



Multistage parameter estimation algorithms for identification of bilinear systems

Fatemeh Shahriari · Mohammad Mehdi Arefi  · Hao Luo · Shen Yin

Received: 26 January 2022 / Accepted: 25 July 2022 / Published online: 11 August 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract In this paper, two methods for parameter estimation of bilinear state-space systems with colored noise, which are expressed by ARMA model, are proposed. Using the hierarchical identification principle and gradient method, to reduce the computational cost, both the four-stage recursive least squares algorithm and the four-stage stochastic gradient algorithm are exploited by which parameter estimation error is reduced and the speed of convergence of parameters is increased. In addition, a bilinear state observer for state estimation is designed to make use of the estimated states in the four-stage recursive least squares and the four-stage stochastic gradient algorithms. Finally, a numerical example and a practical example are provided to indicate the superiority of the

proposed methods. The results show that due to the data length increase, the estimation error of the parameters is reduced. Furthermore, the estimated parameters converge to the actual values in a short time.

Keywords Bilinear systems · Parameter estimation · Gradient search · Hierarchical identification

1 Introduction

System identification has been the focus of many researchers in recent decades in various areas, for modeling linear [1, 2] and nonlinear systems [3–5]. The main goal of system identification is to utilize input/output data to obtain an appropriate model, in a way that the behavior of the identified model is as close as possible to the behavior of the actual system, as per a predefined criterion. The mathematical model thus created can be used for analysis or controller design [6–9].

Mathematical modeling is possible in several ways: analytical modeling (white-box), experimental modeling (black-box) and hybrid modeling (gray-box). In the white-box case, considering the nature of the components of the system and the physical laws governing them, mathematical models are formed between the input and the output of the system. In

F. Shahriari · M. M. Arefi (✉)
Department of Power and Control Engineering, School of
Electrical and Computer Engineering, Shiraz University,
Shiraz, Iran
e-mail: arefi@shirazu.ac.ir

H. Luo
Department of Control Science and Engineering, School
of Astronautics, Harbin Institute of Technology,
Harbin 150001, China
e-mail: hao.luo@hit.edu.cn

S. Yin
Department of Mechanical and Industrial Engineering,
Faculty of Engineering, Norwegian University of Science
and Technology, 7033 Trondheim, Norway
e-mail: shen.yin@ntnu.no

black-box approach, there is no information about the internal components of the system, and only by using the input/output data, a mathematical relationship is obtained between the input and the output. In the gray-box approach, the physical components of the system are distinctive; however, their values are unknown, and by using the input/output data, the unknown parameters can be determined. Several system identification methods have been proposed for linear and nonlinear systems [10, 11]. Since most industrial processes are complex and nonlinear, nonlinear systems identification has attracted a lot of attention in the past two decades. However, nonlinear system identification is much more difficult than linear systems identification [12]. In addition, due to the complexity of the systems, it is difficult to obtain physical models since the exact system model is not always available. Therefore, data-driven methods that are dependent on historical information of the system are recently developed to identify the system behavior. These methods do not require complex mathematical tools and are very useful in practice [13, 14].

Bilinear systems are a class of nonlinear systems that are widely studied and used due to their simplicity, and they can be considered as a suitable model for many physical systems [15–17]. In recent years, bilinear systems have drawn a significant amount of attention due to their intrinsic simplicity and wide applications [18]. The importance of such systems is due to their wide range of applications, not only in engineering but also in biology, economics and chemistry [19]. Such a system can explain many physical phenomena and it has been used in many fields, such as air conditioning control [20], immune system, heart regulator, control of carbon dioxide in the lungs and blood pressure [21–23]. There are a couple of methods that have been proposed for bilinear system identification, such as least-squares methods that are based on minimizing the error, i.e., in other words, the summation of square errors [24, 25], gradient methods [26, 27], maximum likelihood methods [28], iterative methods [29, 30], recursive methods [31–33] and error prediction methods [34]. In addition, interaction matrices approach is utilized for identification and observer design of bilinear systems [35, 36]. In [35], optimal bilinear observers are designed for bilinear state-space models, and a new method for identification of bilinear system is introduced in [36]. As can be seen in this paper, by using

the interaction matrix formulation, the bilinear system state is expressed in terms of input/output measurement. In the iterative methods, to improve the estimation of the parameters, the algorithm uses all the data in every iteration and the results of the previous iteration. In these methods, the algorithm is applied once the input–output dataset is collected. In the recursive algorithms, the algorithm uses only the current data to improve the previous estimation. In [37], recursive extended least squares and maximum likelihood methods have been used to identify the bilinear system parameters. Gibson et al. [28] have provided maximum likelihood parameter estimation algorithm for bilinear system identification. Also, Kalman filtering algorithm for parameter estimation of bilinear systems has been proposed in [38]. In [39], the state-space model of bilinear systems has been converted to a transfer function by removing state variables and the recursive least squares algorithm and multi-innovation theory are used to increase the parameter estimation accuracy. In a multi-innovation identification algorithm, increasing the innovation length can increase parameter estimation accuracy and reduce the algorithm's sensitivity to noise.

Li et al. [25] have presented a least-squares iterative algorithm to identify bilinear system parameters using the maximum likelihood. The maximum likelihood iterative least-squares algorithm can provide a more accurate estimate of bilinear systems than the iterative least-squares algorithm. In [40], an iterative algorithm based on the hierarchical principle is proposed to alleviate the complexity of computational load. The proposed algorithm can improve the accuracy of parameter estimation and reduce the computational load. In [41], to achieve a better accuracy, an algorithm using the Kalman filter and multi-innovation theory has been proposed. Both algorithms work well, and the Kalman filter-based multi-innovation recursive extended least squares algorithm has a higher parameter estimation accuracy than the Kalman filter-based recursive extended least squares algorithm. Using the principle of hierarchical identification [40] and data filtering method, a gradient iterative algorithm and a filtering-based gradient iterative algorithm have been presented in [26]. The proposed algorithms can provide very accurate estimates of bilinear systems. Two-step gradient-based iterative algorithm has less computational cost than gradient-based iterative algorithm and its convergence speed is

faster than other two methods in this paper. The algorithms presented in [26] provide a better parameter estimation accuracy than the methods presented in [27]. In [42], a state filtering-based hierarchical identification algorithm has been proposed. In [43], multi-innovation-based stochastic gradient algorithm is presented by using the decomposition method. Decomposition-based multi-innovation stochastic gradient algorithm has higher accuracy than the decomposition-based stochastic gradient algorithm. Also, pursuant to hierarchical principle and data filtering, a least squares iterative algorithm is proposed for identification of bilinear systems [44]. Ding et al. [45] have provided stochastic gradient algorithm and gradient iterative algorithm to estimate the parameters of bilinear systems using an auxiliary model. Auxiliary model-based gradient iterative algorithm uses all the input–output data measured in each iteration which is more suitable than the auxiliary stochastic gradient algorithm. The auxiliary model-based gradient iterative algorithm is effective for bilinear systems in a white noise environment. With the help of subspace identification method, a data-driven design approach has been presented in [46]. In [47], the stochastic gradient algorithm using hierarchical identification principle and multi-innovation idea is developed. In addition, in [48], considering the data filtering method, an extended stochastic gradient algorithm has been proposed.

Although the recursive least-squares method is the most common estimation method among the various previously published methods and has a high convergence rate, it suffers from some drawbacks such as high computational burden. Therefore, in order to overcome this problem, other identification methods such as the principle of hierarchical identification are proposed that divide the main system into multiple subsystems with smaller dimensions to estimate unknown parameters. For example, in [25, 48], the computational load for identifying the system is high, and the hierarchical identification principle has been utilized to obtain an effective method for parameter estimation with higher computational efficiency. It has been shown that the suggested algorithms have less parameter estimation error compared to other presented algorithms.

Motivated by the above-stated concerns, a four-stage hierarchical identification algorithm is used in this paper to identify a bilinear system based on state-

space equations. In this regard, a four-stage recursive least squares algorithm and a four-stage stochastic gradient algorithm are proposed for the identification of bilinear systems. To improve the computational efficiency using the hierarchical identification principle, the identification model is decomposed into four subsystems and the information vector is separately decomposed into four subvectors with smaller dimensions. In addition, an ARMA colored noise model is used in the presented model. Since only input/output data of the system is available in these algorithms, a state observer is used to estimate the system states, and then, the estimated states are used in the identification algorithm. Finally, the proposed algorithm presented for bilinear system identification is simulated and the convergence of the identified parameters is reported. The main contributions of this paper are listed as follows:

- A four-stage recursive least squares algorithm and a four-stage stochastic gradient are proposed using the hierarchical identification principle to reduce the computational efficiency. The principle of hierarchical identification divides the main system into several subsystems with small dimensions. Also, the information vector is broken down into several information subdivisions.
- A bilinear state observer is presented based on the Kalman filter algorithm for bilinear state-space estimation.
- To show the high efficiency of the four-stage recursive least squares algorithm, a comparison of the computational efficiency between two recursive least squares and four-stage recursive least squares algorithm is provided.

The rest of this paper is organized as follows: In Sect. 2, the preliminary definitions, problem statement and the bilinear state-space system is presented. In Sect. 3, a four-stage recursive least squares algorithm is described. Section 4 shows the computational efficiency of the 4S-RLS algorithm. Section 5 presents a four-stage stochastic gradient algorithm. A numerical example and a practical example are presented in Sect. 6 to show the effectiveness of the proposed algorithm. Finally, the paper is ended by Sect. 7 with some concluding points.

2 Problem statement

In this section, first, a number of notations are explained. Superscript T represents the matrix transpose. $\hat{\rho}(t)$ is the estimation of the parameter ρ in time. $I(I_n)$ represents identity matrix $n \times n$, and q is the unit shift operator as

$$qz(t) = z(t + 1), \quad q^{-1}z(t) = z(t - 1)$$

Figure 1 shows the state-space representation of bilinear systems. According to this figure, the bilinear system state-space model is defined as follows:

$$z(t + 1) = Az(t) + Bz(t)\bar{u}(t) + f\bar{u}(t) \tag{1}$$

$$\bar{y}(t) = hz(t) + \omega(t) \tag{2}$$

where $z(t) = [z_1(t), z_2(t), \dots, z_n(t)]^T \in \mathbb{R}^n$ is the state vector, $\bar{u}(t)$ is the system input, $\bar{y}(t)$ is the system output, $\omega(t) = \frac{D(q)}{C(q)}v(t)$ is a colored noise and $v(t) \in \mathbb{R}$ is a zero mean white noise. $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times n}$, $f \in \mathbb{R}^n$ and $h \in \mathbb{R}^{1 \times n}$ are the system matrices and vectors with an appropriate dimension as follows:

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

$$B = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} \in \mathbb{R}^{n \times n} \quad \mathbf{b}_i \in \mathbb{R}^{1 \times n}$$

$$f = [f_1, f_2, \dots, f_n]^T \in \mathbb{R}^n, \quad h = [1, 0, \dots, 0] \in \mathbb{R}^{1 \times n}$$

Using the shift operator, the polynomials $D(q)$ and $C(q)$ are defined as

$$D(q) = 1 + d_1q^{-1} + d_2q^{-2} + \dots + d_pq^{-p}$$

$$C(q) = 1 + c_1q^{-1} + c_2q^{-2} + \dots + c_mq^{-m}$$

According to the method presented in [48] and also using Eqs. (1) and (2), it can be concluded that

$$z_1(t) = - \sum_{i=1}^n a_i z_1(t - i) + \sum_{i=1}^n \mathbf{b}_i z(t - i) \bar{u}(t - i) + \sum_{i=1}^n f_i \bar{u}(t - i) \tag{3}$$

The parameter vector ρ is defined as:

$$\rho = [\rho_s, \rho_n]^T \in \mathbb{R}^{n^2 + 2n + m + p}$$

where

$$\rho_s = [\rho_a^T, \rho_b^T, \rho_f^T]^T \in \mathbb{R}^{n^2 + 2n}$$

$$\rho_n = [c_1, c_2, \dots, c_m, d_1, d_2, \dots, d_p]^T \in \mathbb{R}^{m+p}$$

and

$$\rho_a = [a_1, a_2, \dots, a_n]^T \in \mathbb{R}^n$$

$$\rho_b = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n]^T \in \mathbb{R}^{n^2}$$

$$\rho_f = [f_1, f_2, \dots, f_n]^T \in \mathbb{R}^n$$

In addition, the information vector $\varphi(t)$ is defined as follows:

$$\varphi(t) = [\varphi_z^T(t), \varphi_{z\bar{u}}^T(t), \varphi_{\bar{u}}^T(t), \varphi_n^T(t)]^T \in \mathbb{R}^{n_0},$$

$$n_0 := n^2 + 2n + m + p$$

$$\varphi_s(t) = [\varphi_z^T(t), \varphi_{z\bar{u}}^T(t), \varphi_{\bar{u}}^T(t)]^T \in \mathbb{R}^{n_1},$$

$$n_1 := n^2 + 2n$$

$$\varphi_z(t) = [-z_1(t - 1), -z_1(t - 1), \dots, -z_1(t - n)]^T \in \mathbb{R}^{n^2},$$

$$n_2 := n$$

$$\varphi_{z\bar{u}} = [z^T(t - 1)\bar{u}(t - 1), z^T(t - 2)\bar{u}(t - 2), \dots, z^T(t - n)\bar{u}(t - n)]^T \in \mathbb{R}^{n^2},$$

$$n_3 := n^2$$

$$\varphi_{\bar{u}}(t) = [\bar{u}(t - 1), \bar{u}(t - 2), \dots, \bar{u}(t - n)]^T \in \mathbb{R}^{n^2},$$

$$n_4 := n$$

From (2), the colored noise equation can be written as

$$\omega(t) = [1 - C(z)]\omega(t) + D(z)v(t)$$

$$= -c_1\omega(t - 1) - c_2\omega(t - 2) - \dots - c_m\omega(t - m)$$

$$+ v(t) + d_1v(t - 1) + d_2v(t - 2) + \dots$$

$$+ d_pv(t - p) = \varphi_n^T(t)\rho_n + v(t) \tag{4}$$

where the information vector $\varphi_n(t)$ is defined as follows:

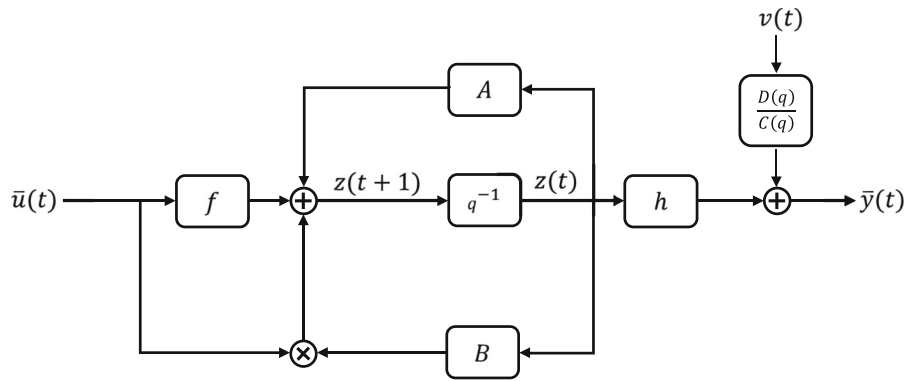


Fig.1 Bilinear state-space system

$$\varphi_n(t) = [-\omega(t-1), -\omega(t-2), \dots, -\omega(t-m), v(t-1), v(t-2), \dots, v(t-p)]^T \in \mathbb{R}^{n_4}, \quad n_4 := m+p$$

Substituting (3) in (2) and according to the definition of information vectors, the bilinear system identification model in (1) and (2) can be expressed as:

$$\bar{y}(t) = \varphi_z^T(t)\rho_a + \varphi_{z\bar{u}}^T(t)\rho_b + \varphi_u^T(t)\rho_f + \varphi_n^T(t)\rho_n + v(t) = \varphi^T(t)\rho + v(t) \tag{5}$$

2.1 Input–output model of bilinear system

The input–output relationship of a bilinear state-space system is obtained for the identification purpose by eliminating the state variables. By removing the state vector in Eqs. (1) and (2), the input–output relationship of the bilinear system is expressed as follows:

$$[A(q) + u(t-n)B(q)]y(t) = [C(q) + u(t-1)D(q)]u(t) + v(t) \tag{6}$$

$$\begin{aligned} A(q) &= 1 + a_1q^{-1} + a_2q^{-2} + \dots + a_{n_a}q^{-n_a} \\ B(q) &= b_1q^{-1} + b_2q^{-2} + \dots + b_{n_b}q^{-n_b} \\ C(q) &= c_1q^{-1} + c_2q^{-2} + \dots + c_mq^{-m} \\ D(q) &= d_2q^{-2} + d_3q^{-3} + \dots + c_wq^{-w} \end{aligned}$$

The following steps have been carried out to obtain (6). Therefore, using (1), one can write

$$\begin{cases} z_1(t+1) = z_2(t) + f_1u(t) \\ z_2(t+1) = z_3(t) + f_2u(t) \\ \vdots \\ z_{n-1}(t+1) = z_n(t) + f_{n-1}u(t) \\ z_n(t+1) = -a_nz_1(t) - a_{n-1}z_2(t) \\ \quad - a_{n-2}z_3(t) - \dots - a_1z_n(t) \\ \quad - [b_nz_1(t) + b_{n-1}z_2(t) + b_{n-2}z_3(t) + \dots \\ \quad + b_1z_n(t)]u(t) + f_nu(t). \end{cases} \tag{7}$$

Then, by using (7), the following equations are obtained directly:

$$\begin{cases} z_2(t) = z_1(t+1) - f_1u(t) \\ z_3(t) = z_2(t+1) - f_2u(t) \\ \quad = z_1(t+2) - f_1u(t+1) - f_2u(t) \\ z_4(t) = z_3(t+1) - f_3u(t) \\ \quad = z_1(t+3) - f_1u(t+2) - f_2u(t+1) - f_3u(t) \\ \vdots \\ z_n(t) = z_{n-1}(t+1) - f_{n-1}u(t) \\ \quad = z_1(t+n-1) - f_1u(t+n-2) - f_2u(t+n-3) \\ \quad - \dots - f_{n-1}u(t) \end{cases} \tag{8}$$

Multiplying both sides of last equation of (8) by operator q , we have

$$z_n(t + 1) = z_1(t + n) - f_1u(t + n - 1) - f_2u(t + n - 2) - \dots - f_{n-1}u(t + 1) \tag{9}$$

Replacing (9) in the last equation of (7) yields

$$- [a_n, a_{n-1}, a_{n-2}, \dots, a_1] \begin{bmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \\ \vdots \\ z_n(t) \end{bmatrix} - [b_n, b_{n-1}, b_{n-2}, \dots, b_1] \begin{bmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \\ \vdots \\ z_n(t) \end{bmatrix} u(t) + f_nu(t) = z_1(t + n) - f_1u(t + n - 1) - f_2u(t + n - 2) - \dots - f_{n-1}u(t + 1) \tag{10}$$

Now, according to matrix representation of (8) and (10), one can write

$$(1 + a_1q^{-1} + a_2q^{-2} + \dots + a_nq^{-n})q^n z_1(t) + [b_1q^{-1} + b_2q^{-2} + \dots + b_nq^{-n}]q^n z_1(t)u(t) = [f_n + a_{n-1}f_1 + a_{n-2}f_2 + \dots + a_1f_{n-1}, f_{n-1} + a_{n-2}f_1 + a_{n-3}f_2 + \dots + a_1f_{n-2}, \dots, f_2 + a_1f_1, f_1] \begin{bmatrix} u(t) \\ u(t+1) \\ \vdots \\ u(t+n-1) \end{bmatrix} + [b_{n-3}f_2 + \dots + b_1f_{n-2}, \dots, b_1f_1, 0] \left. \begin{bmatrix} u(t) \\ u(t+1) \\ \vdots \\ u(t+n-1) \end{bmatrix} \right\} u(t) \tag{11}$$

In order to simplify (11), we define two vectors as follows:

$$[c_n, \dots, c_2, c_1] := [f_n + a_{n-1}f_1 + a_{n-2}f_2 + \dots + a_1f_{n-1}, f_2 + a_1f_1, f_1] \in \mathbb{R}^{1 \times n} \tag{12}$$

$$[d_n, \dots, d_3, d_2] := [b_{n-1}f_1 + b_{n-2}f_2 + \dots + b_1f_{n-1}, b_1f_1] \in \mathbb{R}^{1 \times (n-1)} \tag{13}$$

Then, (11) can be written as

$$A(q)q^n z_1(t) + [B(q)q^n z_1(t)]u(t) = C(q)q^n u(t) + [D(q)]q^n u(t)u(t) \tag{14}$$

where

$$A(q) = 1 + a_1q^{-1} + a_2q^{-2} + \dots + a_nq^{-n}$$

$$B(q) = b_1q^{-1} + b_2q^{-2} + \dots + b_nq^{-n}$$

$$C(q) = c_1q^{-1} + c_2q^{-2} + \dots + c_nq^{-n}$$

$$D(q) = d_2q^{-2} + d_3q^{-3} + \dots + d_nq^{-n}$$

Then, Eq. (14) can be rewritten as

$$A(q)q^n z_1(t) + u(t)[B(q)q^n z_1(t)] = C(q)z^n u(t) + u(t)[D(q)q^n u(t)] \tag{15}$$

By substituting t with $t - n$, we have

$$z_1(t) = \frac{C(q) + u(t-n)D(q)}{A(q) + u(t-n)B(q)} u(t)$$

Replacing $z_1(t)$ in relation (2), the input–output relation of the bilinear state-space system in (1) and (2) is obtained as follows:

$$y(t) = \frac{C(q) + u(t-n)D(q)}{A(q) + u(t-n)B(q)} u(t) + v(t)$$

3 Four-stage recursive least squares algorithm

In this section, a four-stage recursive least squares algorithm is proposed to alleviate computational load, increase the convergence rate of the parameters to actual values and reduce the error simultaneously. According to the hierarchical principle, the main system is broken down into four subsystems; then, an algorithm is presented to estimate the unknown parameters of the bilinear system. Consider the following performance index:

$$J(\rho) = \sum_{j=1}^t [\bar{y}(j) - \varphi^T(j)\rho]^2$$

Using the least-squares principle and minimizing the performance index, the recursive least squares algorithm can be written as

$$\hat{\rho}(t) = \hat{\rho}(t - 1) + K(t) [\bar{y}(t) - \varphi^T(t)\hat{\rho}(t - 1)] \tag{16}$$

$$K(t) = R(t - 1)\varphi(t)[1 + \varphi^T(t)R(t - 1)\varphi(t)]^{-1} \tag{17}$$

$$R(t) = [I - K(t)\varphi^T(t)]R(t - 1) \tag{18}$$

where $R(t)$ is the covariance matrix and $K(t) = R(t)\varphi(t)$ is the gain vector. The first problem of identification is that only the input and output data are available. Since $\varphi(t)$ contains unknown state variables, and $\varphi_n(t)$ consists of the noise variable $(\omega(t - i), i = 1, 2, \dots, m)$, it is not possible to estimate the parameter $\hat{\rho}(t)$ with Eqs. (16)–(18). Therefore, a bilinear state observer must be designed for state estimation.

3.1 Bilinear state observer algorithm

As we know, Kalman filter algorithm is commonly used to estimate the states of linear systems. Here, in order to apply the Kalman filter for bilinear systems, Eqs. (1) and (2) should be written in the following form:

$$\begin{aligned} z(t + 1) &= A_1(t)z(t) + f\bar{u}(t) \\ \bar{y}(t) &= hz(t) + \omega(t) \\ A_1(t) &= A + B\bar{u}(t) \end{aligned}$$

which may be considered as a Linear state-space model. Therefore, the Kalman filter can be used to design a bilinear state observer [48].

$$\hat{z}(t + 1) = A\hat{z}(t) + B\hat{z}(t)\bar{u}(t) + f\bar{u}(t) + K_z(t)[\bar{y}(t) - h\hat{z}(t) - \varphi_n^T(t)\rho_n] \tag{19}$$

$$\begin{aligned} K_z(t) &= AR_z(t)h^T[hR_z(t)h^T + R_v]^{-1}B\bar{u}(t)R_z(t) \\ &\quad \times h^T[hR_z(t)h^T + R_v]^{-1} \end{aligned} \tag{20}$$

$$\begin{aligned} R_z(t + 1) &= [A - K_z(t)h + B\bar{u}(t)]R_z(t) \\ &\quad [A^T - h^TK_z^T(t) + B^T\bar{u}(t)] \\ &\quad + K_z(t)R_vK_z^T(t) \end{aligned} \tag{21}$$

where $\hat{R}_v(t) = \frac{1}{t} \sum_{j=1}^t [\bar{y}(j) - h\hat{z}(j)]^2$, $K_z(t)$ is the optimal vector of state observer and $R_z(t + 1)$ is the covariance matrix state estimation error.

If the vectors and matrices A, B, f and n are unknown, the bilinear state observer in (19)–(21)

cannot be used. Therefore, $\hat{z}(t)$ should be estimated by considering the estimated parameters.

Therefore, the parameter estimation vectors are defined as follows:

$$\begin{aligned} \hat{\rho} &= [\hat{\rho}_s, \hat{\rho}_n]^T \in \mathbb{R}^{n_0} \\ \rho_s &= [\hat{\rho}_a^T, \hat{\rho}_b^T, \hat{\rho}_f^T]^T \in \mathbb{R}^{n_1} \\ \hat{\rho}_n &= [\hat{c}_1, \hat{c}_2, \dots, \hat{c}_m, \hat{d}_1, \hat{d}_2, \dots, \hat{d}_p]^T \in \mathbb{R}^{n_4} \\ \hat{\rho}_a &= [\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n]^T \\ \hat{\rho}_b &= [\hat{b}_1, \hat{b}_2, \dots, \hat{b}_n]^T \in \mathbb{R}^{n_3} \\ \hat{\rho}_f &= [\hat{f}_1, \hat{f}_2, \dots, \hat{f}_n]^T \in \mathbb{R}^{n_2} \end{aligned}$$

$$\hat{A}(t) = \begin{bmatrix} -\hat{a}_1(t) & 1 & 0 & \cdots & 0 \\ -\hat{a}_2(t) & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\hat{a}_{n-1}(t) & 0 & \cdots & 0 & 1 \\ -\hat{a}_n(t) & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{22}$$

$$\hat{B}(t) := \begin{bmatrix} \hat{b}_1(t) \\ \hat{b}_2(t) \\ \vdots \\ \hat{b}_n(t) \end{bmatrix} \quad b_i \in \mathbb{R}^{1 \times n}, \hat{f}(t) := \begin{bmatrix} \hat{f}_1(t) \\ \hat{f}_2(t) \\ \vdots \\ \hat{f}_n(t) \end{bmatrix} \tag{23}$$

By substituting A, B and f in (19)–(21) with the estimated matrices and vectors $\hat{A}(t), \hat{B}(t)$ and $\hat{f}(t)$, we have

$$\hat{z}(t + 1) = \hat{A}\hat{z}(t) + \hat{B}\hat{z}(t)\bar{u}(t) + \hat{\rho}_f\bar{u}(t) + K_z(t)[\bar{y}(t) - h\hat{z}(t) - \hat{\varphi}_n^T(t)\hat{\rho}_n] \tag{24}$$

$$\begin{aligned} K_z(t) &= \hat{A}R_z(t)h^T[hR_z(t)h^T + R_v]^{-1}\hat{B}\bar{u}(t)R_z(t) \\ &\quad \times h^T[hR_z(t)h^T + R_v]^{-1} \end{aligned} \tag{25}$$

$$\begin{aligned} R_z(t + 1) &= [\hat{A} - K_z(t)h + \hat{B}\bar{u}(t)]R_z(t) \\ &\quad [\hat{A}^T - h^TK_z^T(t) + \hat{B}^T\bar{u}(t)] \\ &\quad + K_z(t)R_vK_z^T(t) \end{aligned} \tag{26}$$

Thus, based on the bilinear state observer the estimates $\hat{z}(t)$ of the unknown states $z(t)$ can be calculated.

Replacing an unknown state $z(t - i)$ with estimated $\widehat{z}(t - i)$ and an unknown noise term $\omega(t - i)$ with estimated $\widehat{\omega}(t - i)$, the information estimation vectors are defined as:

$$\widehat{\varphi}(t) = \left[\widehat{\varphi}_z^T(t), \widehat{\varphi}_{z\bar{u}}^T(t), \varphi_{\bar{u}}^T(t), \widehat{\varphi}_n^T(t) \right]^T \in \mathbb{R}^{m_0}$$

$$\widehat{\varphi}_s(t) = \left[\widehat{\varphi}_z^T(t), \widehat{\varphi}_{z\bar{u}}^T(t), \varphi_{\bar{u}}^T(t) \right]^T \in \mathbb{R}^{n_1}$$

$$\widehat{\varphi}_z(t) = [-\widehat{z}_1(t - 1), -\widehat{z}_1(t - 2), \dots, -\widehat{z}_1(t - n)]^T \in \mathbb{R}^{n_2}$$

$$\widehat{\varphi}_{z\bar{u}} = [\widehat{z}^T(t - 1)\bar{u}(t - 1), \widehat{z}^T(t - 2)\bar{u}(t - 2), \dots, \widehat{z}^T(t - n)\bar{u}(t - n)]^T \in \mathbb{R}^{n_3}$$

$$\widehat{\varphi}_n(t) = [-\widehat{\omega}(t - 1), -\widehat{\omega}(t - 2), \dots, -\widehat{\omega}(t - m), \widehat{v}(t - 1), \widehat{v}(t - 2), \dots, \widehat{v}(t - \rho)]^T \in \mathbb{R}^{n_4}$$

It should be noted that the estimations of $\omega(t)$ and $v(t)$ are defined as:

$$\widehat{\omega}(t) = \bar{y}(t) - \widehat{\varphi}_s(t)\widehat{\rho}_s(t - 1) \tag{27}$$

$$\widehat{v}(t) = \bar{y}(t) - \widehat{\varphi}^T(t)\widehat{\rho}(t - 1) \tag{28}$$

The identification models 4 and 5 can be written in four subsystems as follows:

$$\bar{y}_a(t) = \varphi_z^T(t)\rho_a + v(t)$$

$$\bar{y}_b(t) = \varphi_{z\bar{u}}^T(t)\rho_b + v(t)$$

$$\bar{y}_{\bar{u}}(t) = \varphi_{\bar{u}}^T(t)\rho_f + v(t)$$

$$\omega(t) = \varphi_n^T(t)\rho_n + v(t)$$

According to the least squares principle, by defining the criterion function, the recursive relationships are as follows:

$$\begin{aligned} \widehat{\rho}_a(t) &= \widehat{\rho}_a(t - 1) + K_1(t) [\bar{y}_z(t) - \varphi_z^T(t)\widehat{\rho}_a(t - 1)] \\ &= \widehat{\rho}_a(t - 1) + K_1(t) [\bar{y}(t) - \varphi_{z\bar{u}}^T(t)\rho_b \\ &\quad - \varphi_{\bar{u}}^T(t)\rho_f - \varphi_z^T(t)\widehat{\rho}_a(t - 1)] \end{aligned} \tag{29}$$

$$K_1(t) = \frac{R_1(t - 1)\varphi_z(t)}{\beta + \varphi_z^T(t)R_1(t - 1)\varphi_z(t)} \tag{30}$$

$$R_1(t) = \frac{1}{\beta} [I - K_1(t)\varphi_z^T(t)]R_1(t - 1) \tag{31}$$

$$\begin{aligned} \widehat{\rho}_b(t) &= \widehat{\rho}_b(t - 1) + K_2(t) [\bar{y}_{z\bar{u}}(t) - \varphi_{z\bar{u}}^T(t)\widehat{\rho}_b(t - 1)] \\ &= \widehat{\rho}_b(t - 1) + K_2(t) [\bar{y}(t) - \varphi_z^T(t)\rho_a \\ &\quad - \varphi_{\bar{u}}^T(t)\rho_f - \varphi_{z\bar{u}}^T(t)\widehat{\rho}_b(t - 1)] \end{aligned} \tag{32}$$

$$K_2(t) = \frac{R_2(t - 1)\varphi_{z\bar{u}}(t)}{\beta + \varphi_{z\bar{u}}^T(t)R_2(t - 1)\varphi_{z\bar{u}}(t)} \tag{33}$$

$$R_2(t) = \frac{1}{\beta} [I - K_2(t)\varphi_{z\bar{u}}^T(t)]R_2(t - 1) \tag{34}$$

$$\begin{aligned} \widehat{\rho}_f(t) &= \widehat{\rho}_f(t - 1) + K_3(t) [\bar{y}_{\bar{u}}(t) - \varphi_{\bar{u}}^T(t)\widehat{\rho}_f(t - 1)] \\ &= \widehat{\rho}_f(t - 1) + K_3(t) [\bar{y}(t) - \varphi_z^T(t)\rho_a \\ &\quad - \varphi_{z\bar{u}}^T(t)\rho_b - \varphi_{\bar{u}}^T(t)\widehat{\rho}_f(t - 1)] \end{aligned} \tag{35}$$

$$K_3(t) = \frac{R_3(t - 1)\varphi_{\bar{u}}(t)}{\beta + \varphi_{\bar{u}}^T(t)R_3(t - 1)\varphi_{\bar{u}}(t)} \tag{36}$$

$$R_3(t) = \frac{1}{\beta} [I - K_3(t)\varphi_{\bar{u}}^T(t)]R_3(t - 1) \tag{37}$$

$$\widehat{\rho}_n(t) = \widehat{\rho}_n(t - 1) + K_4(t) [\omega(t) - \varphi_n^T(t)\widehat{\rho}_n(t - 1)] \tag{38}$$

$$K_4(t) = \frac{R_4(t - 1)\varphi_n(t)}{\beta + \varphi_n^T(t)R_4(t - 1)\varphi_n(t)} \tag{39}$$

$$R_4(t) = \frac{1}{\beta} [I - K_4(t)\varphi_n^T(t)]R_4(t - 1) \tag{40}$$

The information vectors φ_z , $\varphi_{z\bar{u}}$ and φ_n contain unknown states $z(t)$, and Eqs. (29), (32), (35) and (38) are used to estimate unknown parameters. Therefore, the algorithm (29)–(40) cannot directly estimate the unknown parameters. Consequently, by substituting their estimations, we have the following relations:

$$\begin{aligned} \widehat{\rho}_a(t) &= \widehat{\rho}_a(t - 1) + K_1(t) [\bar{y}(t) - \varphi_{z\bar{u}}^T(t)\widehat{\rho}_b \\ &\quad - \varphi_{\bar{u}}^T(t)\widehat{\rho}_f - \varphi_z^T(t)\widehat{\rho}_a(t - 1)] \end{aligned} \tag{41}$$

$$K_1(t) = \frac{R_1(t-1)\hat{\varphi}_z(t)}{\beta + \hat{\varphi}_z^T(t)R_1(t-1)\hat{\varphi}_z(t)} \tag{42}$$

$$R_1(t) = \frac{1}{\beta} \left[I - K_1(t)\hat{\varphi}_z^T(t) \right] R_1(t-1) \tag{43}$$

$$\hat{\rho}_b(t) = \hat{\rho}_b(t-1) + K_2(t)[\bar{y}(t) - \hat{\varphi}_z^T(t)\hat{\rho}_a - \varphi_u^T(t)\hat{\rho}_f - \hat{\varphi}_{zu}^T(t)\hat{\rho}_b(t-1)] \tag{44}$$

$$K_2(t) = \frac{R_2(t-1)\hat{\varphi}_{zu}(t)}{\beta + \hat{\varphi}_{zu}^T(t)R_2(t-1)\hat{\varphi}_{zu}(t)} \tag{45}$$

$$R_2(t) = \frac{1}{\beta} \left[I - K_2(t)\hat{\varphi}_{zu}^T(t) \right] R_2(t-1) \tag{46}$$

$$\hat{\rho}_f(t) = \hat{\rho}_f(t-1) + K_3(t)[\bar{y}(t) - \hat{\varphi}_z^T(t)\hat{\rho}_a - \hat{\varphi}_{zu}^T(t)\hat{\rho}_b - \varphi_u^T(t)\hat{\rho}_f(t-1)] \tag{47}$$

$$K_3(t) = \frac{R_3(t-1)\varphi_u(t)}{\beta + \varphi_u^T(t)R_3(t-1)\varphi_u(t)} \tag{48}$$

$$R_3(t) = \frac{1}{\beta} \left[I - K_3(t)\varphi_u^T(t) \right] R_3(t-1) \tag{49}$$

$$\hat{\rho}_n(t) = \hat{\rho}_n(t-1) + K_4(t)[\hat{\omega}(t) - \varphi_n^T(t)\hat{\rho}_n(t-1)] \tag{50}$$

$$K_4(t) = \frac{R_4(t-1)\hat{\varphi}_n(t)}{1 + \hat{\varphi}_n^T(t)R_3(t-1)\hat{\varphi}_n(t)} \tag{51}$$

$$R_4(t) = \left[I - K_4(t)\hat{\varphi}_n^T(t) \right] R_4(t-1) \tag{52}$$

$$\hat{\varphi}_z(t) = [-\hat{z}_1(t-1), -\hat{z}_1(t-2), \dots, -\hat{z}_1(t-n)]^T \in \mathbb{R}^n \tag{53}$$

$$\hat{\varphi}_{zu} = [\hat{z}^T(t-1)\bar{u}(t-1), \hat{z}^T(t-2)\bar{u}(t-2), \dots, \hat{z}^T(t-n)\bar{u}(t-n)]^T \in \mathbb{R}^{n^2} \tag{54}$$

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

$$\hat{\varphi}_n(t) = [-\hat{\omega}(t-1), -\hat{\omega}(t-2), \dots, -\hat{\omega}(t-m), \hat{v}(t-1), \hat{v}(t-2), \dots, \hat{v}(t-p)]^T \in \mathbb{R}^{m+p} \tag{56}$$

$$\hat{z}(t+1) = \hat{A}\hat{z}(t) + \hat{B}\hat{z}(t)\bar{u}(t) + \hat{\rho}_f\bar{u}(t) + K_z(t)[\bar{y}(t) - h\hat{z}(t) - \hat{\varphi}_n^T(t)\hat{\rho}_n] \tag{57}$$

$$K_z(t) = \hat{A}R_z(t)h^T [hR_z(t)h^T + R_v]^{-1} \hat{B}\bar{u}(t)R_z(t) \times h^T [hR_z(t)h^T + R_v]^{-1} \tag{58}$$

$$R_z(t+1) = [\hat{A} - K_z(t)h + \hat{B}\bar{u}(t)]R_z(t) [\hat{A}^T - h^TK_z^T(t) + \hat{B}^T\bar{u}(t)] + K_z(t)R_vK_z^T(t) \tag{59}$$

$$\hat{A}(t) = \begin{bmatrix} -\hat{a}_1(t) & 1 & 0 & \dots & 0 \\ -\hat{a}_2(t) & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\hat{a}_{n-1}(t) & 0 & \dots & 0 & 1 \\ -\hat{a}_n(t) & 0 & \dots & 0 & 0 \end{bmatrix} \tag{60}$$

$$\hat{B}(t) := \begin{bmatrix} \hat{b}_1(t) \\ \hat{b}_2(t) \\ \vdots \\ \hat{b}_n(t) \end{bmatrix} \quad \mathbf{b}_i \in \mathbb{R}^{1 \times n}, \hat{f}(t) := \begin{bmatrix} \hat{f}_1(t) \\ \hat{f}_2(t) \\ \vdots \\ \hat{f}_n(t) \end{bmatrix} \tag{61}$$

Equations (41)–(61) consist of four recursive least-squares algorithms for the bilinear system (1) and (2). In summary, the steps of the above algorithm are given in Algorithm 1.

Algorithm 1: Four-stage recursive least squares

- 1: **Initializing**
- 2: Let $t=1, \hat{z}(1) = 1_n/r_0, R_1(0) = r_0 I_n, R_z(1) = I_n, r_0 = 10^8, R_2(0) = r_0 I_{n^2}, R_3(0) = r_0 I_n, R_4(0) = r_0 I_{m+p}, \hat{v}(t-i) = 1/r_0, \hat{\omega}(t-j) = 1/r_0, i = 1, 2, \dots, p, j = 1, 2, \dots, m.$
- 3: **Estimate the parameters ρ**
- 4: Set the initial parameters $\hat{\rho}_a(0) = 1_n/r_0, \hat{\rho}_b(0) = 1_{n^2}/r_0, \hat{\rho}_f(0) = 1_n/r_0, \hat{\rho}_n(0) = 1_{m+p}/r_0$
- 5: **For $t < L$ do**
- 6: Gather input and output data.
- 7: Form $\hat{\varphi}_z(t), \hat{\varphi}_{zu}, \varphi_u(t), \hat{\varphi}_n(t)$ by (53)-(56).
- 8: Calculate gain vector $K_1(t)$ and covariance matrix $R_1(t)$ by (42)-(43).
- 9: Update parameter estimation vector $\hat{\rho}_a(t)$ by (41).
- 10: Calculate gain vector $K_2(t)$ and covariance matrix $R_2(t)$ by (45)-(46).
- 11: Update parameter estimation vector $\hat{\rho}_b(t)$ by (44).
- 12: Calculate $K_3(t)$ and $R_3(t)$ by (48) - (49).
- 13: Update parameter estimation vector $\hat{\rho}_f(t)$ by (47).
- 14: Calculate gain vector $K_4(t)$ and covariance matrix $R_4(t)$ by (51)-(52).
- 15: Update parameter estimation vector $\hat{\rho}_n(t)$ by (50).
- 16: Calculate $\hat{\omega}(t)$ and $\hat{v}(t)$ by (27)-(28).
- 17: **Compute the state estimate $z(t)$**
- 18: Form system matrix $\hat{A}(t), \hat{B}(t)$ and $\hat{f}(t)$ by (60) and (61).
- 19: Calculate the observer gain vector $K_z(t)$ and $R_z(t+1)$ by (58)-(59).
- 20: Update the state estimate $\hat{z}(t+1)$ by (57).
- 21: **if** $\|\hat{\rho}(t) - \hat{\rho}(t-1)\| > \varepsilon$
- 22: $t = t + 1$
- 23: **Else**
- 24: break;
- 25: **End**
- 26: **end for**

3.2 Convergence analysis

Assumption 1 Assume $\{v(t)\}$ is random white noise with zero mean and abounded variance σ^2

$$E[v(t)] = 0 \tag{62}$$

$$E[v^2(t)] = \sigma^2 < \infty \tag{63}$$

Lemma 1 For the four-stage recursive least squares algorithm in (41)–(61), for any $c > 1$, the following inequalities hold:

$$\sum_{t=1}^{\infty} \frac{\hat{\varphi}_z^T(t) R_1(t) \hat{\varphi}_z(t)}{[\ln |R_1^{-1}(t)|]^c} < \infty \tag{64}$$

$$\sum_{t=1}^{\infty} \frac{\hat{\varphi}_{z\bar{u}}^T(t) R_2(t) \hat{\varphi}_{z\bar{u}}(t)}{[\ln |R_2^{-1}(t)|]^c} < \infty \tag{65}$$

$$\sum_{t=1}^{\infty} \frac{\hat{\varphi}_{\bar{u}}^T(t) R_3(t) \hat{\varphi}_{\bar{u}}(t)}{[\ln |R_3^{-1}(t)|]^c} < \infty \tag{66}$$

Table 1 Computational efficiency of the RLS algorithm

Expressions	Multiplications	Additions
$\hat{\rho}(t) = \hat{\rho}(t-1) + K(t)E(t)$	n_0	n_0
$E(t) = [\bar{y}(t) - \varphi^T(t)\hat{\rho}(t-1)]$	n_0	n_0
$K(t) = \xi(t)[\beta + \varphi^T(t)\xi(t)]^{-1}$	$2n_0$	n_0
$\xi(t) = R(t-1)\hat{\varphi}(t)$	n_0^2	$n_0^2 - n_0$
$R(t) = \frac{1}{\beta}R(t-1) - \frac{1}{\beta}K(t)\xi(t)^T$	$3n_0^2$	n_0^2
Total flops	$N_2 := 6n_0^2 + 6n_0$ $n_0 := 2n_2 + n_3 + n_4$	

Table 2 Computational efficiency of the 4S-RLS algorithm

Expressions	Number of multiplications	Number of additions
$\hat{\rho}_a(t) = \hat{\rho}_a(t-1) + K_1(t)E(t)$	n_2	n_2
$E(t) = [\bar{y}(t) - \varphi_{zu}^T(t)\hat{\rho}_b - \varphi_u^T(t)\hat{\rho}_f - \varphi_z^T(t)\hat{\rho}_a(t-1)]$	n_1	n_1
$K_1(t) = \xi_{\rho_a} [\beta + \hat{\varphi}_z^T(t)\xi_{\rho_a}]^{-1}$	$2n_2$	n_2
$\xi_{\rho_a}(t) = R_1(t-1)\hat{\varphi}_z(t)$	n_2^2	$n_2^2 - n_2$
$R_1(t) = \frac{1}{\beta}R_1(t-1) - \frac{1}{\beta}K_1(t)\xi_{\rho_a}^T$	$3n_2^2$	n_2^2
$\hat{\rho}_b(t) = \hat{\rho}_b(t-1) + K_2(t)E_2(t)$	n_3	n_3
$E_2(t) = \bar{y}(t) - \hat{\varphi}_z^T(t)\hat{\rho}_a - \varphi_u^T(t)\hat{\rho}_f - \hat{\varphi}_{zu}^T(t)\hat{\rho}_b(t-1)$	n_1	n_1
$K_2(t) = \xi_{\rho_b} [\beta + \hat{\varphi}_{zu}^T(t)\xi_{\rho_b}]^{-1}$	$2n_3$	n_3
$\xi_{\rho_b}(t) = R_2(t-1)\hat{\varphi}_{zu}(t)$	n_3^2	$n_3^2 - n_3$
$R_2(t) = \frac{1}{\beta}R_2(t-1) - \frac{1}{\beta}K_2(t)\xi_{\rho_b}^T$	$3n_3^2$	n_3^2
$\hat{\rho}_f(t) = \hat{\rho}_f(t-1) + K_3(t)E_3(t)$	n_2	n_2
$E_3(t) = \bar{y}(t) - \hat{\varphi}_z^T(t)\hat{\rho}_a - \hat{\varphi}_{zu}^T(t)\hat{\rho}_b - \varphi_u^T(t)\hat{\rho}_f(t-1)$	n_1	n_1
$K_3(t) = \xi_{\rho_f} [\beta + \hat{\varphi}_u^T(t)\xi_{\rho_f}]^{-1}$	$2n_2$	n_2
$\xi_{\rho_f}(t) = R_3(t-1)\hat{\varphi}_u(t)$	n_2^2	$n_2^2 - n_2$
$R_3(t) = \frac{1}{\beta}R_3(t-1) - \frac{1}{\beta}K_3(t)\xi_{\rho_f}^T$	$3n_2^2$	n_2^2
$\hat{\rho}_n(t) = \hat{\rho}_n(t-1) + K_4(t)E_4(t)$	n_4	n_4
$E_4(t) = \hat{\omega}(t) - \varphi_n^T(t)\hat{\rho}_n(t-1)$	n_1	n_1
$K_4(t) = \xi_{\rho_n} [\beta + \hat{\varphi}_n^T(t)\xi_{\rho_n}]^{-1}$	$2n_4$	n_4
$\xi_{\rho_n}(t) = R_4(t-1)\hat{\varphi}_n(t)$	n_4^2	$n_4^2 - n_4$
$R_4(t) = \frac{1}{\beta}R_4(t-1) - \frac{1}{\beta}K_4(t)\xi_{\rho_n}^T$	$3n_4^2$	n_4^2
Total flops	$N_1 := 6(2n_2^2 + n_3^2 + n_4^2) + 8n_1 + 4n_0$ $n_0 := 2n_2 + n_3 + n_4, n_1 = n_2^2 + 2n_2$	

Table 3 Number of additions and multiplications of the algorithms

Algorithm	Multiplications	Additions
4S-RLS	$8n_2^2 + 4n_3^2 + 4n_4^2 + 4n_1 + 3n_0$	$2(2n_2^2 + n_3^2 + n_4^2) + 4n_1 + n_0$
RLS	$4n_0^2 + 4n_0$	$2n_0^2 + 2n_0$

Table 4 Comparison of total flops of the algorithms with $n_1 = 8, n_2 = 2, n_3 = 4, n_4 = 2$

Algorithm	Number of multiplications	Number of additions	Total flops
4S-RLS	174	98	272
RLS	440	220	660

$$\sum_{t=1}^{\infty} \frac{\widehat{\varphi}_n^T(t)R_4(t)\widehat{\varphi}_n(t)}{[\ln|R_4^{-1}(t)|]^c} < \infty \tag{67}$$

The proof of Lemma 1 can be found in [49].

Theorem 1 For the system in (1)–(2) and the four-stage recursive least squares algorithm in (41)–(61), let

$$M(t) := [\ln|R_1^{-1}(t)|]^c + [\ln|R_2^{-1}(t)|]^c + [\ln|R_3^{-1}(t)|]^c + [\ln|R_4^{-1}(t)|]^c \tag{68}$$

Assume that Conditions (62) and (63) hold, for any $c > 1$, we have

$$\|\widehat{\rho}_a(t) - \rho_a\|^2 = O\left(\frac{M(t)}{\lambda_{\min}[R_1^{-1}(t)]}\right), \tag{69}$$

$$\|\widehat{\rho}_b(t) - \rho_b\|^2 = O\left(\frac{M(t)}{\lambda_{\min}[R_2^{-1}(t)]}\right)$$

$$\|\widehat{\rho}_f(t) - \rho_f\|^2 = O\left(\frac{M(t)}{\lambda_{\min}[R_3^{-1}(t)]}\right), \tag{70}$$

$$\|\widehat{\rho}_n(t) - \rho_n\|^2 = O\left(\frac{M(t)}{\lambda_{\min}[R_4^{-1}(t)]}\right)$$

Proof See the detailed proof in [50]. □

Theorem 2 For the identification model in Eq. (5) and the four-stage recursive least squares algorithm in Eqs. (41)–(61), assume that there are positive constants $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \beta_3, \beta_4$ and large t such that the following persistent excitation conditions hold:

$$\alpha_1 I_{n_2} \leq \frac{1}{t} \sum_{i=1}^t \widehat{\varphi}_z(i)\widehat{\varphi}_z^T(i) \leq \beta_1 I_{n_2} \tag{71}$$

$$\alpha_2 I_{n_3} \leq \frac{1}{t} \sum_{i=1}^t \widehat{\varphi}_{z\bar{u}}(i)\widehat{\varphi}_{z\bar{u}}^T(i) \leq \beta_2 I_{n_3} \tag{72}$$

$$\alpha_3 I_{n_2} \leq \frac{1}{t} \sum_{i=1}^t \widehat{\varphi}_{\bar{u}}(i)\widehat{\varphi}_{\bar{u}}^T(i) \leq \beta_3 I_{n_2} \tag{73}$$

$$\alpha_4 I_{n_4} \leq \frac{1}{t} \sum_{i=1}^t \widehat{\varphi}_n(i)\widehat{\varphi}_n^T(i) \leq \beta_4 I_{n_4} \tag{74}$$

Then, four-stage recursive least squares parameter estimation errors converge to zero as t goes to infinity.

$$\|\widehat{\rho}_a(t) - \rho_a\|^2 \rightarrow 0, \|\widehat{\rho}_b(t) - \rho_b\|^2 \rightarrow 0 \tag{75}$$

$$\|\widehat{\rho}_f(t) - \rho_f\|^2 \rightarrow 0, \|\widehat{\rho}_n(t) - \rho_n\|^2 \rightarrow 0 \tag{76}$$

The proof is expressed in Appendix section.

4 The computational efficiency

Utilizing flops is an useful way to determine computational efficiency [51]. Here, a flop is each operation of addition, multiplication, subtraction or division. In general, a division is presumed as a multiplication and a subtraction is presumed as an addition. Therefore, the algorithm can be represented by additions and multiplications. The number of multiplications and additions of the proposed algorithms are listed in Tables 1 and 2. In order to show the computational efficiency in the 4S-RLS algorithm, an RLS algorithm is presented for the sake of comparison. Tables 3 and 4 show that the computational load of the proposed algorithm is less than the RLS algorithm.

The flop difference between the RLS algorithm and the 4S-RLS algorithm is as follows: $N_2 - N_1 = 6(2n_2 + n_3 + n_4)^2 + 6(2n_2 + n_3 + n_4) - [6(2n_2^2 + n_3^2 + n_4^2) + 8(n_2^2 + 2n_2) + 4(2n_2 + n_3 + n_4)] = 4n_2^2 + 24n_2n_3 + 24n_2n_4 + 12n_3n_4 - 12n_2 + 2n_3 + 2n_4 > 0$. Therefore, $N_1 < N_2$ which means that the 4S-RLS algorithm is more efficient than the RLS algorithm.

5 Four-stage stochastic gradient algorithm

In this part, a four-stage stochastic gradient algorithm is considered to estimate the unknown parameters and reduce the computational burden.

The second-order criterion function is considered as follows:

$$J(\rho) = \frac{1}{2} [\bar{y}(t) - \varphi^T(t)\rho]^2$$

By computing the gradient of J , we have

$$\nabla[J(\rho)] = \frac{\partial(J(\rho))}{\partial\rho} = -\varphi(t) [\bar{y}(t) - \varphi^T(t)\rho]$$

According to the gradient search principle and minimizing the objective function, the stochastic gradient algorithm is obtained as follows:

Assume that $\frac{1}{\mu(t)}$ is the step size.

$$\hat{\rho}(t) = \hat{\rho}(t-1) + \frac{\varphi(t)}{\mu(t)} [\bar{y}(t) - \varphi^T(t)\hat{\rho}(t-1)]$$

$$\mu(t) = \alpha \mu(t-1) + \|\varphi(t)\|^2, \mu(0) = 1$$

where $0 \leq \alpha < 1$ is a forgetting factor that can improve the accuracy of parameter estimation. Hence, the four-stage stochastic gradient algorithm is obtained as follows:

$$\begin{aligned} \hat{\rho}_a(t) &= \hat{\rho}_a(t-1) + \frac{\varphi_z}{\mu_1(t)} [\bar{y}(t) - \varphi_{z\bar{u}}^T(t)\rho_b \\ &\quad - \varphi_{\bar{u}}^T(t)\rho_f - \varphi_z^T(t)\hat{\rho}_a(t-1)] \end{aligned} \tag{77}$$

$$\mu_1(t) = \alpha \mu_1(t-1) + \|\varphi_z(t)\|^2 \tag{78}$$

$$\begin{aligned} \hat{\rho}_b(t) &= \hat{\rho}_b(t-1) + \frac{\hat{\varphi}_{z\bar{u}}}{\mu_2(t)} [\bar{y}(t) - \hat{\varphi}_z^T(t)\hat{\rho}_a \\ &\quad - \varphi_{\bar{u}}^T(t)\hat{\rho}_f - \hat{\varphi}_{z\bar{u}}^T(t)\hat{\rho}_b(t-1)] \end{aligned} \tag{79}$$

$$\mu_2(t) = \alpha \mu_2(t-1) + \|\varphi_{z\bar{u}}(t)\|^2 \tag{80}$$

$$\begin{aligned} \hat{\rho}_f(t) &= \hat{\rho}_f(t-1) + \frac{\varphi_{\bar{u}}}{\mu_3(t)} [y(t) - \hat{\varphi}_z^T(t)\hat{\rho}_a \\ &\quad - \hat{\varphi}_{z\bar{u}}^T(t)\hat{\rho}_b - \varphi_{\bar{u}}^T(t)\hat{\rho}_f(t-1)] \end{aligned} \tag{81}$$

$$\mu_3(t) = \alpha \mu_3(t-1) + \|\varphi_{\bar{u}}(t)\|^2 \tag{82}$$

$$\begin{aligned} \hat{\rho}_n(t) &= \hat{\rho}_n(t-1) + \frac{\varphi_n}{\mu_4(t)} [\hat{\omega}(t) - \varphi_n^T(t)\hat{\rho}_n(t-1)] \end{aligned} \tag{83}$$

$$\mu_4(t) = \alpha \mu_4(t-1) + \|\varphi_n(t)\|^2 \tag{84}$$

$$\hat{\varphi}_z(t) = [-\hat{z}_1(t-1), -\hat{z}_1(t-2), \dots, -\hat{z}_1(t-n)]^T \tag{85}$$

$$\hat{\varphi}_{z\bar{u}} = [\hat{z}^T(t-1)\bar{u}(t-1), \hat{z}^T(t-2)\bar{u}(t-2), \dots, \hat{z}^T(t-n)\bar{u}(t-n)]^T \in \mathbb{R}^{nt} \tag{86}$$

$$\varphi_u(t) = [\bar{u}(t-1), \bar{u}(t-2), \dots, \bar{u}(t-n)]^T \tag{87}$$

$$\hat{\varphi}_n(t) = [-\hat{\omega}(t-1), -\hat{\omega}(t-2), \dots, -\hat{\omega}(t-m), \hat{v}(t-1), \hat{v}(t-2), \dots, \hat{v}(t-p)]^T \in \mathbb{R}^{m+p} \tag{88}$$

$$\begin{aligned} \hat{z}(t+1) &= \hat{A}\hat{z}(t) + \hat{B}\hat{z}(t)\bar{u}(t) + \hat{\rho}_f\bar{u}(t) \\ &\quad + K_z(t) [\bar{y}(t) - h\hat{z}(t) - \hat{\varphi}_n^T(t)\hat{\rho}_n] \end{aligned} \tag{89}$$

$$\begin{aligned} K_z(t) &= \hat{A}R_z(t)h^T [hR_z(t)h^T + R_v]^{-1} \hat{B}\bar{u}(t)R_z(t) \\ &\quad \times h^T [hR_z(t)h^T + R_v]^{-1} \end{aligned} \tag{90}$$

$$\begin{aligned} R_z(t+1) &= [\hat{A} - K_z(t)h + \hat{B}\bar{u}(t)]R_z(t)[\hat{A}^T \\ &\quad - h^TK_z^T(t) + \hat{B}^T\bar{u}(t)] + K_z(t)R_vK_z^T(t) \end{aligned} \tag{91}$$

$$\hat{A}(t) = \begin{bmatrix} -\hat{a}_1(t) & 1 & 0 & \cdots & 0 \\ -\hat{a}_2(t) & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ -\hat{a}_{n-1}(t) & 0 & \cdots & 0 & 1 \\ -\hat{a}_n(t) & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{92}$$

$$\hat{B}(t) := \begin{bmatrix} \hat{b}_1(t) \\ \hat{b}_2(t) \\ \vdots \\ \hat{b}_n(t) \end{bmatrix}, \hat{f}(t) := \begin{bmatrix} \hat{f}_1(t) \\ \hat{f}_2(t) \\ \vdots \\ \hat{f}_n(t) \end{bmatrix} \tag{93}$$

In summary, the steps of the proposed algorithm are presented in *Algorithm 2*.

Algorithm 2: Four-stage stochastic gradient

- 1: **Initializing**
- 2: Let $t=1$, $\hat{z}(1) = 1_n/r_0$, $R_1(0) = r_0 I_n$, $R_z(1) = I_n$, $r_0 = 10^8$, $R_2(0) = r_0 I_{n^2}$, $R_3(0) = r_0 I_n$, $R_4(0) = r_0 I_{m+p}$, $\hat{v}(t-i) = 1/r_0$, $\hat{\omega}(t-j) = 1/r_0$, $i = 1, 2, \dots, p$, $j = 1, 2, \dots, m$.
- 3: **Estimate the parameters ρ**
- 4: Set the initial parameters $\hat{\rho}_a(0) = 1_n/r_0$, $\hat{\rho}_b(0) = 1_{n^2}/r_0$, $\hat{\rho}_f(0) = 1_n/r_0$, $\hat{\rho}_n(0) = 1_{m+p}/r_0$
- 5: **For $t < L$ do**
- 6: Gather input and output data.
- 7: Form $\hat{\varphi}_z(t)$, $\hat{\varphi}_{zu}$, $\varphi_u(t)$, $\hat{\varphi}_n(t)$ by (85)-(88).
- 8: Calculate convergence coefficient $\hat{\mu}_1(t)$ by (78).
- 9: Update the parameter estimation $\hat{\rho}_a(t)$ by (77).
- 10: Calculate convergence coefficient $\hat{\mu}_2(t)$ by (80).
- 11: Update the parameter estimation $\hat{\rho}_b(t)$ by (79).
- 12: Calculate convergence coefficient $\hat{\mu}_3(t)$ by (82).
- 13: Update the parameter estimation $\hat{\rho}_f(t)$ by (81).
- 14: Calculate convergence coefficient $\hat{\mu}_4(t)$ by (84).
- 15: Update the parameter estimation $\hat{\rho}_n(t)$ by (83).
- 16: Calculate $\hat{\omega}(t)$ and $\hat{v}(t)$ by (27) and (28).
- 17: **Compute the state estimate $z(t)$**
- 18: Form system matrix $\hat{A}(t)$, $\hat{B}(t)$ and $\hat{f}(t)$ by (92) and (93).
- 19: Calculate $K_z(t)$ and $R_z(t+1)$ by (90)-(91)
- 20: Update the state estimate $\hat{z}(t+1)$ by (89).
- 21: **if** $\|\hat{\rho}(t) - \hat{\rho}(t-1)\| > \varepsilon$
- 22: $t = t + 1$
- 23: **else**
- 24: break;
- 25: **end**
- 26: **end for**

6 Simulation results

In order to show the efficiency of the proposed method, two numerical examples are provided. The first example is a bilinear state-space system that should be identified, and the second one is a practical example where the system is represented by bilinear state-space model.

6.1 Numerical example

Consider a bilinear state-space system as follows:

$$\begin{aligned} z(t+1) &= \begin{bmatrix} -0.20 & 1 \\ 0.25 & 0 \end{bmatrix} z(t) \\ &+ \begin{bmatrix} 0.08 & 0.17 \\ -0.12 & -0.2 \end{bmatrix} z(t) \bar{u}(t) \\ &+ \begin{bmatrix} 0.4 \\ 2 \end{bmatrix} \bar{u}(t) \\ \bar{y}(t) &= [1, 0] z(t) - c\omega(t-1) + d\nu(t-1) + \nu(t) \end{aligned}$$

The parameter vector for identification is

$$\begin{aligned} \rho &= [a_1, a_2, b_{11}, b_{12}, b_{21}, b_{22}, f_1, f_2, c, d]^T \\ \rho &= [0.20, -0.25, 0.08, 0.17, -0.12, -0.2, 0.40, 2, -0.3, 1] \end{aligned}$$

Fig. 2. 4S-RLS estimation errors

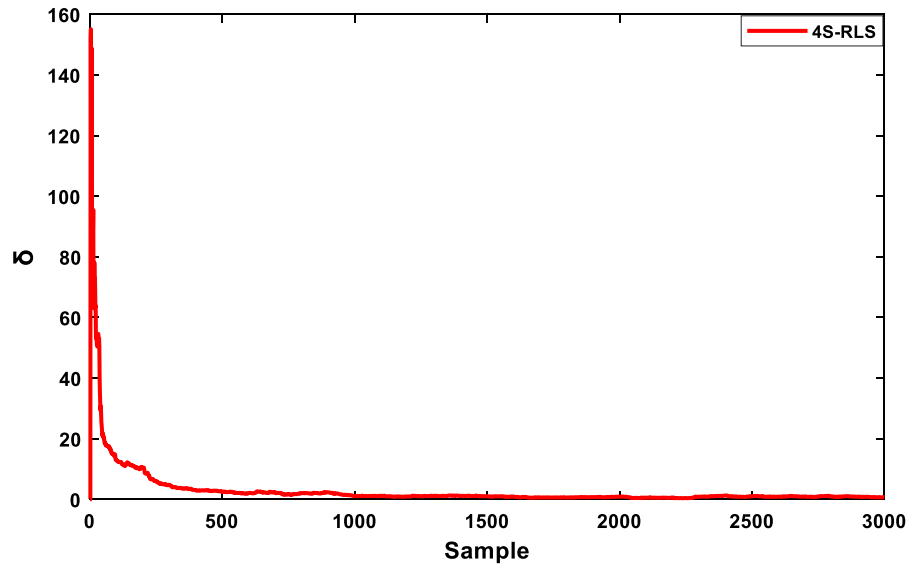


Table 5 4S-RLS estimates and errors ($\sigma^2 = 0.10^2$)

t	a_1	a_2	b_{11}	b_{12}	b_{21}	b_{22}	f_1	f_2	c	d	δ (%)
100	0.2991	- 0.3563	0.2064	0.0787	- 0.0414	0.0663	0.5845	2.2744	- 0.4334	0.8338	23.6133
200	0.2720	- 0.1443	0.1175	0.1519	- 0.0249	- 0.0765	0.4648	2.1403	- 0.3231	1.3230	11.1238
500	0.2402	- 0.1936	0.0927	0.1685	- 0.0629	- 0.1561	0.4066	2.0425	- 0.2711	1.0558	5.4112
1000	0.2151	- 0.2328	0.0822	0.1771	- 0.0916	- 0.1929	0.4054	2.0220	- 0.2575	1.0516	0.4339
2000	0.2033	- 0.2491	0.0739	0.1599	- 0.1192	- 0.2007	0.4018	2.0023	- 0.2984	0.9903	0.6882
3000	0.2012	- 0.2497	0.0808	0.1699	- 0.1151	- 0.2040	0.4046	2.0001	- 0.2944	1.0024	0.4303
True values	0.2000	- 0.2500	0.0800	0.1700	- 0.1200	- 0.2000	0.4000	2.0000	- 0.3000	1.0000	

Fig. 3. 4S-SG estimation errors

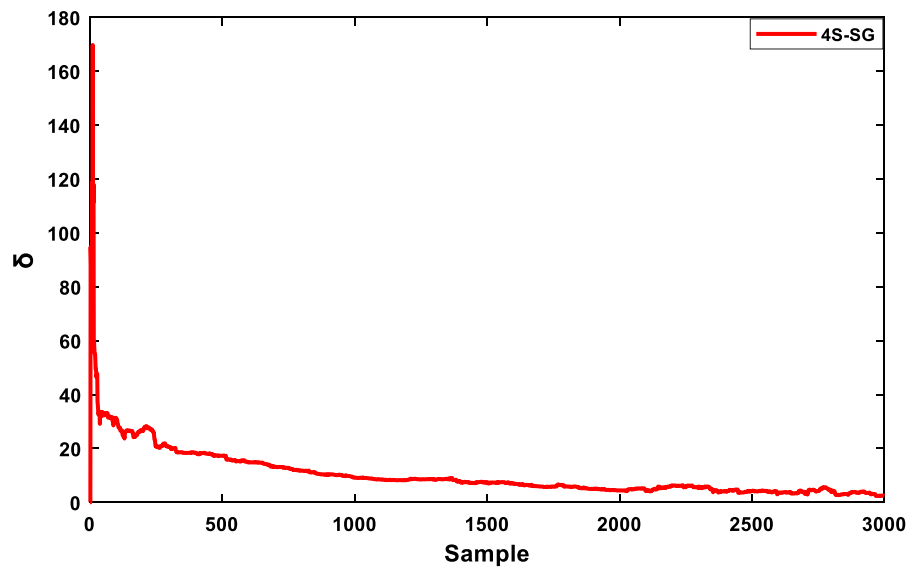


Table 6 4S-SG estimates and errors ($\sigma^2 = 0.10^2$)

t	a_1	a_2	b_{11}	b_{12}	b_{21}	b_{22}	f_1	f_2	c	d	δ (%)
100	0.1173	0.1219	0.1503	0.0993	0.0066	-0.4278	0.5450	1.4749	-0.3178	1.0348	30.9489
200	0.1140	0.1062	0.1461	0.0983	-0.0011	-0.4145	0.5186	1.5957	-0.3144	1.0465	26.5585
500	0.1457	0.0333	0.1315	0.1110	-0.0203	-0.3806	0.3052	1.8643	-0.3249	1.0565	17.3115
1000	0.1889	-0.1082	0.0896	0.1138	-0.0634	-0.3157	0.3337	1.9947	-0.3230	1.0330	9.2041
2000	0.2236	-0.1934	0.0865	0.1418	-0.0943	-0.2608	0.3703	1.9871	-0.2967	0.9744	4.4284
3000	0.2135	-0.2630	0.1006	0.1668	-0.1120	-0.2312	0.4141	1.9806	-0.2902	0.9710	2.4843
True values	0.2000	-0.2500	0.0800	0.1700	-0.1200	-0.2000	0.4000	2.0000	-0.3000	1.0000	

Fig. 4 States $z(1)$ and $z(2)$ and their estimates using 4S-RLS algorithm

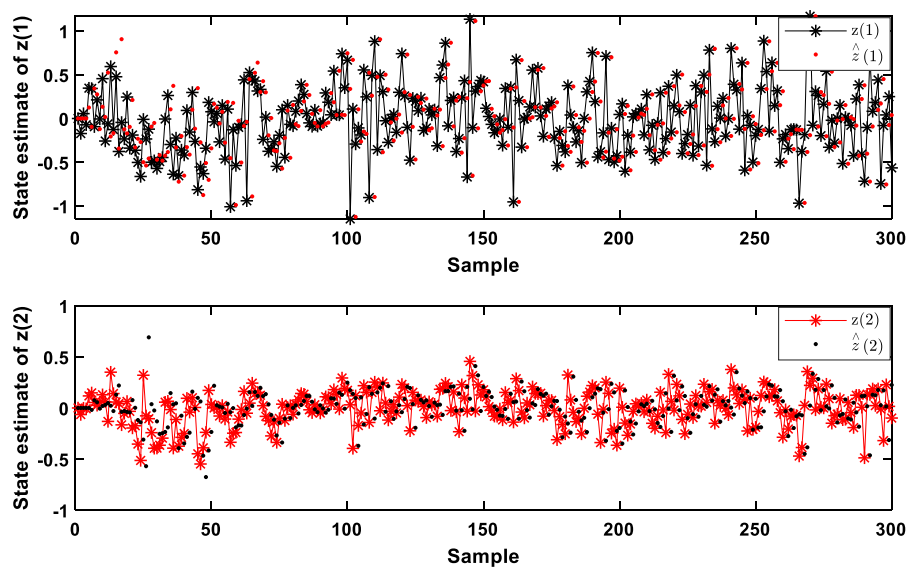


Fig. 5 States $z(1)$ and $z(2)$ and their estimates using 4S-SG algorithm

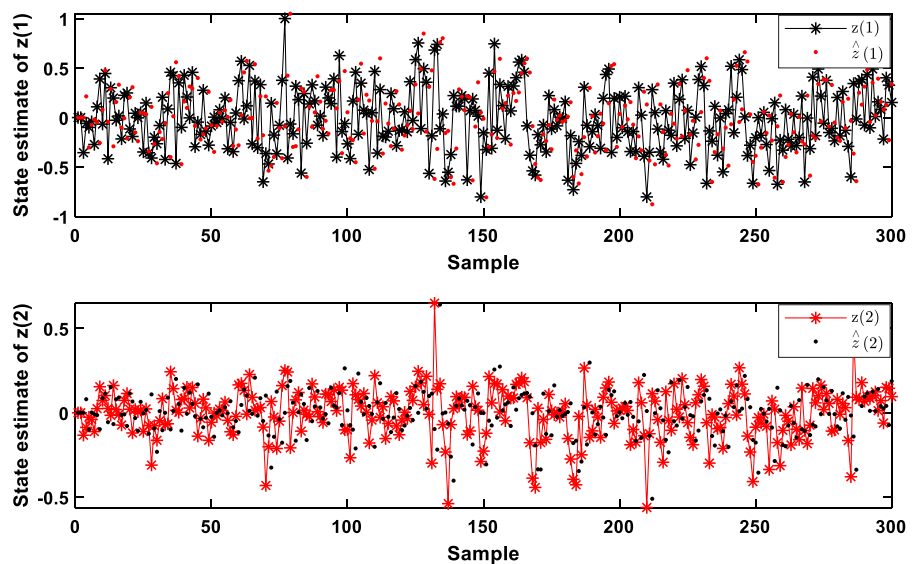
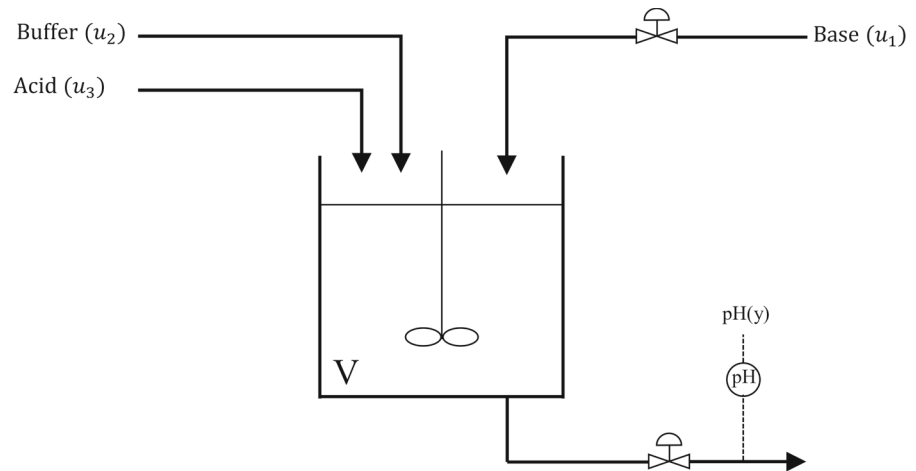
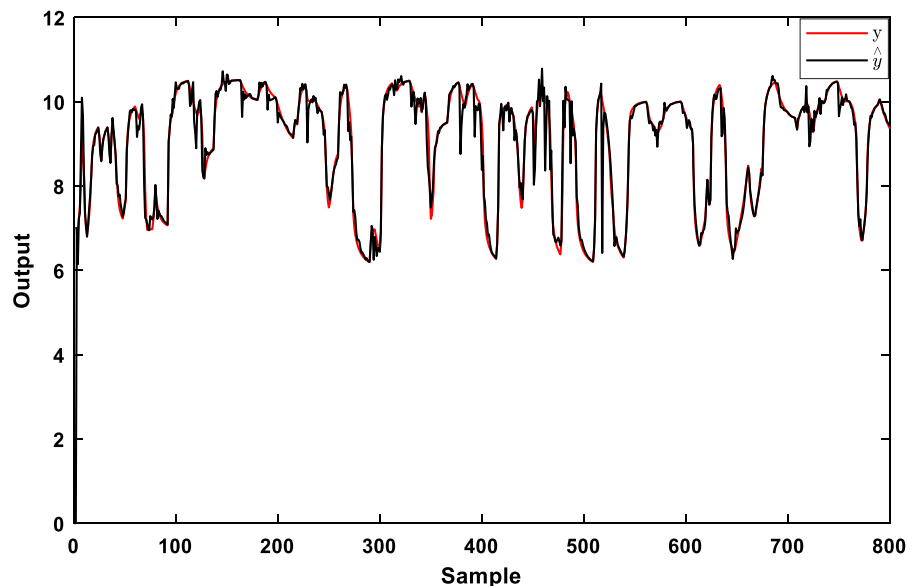
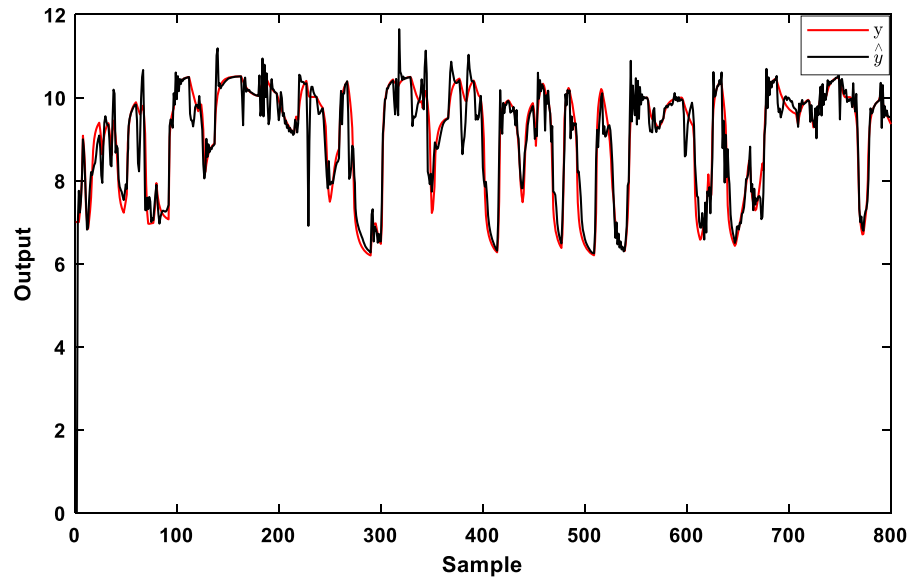


Fig. 6 Structure of pH neutralization process [17]**Fig. 7** Estimated output and real output of pH neutralization process using 4S-RLS algorithm

For simulation studies, consider a persistent excitation input signal $u(t)$, where $v(t)$ is a white noise with zero mean and variance $\sigma^2 = 0.10^2$, $\beta = 0.88$, $\alpha = 0.998$ and the data length is $L = 3000$. The parameter estimation error is calculated by $\delta = \|\rho(t) - \rho\|/\|\rho\|$. The results of the parameters estimations and the error with the four-stage recursive least squares algorithm are presented in Fig. 2 and Table 5. In addition, the results of the four-stage stochastic gradient algorithm are shown in Fig. 3 and Table 6.

From the simulation results presented in the figures and the tables, the following conclusions can be derived: Convergence analysis shows that the proposed algorithms are effective and the estimated parameters by the proposed algorithms can converge to their real values. Figures 2 and 3 show that the estimation error decreases with a suitable speed. Also, the results presented in Tables 5 and 6 show the convergence of the parameters to the real values. The 4S-RLS algorithm can provide an effective parameter estimation compared to the 4S-SG algorithm. Due to

Fig. 8 Estimated output and real output of pH neutralization process using 4S-SG algorithm



the data length and the noise variance, the 4S-RLS algorithm has a smaller estimation error than the 4S-SG algorithm and also has a higher parameter estimation accuracy. The system states and their estimates are shown in Figs. 4 and 5, respectively. The estimated states correspond well to the actual system states, indicating that the bilinear state observer is effective.

6.2 Practical example

The pH neutralization would be a case study to demonstrate the effectiveness of the proposed methods. For this process, the input/output data are gathered using the GMN test signal [32]. In this process, which is a highly nonlinear process, acid flow (HNO₃), base flow (NaOH) and buffer (NaHCO₃) are considered as input to the system, and the pH level is considered as the system output. The inputs can be regarded as (u₁), (u₂), (u₃). In this process, it can be assumed that acid flow rate and tank capacity are constant. The structure of the pH neutralization process is shown in Fig. 6.

In this example, it is assumed that only input/output data is available and the proposed methods are used to identify the system with bilinear state-space models.

To identify the system, 4S-RLS and 4S-SG algorithms are used to estimate the vector parameter $\hat{\rho}(t)$ with data length $t = N = 1280$ and variance $\sigma^2 = 0.10^2$. To confirm the results, the estimated output and the actual output data for the test dataset are shown in Figs. 7 and 8. The estimation error is calculated as $e := \frac{\|\hat{y}(t) - y\|}{\|y\|} \times 100$. According to the simulation results, the obtained error value for the 4S-RLS algorithm is 3.9734% and for the 4S-SG algorithm is 5.5564%.

7 Conclusion

In this paper, the parameter identification of bilinear state-space systems with colored noise which is expressed by the ARMA model was investigated. The proposed methods are based on the hierarchy principle. Since the states of the system need to be used for identification and only the system input/output data are available, a bilinear state observer was designed to estimate the system states. By using the hierarchical identification principle and the gradient search, a four-stage recursive least-squares algorithm and a four-stage stochastic gradient algorithm were presented to reduce the computational burden. The

simulation results have demonstrated that 4S-SG algorithm is efficient for identifying the bilinear systems, and the 4S-RLS algorithm outperforms the 4S-SG algorithm, with less estimation error. In addition, with increasing data length for different noise variances, the accuracy of the proposed method increases.

Funding The authors have not disclosed any funding.

Data availability Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

Proof of Theorem 2 From the covariance matrix relations, we have

$$\begin{aligned}
 R_1^{-1}(t) &= R_1^{-1}(t-1) + \hat{\varphi}_z(t)\hat{\varphi}_z^T(t) \\
 &= R_1^{-1}(0) + \sum_{i=1}^t \hat{\varphi}_z(i)\hat{\varphi}_z^T(i)
 \end{aligned}
 \tag{A1}$$

$$\begin{aligned}
 R_2^{-1}(t) &= R_2^{-1}(t-1) + \hat{\varphi}_{z\bar{u}}(t)\hat{\varphi}_{z\bar{u}}^T(t) \\
 &= R_2^{-1}(0) + \sum_{i=1}^t \hat{\varphi}_{z\bar{u}}(i)\hat{\varphi}_{z\bar{u}}^T(i)
 \end{aligned}
 \tag{A2}$$

$$\begin{aligned}
 R_3^{-1}(t) &= R_3^{-1}(t-1) + \hat{\varphi}_{\bar{u}}(t)\hat{\varphi}_{\bar{u}}^T(t) \\
 &= R_3^{-1}(0) + \sum_{i=1}^t \hat{\varphi}_{\bar{u}}(i)\hat{\varphi}_{\bar{u}}^T(i)
 \end{aligned}
 \tag{A3}$$

$$\begin{aligned}
 R_4^{-1}(t) &= R_4^{-1}(t-1) + \hat{\varphi}_n(t)\hat{\varphi}_n^T(t) \\
 &= R_4^{-1}(0) + \sum_{i=1}^t \hat{\varphi}_n(i)\hat{\varphi}_n^T(i)
 \end{aligned}
 \tag{A4}$$

Considering the left side of inequality in (71) and the covariance relation in (A1), we have

$$\alpha_1 I_{n_2} \leq \frac{1}{t} \sum_{i=1}^t \hat{\varphi}_z(i)\hat{\varphi}_z^T(i)
 \tag{A5}$$

Define

$$R_1(0) = r_0 I_{n_2} R_1^{-1}(0) = \frac{1}{r_0} I_{n_2}
 \tag{A6}$$

where r_0 is positive constant. As the elements of $R_1^{-1}(0)$ are nonnegative, by adding $R_1(0)$ to the right side of inequality in (A5), one can write

$$\alpha_1 I_{n_2} \leq \frac{1}{t} \sum_{i=1}^t \hat{\varphi}_z(i)\hat{\varphi}_z^T(i) + R_1^{-1}(0)$$

Therefore, we have

$$t\alpha_1 I_{n_2} \leq R_1^{-1}(t)$$

and

$$t\alpha_1 \leq \lambda_{\min}[R_1^{-1}(t)]
 \tag{A7}$$

Similarly

$$t\alpha_2 \leq \lambda_{\min}[R_2^{-1}(t)]
 \tag{A8}$$

$$t\alpha_3 \leq \lambda_{\min}[R_3^{-1}(t)]
 \tag{A9}$$

$$t\alpha_4 \leq \lambda_{\min}[R_4^{-1}(t)]
 \tag{A10}$$

Now, consider the right side inequalities in (71) and Theorem 1.

$$\frac{1}{t} \sum_{i=1}^t \hat{\varphi}_z(i)\hat{\varphi}_z^T(i) \leq \beta_1 I_{n_2}
 \tag{A11}$$

We add $R_1^{-1}(0)$ to the both sides of inequality (A11):

$$R_1^{-1}(t) \leq t\beta_1 I_{n_2} + R_1^{-1}(0)$$

$$|R_1^{-1}(t)| \leq |t\beta_1 I_{n_2} + R_1^{-1}(0)| = [t\beta_1 + R_1^{-1}(0)]^{n_2}$$

$$\ln|R_1^{-1}(t)| \leq n_2 \ln(t\beta_1 + R_1^{-1}(0))
 \tag{A12}$$

Similarly

$$\ln|R_2^{-1}(t)| \leq n_3 \ln(t\beta_2 + R_2^{-1}(0))
 \tag{A13}$$

$$\ln|R_3^{-1}(t)| \leq n_2 \ln(t\beta_3 + R_3^{-1}(0))
 \tag{A14}$$

$$\ln|R_4^{-1}(t)| \leq n_4 \ln(t\beta_4 + R_4^{-1}(0))
 \tag{A15}$$

According to Theorem 1, we have

Now, the proof is complete. □

$$\begin{aligned}
\|\hat{\rho}_a(t) - \rho_a\|^2 &= O\left(\frac{M(t)}{\lambda_{\min}[R_1^{-1}(t)]}\right) \\
&= O\left(\frac{[\ln|R_1^{-1}(t)|]^c + [\ln|R_2^{-1}(t)|]^c + [\ln|R_3^{-1}(t)|]^c + [\ln|R_4^{-1}(t)|]^c}{t\alpha_1}\right) \\
&= O\left(\frac{n_2^c [\ln(t\beta_1 + R_1^{-1}(0))]^c + n_3^c [\ln(t\beta_2 + R_2^{-1}(0))]^c + n_2^c [\ln(t\beta_3 + R_3^{-1}(0))]^c + n_4^c [\ln(t\beta_4 + R_4^{-1}(0))]^c}{t\alpha_1}\right) \\
&= O\left(\frac{[\ln t]^c}{t}\right) \rightarrow 0
\end{aligned}$$

References

- Ding, F., Wang, F., Xu, L., Wu, M.: 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)
- Ding, F., Xu, L., Zhu, Q.: Performance analysis of the generalised projection identification for time-varying systems. *IET Control Theory Appl.* **10**(18), 2506–2514 (2016)
- Wang, D.: Hierarchical parameter estimation for a class of MIMO Hammerstein systems based on the reframed models. *Appl. Math. Lett.* **57**, 13–19 (2016)
- Kazemi, M., Arefi, M.M., Alipouri, Y.: Wiener model based GMVC design considering sensor noise and delay. *ISA Trans.* **88**, 73–81 (2019)
- Wang, D., Zhang, W.: Improved least squares identification algorithm for multivariable Hammerstein systems. *J. Frankl. Inst.* **352**(11), 5292–5307 (2015)
- Zhang, C., Li, J.: Adaptive iterative learning control for nonlinear pure-feedback systems with initial state error based on fuzzy approximation. *J. Frankl. Inst.* **351**(3), 1483–1500 (2014)
- Zhang, H., Luo, Y., Liu, D.: Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints. *IEEE Trans. Neural Netw.* **20**(9), 1490–1503 (2009)
- Xu, L.: A proportional differential control method for a time-delay system using the Taylor expansion approximation. *Appl. Math. Comput.* **236**, 391–399 (2014)
- Xu, L.: The parameter estimation algorithms based on the dynamical response measurement data. *Adv. Mech. Eng.* **9**(11), 1687814017730003 (2017)
- Beauduin, T., Fujimoto, H.: Identification of system dynamics with time delay: a two-stage frequency domain approach. *IFAC-PapersOnLine* **50**(1), 10870–10875 (2017)
- Bianchi, F., Prandini, M., Piroddi, L.: A randomized two-stage iterative method for switched nonlinear systems identification. *Nonlinear Anal. Hybrid Syst.* **35**, 100818 (2020)
- Liu, Q., Ding, F.: Auxiliary model-based recursive generalized least squares algorithm for multivariate output-error autoregressive systems using the data filtering. *Circuits Syst. Signal Process.* **38**(2), 590–610 (2019)
- Yin, S., Gao, H., Kaynak, O.: Data-driven control and process monitoring for industrial applications—part I. *IEEE Trans. Ind. Electron.* **61**(11), 6356–6359 (2014)
- Yin, S., Gao, H., Qiu, J., Kaynak, O.: Fault detection for nonlinear process with deterministic disturbances: a just-in-time learning based data driven method. *IEEE Trans. Cybern.* **47**(11), 3649–3657 (2016)
- Zhang, X., Ding, F., Xu, L., Yang, E.: State filtering-based least squares parameter estimation for bilinear systems using the hierarchical identification principle. *IET Control Theory Appl.* **12**(12), 1704–1713 (2018)
- Bruni, C., Dipillo, G., Koch, G.: Bilinear systems: an appealing class of “nearly linear” systems in theory and applications. *IEEE Trans. Autom. Control* **19**(4), 334–348 (1974)
- Hafezi, Z., Arefi, M.M.: Recursive generalized extended least squares and RML algorithms for identification of bilinear systems with ARMA noise. *ISA Trans.* **88**, 50–61 (2019)
- Hwang, C., Chen, M.-Y.: Parameter identification of bilinear systems using the Galerkin method. *Int. J. Syst. Sci.* **16**(5), 641–648 (1985)
- Balatif, O., Abdelbaki, I., Rachik, M., Rachik, Z.: Optimal control for multi-input bilinear systems with an application in cancer chemotherapy. *Int. J. Sci. Innov. Math. Res. (IJSIMR)* **3**(2), 22–31 (2015)
- Arguello-Serrano, B., Velez-Reyes, M.: Nonlinear control of a heating, ventilating, and air conditioning system with thermal load estimation. *IEEE Trans. Control Syst. Technol.* **7**(1), 56–63 (1999)
- Tsai, S.-H., Hsiao, M.-Y., Tsai, K.-L.: LMI-based fuzzy control for a class of time-delay discrete fuzzy bilinear system. In *IEEE International Conference on Fuzzy Systems*, pp. 796–801 (2009)
- Figalli, G., Cava, M.L., Tomasi, L.: An optimal feedback control for a bilinear model of induction motor drives. *Int. J. Control* **39**(5), 1007–1016 (1984)

23. Mohler, R.R.: *Nonlinear Systems (vol. 2) Applications to Bilinear Control*. Prentice-Hall, Inc., Englewood Cliffs (1991)
24. Dai, H., Sinha, N.: Robust recursive least-squares method with modified weights for bilinear system identification. *IEE Proc. Control Theory Appl.* **136**(3), 122–126 (1989)
25. Li, M., Liu, X., Ding, F.: The maximum likelihood least squares based iterative estimation algorithm for bilinear systems with autoregressive moving average noise. *J. Frankl. Inst.* **354**(12), 4861–4881 (2017)
26. Li, M., Liu, X., Ding, F.: The gradient-based iterative estimation algorithms for bilinear systems with autoregressive noise. *Circuits Syst. Signal Process.* **36**(11), 4541–4568 (2017)
27. Li, M., Liu, X., Ding, F.: Least-squares-based iterative and gradient-based iterative estimation algorithms for bilinear systems. *Nonlinear Dyn.* **89**(1), 197–211 (2017)
28. Gibson, S., Wills, A., Ninness, B.: Maximum-likelihood parameter estimation of bilinear systems. *IEEE Trans. Autom. Control* **50**(10), 1581–1596 (2005)
29. Xu, L., Ding, F.: Iterative parameter estimation for signal models based on measured data. *Circuits Syst. Signal Process.* **37**(7), 3046–3069 (2018)
30. Ding, F., Liu, P.X., Liu, G.: Gradient based and least-squares based iterative identification methods for OE and OEMA systems. *Digital Signal Process.* **20**(3), 664–677 (2010)
31. Xu, H., Ding, F., Yang, E.: Modeling a nonlinear process using the exponential autoregressive time series model. *Nonlinear Dyn.* **95**(3), 2079–2092 (2019)
32. Kazemi, M., Arefi, M.M.: A fast iterative recursive least squares algorithm for Wiener model identification of highly nonlinear systems. *ISA Trans.* **67**, 382–388 (2017)
33. 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)
34. Fnaiech, F., Ljung, L.: Recursive identification of bilinear systems. *Int. J. Control* **45**(2), 453–470 (1987)
35. Phan, M.Q., Vicario, F., Longman, R.W., Betti, R.: Optimal bilinear observers for bilinear state-space models by interaction matrices. *Int. J. Control* **88**(8), 1504–1522 (2015)
36. Phan, M.Q., Çelik, H.: A superspace method for discrete-time bilinear model identification by interaction matrices. *J. Astronaut. Sci.* **59**(1), 421–440 (2012)
37. Gabr, M.: A recursive (on-line) identification of bilinear systems. *Int. J. Control* **44**(4), 911–917 (1986)
38. Hizir, N.B., Phan, M.Q., Betti, R., Longman, R.W.: Identification of discrete-time bilinear systems through equivalent linear models. *Nonlinear Dyn.* **69**(4), 2065–2078 (2012)
39. Meng, D.: Recursive least squares and multi-innovation gradient estimation algorithms for bilinear stochastic systems. *Circuits Syst. Signal Process.* **36**(3), 1052–1065 (2017)
40. Liu, S., Ding, F., Xu, L., Hayat, T.: Hierarchical principle-based iterative parameter estimation algorithm for dual-frequency signals. *Circuits Syst. Signal Process.* **38**(7), 3251–3268 (2019)
41. Cui, T., Ding, F., Jin, X.-B., Alsaedi, A., Hayat, T.: Joint multi-innovation recursive extended least squares parameter and state estimation for a class of state-space systems. *Int. J. Control Autom. Syst.* 1–13 (2019)
42. Tsai, S.-H., Li, T.-H.S.: Robust fuzzy control of a class of fuzzy bilinear systems with time-delay. *Chaos, Solitons Fractals* **39**(5), 2028–2040 (2009)
43. Lu, X., Ding, F., Alsaedi, A., Hayat, T.: Decomposition-based gradient estimation algorithms for multivariable equation-error systems. *Int. J. Control Autom. Syst.* **17**(8), 2037–2045 (2019)
44. Li, M., Liu, X.: The least squares based iterative algorithms for parameter estimation of a bilinear system with autoregressive noise using the data filtering technique. *Signal Process.* **147**, 23–34 (2018)
45. Ding, F., Xu, L., Meng, D., Jin, X.-B., Alsaedi, A., Hayat, T.: Gradient estimation algorithms for the parameter identification of bilinear systems using the auxiliary model. *J. Comput. Appl. Math.* 112575 (2019)
46. Luo, H., Li, K., Huo, M., Yin, S., Kaynak, O.: A data-driven process monitoring approach with disturbance decoupling. In: *Data Driven Control and Learning Systems Conference*, pp. 569–574 (2018)
47. Zhang, X., Ding, F., Xu, L.: Recursive parameter estimation methods and convergence analysis for a special class of nonlinear systems. *Int. J. Robust Nonlinear Control* **30**(4), 1373–1393 (2020)
48. Zhang, X., Xu, L., Ding, F., Hayat, T.: Combined state and parameter estimation for a bilinear state space system with moving average noise. *J. Frankl. Inst.* **355**(6), 3079–3103 (2018)
49. Chen, H., Ding, F.: Hierarchical least squares identification for Hammerstein nonlinear controlled autoregressive systems. *Circuits Syst. Signal Process.* **34**(1), 61–75 (2015)
50. Zhang, X., Ding, F.: Recursive parameter estimation and its convergence for bilinear systems. *IET Control Theory Appl.* **14**(5), 677–688 (2019)
51. Ding, F., Chen, H., Xu, L., Dai, J., Li, Q., Hayat, T.: A hierarchical least squares identification algorithm for Hammerstein nonlinear systems using the key term separation. *J. Franklin Inst.* **355**(8), 3737–3752 (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.