multi-innovation identification theory and expand the scalar innovation $e(t)$ to an innovation vector (p represents the innovation length):

$$E(p,t) := \begin{bmatrix} y(t) - \varphi^\mathsf{T}(t)\hat{\theta}(t-1) \\ y(t-1) - \varphi^\mathsf{T}(t-1)\hat{\theta}(t-1) \\ \vdots \\ y(t-p+1) - \varphi^\mathsf{T}(t-p+1)\hat{\theta}(t-1) \end{bmatrix} \in \mathbb{R}^p.$$

Define the stacked output vector $Y(p,t)$ and the stacked information matrix $\Phi(p,t)$ as

$$Y(p,t) := [y(t), y(t-1), \ldots, y(t-p+1)]^\mathsf{T} \in \mathbb{R}^p,$$
$$\Phi(p,t) := [\varphi(t), \varphi(t-1), \ldots, \varphi(t-p+1)] \in \mathbb{R}^{(n^2+2n) \times p}.$$

Then the innovation vector $E(p,t)$ can be expressed as

$$E(p,t) = Y(p,t) - \Phi^\mathsf{T}(p,t)\hat{\theta}(t-1).$$

In the case of $p = 1$, we obtain $E(1,t) = e(t)$, $\Phi(1,t) = \varphi(t)$, $Y(1,t) = y(t)$. Equation (20) is equivalent to

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\Phi(t)}{r(t)}[Y(1,t) - \Phi^\mathsf{T}(1,t)\hat{\theta}(t-1)]. \tag{23}$$

Replacing the "1" in $\Phi(1,t)$ and $Y(1,t)$ with p obtains

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{\Phi(t)}{r(t)}[Y(p,t) - \Phi^\mathsf{T}(p,t)\hat{\theta}(t-1)],$$
$$= \hat{\theta}(t-1) + \frac{\Phi(t)}{r(t)}E(p,t). \tag{24}$$

Because the information vector $\varphi(t)$ contains the unmeasurable state vector $x(t)$, the algorithm in (20) to (22) cannot be realized. The scheme here is to replace $x(t)$ in $\varphi(t)$ with its estimate $\hat{x}(t)$ and to define

$$\hat{\varphi}(t) := [\hat{\varphi}_x^\mathsf{T}(t), \hat{\varphi}_{xu}^\mathsf{T}(t), \varphi_u^\mathsf{T}(t)]^\mathsf{T},$$
$$\hat{\varphi}_x(t) := [-\hat{x}_1(t-1), -\hat{x}_1(t-2), \ldots, -\hat{x}_1(t-n)]^\mathsf{T},$$
$$\hat{\varphi}_{xu}(t) := [\hat{x}^\mathsf{T}(t-1)u(t-1), \hat{x}^\mathsf{T}(t-2)u(t-2), \ldots, \hat{x}^\mathsf{T}(t-n)u(t-n)]^\mathsf{T}. \tag{25}$$

By using the estimate $\hat{\varphi}(t)$ to replace $\varphi(t)$ and using the estimate $\hat{\Phi}(t)$ to replace $\Phi(t)$, the O-MISG algorithm can be derived as

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}(t)}{r(t)} \boldsymbol{E}(p, t), \tag{26}$$

$$\boldsymbol{E}(p, t) = \boldsymbol{Y}(p, t) - \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(p, t)\hat{\boldsymbol{\theta}}(t-1), \tag{27}$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\varphi}}(t)\|^2, \quad r(0) = 1, \tag{28}$$

$$\boldsymbol{Y}(p, t) = [y(t), y(t-1), \ldots, y(t-p+1)]^{\mathrm{T}}, \tag{29}$$

$$\hat{\boldsymbol{\Phi}}(p, t) = [\hat{\boldsymbol{\varphi}}(t), \hat{\boldsymbol{\varphi}}(t-1), \ldots, \hat{\boldsymbol{\varphi}}(t-p+1)], \tag{30}$$

$$\hat{\boldsymbol{\varphi}}(t) = [\hat{\boldsymbol{\varphi}}_x^{\mathrm{T}}(t), \hat{\boldsymbol{\varphi}}_{xu}^{\mathrm{T}}(t), \boldsymbol{\varphi}_u^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{31}$$

$$\hat{\boldsymbol{\varphi}}_x(t) = [-\hat{x}_1(t-1), -\hat{x}_1(t-2), \ldots,$$
$$-\hat{x}_1(t-n)]^{\mathrm{T}}, \tag{32}$$

$$\hat{\boldsymbol{\varphi}}_{xu}(t) = [\hat{\boldsymbol{x}}^{\mathrm{T}}(t-1)u(t-1), \hat{\boldsymbol{x}}^{\mathrm{T}}(t-2)u(t$$
$$-2), \ldots, \hat{\boldsymbol{x}}^{\mathrm{T}}(t-n)u(t-n)]^{\mathrm{T}}, \tag{33}$$

$$\boldsymbol{\varphi}_u(t) = [u(t-1), u(t-2), \ldots, u(t-n)]^{\mathrm{T}}, \tag{34}$$

$$\hat{\boldsymbol{x}}(t+1) = \hat{\boldsymbol{A}}(t)\hat{\boldsymbol{x}}(t) + \hat{\boldsymbol{B}}(t)\hat{\boldsymbol{x}}(t)u(t)$$
$$+ \hat{\boldsymbol{f}}(t)u(t), \quad \hat{\boldsymbol{x}}(1) = \mathbf{1}_n/p_0, \tag{35}$$

$$\hat{\boldsymbol{A}}(t) = \begin{bmatrix} -\hat{a}_1(t) & 1 & 0 & \cdots & 0 \\ -\hat{a}_2(t) & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -\hat{a}_{n-1}(t) & 0 & \cdots & 0 & 1 \\ -\hat{a}_n(t) & 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{36}$$

$$\hat{\boldsymbol{B}}(t) = \begin{bmatrix} \hat{\boldsymbol{b}}_1(t) \\ \hat{\boldsymbol{b}}_2(t) \\ \vdots \\ \hat{\boldsymbol{b}}_n(t) \end{bmatrix}, \quad \hat{\boldsymbol{b}}_i(t)$$
$$= [\hat{b}_{i1}(t), \hat{b}_{i2}(t), \ldots, \hat{b}_{in}(t)], \tag{37}$$

$$\hat{\boldsymbol{f}}(t) = [\hat{f}_1(t), \hat{f}_2(t), \ldots, \hat{f}_{n-1}(t), \hat{f}_n(t)]^{\mathrm{T}}. \tag{38}$$

When $p = 1$, the O-MISG algorithm reduces to the state observer-based stochastic gradient (O-SG) algorithm. Then we study the convergence of the O-SG algorithm by establishing a recursive relation about the parameter estimation error and by using the martingale convergence theorem.

In identification, the output is linear about the parameter space, the algorithms are not sensitive to the initial values and can achieve global minima, and the final parameter estimates do not depend on the initial values. We have the following theorem about the convergence of the proposed algorithms.

Theorem 1 *For the system in* (17) *and the O-SG algorithm in* (26) *to* (38) *($p = 1$), assume that $\{v(t)\}$ is a white noise sequence with zero mean and variance σ^2, i.e.,*

$$\mathrm{E}[v(t)] = 0, \quad \mathrm{E}[v^2(t)] = \sigma^2, \quad \mathrm{E}[v(t)v(i)] = 0, \quad i \neq t,$$

and $r(t) \to \infty$, and there exist an integer N and a positive constant c independent of t such that the following strong excitation condition holds:

$$\sum_{j=0}^{N-1} \frac{\hat{\boldsymbol{\varphi}}(j)\hat{\boldsymbol{\varphi}}^{\mathrm{T}}(j)}{r(j)} \geqslant c\boldsymbol{I}, \quad \text{a.s.}$$

Then the parameter estimation error given by the O-SG algorithm converges to zero.

The proofs of Theorems 1 and 2 in the sequel can be done in a similar way to those in [47,48].

4.3 The state observer-based recursive least squares algorithm

In order to improve the convergence rate and parameter estimation accuracy of the SG algorithm, we define the output vector $\boldsymbol{Y}(t)$ and the information matrix $\boldsymbol{H}(t)$ as

$$\boldsymbol{Y}(t) := \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(t) \end{bmatrix} \in \mathbb{R}^t,$$

$$\boldsymbol{H}(t) := \begin{bmatrix} \boldsymbol{\varphi}^{\mathrm{T}}(1) \\ \boldsymbol{\varphi}^{\mathrm{T}}(2) \\ \vdots \\ \boldsymbol{\varphi}^{\mathrm{T}}(t) \end{bmatrix} \in \mathbb{R}^{t \times (n^2+2n)}. \tag{39}$$

Based on the identification model in (17), we define the criterion function as

$$J(\boldsymbol{\theta}) := \sum_{j=1}^{t} [y(j) - \boldsymbol{\varphi}^{\mathrm{T}}(j)\boldsymbol{\theta}]^2$$
$$= [\boldsymbol{Y}(t) - \boldsymbol{H}(t)\boldsymbol{\theta}]^{\mathrm{T}}[\boldsymbol{Y}(t) - \boldsymbol{H}(t)\boldsymbol{\theta}]$$
$$= \|\boldsymbol{Y}(t) - \boldsymbol{H}(t)\boldsymbol{\theta}\|^2. \tag{40}$$

According to the least squares principle, and minimizing $J(\boldsymbol{\theta})$, we can obtain the following recursive least squares algorithm:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{P}(t)\boldsymbol{\varphi}(t)[y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)], \tag{41}$$

$$\boldsymbol{P}^{-1}(t) = \boldsymbol{P}^{-1}(t-1) + \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t), \quad \boldsymbol{P}(0)$$
$$= p_0 \boldsymbol{I}_{n^2+2n} > 0. \tag{42}$$

In order to avoid computing the inversion of the covariance matrix $\boldsymbol{P}(t)$, applying the matrix inversion lemma

$$(A + BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}$$

to (42) gives

$$P(t) = P(t-1) - \frac{P(t-1)\varphi(t)\varphi^{\mathrm{T}}(t)P(t-1)}{1 + \varphi^{\mathrm{T}}(t)P(t-1)\varphi(t)}. \tag{43}$$

Defining the gain vector $L(t) := P(t)\varphi(t) \in \mathbb{R}^{n^2+2n}$, we obtain

$$L(t) = P(t-1)\varphi(t) - \frac{P(t-1)\varphi(t)\varphi^{\mathrm{T}}(t)P(t-1)\varphi(t)}{1 + \varphi^{\mathrm{T}}(t)P(t-1)\varphi(t)}$$

$$= \frac{P(t-1)\varphi(t)}{1 + \varphi^{\mathrm{T}}(t)P(t-1)\varphi(t)}. \tag{44}$$

Combining (43) and (44) gives

$$P(t) = [I - L(t)\varphi^{\mathrm{T}}(t)]P(t-1). \tag{45}$$

Therefore, the recursive least squares algorithms in (41) and (42) can be equivalent to

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + L(t)[y(t) - \varphi^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)], \quad \hat{\boldsymbol{\theta}}(0)$$

$$= \mathbf{1}_{n^2+2n}/p_0, \tag{46}$$

$$L(t) = P(t-1)\varphi(t)[1 + \varphi^{\mathrm{T}}(t)P(t-1)\varphi(t)]^{-1}, \tag{47}$$

$$P(t) = [I - L(t)\varphi^{\mathrm{T}}(t)]P(t-1). \tag{48}$$

Aimed at the unmeasurable state variables included in the information vector $\varphi(t)$, we replace $\varphi(t)$ with its estimates $\hat{\varphi}(t)$ in the identification algorithm. Then, combined with the state observer, we can obtain the state observer-based recursive least squares (O-RLS) algorithm:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + L(t)[y(t) - \hat{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\theta}}(t-1)], \quad \hat{\boldsymbol{\theta}}(0)$$

$$= \mathbf{1}_{n^2+2n}/p_0, \tag{49}$$

$$L(t) = P(t-1)\hat{\varphi}(t)[1 + \hat{\varphi}^{\mathrm{T}}(t)P(t-1)\hat{\varphi}(t)]^{-1}, \tag{50}$$

$$P(t) = [I - L(t)\hat{\varphi}^{\mathrm{T}}(t)]P(t-1), \quad P(0) = p_0 I_{n^2+2n}, \tag{51}$$

$$\hat{\varphi}(t) = [\hat{\varphi}_x^{\mathrm{T}}(t), \hat{\varphi}_{xu}^{\mathrm{T}}(t), \varphi_u^{\mathrm{T}}(t)]^{\mathrm{T}}, \tag{52}$$

$$\hat{\varphi}_x(t) = [-\hat{x}_1(t-1), -\hat{x}_1(t-2), \ldots, -\hat{x}_1(t-n)]^{\mathrm{T}}, \tag{53}$$

$$\hat{\varphi}_{xu}(t) = [\hat{x}^{\mathrm{T}}(t-1)u(t-1), \hat{x}^{\mathrm{T}}(t-2)u(t-2), \ldots, \hat{x}^{\mathrm{T}}(t-n)u(t-n)]^{\mathrm{T}}, \tag{54}$$

$$\varphi_u(t) = [u(t-1), u(t-2), \ldots, u(t-n)]^{\mathrm{T}}, \tag{55}$$

$$\hat{x}(t+1) = \hat{A}(t)\hat{x}(t) + \hat{B}(t)\hat{x}(t)u(t) + \hat{f}(t)u(t), \tag{56}$$

$$\hat{A}(t) = \begin{bmatrix} -\hat{a}_1(t) & 1 & 0 & \cdots & 0 \\ -\hat{a}_2(t) & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -\hat{a}_{n-1}(t) & 0 & \cdots & 0 & 1 \\ -\hat{a}_n(t) & 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{57}$$

$$\hat{B}(t) = \begin{bmatrix} \hat{b}_1(t) \\ \hat{b}_2(t) \\ \vdots \\ \hat{b}_n(t) \end{bmatrix}, \quad \hat{b}_i(t)$$

$$= [\hat{b}_{i1}(t), \hat{b}_{i2}(t), \ldots, \hat{b}_{in}(t)], \tag{58}$$

$$\hat{f}(t) = [\hat{f}_1(t), \hat{f}_2(t), \ldots, \hat{f}_{n-1}(t), \hat{f}_n(t)]^{\mathrm{T}}. \tag{59}$$

About the convergence of the parameter estimate $\boldsymbol{\theta}(t)$, we have the following theorem.

Theorem 2 *Provided that the controllable and observable system in* (1) *and* (2) *is stable (i.e.,* A *is stable), for the identification model in* (17) *and the state observer-based RLS parameter identification algorithm in* (49) *to* (59), *suppose that* $v(t)$ *is a white noise sequence with zero mean and variance* σ^2, *i.e.,* $\mathrm{E}[v(t)] = 0$, $\mathrm{E}[v^2(t)] = \sigma^2$, $\mathrm{E}[v(t)v(i)] = 0(i \neq t)$, *and that there exist positive constants* α *and* β *such that the following persistent excitation conditions holds:*

$$\alpha I \leqslant \frac{1}{t}\sum_{j=1}^{t} \hat{\varphi}(j)\hat{\varphi}^{\mathrm{T}}(j) \leqslant \beta I, \quad \text{a.s.} \tag{60}$$

Then the parameter estimation error $\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|$ *converges to zero.*

The identification method proposed in this paper can be extended to multi-input multi-output systems through expanding the single input to multiple inputs and the single output to multiple outputs.

5 Numerical example

Consider the following observer canonical bilinear state space system:

$$x(t+1) = \begin{bmatrix} 0.19 & 1 \\ 0.23 & 0 \end{bmatrix} x(t)$$

$$+ \begin{bmatrix} 0.01 & 0.06 \\ 0.15 & 0.14 \end{bmatrix} x(t)u(t) + \begin{bmatrix} 1.17 \\ 1.14 \end{bmatrix} u(t),$$

$$y(t) = [1, 0]x(t) + v(t).$$

The parameter vector to be identified is

$$\boldsymbol{\theta} = [a_1, a_2, b_{11}, b_{12}, b_{21}, b_{22}, f_1, f_2]^{\mathrm{T}}$$

$$= [-0.19, -0.23, 0.01, 0.06,$$

$$0.15, 0.14, 1.17, 1.14]^{\mathrm{T}}.$$

In simulation, the input $\{u(t)\}$ is taken as an uncorrelated persistent excitation signal sequence with zero

Table 1 O-MISG parameter estimates and errors ($\sigma^2 = 0.50^2$)

p	t	a_1	a_2	b_{11}	b_{12}	b_{21}	b_{22}	f_1	f_2	$\delta(\%)$
1	100	−0.42336	−0.03816	0.13625	0.22886	0.09850	0.13462	0.30107	0.43594	70.39445
	200	−0.43245	−0.03827	0.10101	0.21395	0.08006	0.12783	0.33143	0.45079	68.38029
	500	−0.42761	−0.01933	0.08522	0.20813	0.08533	0.12578	0.36438	0.46541	66.42583
	1000	−0.43772	−0.02189	0.06817	0.19925	0.08645	0.12117	0.39041	0.47874	64.79334
	2000	−0.43944	−0.01731	0.05605	0.19228	0.09117	0.12173	0.41048	0.48887	63.53782
	3000	−0.44028	−0.01582	0.05127	0.18895	0.09444	0.12241	0.42212	0.49389	62.84131
	4000	−0.44059	−0.01403	0.04733	0.18626	0.09518	0.12189	0.43018	0.49733	62.36906
	5000	−0.44104	−0.01262	0.04572	0.18459	0.09671	0.12209	0.43641	0.50036	61.99831
2	100	−0.44300	−0.01954	0.02983	0.17082	0.11197	0.16130	0.72782	0.53262	49.51547
	200	−0.44627	−0.01329	0.00372	0.15583	0.08866	0.15164	0.75849	0.55009	47.89669
	500	−0.43403	0.00607	0.00150	0.15173	0.10549	0.14850	0.79382	0.56957	45.98234
	1000	−0.44611	−0.00494	0.00250	0.14705	0.11439	0.14311	0.81880	0.58761	44.44015
	2000	−0.44445	−0.00071	−0.00134	0.14052	0.12370	0.14219	0.83570	0.60172	43.32238
	3000	−0.44342	−0.00145	−0.00148	0.13774	0.12768	0.14250	0.84640	0.60846	42.66932
	4000	−0.44237	−0.00018	−0.00141	0.13536	0.12837	0.14114	0.85354	0.61311	42.25577
	5000	−0.44203	0.00018	−0.00028	0.13408	0.13032	0.14126	0.85914	0.61716	41.91425
3	100	−0.40900	−0.03681	−0.00094	0.11430	0.13205	0.17215	0.89224	0.67379	37.02095
	200	−0.41114	−0.03335	−0.01472	0.10160	0.10343	0.15962	0.92181	0.69000	35.66444
	500	−0.39643	−0.02121	−0.00963	0.10307	0.12273	0.15629	0.95746	0.70801	33.85205
	1000	−0.40749	−0.03823	0.00218	0.10390	0.13090	0.15018	0.97986	0.72613	32.38322
	2000	−0.40249	−0.03420	−0.00014	0.09921	0.13900	0.14828	0.99248	0.73995	31.42717
	3000	−0.39977	−0.03704	−0.00006	0.09736	0.14071	0.14830	1.00155	0.74611	30.82762
	4000	−0.39803	−0.03667	0.00130	0.09552	0.14029	0.14649	1.00747	0.75054	30.46773
	5000	−0.39724	−0.03729	0.00317	0.09487	0.14145	0.14661	1.01202	0.75448	30.16288
5	100	−0.29796	−0.12161	−0.00248	0.04849	0.13863	0.16668	1.05557	0.95267	16.10621
	200	−0.30001	−0.12711	0.00269	0.04259	0.10583	0.14992	1.07813	0.95743	15.44434
	500	−0.28758	−0.12514	0.00247	0.05275	0.13239	0.15216	1.10736	0.96213	14.21503
	1000	−0.29439	−0.14208	0.02153	0.06302	0.13686	0.14867	1.12142	0.97221	13.29218
	2000	−0.28902	−0.13737	0.01400	0.06162	0.14427	0.14797	1.12344	0.97974	12.85860
	3000	−0.28575	−0.14245	0.01088	0.06074	0.14372	0.14871	1.12856	0.98169	12.48979
	4000	−0.28396	−0.14223	0.01223	0.05983	0.14244	0.14704	1.13185	0.98363	12.32016
	5000	−0.28362	−0.14356	0.01428	0.06029	0.14316	0.14751	1.13387	0.98594	12.15214
8	100	−0.24866	−0.17083	−0.00559	0.03692	0.13372	0.16874	1.08400	1.07716	8.47922
	200	−0.24443	−0.17811	0.01278	0.03450	0.09384	0.14527	1.11097	1.07067	7.96532
	500	−0.23104	−0.17895	0.00271	0.04637	0.13959	0.15443	1.14479	1.06481	6.30229
	1000	−0.23729	−0.19683	0.02725	0.06207	0.13906	0.14916	1.15736	1.07076	5.60115
	2000	−0.23339	−0.19010	0.01370	0.06002	0.14744	0.14820	1.15261	1.07547	5.35180
	3000	−0.22868	−0.19577	0.00842	0.05846	0.14435	0.14870	1.15694	1.07448	5.08305
	4000	−0.22703	−0.19477	0.01114	0.05783	0.14234	0.14664	1.15969	1.07507	5.01272
	5000	−0.22786	−0.19653	0.01435	0.05889	0.14349	0.14739	1.16076	1.07689	4.90325
True values		−0.19000	−0.23000	0.01000	0.06000	0.15000	0.14000	1.17000	1.14000	

Fig. 1 O-MISG parameter estimation errors δ versus t ($\sigma^2 = 0.50^2$)
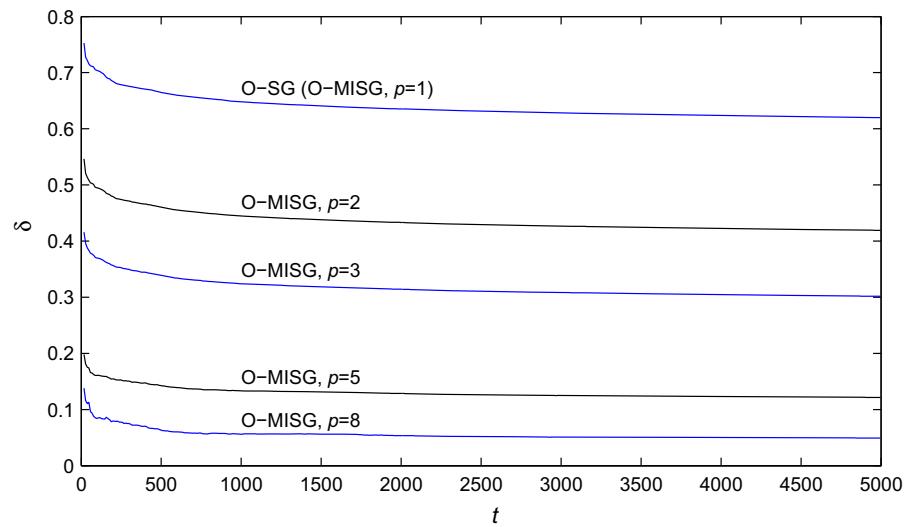
Table 2 O-RLS parameter estimates and errors ($\sigma^2 = 0.50^2$)

t	a_1	a_2	b_{11}	b_{12}	b_{21}	b_{22}	f_1	f_2	δ (%)
100	−0.27450	−0.15201	0.01381	0.03002	0.11895	0.21139	1.15726	1.02358	10.99897
200	−0.26253	−0.16300	0.03821	0.00753	0.10099	0.17483	1.19447	1.03358	10.14256
500	−0.22818	−0.18698	−0.00560	0.07815	0.15564	0.11541	1.21555	1.06048	6.78754
1000	−0.20153	−0.21511	0.01724	0.07260	0.14613	0.13408	1.19948	1.10811	2.98734
2000	−0.20310	−0.20987	0.00707	0.07835	0.15484	0.12785	1.16956	1.11707	2.40325
3000	−0.19185	−0.22254	0.00534	0.07252	0.14785	0.14275	1.17498	1.12497	1.33590
4000	−0.19540	−0.21915	0.01071	0.06280	0.14747	0.14058	1.17796	1.12053	1.46851
5000	−0.19315	−0.22129	0.01360	0.06049	0.14694	0.14272	1.17709	1.12612	1.13105
True values	−0.19000	−0.23000	0.01000	0.06000	0.15000	0.14000	1.17000	1.14000	

Fig. 2 O-RLS parameter estimation errors δ versus t ($\sigma^2 = 0.50^2$)
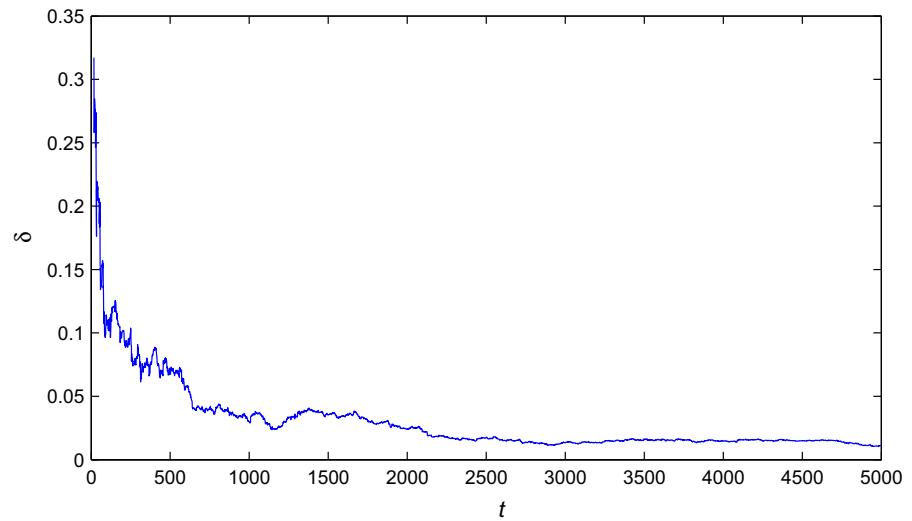
Fig. 3 State $x_1(t)$ and the estimated state $\hat{x}_1(t)$ against t ($p = 8$, $\sigma^2 = 0.50^2$)
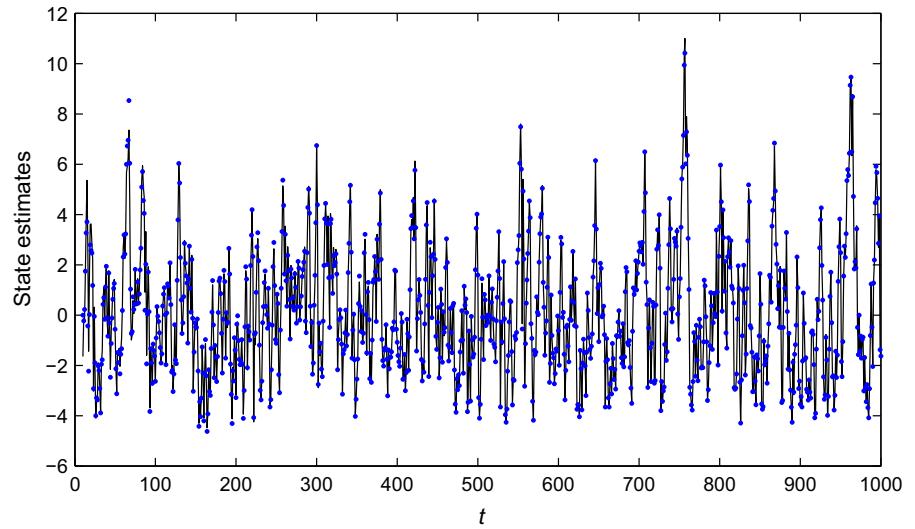

Fig. 4 State $x_2(t)$ and the estimated state $\hat{x}_2(t)$ against t ($p = 8$, $\sigma^2 = 0.50^2$)
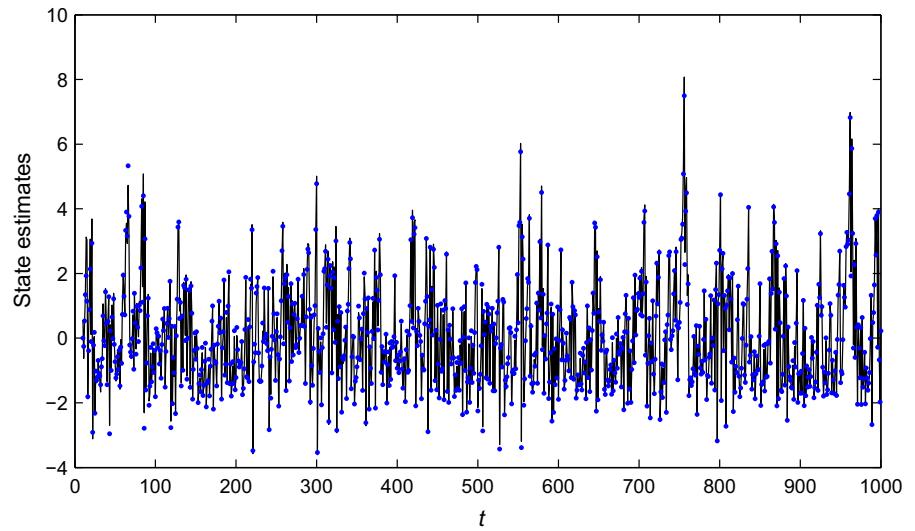
mean and unit variance, and $\{v(t)\}$ as a white noise sequence with zero mean and variance σ^2. Take the data length $L = 5000$, and apply the O-MISG algorithm in (26) to (38) and the O-RLS algorithm in (49) to (59) to estimate the states and parameters of this bilinear system. The O-MISG parameter estimates and errors $\delta = \|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}\|/\|\boldsymbol{\theta}\|$ with different innovation length p are shown in Table 1 and Fig. 1 with $\sigma^2 = 0.50^2$. The O-RLS parameter estimates and errors $\sigma^2 = 0.50^2$ are shown in Table 2 and Fig. 2. The states $x_i(t)$ and their estimates $\hat{x}_i(t)$ against t are shown in Figs. 3 and 4.

For comparison with the noise variance $\sigma^2 = 0.50^2$, Tables 3, 4 and Figs. 5, 6 give the O-MISG and O-RLS parameter estimates and the parameter estimation error curves with $\sigma^2 = 0.10^2$. The states $x_i(t)$ and their estimates $\hat{x}_i(t)$ against t are shown in Figs. 7 and 8.

From Tables 1, 2, 3, 4 and Figs. 1, 2, 3, 4, 5, 6, 7, 8, we can draw the following conclusions.

(1) The O-RLS algorithm is superior to the O-SG algorithm in the parameter estimation accuracy and convergence rates.
(2) The parameter estimation errors δ of the O-MISG algorithm become smaller with the innovation length p increasing and converge to zero fast if the innovation length p is large enough, and the data length tends to infinity.

Table 3 O-MISG parameter estimates and errors ($\sigma^2 = 0.10^2$)

p	t	a_1	a_2	b_{11}	b_{12}	b_{21}	b_{22}	f_1	f_2	$\delta(\%)$
1	100	−0.41225	−0.05193	0.12058	0.23756	0.08789	0.11763	0.30070	0.57027	65.68515
	200	−0.41810	−0.05079	0.08500	0.22342	0.07709	0.11367	0.33100	0.58218	63.70229
	500	−0.41184	−0.03366	0.07205	0.21654	0.08390	0.11315	0.36324	0.59397	61.78741
	1000	−0.41977	−0.03597	0.05368	0.20699	0.08803	0.11022	0.38934	0.60431	60.17515
	2000	−0.42007	−0.03315	0.04310	0.20001	0.09312	0.11153	0.41067	0.61174	58.88622
	3000	−0.42002	−0.03148	0.03931	0.19679	0.09686	0.11228	0.42242	0.61560	58.19410
	4000	−0.41987	−0.03014	0.03559	0.19424	0.09819	0.11217	0.43054	0.61818	57.72220
	5000	−0.41989	−0.02896	0.03399	0.19261	0.09996	0.11243	0.43687	0.62047	57.35215
2	100	−0.41792	−0.05539	0.01601	0.19645	0.10905	0.14058	0.66216	0.70511	44.28592
	200	−0.41245	−0.04580	−0.01346	0.18314	0.10060	0.13583	0.69531	0.71623	42.50653
	500	−0.39725	−0.02902	−0.01002	0.17658	0.11327	0.13461	0.73335	0.72917	40.42732
	1000	−0.40548	−0.03898	−0.01381	0.16927	0.12319	0.13110	0.76175	0.74094	38.80544
	2000	−0.40210	−0.03816	−0.01525	0.16205	0.13034	0.13116	0.78403	0.74946	37.52788
	3000	−0.39994	−0.03815	−0.01395	0.15913	0.13414	0.13127	0.79639	0.75382	36.83365
	4000	−0.39855	−0.03774	−0.01419	0.15675	0.13523	0.13048	0.80460	0.75676	36.38399
	5000	−0.39774	−0.03758	−0.01329	0.15534	0.13699	0.13054	0.81114	0.75935	36.02166
3	100	−0.37704	−0.08511	−0.01466	0.14857	0.13415	0.15151	0.83678	0.82888	31.18628
	200	−0.36804	−0.07576	−0.03350	0.13840	0.12476	0.14571	0.86852	0.83787	29.64566
	500	−0.35179	−0.06599	−0.02070	0.13545	0.13408	0.14340	0.90646	0.84880	27.64423
	1000	−0.35901	−0.07972	−0.01689	0.13112	0.14114	0.13910	0.93268	0.85954	26.17008
	2000	−0.35337	−0.07964	−0.01533	0.12503	0.14546	0.13790	0.95216	0.86703	25.06772
	3000	−0.35021	−0.08068	−0.01308	0.12275	0.14703	0.13736	0.96294	0.87079	24.46580
	4000	−0.34862	−0.08110	−0.01227	0.12081	0.14712	0.13623	0.96994	0.87343	24.08610
	5000	−0.34761	−0.08163	−0.01088	0.11978	0.14805	0.13616	0.97550	0.87577	23.78034
5	100	−0.25276	−0.17749	−0.00894	0.08525	0.14872	0.15127	1.03398	1.08118	10.30728
	200	−0.24532	−0.17737	−0.01431	0.08101	0.14094	0.14559	1.05557	1.07953	9.20002
	500	−0.23528	−0.17865	−0.00496	0.08370	0.14466	0.14353	1.08125	1.07823	7.83413
	1000	−0.23785	−0.18830	0.00020	0.08354	0.14685	0.14150	1.09690	1.08007	6.97197
	2000	−0.23392	−0.18856	0.00015	0.08090	0.14885	0.14086	1.10674	1.08110	6.44755
	3000	−0.23196	−0.19023	0.00059	0.07979	0.14901	0.14073	1.11240	1.08139	6.14245
	4000	−0.23106	−0.19073	0.00103	0.07884	0.14879	0.14011	1.11602	1.08178	5.96310
	5000	−0.23061	−0.19134	0.00195	0.07857	0.14907	0.14017	1.11866	1.08238	5.82491
8	100	−0.18021	−0.23990	−0.00625	0.07619	0.14691	0.15107	1.08344	1.19675	6.42298
	200	−0.17259	−0.24196	−0.00754	0.07337	0.14090	0.14555	1.10447	1.18694	5.18775
	500	−0.16877	−0.24409	−0.00245	0.07510	0.14717	0.14471	1.12770	1.17762	3.90131
	1000	−0.17113	−0.24849	0.00321	0.07512	0.14714	0.14348	1.13994	1.17397	3.29856
	2000	−0.17147	−0.24712	0.00284	0.07314	0.14881	0.14293	1.14571	1.17134	2.95188
	3000	−0.17094	−0.24754	0.00259	0.07171	0.14849	0.14300	1.14943	1.16941	2.77712
	4000	−0.17096	−0.24711	0.00311	0.07094	0.14817	0.14249	1.15186	1.16836	2.64795
	5000	−0.17144	−0.24708	0.00418	0.07083	0.14825	0.14246	1.15327	1.16789	2.57161
True values		−0.19000	−0.23000	0.01000	0.06000	0.15000	0.14000	1.17000	1.14000	

Table 4 O-RLS parameter estimates and errors ($\sigma^2 = 0.10^2$)

t	a_1	a_2	b_{11}	b_{12}	b_{21}	b_{22}	f_1	f_2	δ (%)
100	−0.20646	−0.22561	−0.00134	0.06605	0.14145	0.16103	1.16809	1.12804	1.99646
200	−0.20192	−0.22656	0.00710	0.06032	0.13947	0.15223	1.17405	1.12761	1.45425
500	−0.19641	−0.22960	0.00175	0.06964	0.15303	0.13440	1.17951	1.12776	1.31285
1000	−0.19106	−0.23399	0.00799	0.06616	0.14999	0.13761	1.17586	1.13621	0.63599
2000	−0.19166	−0.23248	0.00693	0.06603	0.15201	0.13555	1.17025	1.13733	0.55302
3000	−0.18916	−0.23485	0.00712	0.06425	0.15041	0.13841	1.17128	1.13888	0.44786
4000	−0.18993	−0.23408	0.00847	0.06204	0.15029	0.13805	1.17182	1.13790	0.35223
5000	−0.18951	−0.23435	0.00918	0.06139	0.15023	0.13825	1.17146	1.13868	0.31991
True values	−0.19000	−0.23000	0.01000	0.06000	0.15000	0.14000	1.17000	1.14000	

Fig. 5 O-MISG parameter estimation errors δ versus t ($\sigma^2 = 0.10^2$)
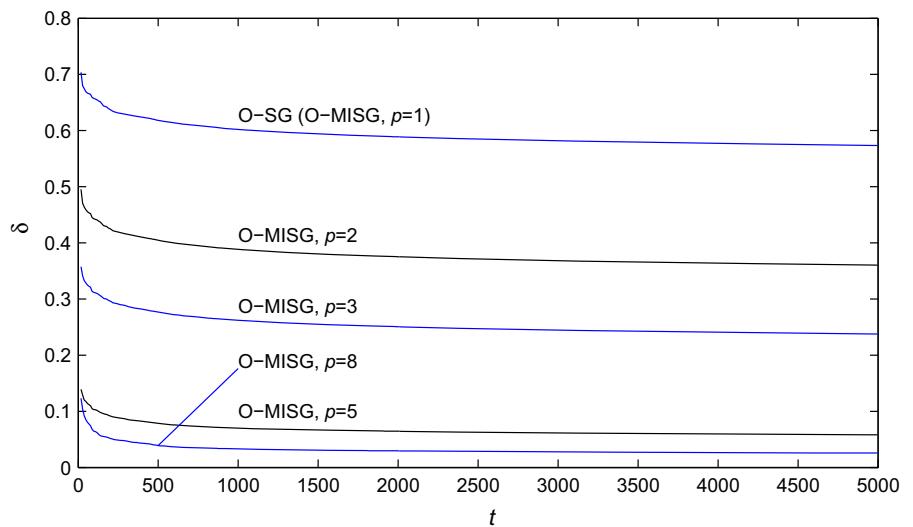
Fig. 6 O-RLS parameter estimation errors δ versus t ($\sigma^2 = 0.10^2$)
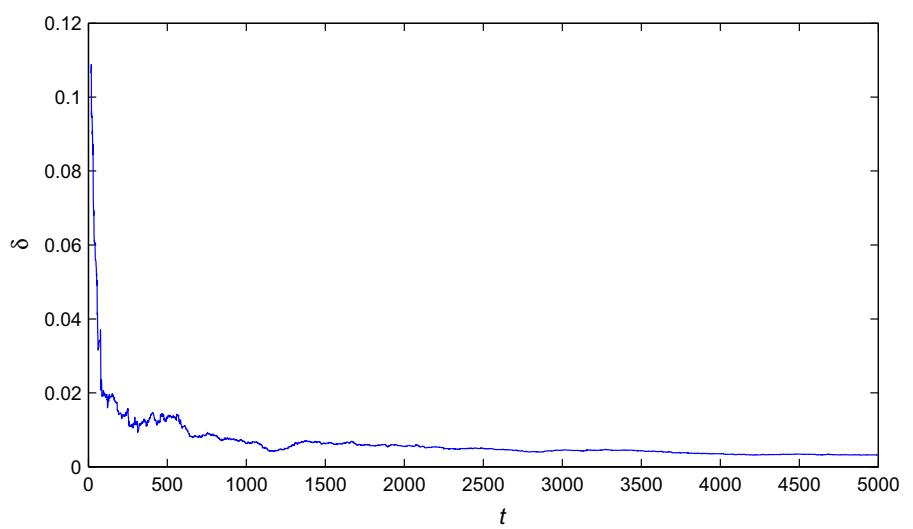
Fig. 7 State $x_1(t)$ and the estimated state $\hat{x}_1(t)$ against t ($p = 8$, $\sigma^2 = 0.10^2$)
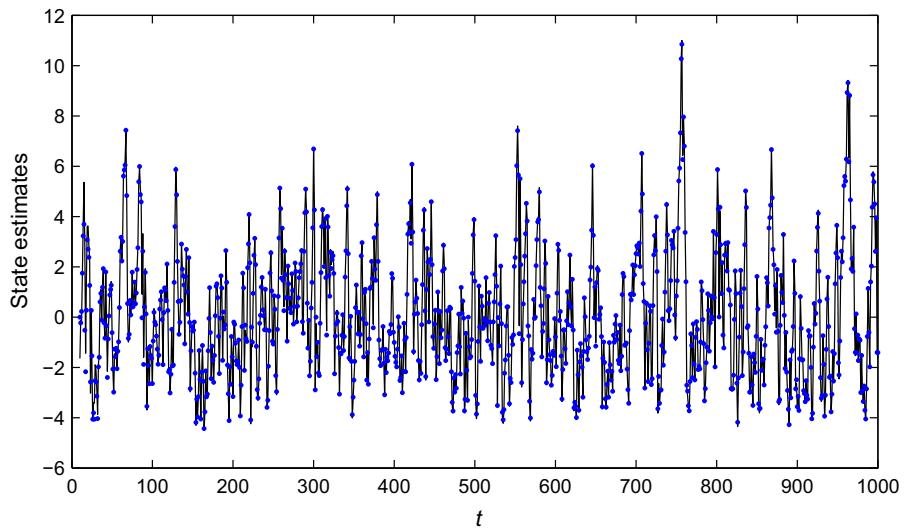

Fig. 8 State $x_2(t)$ and the estimated state $\hat{x}_2(t)$ against t ($p = 8$, $\sigma^2 = 0.10^2$)
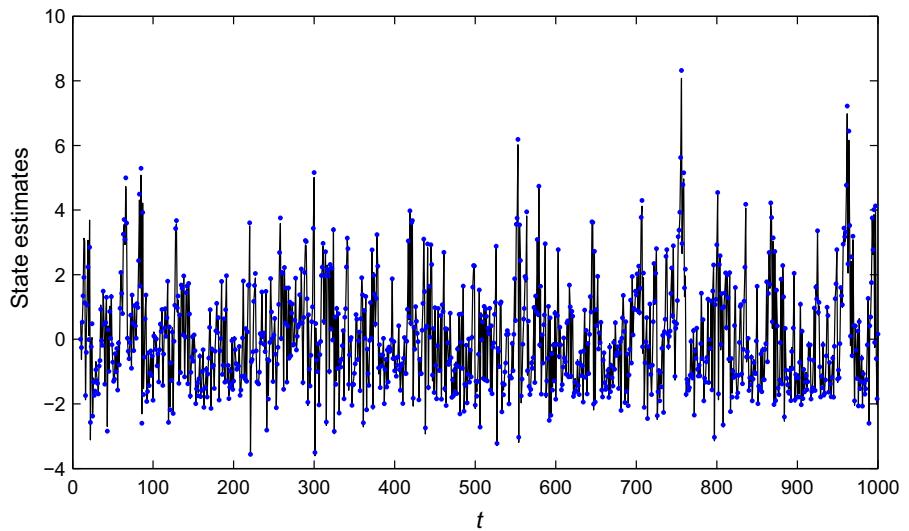
(3) Under the same innovation length, a smaller noise variance leads to higher parameter estimation accuracy and a faster convergence rate.

(4) The estimated states are very close to the actual states of the system.

6 Conclusions

This paper considers the parameter identification problems of bilinear state space systems using the multi-innovation theory and the least squares principle. The utilization of the non-singular linear transformation for general bilinear systems reduces the number of the parameters of the identification model, a state observer-based multi-innovation stochastic gradient algorithm and a state observer-based recursive least squares algorithm are derived for observer canonical bilinear state space systems. The convergence analysis indicates that the algorithms proposed are effective and the parameter estimates given by the proposed algorithms can converge to their true values. The numerical simulation results indicate that the designed state observer makes the estimated states close to the actual states of the systems. The O-MISG and O-RLS algorithm can give effective parameter estimates compared with the O-SG

algorithm. The multi-innovation identification method can improve the parameter estimate accuracy and the convergence rates compared with the single-innovation identification methods. The proposed algorithms can realize the interactive estimation for unknown states and parameters for bilinear systems.

Although the algorithms in the paper are developed for single-input single-output systems with white noise disturbance, the methods can be extended to identify multiple-input multiple-output systems by expanding the single input to multiple inputs and the single output to multiple outputs. They can also be extended to study identification problems of other linear and nonlinear multivariable systems with colored noises and applied to other fields [49–51].

References

1. Xu, L.: A proportional differential control method for a time-delay system using the Taylor expansion approximation. Appl. Math. Comput. **236**, 391–399 (2014)
2. Xu, L., Chen, L., Xiong, W.L.: Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. Nonlinear Dyn. **79**(3), 2155–2163 (2015)
3. Xu, L.: Application of the Newton iteration algorithm to the parameter estimation for dynamical systems. J. Comput. Appl. Math. **288**, 33–43 (2015)
4. Ding, F., Wang, F.F., Xu, L., Wu, M.H.: Decomposition based least squares iterative identification algorithm for multivariate pseudo-linear ARMA systems using the data filtering. J. Frankl. Inst. **354**(3), 1321–1339 (2017)
5. Ding, F., Xu, L., Zhu, Q.M.: Performance analysis of the generalised projection identification for time-varying systems. IET Control Theory Appl. **10**(18), 2506–2514 (2016)
6. Wang, D.Q., Zhang, W.: Improved least squares identification algorithm for multivariable Hammerstein systems. J. Frankl. Inst. **352**(11), 5292–5307 (2015)
7. Wang, D.Q.: Hierarchical parameter estimation for a class of MIMO Hammerstein systems based on the reframed models. Appl. Math. Lett. **57**, 13–19 (2016)
8. Mao, Y.W., Ding, F.: A novel parameter separation based identification algorithm for Hammerstein systems. Appl. Math. Lett. **60**, 21–27 (2016)
9. Mao, Y.W., Ding, F.: Multi-innovation stochastic gradient identification for Hammerstein controlled autoregressive autoregressive systems based on the filtering technique. Nonlinear Dyn. **79**(3), 1745–1755 (2015)
10. Xu, L.: The damping iterative parameter identification method for dynamical systems based on the sine signal measurement. Signal Process. **120**, 660–667 (2016)

11. Wang, C.N., He, Y.J., Ma, J.: Parameters estimation, mixed synchronization, and antisynchronization in chaotic systems. Complexity **20**(1), 64–73 (2014)
12. Ma, J., Zhang, A.H., Xia, Y.F.: Optimize design of adaptive synchronization controllers and parameter observers in different hyperchaotic systems. Appl. Math. Comput. **215**(9), 3318–3326 (2010)
13. Wang, D.Q., Ding, F.: Parameter estimation algorithms for multivariable Hammerstein CARMA systems. Inf. Sci. **355**, 237–248 (2016)
14. Van Overschee, P., De Moor, B.: Subspace Identification for Linear Systems: Theory, Implementation, Applications. Springer Science Business Media, New York (1996)
15. Wang, D.Q., Ding, F., Liu, X.M.: Least squares algorithm for an input nonlinear system with a dynamic subspace state space model. Nonlinear Dyn. **75**(1–2), 49–61 (2014)
16. Xu, L., Ding, F.: The parameter estimation algorithms for dynamical response signals based on the multi-innovation theory and the hierarchical principle. IET Signal Process. **11**(2), 228–237 (2017)
17. Wang, Y.J., Ding, F.: The auxiliary model based hierarchical gradient algorithms and convergence analysis using the filtering technique. Signal Process. **128**, 212–221 (2016)
18. Vörös, J.: Identification of nonlinear cascade systems with output hysteresis based on the key term separation principle. Appl. Math. Model. **39**(18), 5531–5539 (2015)
19. Vanbeylen, L., Pintelon, R., Schoukens, J.: Blind maximum likelihood identification of Hammerstein systems. Automatica **44**(12), 3139–3146 (2008)
20. Ase, H., Katayama, T.: A subspace-based identification of Wiener–Hammerstein benchmark model. Control Eng. Pract. **44**, 126–137 (2015)
21. Juang, J.N.: Applied System Identification. Prentice Hall, New Jersey (1994)
22. Juang, J.N., Phan, M., Horta, L.G.: Identification of observer/Kalman filter Markov parameters: theory and experiments. J. Guid. Control Dyn. **16**(2), 320–329 (1993)
23. Ma, J., Li, F., Huang, L., et al.: Complete synchronization, phase synchronization and parameters estimation in a realistic chaotic system. Commun. Nonlinear Sci. Numer. Simul. **16**(9), 3770–3785 (2011)
24. Hara, S., Furuta, K.: Minimal order state observers for bilinear systems. Int. J. Control **24**(5), 705–718 (1976)
25. Phan, M.Q., Vicario, F., Longman, R.W.: Optimal bilinear observers for bilinear state-space models by interaction matrices. Int. J. Control **88**(8), 1504–1522 (2015)
26. Ma, X.Y., Ding, F.: Recursive and iterative least squares parameter estimation algorithms for observability canonical state space systems. J. Franklin Inst. – Engineering. Appl. Math. **352**(1), 248–258 (2015)
27. Juang, J.N.: Continuous-time bilinear system identification. Nonlinear Dyn. **39**(1–2), 79–94 (2005)
28. Lee, C.H., Juang, J.N.: System identification for a general class of observable and reachable bilinear systems. J. Vib. Control **20**(10), 1538–1551 (2013)
29. Hizir, N.B., Phan, M.Q., Betti, R.: Identification of discrete-time bilinear systems through equivalent linear models. Nonlinear Dyn. **69**(4), 2065–2078 (2012)
30. Vicario, F., Phan, M.Q., Longman, R.W.: A linear-time-varying approach for exact identification of bilinear discrete-

time systems by interaction matrices. Astronaut. Sci. **150**, 1057–1076 (2014)

31. Jan, Y.G., Wong, K.M.: Bilinear system identification by block pulse functions. J. Frankl. Inst. **312**(5), 349–359 (1981)

32. Dai, H., Sinha, N.K.: Robust recursive least-squares method with modified weights for bilinear system identification. IEEE Proc. D Control Theory Appl. **163**(3), 122–126 (1989)

33. dos Santos, P.L., Ramos, J.A., de Carvalho, J.L.M.: Identification of bilinear systems with white noise inputs: an iterative deterministic–stochastic subspace approach. IEEE Trans. Control Syst. Techn. **17**(5), 1145–1153 (2009)

34. Li, B.B.: State estimation with partially observed inputs: a unified Kalman filtering approach. Automatica **49**(3), 816–820 (2013)

35. Pan, J., Yang, X.H., Cai, H.F., Mu, B.X.: Image noise smoothing using a modified Kalman filter. Neurocomputing **173**, 1625–1629 (2016)

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

37. Schön, T.B., Wills, A., Ninness, B.: System identification of nonlinear state-space models. Automatica **47**(1), 39–49 (2011)

38. Marconato, A., Sjöberg, J., Suykens, J., Schoukens, J.: Identification of the Silverbox benchmark using nonlinear state-space models. IFAC Proc. **45**(16), 632–637 (2012)

39. Ljung, L.: System Identification: Theory for the User. Englewood Cliffs, New Jersey (1987)

40. Söderström, T., Stoica, P.: System Identification. Prentice-Hall Inc, New Jersey (1988)

41. Xu, L., Ding, F., Gu, Y., Alsaedi, A., Hayat, T.: A multi-innovation state and parameter estimation algorithm for a state space system with d-step state-delay. Signal Process. **140**, 97–103 (2017)

42. Wang, F.F., Liu, Y.J., Yang, E.: Least squares-based iterative identification methods for linear-in-parameters systems using the decomposition technique. Circuits Syst. Signal Process. **35**(11), 3863–3881 (2015)

43. Wang, Y.J., Ding, F.: The filtering based iterative identification for multivariable systems. IET Control Theory Appl. **10**(8), 894–902 (2016)

44. Arablouei, R., Doğancay, K., Adalı, T.: Unbiased recursive least-squares estimation utilizing dichotomous coordinate-descent iterations. IEEE Trans. Signal Process. **62**(11), 2973–2983 (2014)

45. Xu, L., Ding, F.: Recursive least squares and multi-innovation stochastic gradient parameter estimation methods for signal modeling. Circuits Syst. Signal Process. **36**(4), 1735–1753 (2017)

46. Wan, X.K., Li, Y., Xia, C., Wu, M.H., Liang, J., Wang, N.: A T-wave alternans assessment method based on least squares curve fitting technique. Measurement **86**, 93–100 (2016)

47. Ding, F., Gu, Y.: Performance analysis of the auxiliary model-based stochastic gradient parameter estimation algorithm for state space systems with one-step state delay. Circuits Syst. Signal Process. **32**(2), 585–599 (2013)

48. Ding, F.: State filtering and parameter estimation for state space systems with scarce measurements. Signal Process. **104**, 369–380 (2014)

49. Feng, L., Wu, M.H., Li, Q.X., et al.: Array factor forming for image reconstruction of one-dimensional nonuniform aperture synthesis radiometers. IEEE Geosci. Remote Sens. Lett. **13**(2), 237–241 (2016)

50. Wang, T.Z., Qi, J., Xu, H., et al.: Fault diagnosis method based on FFT-RPCA-SVM for cascaded-multilevel inverter. ISA Trans. **60**, 156–163 (2016)

51. Wang, T.Z., Wu, H., Ni, M.Q., et al.: An adaptive confidence limit for periodic non-steady conditions fault detection. Mech. Syst. Signal Process. **72–73**, 328–345 (2016)