

# Recursive Least Squares and Multi-innovation Gradient Estimation Algorithms for Bilinear Stochastic Systems

Dandan Meng<sup>1</sup>

Received: 19 September 2015 / Revised: 3 May 2016 / Accepted: 4 May 2016 /  
Published online: 25 May 2016  
© Springer Science+Business Media New York 2016

**Abstract** Bilinear systems are a special class of nonlinear systems. Some systems can be described by using bilinear models. This paper considers the parameter identification problems of bilinear stochastic systems. The difficulty of identification is that the model structure of the bilinear systems includes the products of the states and inputs. To this point, this paper gives the input–output representation of the bilinear systems through eliminating the state variables in the model and derives a least squares algorithm and a multi-innovation stochastic gradient algorithm for identifying the parameters of bilinear systems based on the least squares principle and the multi-innovation identification theory. The simulation results indicate that the proposed algorithms are effective for identifying bilinear systems.

**Keywords** Parameter estimation · Least squares · Multi-innovation identification · Bilinear system

## 1 Introduction

The parameter identification of linear systems has been mature [31,32], and many identification methods have been developed, e.g., the recursive identification methods [25] and the iterative identification methods [34,44]. By comparison, there is a lot of research space for the parameter identification for bilinear stochastic systems. The identification of bilinear systems has been studied for a long history, at least more than four decades. Earlier work on identification for bilinear systems exists: Karanam

---

✉ Dandan Meng  
ddmeng12@126.com

<sup>1</sup> Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, People's Republic of China

et al. [26] utilized the Walsh functions for bilinear system identification; Fnaiech and Ljung [19] discussed recursive identification methods for bilinear systems. In respect of nonlinear system identification, the heuristic optimization techniques seem to be very helpful and have been used in system identification [3,5], control engineering [6,12,30], parameter estimation [2,4] and the Lorenz chaotic systems [1]. By means of the heuristic algorithms, Modares et al. [29] proposed an adaptive particle swarm optimization (PSO) algorithm to estimate the parameters of bilinear systems, increasing the convergence speed and accuracy compared with the basic PSO algorithm. Gibson et al. [20] developed a maximum-likelihood parameter estimation algorithm of bilinear systems. Inspired by the fixed point theory, Li et al. [27] presented an iterative algorithm to recursively identify bilinear models. Lopes dos Santos et al. [28] identified the Wiener–Hammerstein bilinear system using an iterative bilinear subspace identification algorithm.

In the field of nonlinear system identification, much work has been performed for decades [41]. A class of nonlinear systems can be described by a static nonlinearity followed by a linear subsystem. For such nonlinear systems, a decomposition-based Newton iterative identification method [14] and a hierarchical parameter estimation algorithm [33] were developed. Also, Chen et al. [9] presented a hierarchical gradient parameter estimation algorithm for Hammerstein nonlinear systems using the key term separation principle. Wang and Zhang [43] derived an improved least squares identification algorithm for multivariable Hammerstein systems.

The least squares (LS) methods are basic in system identification [18,39]. For example, for a class of linear-in-parameter output error moving average systems, Wang and Tang proposed an auxiliary model-based recursive least squares (RLS) algorithm [42]; Hu [21] developed iterative and recursive LS algorithms to estimate the parameters of moving average systems; Cao et al. [7] have presented a novel two-dimensional RLS identification approach with soft constraint for batch processes, which can improve the identification performance by exploiting information not only from time direction within a batch but also along batches. Recently, Ding et al. [15] presented an auxiliary model-based least squares algorithm for a dual-rate state space system with time delay using the data filtering and a Kalman state filtering-based least squares iterative parameter estimation algorithm for observer canonical state space systems using decomposition [17].

Multi-innovation identification is an important branch of system identification [13]. The innovation is the valuable information which can improve parameter or state estimation accuracy. In this respect, Ding [13] proposed the innovation-based identification theory and methods, by expanding a scalar innovation to an innovation vector and then to an innovation matrix. Based on this theory, Hu et al. [22] suggested a multi-innovation generalized extended stochastic gradient algorithm for output nonlinear autoregressive moving average systems.

This paper focuses on the parameter identification problems of bilinear stochastic systems. It is organized as follows. Section 2 derives the identification model for bilinear systems. Sections 3 and 4 propose a RLS identification algorithm and a multi-innovation stochastic gradient (MISG) identification algorithm. Section 5 provides two simulation examples to prove the validity of the proposed algorithms. Finally, some concluding remarks are made in Sect. 6.

## 2 The Identification Model for Bilinear Systems

In this section, we describe a bilinear system and derive its identification model for parameter estimation. Let us define some notation. “ $A =: X$ ” or “ $X := A$ ” represents “ $A$  is defined as  $X$ ,” and  $z^{-1}$  stands for a unit forward shift operator with the following properties:

$$\begin{aligned} z^n f(t) &= f(t+n), \\ z^n f(t)g(t) &= [z^n f(t)]g(t) = f(t+n)g(t), \\ z^n [f(t)g(t)] &= f(t+n)z^n g(t) = f(t+n)g(t+n). \end{aligned}$$

Assume that the bilinear systems under consideration have the following observability canonical form [11]:

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

$$y(t) = \mathbf{h}\mathbf{x}(t), \quad (2)$$

where  $\mathbf{x}(t) := [x_1(t), x_2(t), \dots, x_n(t)]^T$  is the  $n$ -dimensional state vector,  $u(t) \in \mathbb{R}$  and  $y(t) \in \mathbb{R}$  are the input and output of the system, respectively, and  $\mathbf{A} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{B} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{f} \in \mathbb{R}^n$  and  $\mathbf{h} \in \mathbb{R}^{1 \times n}$  are constant matrices and vectors:

$$\begin{aligned} \mathbf{A} &:= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}, \\ \mathbf{B} &:= \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \\ -b_n & -b_{n-1} & -b_{n-2} & \cdots & -b_1 \end{bmatrix}, \\ \mathbf{f} &:= [f_1, f_2, \dots, f_{n-1}, f_n]^T, \quad \mathbf{h} := [1, 0, \dots, 0, 0]. \end{aligned}$$

Define the following polynomials:

$$\begin{aligned} A(z) &:= 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}, \\ B(z) &:= b_1 z^{-1} + b_2 z^{-2} + \cdots + b_n z^{-n}, \\ C(z) &:= c_1 z^{-1} + c_2 z^{-2} + \cdots + c_n z^{-n}, \\ D(z) &:= d_2 z^{-2} + d_3 z^{-3} + \cdots + d_n z^{-n}. \end{aligned}$$

The coefficients  $c_i$  and  $d_i$  can be determined using  $a_i$ ,  $b_i$  and  $f_i$ . Referring to the method in [11], eliminating the state vector  $\mathbf{x}(t)$  in (1)–(2), we have the following input–output relation,

$$[A(z) + u(t - n)B(z)]y(t) = [C(z) + u(t - n)D(z)]u(t).$$

Since the practical systems are always interfered with various factors, e.g., stochastic noise, introducing a noise term  $v(t) \in \mathbb{R}$ , we have

$$[A(z) + u(t - n)B(z)]y(t) = [C(z) + u(t - n)D(z)]u(t) + v(t). \tag{3}$$

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

$$\begin{aligned} \theta &:= [\mathbf{a}^T, \mathbf{b}^T, \mathbf{c}^T, \mathbf{d}^T]^T \in \mathbb{R}^{4n-1}, \\ \varphi(t) &:= \left[ \varphi_y^T(t), u(t - n)\varphi_y^T(t), \varphi_u^T(t), u(t - n)\varphi_u^T(t) \right]^T \in \mathbb{R}^{4n-1}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{a} &:= [a_1, a_2, \dots, a_n]^T \in \mathbb{R}^n, & \mathbf{b} &:= [b_1, b_2, \dots, b_n]^T \in \mathbb{R}^n, \\ \mathbf{c} &:= [c_1, c_2, \dots, c_n]^T \in \mathbb{R}^n, & \mathbf{d} &:= [d_2, d_3, \dots, d_n]^T \in \mathbb{R}^{n-1}, \\ \varphi_y(t) &:= [-y(t - 1), -y(t - 2), \dots, -y(t - n)]^T \in \mathbb{R}^n, \\ \varphi_u(t) &:= [u(t - 1), u(t - 2), \dots, u(t - n)]^T \in \mathbb{R}^n, \\ \varphi_{\bar{u}}(t) &:= [u(t - 2), u(t - 3), \dots, u(t - n)]^T \in \mathbb{R}^{n-1}. \end{aligned}$$

Inserting  $A(z)$ ,  $B(z)$ ,  $C(z)$  and  $D(z)$  into (3) leads to

$$y(t) = \varphi^T(t)\theta + v(t). \tag{4}$$

Equation (4) is the identification model of the bilinear system with white noise.

The objective of the paper is to develop new identification methods for estimating the parameters  $a_i$ ,  $b_i$ ,  $c_i$  and  $d_i$  or the parameter vectors  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  and  $\mathbf{d}$ , from the observation data  $\{u(t), y(t) : t = 1, 2, 3, \dots\}$ .

### 3 The Recursive Least Squares Identification Algorithm

In this part, aiming at the model in (4), we propose the RLS estimation algorithm so that online identification can be carried out. Using the data with  $t = 1, 2, \dots, t$ , we define

$$\mathbf{Y}_t := \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(t) \end{bmatrix} \in \mathbb{R}^t, \quad \mathbf{F}_t := \begin{bmatrix} \varphi^T(1) \\ \varphi^T(2) \\ \vdots \\ \varphi^T(t) \end{bmatrix} \in \mathbb{R}^{t \times (4n-1)}, \quad \mathbf{V}_t := \begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(t) \end{bmatrix} \in \mathbb{R}^t.$$

According to (4), define a quadratic criterion function

$$\begin{aligned} J(\boldsymbol{\theta}) &:= (\mathbf{Y}_t - \boldsymbol{\Gamma}_t \boldsymbol{\theta})^\top (\mathbf{Y}_t - \boldsymbol{\Gamma}_t \boldsymbol{\theta}) \\ &= \sum_{j=1}^t [\mathbf{y}(j) - \boldsymbol{\varphi}^\top(j) \boldsymbol{\theta}]^2 \end{aligned}$$

Minimizing the criterion function  $J(\boldsymbol{\theta})$  and letting the partial derivative of  $J(\boldsymbol{\theta})$  with regard to  $\boldsymbol{\theta}$  be zero, we obtain the least squares estimate of  $\boldsymbol{\theta}$ :

$$\hat{\boldsymbol{\theta}}(t) = (\boldsymbol{\Gamma}_t^\top \boldsymbol{\Gamma}_t)^{-1} \boldsymbol{\Gamma}_t^\top \mathbf{Y}_t, \quad (5)$$

where  $\hat{\boldsymbol{\theta}}(t)$  is the estimate of  $\boldsymbol{\theta}$  at time  $t$ ,  $\hat{\boldsymbol{\theta}}(0)$  is a very small real vector. To derive the recursive least squares algorithm conveniently, define the covariance matrix  $\mathbf{P}(t)$  as

$$\mathbf{P}^{-1}(t) := \boldsymbol{\Gamma}_t^\top \boldsymbol{\Gamma}_t \in \mathbb{R}^{4n-1}. \quad (6)$$

Using the definition of  $\boldsymbol{\Gamma}_t$ , it follows that

$$\mathbf{P}^{-1}(t) = \sum_{j=1}^t \boldsymbol{\varphi}(j) \boldsymbol{\varphi}^\top(j) = \mathbf{P}^{-1}(t-1) + \boldsymbol{\varphi}(t) \boldsymbol{\varphi}^\top(t), \quad \mathbf{P}(0) = p_0 \mathbf{I}_{4n-1} > 0. \quad (7)$$

To avoid computing the matrix inversion of  $\mathbf{P}(t)$ , applying the matrix inverse lemma [18]

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

to (7) gives

$$\mathbf{P}(t) = \mathbf{P}(t-1) - \mathbf{P}(t-1) \boldsymbol{\varphi}(t) [1 + \boldsymbol{\varphi}^\top(t) \mathbf{P}(t-1) \boldsymbol{\varphi}(t)]^{-1} \boldsymbol{\varphi}^\top(t) \mathbf{P}(t-1). \quad (8)$$

Using (6) and from (5), we have

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t) &= \mathbf{P}(t) \boldsymbol{\Gamma}_t^\top \mathbf{Y}_t \\ &= \mathbf{P}(t) \left[ \boldsymbol{\Gamma}_{t-1}^\top, \boldsymbol{\varphi}(t) \right] \begin{bmatrix} \mathbf{Y}_{t-1} \\ \mathbf{y}(t) \end{bmatrix} \\ &= \mathbf{P}(t) \left[ \boldsymbol{\Gamma}_{t-1}^\top \mathbf{Y}_{t-1} + \boldsymbol{\varphi}(t) \mathbf{y}(t) \right] \\ &= \mathbf{P}(t) \left[ \mathbf{P}^{-1}(t-1) \mathbf{P}(t-1) \boldsymbol{\Gamma}_{t-1}^\top \mathbf{Y}_{t-1} + \boldsymbol{\varphi}(t) \mathbf{y}(t) \right] \\ &= \mathbf{P}(t) \left[ \mathbf{P}^{-1}(t-1) \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{\varphi}(t) \mathbf{y}(t) \right] \\ &= \mathbf{P}(t) \left[ \mathbf{P}^{-1}(t) - \boldsymbol{\varphi}(t) \boldsymbol{\varphi}^\top(t) \right] \hat{\boldsymbol{\theta}}(t-1) + \mathbf{P}(t) \boldsymbol{\varphi}(t) \mathbf{y}(t) \\ &= \hat{\boldsymbol{\theta}}(t-1) + \mathbf{P}(t) \boldsymbol{\varphi}(t) \left[ \mathbf{y}(t) - \boldsymbol{\varphi}^\top(t) \hat{\boldsymbol{\theta}}(t-1) \right]. \end{aligned} \quad (9)$$

Combining (8) and (9), we can obtain the RLS algorithm for estimating the parameter vector  $\theta$  in (4):

$$\hat{\theta}(t) = \hat{\theta}(t - 1) + \mathbf{P}(t)\varphi(t) \left[ y(t) - \varphi^T(t)\hat{\theta}(t - 1) \right], \tag{10}$$

$$\mathbf{P}(t) = \mathbf{P}(t - 1) - \mathbf{P}(t - 1)\varphi(t) \left[ 1 + \varphi^T(t)\mathbf{P}(t - 1)\varphi(t) \right]^{-1} \varphi^T(t)\mathbf{P}(t - 1), \tag{11}$$

$$\varphi^T(t) = \left[ \varphi_y^T(t), u(t - n)\varphi_y^T(t), \varphi_u^T(t), u(t - n)\varphi_u^T(t) \right], \tag{12}$$

$$\varphi_u^T(t) = [u(t - 1), u(t - 2), \dots, u(t - n)], \tag{13}$$

$$\varphi_{\bar{u}}(t) = [u(t - 2), u(t - 3), \dots, u(t - n)]^T, \tag{14}$$

$$\varphi_y^T(t) = [-y(t - 1), -y(t - 2), \dots, -y(t - n)], \tag{14}$$

$$\hat{\theta} = \left[ \hat{a}^T(t), \hat{b}^T(t), \hat{c}^T(t), \hat{d}^T(t) \right]^T,$$

$$\hat{a}(t) = [\hat{a}_1(t), \hat{a}_2(t), \dots, \hat{a}_n(t)]^T, \quad \hat{b}(t) = [\hat{b}_1(t), \hat{b}_2(t), \dots, \hat{b}_n(t)]^T,$$

$$\hat{c}(t) = [\hat{c}_1(t), \hat{c}_2(t), \dots, \hat{c}_n(t)]^T, \quad \hat{d}(t) = [\hat{d}_2(t), \hat{d}_3(t), \dots, \hat{d}_n(t)]^T. \tag{15}$$

The computational amount of the RLS algorithm for bilinear systems is given in Table 1 and its has  $(64n^2 - 8n - 2)$  flops. The RLS algorithm is simple and can be implemented online.

### 4 The Multi-Innovation Stochastic Gradient Identification Algorithm

Based on the identification model in (4), we can acquire the stochastic gradient (SG) identification algorithm;

$$\hat{\theta}(t) = \hat{\theta}(t - 1) + \frac{\varphi(t)}{r(t)}e(t), \tag{16}$$

$$e(t) = y(t) - \varphi^T(t)\hat{\theta}(t - 1), \tag{17}$$

$$r(t) = r(t - 1) + \|\varphi(t)\|^2, \quad r(0) = 1, \tag{18}$$

**Table 1** Computation load of the RLS algorithm

Variables	Numbers of multiplications	Numbers of additions
$\hat{\theta}(t) = \hat{\theta}(t - 1) + \mathbf{L}(t)e(t)$	$4n - 1$	$4n - 1$
$e(t) := y(t) - \varphi^T(t)\hat{\theta}(t - 1)$	$4n - 1$	$4n - 1$
$\zeta(t) := \mathbf{P}(t - 1)\varphi(t)$	$(4n - 1)^2$	$(4n - 1)^2 - (4n - 1)$
$\mathbf{L}(t) = \zeta(t)/[1 + \varphi^T(t)\zeta(t)]$	$2(4n - 1)$	$4n - 1$
$\mathbf{P}(t) = \mathbf{P}(t - 1) - \mathbf{L}(t)\zeta^T(t)$	$(4n - 1)^2$	$(4n - 1)^2$
Sum	$32n^2 - 2$	$32n^2 - 8n$
Total flops	$N_1 := 64n^2 - 8n - 2$	

where the norm of a matrix  $X$  is defined by  $\|X\|^2 := \text{tr}[XX^T]$ . It is well known that the SG algorithm has slow convergence rates, in order to improve the situation, expanding the scalar innovation  $e(t)$  to an innovation vector [40]:

$$E(p, t) := \begin{bmatrix} y(t) - \boldsymbol{\varphi}^T(t)\hat{\boldsymbol{\theta}}(t-1) \\ y(t-1) - \boldsymbol{\varphi}^T(t-1)\hat{\boldsymbol{\theta}}(t-1) \\ \vdots \\ y(t-p+1) - \boldsymbol{\varphi}^T(t-p+1)\hat{\boldsymbol{\theta}}(t-1) \end{bmatrix} \in \mathbb{R}^p. \tag{19}$$

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

$$\begin{aligned} \Phi(p, t) &:= [\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(t-1), \dots, \boldsymbol{\varphi}(t-p+1)] \in \mathbb{R}^{(4n-1) \times p}, \\ Y(p, t) &:= [y(t), y(t-1), \dots, y(t-p+1)]^T \in \mathbb{R}^p. \end{aligned}$$

Then the innovation vector  $E(p, t)$  in (19) can be expressed as

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

We can obtain the MISG algorithm for bilinear systems:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\Phi(p, t)}{r(t)} E(p, t), \quad \hat{\boldsymbol{\theta}}(0) = \mathbf{1}_{4n-1}/p_0, \tag{20}$$

$$E(p, t) = Y(p, t) - \Phi^T(p, t)\hat{\boldsymbol{\theta}}(t-1), \tag{21}$$

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

$$Y(p, t) = [y(t), y(t-1), \dots, y(t-p+1)]^T, \tag{23}$$

$$\Phi(p, t) = [\boldsymbol{\varphi}(t), \boldsymbol{\varphi}(t-1), \dots, \boldsymbol{\varphi}(t-p+1)], \tag{24}$$

$$\boldsymbol{\varphi}^T(t) = [\boldsymbol{\varphi}_y^T(t), u(t-n)\boldsymbol{\varphi}_y^T(t), \boldsymbol{\varphi}_u^T(t), u(t-n)\boldsymbol{\varphi}_u^T(t)], \tag{25}$$

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

$$\boldsymbol{\varphi}_u^T(t) = [u(t-2), u(t-3), \dots, u(t-n)]^T, \tag{26}$$

$$\boldsymbol{\varphi}_y^T(t) = [-y(t-1), -y(t-2), \dots, -y(t-n)], \tag{27}$$

$$\hat{\boldsymbol{\theta}}(t) = [\hat{\mathbf{a}}^T(t), \hat{\mathbf{b}}^T(t), \hat{\mathbf{c}}^T(t), \hat{\mathbf{d}}^T(t)]^T. \tag{28}$$

In this algorithm,  $E(p, t) \in \mathbb{R}^p$  is an innovation vector (i.e., multi-innovation). When  $p = 1$ , the MISG algorithm reduces to the SG algorithm.

The advantage of the multi-innovation identification method is that it makes full use of the identification innovations and can raise the convergence speed to a certain degree, and the stationarity of data can improve identification accuracy. The computation amount of the MISG algorithm for bilinear systems is  $(24pn - 5p)$  flops as given in Table 2.

Equation (22) can be modified as

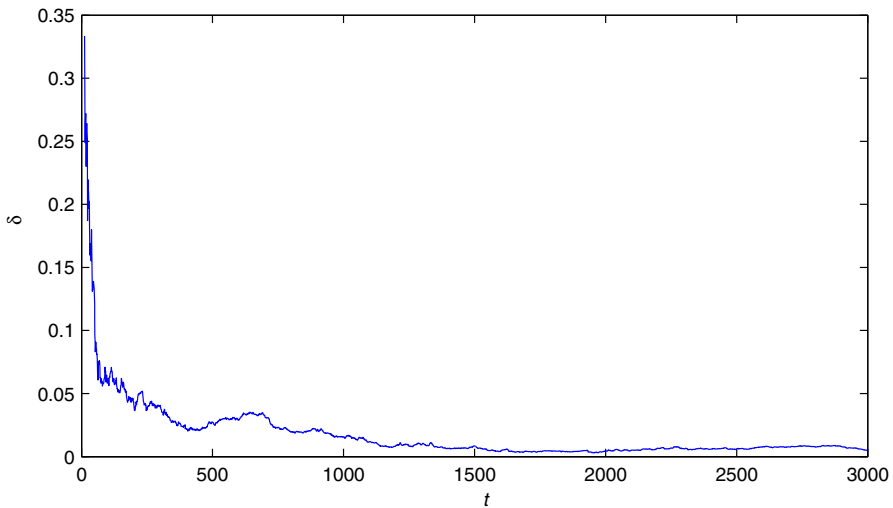
$$r(t) = r(t-1) + \|\boldsymbol{\varphi}(t)\|^2, \quad r(0) = 1. \tag{29}$$

**Table 2** Computation load of the MISG algorithm

Variables	Numbers of multiplications	Numbers of additions
$\hat{\theta}(t) = \hat{\theta}(t - 1) + \frac{\Phi(p,t)}{r(t)} E(p, t)$	$p(4n - 1) + p$	$p(4n - 1)$
$E(p, t) = Y(p, t) - \Phi^T(p, t)\hat{\theta}(t - 1)$	$p(4n - 1)$	$p(4n - 1)$
$r(t) = r(t - 1) + \ \Phi(p, t)\ ^2$	$p(4n - 1)$	$p(4n - 1)$
Sum	$12pn - 2p$	$12pn - 3p$
Total flops	$N_2 := 24pn - 5p$	

**Table 3** RLS parameter estimates and errors

$t$	$-a_1$	$-a_2$	$-b_1$	$-b_2$	$c_1$	$c_2$	$d_2$	$\delta$ (%)
100	0.89217	0.49346	-0.79428	-0.30942	0.24393	-0.58939	-0.26156	5.87185
200	0.89513	0.49579	-0.79304	-0.31000	0.26787	-0.60345	-0.23819	4.13730
500	0.89354	0.49763	-0.79764	-0.31021	0.32176	-0.60339	-0.25662	2.71974
1000	0.89228	0.50774	-0.79775	-0.31463	0.32659	-0.62217	-0.25235	1.57060
2000	0.89747	0.51815	-0.79345	-0.31142	0.32088	-0.63723	-0.24934	0.43879
3000	0.90053	0.52128	-0.79461	-0.31199	0.31321	-0.64414	-0.24820	0.53781
True values	0.90000	0.52000	-0.79300	-0.31200	0.31500	-0.63650	-0.24980	
Mean values	0.89597	0.51285	-0.79498	-0.31378	0.3101	-0.62406	-0.25146	
Standard deviation	0.024515	0.03157	0.032438	0.059533	0.042088	0.045866	0.044199	



**Fig. 1** RLS parameter estimation errors  $\delta$  versus  $t$

In order to track the time-varying parameters, we must introducing a forgetting factor  $\lambda$  in (22) and get

$$r(t) = \lambda r(t - 1) + \|\Phi(p, t)\|^2, \quad r(0) = 1, \tag{30}$$



**Table 4** SG parameter estimates and errors

$t$	$-a_1$	$-a_2$	$-b_1$	$-b_2$	$c_1$	$c_2$	$d_2$	$\delta$ (%)
100	0.57860	0.34192	-0.67618	-0.26261	-0.17343	-0.21990	-0.01173	40.94876
200	0.61708	0.34365	-0.68975	-0.26434	-0.16846	-0.22591	-0.01799	39.02828
500	0.65672	0.34747	-0.70040	-0.26039	-0.15787	-0.23694	-0.02676	36.89670
1000	0.66891	0.35149	-0.70082	-0.26111	-0.15119	-0.24285	-0.03053	36.07331
2000	0.69398	0.35964	-0.71040	-0.26120	-0.14445	-0.24981	-0.03495	34.66370
3000	0.70161	0.36168	-0.71158	-0.26163	-0.14058	-0.25343	-0.03751	34.16188
True values	0.90000	0.52000	-0.79300	-0.31200	0.31500	-0.63650	-0.24980	
Mean values	0.66372	0.26183	-0.59899	-0.07633	-0.12658	-0.30271	0.17034	
Standard deviation	0.042447	0.037624	0.017994	0.034471	0.0085074	0.01019	0.011887	

or

$$r(t) = \lambda r(t-1) + \|\varphi(t)\|^2, \quad r(0) = 1. \quad (31)$$

## 5 Examples

*Example 1* Consider the following bilinear system:

$$\begin{aligned} [A(z) + u(t-n)B(z)]y(t) &= [C(z) + u(t-n)D(z)]u(t) + v(t), \\ A(z) &= 1 + a_1z^{-1} + a_2z^{-2} = 1 + 0.90z^{-1} + 0.52z^{-2}, \\ B(z) &= b_1z^{-1} + b_2z^{-2} = -0.793z^{-1} - 0.312z^{-2}, \\ C(z) &= c_1z^{-1} + c_2z^{-2} = 0.315z^{-1} - 0.6365z^{-2}, \\ D(z) &= d_2z^{-2} = -0.2498z^{-2}. \end{aligned}$$

The parameter vector to be estimated is given by

$$\begin{aligned} \theta &= [a_1, a_2, b_1, b_2, c_1, c_2, d_2]^T \\ &= [0.90, 0.52, -0.793, -0.312, 0.315, -0.6365, -0.2498]^T. \end{aligned}$$

In simulation, we adopt an independent persistent excitation signal sequence with zero mean and unit variance as the input  $\{u(t)\}$ , and a white noise sequence with zero mean and variance  $\sigma^2 = 0.50^2$  as  $\{v(t)\}$ . Applying the recursive least squares algorithm in (10)–(15) to compute the parameters estimates  $\hat{\theta}(t)$  of the system, the parameter estimates and the estimation errors are given in Table 3, the estimation errors  $\delta$  versus the data length  $t$  is shown in Fig. 1. The parameter estimation errors are computed by  $\delta := \|\hat{\theta}(t) - \theta\| / \|\theta\|$ .

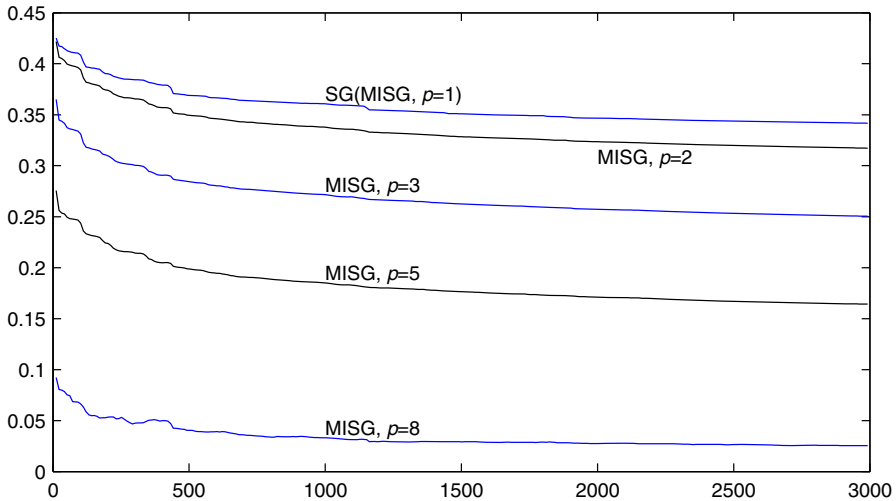
From Table 3 and Fig. 1, we can see that the RLS parameter estimation errors become smaller with the increase of  $t$ .

**Table 5** MISG gradient parameter estimates and errors

$p$	$t$	$-a_1$	$-a_2$	$-b_1$	$-b_2$	$c_1$	$c_2$	$d_2$	$\delta$ (%)
2	100	0.69647	0.40709	-0.71296	-0.34540	-0.28709	-0.09565	-0.22859	39.49122
	200	0.73778	0.41109	-0.73818	-0.33043	-0.27404	-0.11074	-0.23262	37.40836
	500	0.78069	0.41711	-0.76013	-0.30688	-0.24790	-0.13491	-0.23513	34.94728
	1000	0.79156	0.42631	-0.75982	-0.30860	-0.23192	-0.14914	-0.23317	33.77682
	2000	0.81297	0.43955	-0.77188	-0.30298	-0.21597	-0.16408	-0.23326	32.35225
	3000	0.82044	0.44232	-0.77236	-0.30283	-0.20717	-0.17227	-0.23288	31.70967
	True values	0.90000	0.52000	-0.79300	-0.31200	0.31500	-0.63650	-0.24980	
	Mean values	0.79364	0.43052	-0.76277	-0.30824	-0.22931	-0.15148	-0.23345	
	Standard deviation	0.03998	0.013967	0.015365	0.016953	0.022147	0.020882	0.0055945	
3	100	0.76015	0.44566	-0.72335	-0.33550	-0.37312	-0.17533	-0.16236	33.20983
	200	0.79993	0.44718	-0.75090	-0.31819	-0.35001	-0.19419	-0.17209	31.01859
	500	0.83719	0.45365	-0.77277	-0.29443	-0.30772	-0.22432	-0.18134	28.41751
	1000	0.84429	0.46515	-0.76855	-0.30104	-0.28152	-0.24106	-0.18267	27.15597
	2000	0.85939	0.47772	-0.77923	-0.29884	-0.25561	-0.25871	-0.18680	25.72476
	3000	0.86556	0.47997	-0.77814	-0.30007	-0.24138	-0.26814	-0.18857	25.04032
	Mean values	0.84422	0.46857	-0.77132	-0.30188	-0.27759	-0.24365	-0.18363	
	Standard deviation	0.03767	0.014211	0.013958	0.015113	0.035861	0.024927	0.007844	
	100	0.83243	0.49453	-0.73351	-0.30860	-0.35780	-0.28631	-0.18442	24.42510
5	200	0.86397	0.48443	-0.75968	-0.29995	-0.32007	-0.30867	-0.19615	22.40645
	500	0.88396	0.48904	-0.78322	-0.28302	-0.25531	-0.34379	-0.20624	19.86344
	1000	0.88433	0.50172	-0.77579	-0.29776	-0.21561	-0.36265	-0.20693	18.50900
	2000	0.89177	0.50860	-0.78715	-0.30004	-0.17747	-0.38212	-0.21089	17.11273

Table 5 continued

$p$	$t$	$-a_1$	$-a_2$	$-b_1$	$-b_2$	$c_1$	$c_2$	$d_2$	$\delta$ (%)
	3000	0.89586	0.50950	-0.78423	-0.30270	-0.15708	-0.39236	-0.21253	16.41585
	Mean values	0.88383	0.5023	-0.77916	-0.29842	-0.21084	-0.36513	-0.20783	
	Standard deviation	0.029063	0.01317	0.014383	0.013953	0.054703	0.028537	0.0079798	
8	100	0.88403	0.51407	-0.75805	-0.28742	0.02339	-0.68155	-0.17152	6.72093
	200	0.89988	0.49219	-0.77959	-0.28646	0.04570	-0.67697	-0.19338	5.31592
	500	0.88976	0.50428	-0.79041	-0.28799	0.08687	-0.67131	-0.21031	4.03711
	1000	0.88701	0.50961	-0.78520	-0.29817	0.11299	-0.66632	-0.21706	3.30890
	2000	0.89175	0.51425	-0.79169	-0.30473	0.13702	-0.66362	-0.22109	2.74642
	3000	0.89254	0.51546	-0.78653	-0.30767	0.14909	-0.66202	-0.22350	2.55075
	Mean values	0.89004	0.51019	-0.78656	-0.30044	0.11558	-0.66696	-0.21547	
	Standard deviation	0.0096676	0.016343	0.0094675	0.018379	0.035059	0.0083604	0.013777	



**Fig. 2** Parameter estimation errors  $\delta$  versus  $t$

*Example 2* Consider the system in Example 1 and the same simulation conditions. Applying the MISG algorithm to estimate the parameters of this system, the parameter estimates and their errors are shown in Tables 4, 5 and Fig. 2, where Table 4 lists the parameter estimates and errors given by the MISG algorithm when  $p = 1$ , i.e., the stochastic gradient algorithm.

From Table 5 and Fig. 2, it is observed that the MISG algorithm can identify the parameters of the bilinear system. Meanwhile, the estimation errors become smaller as the innovation length increases. These results show that the proposed algorithms work well for bilinear systems with white noise.

## 6 Conclusions

This paper derives the identification model of bilinear systems. On the basis of the obtained model, we present a RLS algorithm and an MISG algorithm for bilinear stochastic systems and give the state estimation algorithm. The simulation results indicate that for the multi-innovation identification algorithm, increasing the innovation length can enhance the parameter estimation accuracy, reducing the sensitivity of the algorithm to noise. The algorithms in this paper can be extended to study the identification problems of other bilinear systems, linear and nonlinear systems with colored noise [16, 33, 35], stochastic multivariable systems [36–38], uncertain chaotic delayed nonlinear systems [23] and hybrid switching–impulsive dynamical networks [24], and applied to other fields [8, 10].

When the uncertainty in parameters is small, the obtained parameter estimates are close to their true values. The estimates can be taken as the real parameters for controller design when the errors are tolerable.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (No. 61273194). The author is grateful to her Master Supervisor Professor Feng Ding and the main idea of this paper comes from him and his books “System Identification—New Theory and Methods, Science Press, Beijing, 2013,” “System Identification—Performances Analysis for Identification Methods, Science Press, Beijing, 2014” and “System Identification—Multi-Innovation Identification Theory and Methods, Science Press, Beijing, 2016.”

## References

1. A. Alfi, Particle swarm optimization algorithm with dynamic inertia weight for online parameter identification applied to Lorenz chaotic system. *Int. Innov. Comput. Inf. Control* **8**(2), 1191–1203 (2012)
2. A. Alfi, PSO with adaptive mutation and inertia weight and its application in parameter estimation of dynamic systems. *Acta Autom. Sin.* **37**(5), 541–549 (2011)
3. A. Alfi, M.M. Fateh, Identification of nonlinear systems using modified particle swarm optimization: a hydraulic suspension system. *J. Veh. Syst. Dyn.* **46**(6), 871–887 (2011)
4. A. Alfi, M.M. Fateh, Intelligent identification and control using improved fuzzy particle swarm optimization. *Expert Syst. Appl.* **38**(10), 12312–12317 (2011)
5. A. Alfi, H. Modares, System identification and control using adaptive particle swarm optimization. *Appl. Math. Model.* **35**, 1210–1221 (2011)
6. A. Arab, A. Alfi, An adaptive gradient descent-based local search in memetic algorithm for solving engineering optimization problems. *Inf. Sci.* **299**, 117–142 (2015)
7. Z.X. Cao, Y. Yang, J.Y. Lu, F.R. Gao, Constrained two dimensional recursive least squares model identification for batch processes. *J. Process Control* **24**(6), 871–879 (2014)
8. X. Cao, D.Q. Zhu, S.X. Yang, Multi-AUV target search based on bioinspired neurodynamics model in 3-D underwater environments. *IEEE Trans. Neural Netw. Learn. Syst.* (2016). doi:[10.1109/TNNLS.2015.2482501](https://doi.org/10.1109/TNNLS.2015.2482501)
9. H.B. Chen, Y.S. Xiao, Hierarchical gradient parameter estimation algorithm for Hammerstein nonlinear systems using the key term separation principle. *Appl. Math. Comput.* **247**, 1202–1210 (2014)
10. Z.Z. Chu, D.Q. Zhu, S.X. Yang, Observer-based adaptive neural network trajectory tracking control for remotely operated Vehicle. *IEEE Trans. Neural Netw. Learn. Syst.* (2016). doi:[10.1109/TNNLS.2016.2544786](https://doi.org/10.1109/TNNLS.2016.2544786)
11. H. Dai, N.K. Sinha, Robust recursive least-squares method with modified weights for bilinear system identification. *IEE Proc. Part D* **136**(3), 122–126 (1989)
12. A. Darabi, A. Alfi, B. Kiumarsi, H. Modares, Employing adaptive PSO algorithm for parameter estimation of an exciter machine. *J. Dyn. Syst-T ASME* **134**(1). doi:[10.1115/1.4005371](https://doi.org/10.1115/1.4005371), (2012)
13. F. Ding, *System Identification—New Theory and Methods* (Science Press, Beijing, 2013)
14. F. Ding, K.P. Deng, X.M. Liu, Decomposition based Newton iterative identification method for a Hammerstein nonlinear FIR system with ARMA noise. *Circuits Syst. Signal Process.* **33**(9), 2881–2893 (2014)
15. F. Ding, X.M. Liu, Y. Gu, An auxiliary model based least squares algorithm for a dual-rate state space system with time-delay using the data filtering. *J. Franklin Inst.* **353**(2), 398–408 (2016)
16. F. Ding, X.M. Liu, M.M. Liu, The recursive least squares identification algorithm for a class of Wiener nonlinear systems. *J. Franklin Inst.* **353**(7), 1518–1526 (2016)
17. F. Ding, X.M. Liu, X.Y. Ma, Kalman state filtering based least squares iterative parameter estimation for observer canonical state space systems using decomposition. *J. Comput. Appl. Math.* **301**, 135–143 (2016)
18. F. Ding, X.H. Wang, Q.J. Chen, Y.S. Xiao, Recursive least squares parameter estimation for a class of output nonlinear systems based on the model decomposition. *Circuits Syst. Signal Process.* **35**, (2016). doi:[10.1007/s00034-015-0190-6](https://doi.org/10.1007/s00034-015-0190-6)
19. F. Fnaiech, L. Ljung, Recursive identification of bilinear systems. *Int. J. Control* **45**(2), 453–470 (1987)
20. S. Gibson, A. Wills, B. Ninness, Maximum-likelihood parameter estimation of bilinear systems. *IEEE Trans. Autom. Control* **50**(10), 1581–1596 (2005)
21. Y.B. Hu, Iterative and recursive least squares estimation algorithms for moving average systems. *Simul. Model. Pract. Theory* **34**, 12–19 (2013)

22. Y.B. Hu, B.L. Liu, Q. Zhou, A multi-innovation generalized extended stochastic gradient algorithm for output nonlinear autoregressive moving average systems. *Appl. Math. Comput.* **247**, 218–224 (2014)
23. Y. Ji, X.M. Liu, New criteria for the robust impulsive synchronization of uncertain chaotic delayed nonlinear systems. *Nonlinear Dyn.* **79**(1), 1–9 (2015)
24. Y. Ji, X.M. Liu, Unified synchronization criteria for hybrid switching-impulsive dynamical networks. *Circuits Syst. Signal Process.* **34**(5), 1499–1517 (2015)
25. S.X. Jing, T.H. Pan, Z.M. Li, Recursive bayesian algorithm with covariance resetting for identification of Box–Jenkins systems with non-uniformly sampled input data. *Circuits Syst. Signal Process.* **16**, 1–14 (2015)
26. V.R. Karanam, P.A. Frick, R.R. Mohler, Bilinear system identification by Walsh functions. *IEEE Trans. Autom. Control* **23**, 709–713 (1978)
27. G.Q. Li, C.Y. Wen, A.M. Zhang, Fixed point iteration in identifying bilinear models. *Syst. Control Lett.* **83**, 28–37 (2015)
28. P. Lopes dos Santos, J.A. Ramos, J.L. Martins de Carvalho, Identification of a Benchmark Wiener–Hammerstein: a bilinear and Hammerstein–Bilinear model approach. *Control Eng. Pract.* **20**, 1156–1164 (2012)
29. H. Modares, A. Alfi, M.N. Sistani, Parameter estimation of bilinear systems based on an adaptive particle swarm optimization. *Eng. Appl. Artif. Intell.* **23**, 1105–1111 (2010)
30. Y. Mousavi, A. Alfi, A memetic algorithm applied to trajectory control by tuning of fractional order proportional–integral–derivative controllers. *Appl. Soft. Comput.* **36**, 599–617 (2015)
31. B.Q. Mu, J. Guo, L.Y. Wang, G. Yin, L.J. Xu, W.X. Zheng, Identification of linear continuous-time systems under irregular and random output sampling. *Automatica* **60**, 100–114 (2015)
32. J. Na, X.M. Ren, Y.Q. Xia, Adaptive parameter identification of linear SISO systems with unknown time-delay. *Syst. Control Lett.* **66**, 43–50 (2014)
33. D.Q. Wang, Hierarchical parameter estimation for a class of MIMO Hammerstein systems based on the reframed models. *Appl. Math. Lett.* **57**, 13–19 (2016)
34. D.Q. Wang, Least squares-based recursive and iterative estimation for output error moving average systems using data filtering. *IET Control Theory Appl.* **5**(14), 1648–1657 (2011)
35. D.W. Wang, F. Ding, Parameter estimation algorithms for multivariable Hammerstein CARMA systems. *Inf. Sci.* **355**, 237–248 (2016)
36. Y.J. Wang, F. Ding, Novel data filtering based parameter identification for multiple-input multiple-output systems using the auxiliary model. *Automatica* (2016). doi:[10.1016/j.automatica.2016.05.024](https://doi.org/10.1016/j.automatica.2016.05.024)
37. Y.J. Wang, F. Ding, The filtering based iterative identification for multivariable systems. *IET Control Theory Appl.* **10**(8), 894–902 (2016)
38. Y.J. Wang, F. Ding, The auxiliary model based hierarchical gradient algorithms and convergence analysis using the filtering technique. *Signal Process.* **128**, 212–221 (2016)
39. Y.J. Wang, F. Ding, Recursive least squares algorithm and gradient algorithm for Hammerstein–Wiener systems using the data filtering. *Nonlinear Dyn.* **84**(2), 1045–1053 (2016)
40. X.H. Wang, F. Ding, Modelling and multi-innovation parameter identification for Hammerstein nonlinear state space systems using the filtering technique. *Math. Comput. Model. Dyn. Syst.* **22**(2), 113–140 (2016)
41. X.H. Wang, F. Ding, Recursive parameter and state estimation for an input nonlinear state space system using the hierarchical identification principle. *Signal Process.* **117**, 208–218 (2015)
42. C. Wang, T. Tang, Recursive least squares estimation algorithm applied to a class of linear-in-parameters output error moving average systems. *Appl. Math. Lett.* **29**, 36–41 (2014)
43. D.Q. Wang, W. Zhang, Improved least squares identification algorithm for multivariable Hammerstein systems. *J. Franklin Inst. Eng. Appl. Math.* **352**(11), 5292–5370 (2015)
44. W.G. Zhang, Decomposition based least squares iterative estimation for output error moving average systems. *Eng. Comput.* **31**(4), 709–725 (2014)