Hierarchical multi-innovation stochastic gradient identification algorithm for estimating a bilinear state-space model with moving average noise[☆]

Ya Gu[*,a], Wei Dai[a], Quanmin Zhu[b], Hassan Nouri[b]

[a] *College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, 201418, Shanghai, PR China*

[b] *Department of Engineering Design and Mathematics, University of the West of England, Bristol BS16 1QY, U.K.*

Abstract

This paper considers the combined parameter and state estimation problem of a bilinear state space system with moving average noise. There are product terms of state variables and control variables in bilinear systems, which brings difficulties to parameter and state estimation. By designing a bilinear state estimator based on Kalman filter and using input-output data to estimate the state, a hierarchical multi-innovation stochastic gradient (i.e., H-MISG) algorithm based on the state estimator is proposed to jointly estimate unknown states and parameters. In addition, compared with the hierarchical stochastic gradient algorithm, H-MISG algorithm introduces the innovation length parameter, makes full use of the system input and output data information, and improves the accuracy of parameter estimation. Numerical simulation examples verify the effectiveness of the proposed algorithm.

Key words: Bilinear system, Multi-innovation identification, Kalman filtering, Parameter estimation, State estimation

1. Introduction

The implementation of control strategies mainly relies on the accurate mathematical model of a system. System identification is the theory and methods of establishing the mathematical models of systems [1, 2, 3, 4, 5]. The basic method of the current common identification algorithm is to transform the identification problem into a parameter estimation problem by establishing the parameter model of a system [6, 7, 8]. This kind of algorithms can better solve the identification problem of linear systems or intrinsically linear systems, but it is more difficult to apply to intrinsically nonlinear systems [9, 10, 11]. However, the models in the real world are not strictly linear. They show nonlinear characteristics more or less [12, 13]. Therefore, more and more nonlinear phenomena and nonlinear models have attracted people's attention.

Nonlinear systems widely exist in people's production and life. With the development and progress of human society, more and more nonlinear phenomena and nonlinear systems have attracted wide attention from researchers. The discovery of chaotic phenomena is hailed as "One of the Three Great Discoveries of the Twentieth Century". At present, the research on nonlinear theory is in the development stage. Establishing models describing nonlinear phenomena and nonlinear systems is the basis for studying nonlinear problems [14, 15, 16, 17]. Linear system identification theory has become mature, but the general linear model is actually an approximate mathematical description of the real system obtained after some nonlinearities are ignored or replaced by linear relations. With the rapid development of science and technology, control systems are becoming more and more complex, and the requirements for control accuracy are getting higher and higher. Systems with complex nonlinearities

[☆]This work was supported by the National Natural Science Foundation of China (No. 61903050), the Natural Science Foundation of Shanghai (No. 22ZR1445300) and the Taizhou science and technology support plan (Social Development) project (No. SSF20210004).

[*]Corresponding author

Email address: guya@shnu.edu.cn (Ya Gu)

cannot be approximated by linear models, so the study of nonlinear system identification theory has very important practical significance [18, 19, 20, 21].

The study of bilinear systems began in the 1960s, and has received extensive attention and rapid development since the 1970s, becoming one of the more mature branches in the study of nonlinear systems. Bilinear system has some characteristics of variable structure system, so it has certain adaptability [22, 23]. For the situation where the control variable is restricted (that is, the size of the control variable must be within a certain limit), the stability condition expressed in frequency domain language has been found. Bilinear systems have better controllability than linear systems [24]. Even if the control variables are restricted, the system may still be fully controllable. Some sufficient conditions for the system to be fully controllable have been obtained. The use of dynamic programming or maximum value principle has been able to solve some of the optimal control problems of bilinear systems, such as the fastest control, the most fuel-efficient control, and the optimal control of discrete bilinear systems and stochastic bilinear systems. There are many practical examples of bilinear system theory. For example, optimal control of nuclear fission and heat exchange processes in nuclear power plants and nuclear power plants, population prediction and control, etc.

Combining the gradient search with the hierarchical identification principle and the multi-innovation identification theory, this paper intends to study recursive identification algorithms for the bilinear state space model with moving average noise. On the basis of gradient search, this paper aims to use useful information of system accuracy to update the collected data to identify bilinear systems. Therefore, this paper studies a H-MISG identification algorithm to realize joint parameter and state estimation and reduce the influence of colored noise on parameter estimation. The main contributions are summarized as follows. Compared with the previous work in [25], the proposed algorithm gives the joint state and parameter estimation to improve the convergence speed, and gives the convergence analyzes to ensure the stability of the algorithm.

- A moving average noise identification model is derived for bilinear state space systems. By introducing the length of new information, expanding the identification of new information, fully mining useful information from the available measurement data, by extending the new information, rolling and reusing information, to make up for the impact of lost data and improve system parameter estimation accuracy.
- A state estimator is designed to obtain the system states on the basis of the Kalman filtering, and a state estimator-based multi-innovation stochastic gradient algorithm is proposed to estimate the unknown system parameters by the hierarchical identification principle.

The rest of this paper is organized as follows. Section 2 derives the identification model of bilinear state-space systems with moving average noise. Section 3 proposes the stochastic gradient algorithm using the hierarchical identification principle. Section 4 derives the multi-innovation stochastic gradient algorithm as a comparison. Section 5 gives the state estimation for bilinear systems. Section 6 gives the convergence analysis. Section 7 provides two examples to verify the effectiveness of the proposed algorithm. Finally, some concluding remarks are given in Section 8.

2. System description and identification model

First of all, let us introduce some notation. "$X := A$" stands for "A is defined as X"; the symbol $\boldsymbol{I}$ ($\boldsymbol{I}_n$) means the identity matrix of appropriate size ($n \times n$); z denotes a unit forward shift operator like $z\boldsymbol{x}_k = \boldsymbol{x}_{k+1}$ and $z^{-1}\boldsymbol{x}_k = \boldsymbol{x}_{k-1}$; the superscript T symbolizes the vector/matrix transpose; $\hat{\boldsymbol{\vartheta}}_k$ denotes the estimate of $\boldsymbol{\vartheta}$ at time k; $\mathbf{1}_n$ represents an n-dimensional column vector whose entries are all 1.

Consider a bilinear state-space system in the observability canonical form:

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{x}_k u_k + \boldsymbol{f}u_k, \tag{2.1}$$

$$y_k = \boldsymbol{h}\boldsymbol{x}_k + \frac{D(z)}{C(z)}v_k, \tag{2.2}$$

where $\boldsymbol{x}_k := [x_{1,k}, x_{2,k}, \ldots, x_{n,k}]^{\mathrm{T}} \in \mathbb{R}^n$ is the state vector, $u_k \in \mathbb{R}$ and $y_k \in \mathbb{R}$ are the control variable and output variable, and $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n}$, $\boldsymbol{f} \in \mathbb{R}^{n\times 1}$ and $\boldsymbol{h} \in \mathbb{R}^{1\times n}$ are the system parameter matrices and vectors,

$$\boldsymbol{A} := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ a_n & a_{n-1} & a_{n-2} & \cdots & a_1 \end{bmatrix} \in \mathbb{R}^{n\times n}, \quad a_i \in \mathbb{R}, \quad i = 1, 2, \ldots, n,$$

$$\boldsymbol{B} := \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_n \end{bmatrix} \in \mathbb{R}^{n\times n}, \quad \boldsymbol{f} := \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \in \mathbb{R}^{n\times 1},$$

$$\boldsymbol{h} := [1, 0, \ldots, 0] \in \mathbb{R}^{1\times n},$$

$w_k := \frac{D(z)}{C(z)} v_k$ is the measurement noise with $C(z) := 1 + c_1 z^{-1} + c_2 z^{-2} + \ldots + c_{n_c} z^{-n_c}$ and $D(z) := 1 + d_1 z^{-1} + d_2 z^{-2} + \ldots + d_{n_d} z^{-n_d}$, $\mathrm{E}[v_k] = 0$, $\mathrm{E}[v_k^2] = \sigma^2$, $\mathrm{E}[v_k v_i] = 0$, $t \neq i$. Without loss of generality, this paper makes the following assumptions. The dimension n of the system state vector is known, $u_k = 0$, $\boldsymbol{x}_k = \boldsymbol{0}$, $y_k = 0$ and $w_k = 0$ for $t \leqslant 0$. The bilinear system in (2.1)–(2.2) is observable and controllable.

The system in (2.1)–(2.2) is written as

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ \vdots \\ x_{n-1,k+1} \\ x_{n,k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ a_n & a_{n-1} & a_{n-2} & \cdots & a_1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n-1,k} \\ x_{n,k} \end{bmatrix} + \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_{n-1} \\ \boldsymbol{b}_n \end{bmatrix} \boldsymbol{x}_k u_k + \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{n-1} \\ f_n \end{bmatrix} u_k, \tag{2.3}$$

$$y_k = [1, 0, \ldots, 0]\boldsymbol{x}_k + \frac{D(z)}{C(z)} v_k. \tag{2.4}$$

From (2.3), we have the following relations:

$$\begin{aligned} x_{i,k+1} &= x_{i+1,k} + \boldsymbol{b}_i \boldsymbol{x}_k u_k + f_i u_k, \quad i = 1, 2, \ldots, n-1, \\ x_{n,k+1} &= \boldsymbol{a}\boldsymbol{x}_{k-n} + \boldsymbol{b}_n \boldsymbol{x}_k u_k + f_n u_k. \end{aligned} \tag{2.5}$$

Multiplying both sides of (2.5) by z^{-i}, and summing it for $i = 1$ to $i = n-1$ yields:

$$\begin{aligned} x_{1,k} = {} & \boldsymbol{a}\boldsymbol{x}_{k-n} + \boldsymbol{b}_1 \boldsymbol{x}_{k-1} u_{k-1} + \boldsymbol{b}_2 \boldsymbol{x}_{k-2} u_{k-2} + \ldots + \boldsymbol{b}_{n-1} \boldsymbol{x}_{k-n+1} u_{k-n+1} + \boldsymbol{b}_n \boldsymbol{x}_{k-n} u_{k-n} \\ & + f_1 u_{k-1} + f_2 u_{k-2} + \ldots + f_n u_{k-n}. \end{aligned}$$

Define the parameter vector $\boldsymbol{\vartheta}_a$, $\boldsymbol{\vartheta}_b$, $\boldsymbol{\vartheta}_f$, $\boldsymbol{\vartheta}_2$ and the information vector $\boldsymbol{\phi}_{a,k}$, $\boldsymbol{\phi}_{b,k}$, $\boldsymbol{\phi}_{u,k}$, $\boldsymbol{\phi}_{2,k}$ as

$$\begin{aligned} \boldsymbol{\vartheta}_a &:= [a_1, a_2, \ldots, a_n]^{\mathrm{T}} \in \mathbb{R}^n, \\ \boldsymbol{\vartheta}_b &:= [\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n]^{\mathrm{T}} \in \mathbb{R}^{n^2}, \\ \boldsymbol{\vartheta}_f &:= [f_1, f_2, \ldots, f_n]^{\mathrm{T}} \in \mathbb{R}^n, \\ \boldsymbol{\vartheta}_2 &:= [c_1, c_2, \ldots, c_{n_c}, d_1, d_2, \ldots, d_{n_d}]^{\mathrm{T}} \in \mathbb{R}^{n_c+n_d}, \\ \boldsymbol{\phi}_{a,k} &:= [x_{k-1}, x_{k-2}, \ldots, x_{k-n}]^{\mathrm{T}} \in \mathbb{R}^n, \\ \boldsymbol{\phi}_{b,k} &:= [\boldsymbol{x}_{k-1}u_{k-1}, \ldots, \boldsymbol{x}_{k-n+1}u_{k-n+1}, \boldsymbol{x}_{k-n}u_{k-n}]^{\mathrm{T}} \in \mathbb{R}^{n^2}, \\ \boldsymbol{\phi}_{f,k} &:= [u_{k-1}, u_{k-2}, \ldots, u_{k-n}]^{\mathrm{T}} \in \mathbb{R}^n, \\ \boldsymbol{\phi}_{2,k} &:= [-w_{k-1}, -w_{k-2}, \ldots, -w_{k-n_c}, v_{k-1}, v_{k-2}, \ldots, v_{k-n_d})]^{\mathrm{T}} \in \mathbb{R}^{n_c+n_d}. \end{aligned}$$

Mathematically, Equation (2.4) implies that:

$$y_k = x_{1,k} + \frac{D(z)}{C(z)} v_k$$

$$
\begin{aligned}
&= \boldsymbol{\vartheta}_a^{\mathrm{T}}\boldsymbol{\phi}_{a,k} + \boldsymbol{\vartheta}_b^{\mathrm{T}}\boldsymbol{\phi}_{b,k} + \boldsymbol{\vartheta}_f^{\mathrm{T}}\boldsymbol{\phi}_{f,k} + \boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\phi}_{2,k} + v_k \\
&= y_{a,k} + y_{b,k} + y_{f,k} + w_k.
\end{aligned}
$$

The output y_k is decomposed into the following identification models of the system in (2.1) and (2.2):

$$y_{a,k} = \boldsymbol{\vartheta}_a^{\mathrm{T}}\boldsymbol{\phi}_{a,k} + v_k, \tag{2.6}$$
$$y_{b,k} = \boldsymbol{\vartheta}_b^{\mathrm{T}}\boldsymbol{\phi}_{b,k} + v_k, \tag{2.7}$$
$$y_{f,k} = \boldsymbol{\vartheta}_f^{\mathrm{T}}\boldsymbol{\phi}_{f,k} + v_k, \tag{2.8}$$
$$w_k = \boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\phi}_{2,k} + v_k. \tag{2.9}$$

The proposed parameter estimation algorithms in this paper are based on this identification model. Many identification methods are derived based on the identification models of the systems [26, 27, 28, 29, 30] and these methods can be used to estimate the parameters of other linear systems and nonlinear systems [31, 32, 33, 34, 35] and can be applied to other fields [36, 37, 38, 39, 40] such as chemical process control systems.

Remark 2.1: Previous study has developed an expectation maximization algorithm for state space systems with time delay [41]. While, due to the removal of state variables, the calculation burden is heavy and state estimation is not considered. Besides, less work is focused on theoretical analysis of the performance of the proposed method. This leads us to explore a new combination of state and parameter estimation methods and provide theoretical analysis to prove the convergence of the proposed algorithm.

The purpose of this paper is to acquire estimates of unavailable parameters and states of bilinear systems using measured information. In view of the state space identification model with moving average noise, this problem will be solved in the next section.

3. Hierarchical stochastic gradient identification algorithm

This section presents an hierarchical stochastic gradient (H-SG) identification algorithm for the considered system by decomposing the identification model into several sub-identification models based on the negative gradient search and the hierarchical identification principle. Let $\hat{\boldsymbol{\vartheta}}_{a,k} := [\hat{a}_{1,k}, \hat{a}_{2,k}, \ldots, \hat{a}_{n,k}]^{\mathrm{T}}$, $\hat{\boldsymbol{\vartheta}}_{b,k} := [\hat{\boldsymbol{b}}_{1,k}, \hat{\boldsymbol{b}}_{2,k}, \ldots, \hat{\boldsymbol{b}}_{n,k}]^{\mathrm{T}}$, $\hat{\boldsymbol{\vartheta}}_{f,k} := [\hat{f}_{1,k}, \hat{f}_{2,k}, \ldots, \hat{f}_{n,k}]^{\mathrm{T}}$, $\hat{\boldsymbol{\vartheta}}_{2,k} := [\hat{c}_{1,k}, \hat{c}_{2,k}, \ldots, \hat{c}_{n_c,k}, \hat{d}_{1,k}, \hat{d}_{2,k}, \ldots, \hat{d}_{n_d,k}]^{\mathrm{T}}$ denote the estimates of the parameter vectors $\boldsymbol{\vartheta}_{a,k}$, $\boldsymbol{\vartheta}_{b,k}$, $\boldsymbol{\vartheta}_{f,k}$, $\boldsymbol{\vartheta}_{2,k}$ at time k. The estimates of the intermediate variables are used to improve the parameter estimation algorithm:

$$
\begin{aligned}
\hat{\boldsymbol{\phi}}_{1,k} &:= [\hat{\boldsymbol{x}}_{k-n}, \hat{\boldsymbol{x}}_{k-1}u_{k-1}, \hat{\boldsymbol{x}}_{k-2}u_{k-2}, \ldots, \hat{\boldsymbol{x}}_{k-n+1}u_{k-n+1}, \hat{\boldsymbol{x}}_{k-n}u_{k-n}, u_{k-1}, \ldots, u_{k-n}]^{\mathrm{T}}, \\
\hat{\boldsymbol{\phi}}_{2,k} &:= [-\hat{w}_{k-1}, -\hat{w}_{k-2}, \ldots, -\hat{w}_{k-n_c}, v_{k-1}, v_{k-2}, \ldots, v_{k-n_d}]^{\mathrm{T}}.
\end{aligned}
$$

By means of the negative gradient search, we derive the H-SG parameter estimation algorithm for estimating the parameters of the bilinear system:

$$\hat{\boldsymbol{\vartheta}}_{a,k} = \hat{\boldsymbol{\vartheta}}_{a,k-1} + \frac{\hat{\boldsymbol{\phi}}_{a,k}}{r_{a,k}}[y_{a,k} - \hat{\boldsymbol{\vartheta}}_{a,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{a,k}], \tag{3.10}$$
$$r_{a,k} = r_{a,k-1} + \|\hat{\boldsymbol{\phi}}_{a,k}\|^2, \tag{3.11}$$
$$\hat{\boldsymbol{\phi}}_{a,k} = [\hat{x}_{k-1}, \hat{x}_{k-2}, \ldots, \hat{x}_{k-n}]^{\mathrm{T}}, \tag{3.12}$$
$$\hat{\boldsymbol{\vartheta}}_{b,k} = \hat{\boldsymbol{\vartheta}}_{b,k-1} + \frac{\hat{\boldsymbol{\phi}}_{b,k}}{r_{b,k}}[y_{b,k} - \hat{\boldsymbol{\vartheta}}_{b,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{b,k}], \tag{3.13}$$
$$r_{b,k} = r_{b,k-1} + \|\hat{\boldsymbol{\phi}}_{b,k}\|^2, \tag{3.14}$$
$$\hat{\boldsymbol{\phi}}_{b,k} = [\hat{\boldsymbol{x}}_{k-1}u_{k-1}, \ldots, \hat{\boldsymbol{x}}_{k-n+1}u_{k-n+1}, \hat{\boldsymbol{x}}_{k-n}u_{k-n}]^{\mathrm{T}}, \tag{3.15}$$
$$\hat{\boldsymbol{\vartheta}}_{f,k} = \hat{\boldsymbol{\vartheta}}_{f,k-1} + \frac{\hat{\boldsymbol{\phi}}_{f,k}}{r_{f,k}}[y_{f,k} - \hat{\boldsymbol{\vartheta}}_{f,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{f,k}], \tag{3.16}$$
$$r_{f,k} = r_{f,k-1} + \|\hat{\boldsymbol{\phi}}_{f,k}\|^2, \tag{3.17}$$

$$\hat{\boldsymbol{\phi}}_{f,k} = [u_{k-1}, u_{k-2}, \ldots, u_{k-n}]^{\mathrm{T}}, \tag{3.18}$$

$$\hat{\boldsymbol{\vartheta}}_{2,k} = \hat{\boldsymbol{\vartheta}}_{2,k-1} + \frac{\hat{\boldsymbol{\phi}}_{2,k}}{r_{2,k}}[\hat{w}_k - \hat{\boldsymbol{\vartheta}}_{2,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{2,k}], \tag{3.19}$$

$$r_{2,k} = r_{2,k-1} + \|\hat{\boldsymbol{\phi}}_{2,k}\|^2, \tag{3.20}$$

$$\hat{\boldsymbol{\phi}}_{2,k} = [-\hat{w}_{k-1}, -\hat{w}_{k-2}, \ldots, -\hat{w}_{k-n_c}, v_{k-1}, v_{k-2}, \ldots, v_{k-n_d}]^{\mathrm{T}}. \tag{3.21}$$

The steps of computing the parameter estimates of the nonlinear sandwich system using the proposed H-SG method in (3.10) to (3.21) are as follows:

1. To initiate: Let $k = 1$; let $r_{a,0} = 1$, $r_{b,0} = 1$, $r_{f,0} = 1$ and $r_{2,0} = 1$, $\hat{\boldsymbol{\vartheta}}_{a,0} = [\hat{a}_{1,0}, \hat{a}_{2,0}, \ldots, \hat{a}_{n,0}]^{\mathrm{T}} = \mathbf{1}_n/p_0$, $\hat{\boldsymbol{\vartheta}}_{b,0} = [\hat{\boldsymbol{b}}_{1,0}, \hat{\boldsymbol{b}}_{2,0}, \ldots, \hat{\boldsymbol{b}}_{n,0}]^{\mathrm{T}} = \mathbf{1}_n/p_0$, $\hat{\boldsymbol{\vartheta}}_{f,0} = [\hat{f}_{1,0}, \hat{f}_{2,0}, \ldots, \hat{f}_{n,0}]^{\mathrm{T}} = \mathbf{1}_n/p_0$, $\hat{\boldsymbol{\vartheta}}_{2,0} = [\hat{c}_{1,0}, \hat{c}_{2,0}, \ldots, \hat{c}_{n_c,0}, \hat{d}_{1,0}, \hat{d}_{2,0}, \ldots, \hat{d}_{n_d,0}]^{\mathrm{T}} = \mathbf{1}_n/p_0$, $P_0 = 10^6$, $\mathbf{1}_n \in \mathbb{R}^n$ with element 1. Set the estimation accuracy ε.
2. Collect the observed data u_k and y_k, form $\hat{\boldsymbol{\phi}}_{f,k}$ using (3.18).
3. Compute $\hat{\boldsymbol{\phi}}_{2,k}$ using (3.21), form $\hat{\boldsymbol{\phi}}_{a,k}$ using (3.12) and form $\hat{\boldsymbol{\phi}}_{b,k}$ using (3.15).
4. Compute $r_{a,k}$, $r_{b,k}$, $r_{f,k}$ and $r_{2,k}$ using (3.11), (3.14), (3.17) and (3.20), respectively.
5. Update the parameter estimates $\hat{\boldsymbol{\vartheta}}_{a,k}$, $\hat{\boldsymbol{\vartheta}}_{b,k}$, $\hat{\boldsymbol{\vartheta}}_{f,k}$ and $\hat{\boldsymbol{\vartheta}}_{2,k}$ using (3.10), (3.13), (3.16) and (3.19), respectively.
6. If $\|\hat{\boldsymbol{\vartheta}}_{a,k} - \hat{\boldsymbol{\vartheta}}_{a,k}\| > \varepsilon$, $\|\hat{\boldsymbol{\vartheta}}_{b,k} - \hat{\boldsymbol{\vartheta}}_{b,k}\| > \varepsilon$, $\|\hat{\boldsymbol{\vartheta}}_{f,k} - \hat{\boldsymbol{\vartheta}}_{f,k}\| > \varepsilon$, $\|\hat{\boldsymbol{\vartheta}}_{2,k} - \hat{\boldsymbol{\vartheta}}_{2,k}\| > \varepsilon$, then $k := k + 1$ and go to Step 2, else terminate the computation process.

Remark 3.1: For the dynamic sampling data of nonlinear systems with moving average noise, the proposed H-MISG identification algorithm can capture the real-time information of the system to be identified. Complex nonlinear systems with time delays are solved by gradient search to obtain variables that cannot be measured by internal parameter estimation. It should be noted that the H-SG identification method only uses the current time base and cannot supplement the high-precision parameter estimation. In the next section, we will extend the identified regression new information vector to the new information matrix, reuse the data information, improve the convergence speed of the identification algorithm, and improve the accuracy of parameter estimation.

4. Hierarchical multi-innovation stochastic gradient identification algorithm

The main idea of hierarchical identification is the principle of hierarchical identification based on identification model decomposition. Hierarchical identification is different from subsystem decomposition identification methods. It can not only solve the problem of large-scale systems with a large amount of calculation, but also solve the identification problems of complex identification models, including the hierarchical iterative solution of the Lyapunov matrix equation and the Sylvester matrix equation.

In the recursive formula of parameter estimation, we use more measured data and information to deduce a parameter estimation algorithm with high estimation accuracy. Dynamic batch data is designed and expressed in $y_k, y_{k-1}, \ldots, y_{k-p+1}$, where the length of dynamic data is p, and the current time is k. Obviously, dynamic data can be regarded as a window data with width p, and these dynamic data will be updated over time. The data window contains not only the current time data, but also the previous data. Compared with the dynamic observation data used in the stochastic gradient algorithm, the dynamic window data can obtain more system dynamic information.

Define the stacked information matrix

$$\boldsymbol{\Phi}_a(p_a, k) := \begin{bmatrix} \boldsymbol{\phi}_{a,k}^{\mathrm{T}} \\ \boldsymbol{\phi}_{a,k-1}^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\phi}_{a,k-p_a+1}^{\mathrm{T}} \end{bmatrix} \in \mathbb{R}^{n \times p_a}, \quad \boldsymbol{Y}_a(p_a, k) := \begin{bmatrix} y_{a,k} \\ y_{a,k-1} \\ \vdots \\ y_{a,k-p_a+1} \end{bmatrix} \in \mathbb{R}^{p_a},$$

$$\boldsymbol{\Phi}_b(p_b,k) := \begin{bmatrix} \boldsymbol{\phi}_{b,k}^{\mathrm{T}} \\ \boldsymbol{\phi}_{b,k-1}^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\phi}_{b,k-p_b+1}^{\mathrm{T}} \end{bmatrix} \in \mathbb{R}^{n^2\times p_b}, \quad \boldsymbol{Y}_b(p_b,k) := \begin{bmatrix} y_{b,k} \\ y_{b,k-1} \\ \vdots \\ y_{b,k-p_b+1} \end{bmatrix} \in \mathbb{R}^{p_b},$$

$$\boldsymbol{\Phi}_f(p_f,k) := \begin{bmatrix} \boldsymbol{\phi}_{f,k}^{\mathrm{T}} \\ \boldsymbol{\phi}_{f,k-1}^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\phi}_{f,k-p_f+1}^{\mathrm{T}} \end{bmatrix} \in \mathbb{R}^{n\times p_f}, \quad \boldsymbol{Y}_f(p_f,k) := \begin{bmatrix} y_{f,k} \\ y_{f,k-1} \\ \vdots \\ y_{f,k-p_f+1} \end{bmatrix} \in \mathbb{R}^{p_f},$$

$$\boldsymbol{W}(p_2,k) := \begin{bmatrix} w_k \\ w_{k-1} \\ \vdots \\ w_{k-p_2+1} \end{bmatrix} \in \mathbb{R}^{p_2}, \quad \boldsymbol{V}_i(p_i,k) := \begin{bmatrix} v_k \\ v_{k-1} \\ \vdots \\ v_{k-p_i+1} \end{bmatrix} \in \mathbb{R}^{p_i}, \quad i = a,b,f,2.$$

From Equations (2.6)–(2.9), we have

$$\begin{aligned}
\boldsymbol{Y}_a(p_a,k) &= \boldsymbol{\vartheta}_a^{\mathrm{T}}\boldsymbol{\Phi}_a(p_a,k) + \boldsymbol{V}(p_a,k),\\
\boldsymbol{Y}_b(p_b,k) &= \boldsymbol{\vartheta}_b^{\mathrm{T}}\boldsymbol{\Phi}_b(p_b,k) + \boldsymbol{V}(p_b,k),\\
\boldsymbol{Y}_f(p_f,k) &= \boldsymbol{\vartheta}_f^{\mathrm{T}}\boldsymbol{\Phi}_f(p_f,k) + \boldsymbol{V}(p_f,k),\\
\boldsymbol{W}(p_2,k) &= \boldsymbol{\vartheta}_2^{\mathrm{T}}\boldsymbol{\Phi}_2(p_2,k) + \boldsymbol{V}_2(p_2,k).
\end{aligned}$$

Then, the following recursive relations are given by

$$\hat{\boldsymbol{\vartheta}}_{a,k} = \hat{\boldsymbol{\vartheta}}_{a,k-1} - \mu_{a,k}\boldsymbol{\Phi}_a(p_a,k)[\boldsymbol{Y}_a(p_a,k) - \boldsymbol{\vartheta}_{a,k-1}^{\mathrm{T}}\boldsymbol{\Phi}_a(p_a,k)], \tag{4.22}$$

$$\hat{\boldsymbol{\vartheta}}_{b,k} = \hat{\boldsymbol{\vartheta}}_{b,k-1} - \mu_{b,k}\boldsymbol{\Phi}_b(p_b,k)[\boldsymbol{Y}_b(p_b,k) - \boldsymbol{\vartheta}_{b,k-1}^{\mathrm{T}}\boldsymbol{\Phi}_b(p_b,k)], \tag{4.23}$$

$$\hat{\boldsymbol{\vartheta}}_{f,k} = \hat{\boldsymbol{\vartheta}}_{f,k-1} - \mu_{f,k}\boldsymbol{\Phi}_f(p_f,k)[\boldsymbol{Y}_f(p_f,k) - \boldsymbol{\vartheta}_{f,k-1}^{\mathrm{T}}\boldsymbol{\Phi}_f(p_f,k)], \tag{4.24}$$

$$\hat{\boldsymbol{\vartheta}}_{2,k} = \hat{\boldsymbol{\vartheta}}_{2,k-1} - \mu_{2,k}\boldsymbol{\Phi}_2(p_2,k)[\boldsymbol{W}(p_2,k) - \boldsymbol{\vartheta}_{2,k-1}^{\mathrm{T}}\boldsymbol{\Phi}_2(p_2,k)]. \tag{4.25}$$

where $\mu_{a,k}$, $\mu_{b,k}$, $\mu_{f,k}$ and $\mu_{2,k}$ are the step-sizes.

However, these recursion formulas in (4.22)–(4.25) cannot be accomplished for the reason that there are unknown internal variables x_{k-i}, $\boldsymbol{x}_{k-i}u_{k-i}$ $(i = 1,2,\ldots,n)$, $-w_{k-j}$ $(j = 1,2,\ldots,n_c)$ lying in the information vectors $\boldsymbol{\phi}_{a,k}$, $\boldsymbol{\phi}_{b,k}$, $\boldsymbol{\phi}_{f,k}$ and $\boldsymbol{\phi}_{2,k}$. In order to accomplish the parameter estimation algorithm, we use the estimated values to replace the unknown variables which can be participated in the recursive computation.

By using the window data with length p, the innovation scalars are expanded into the innovation vectors as follows:

$$\boldsymbol{E}_a(p_a,k) := \boldsymbol{Y}_a(p_a,k) - \boldsymbol{\vartheta}_{a,k-1}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_a(p_a,k) = \begin{bmatrix} y_{a,k} - \hat{\boldsymbol{\vartheta}}_{a,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{a,k} \\ y_{a,k-1} - \hat{\boldsymbol{\vartheta}}_{a,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{a,k-1} \\ \vdots \\ y_{a,k-p_a+1} - \hat{\boldsymbol{\vartheta}}_{a,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{a,k-p_a+1} \end{bmatrix},$$

$$\boldsymbol{E}_b(p_b,k) := \boldsymbol{Y}_b(p_b,k) - \boldsymbol{\vartheta}_{b,k-1}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_b(p_b,k) = \begin{bmatrix} y_{b,k} - \hat{\boldsymbol{\vartheta}}_{b,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{b,k} \\ y_{b,k-1} - \hat{\boldsymbol{\vartheta}}_{b,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{b,k-1} \\ \vdots \\ y_{b,k-p_b+1} - \hat{\boldsymbol{\vartheta}}_{b,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{b,k-p_b+1} \end{bmatrix},$$

$$\boldsymbol{E}_f(p_f,k) := \boldsymbol{Y}_f(p_f,k) - \boldsymbol{\vartheta}_{f,k-1}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_f(p_f,k) = \begin{bmatrix} y_{f,k} - \hat{\boldsymbol{\vartheta}}_{f,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{f,k} \\ y_{f,k-1} - \hat{\boldsymbol{\vartheta}}_{f,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{f,k-1} \\ \vdots \\ y_{f,k-p_f+1} - \hat{\boldsymbol{\vartheta}}_{f,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{f,k-p_f+1} \end{bmatrix},$$

$$\boldsymbol{E}_2(p_2,k) := \boldsymbol{W}(p_2,k) - \boldsymbol{\vartheta}_{2,k-1}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_2(p_2,k) = \begin{bmatrix} w_k - \hat{\boldsymbol{\vartheta}}_{2,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{2,k} \\ w_{k-1} - \hat{\boldsymbol{\vartheta}}_{2,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{2,k-1} \\ \vdots \\ w_{k-p_2+1} - \hat{\boldsymbol{\vartheta}}_{2,k-1}^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{2,k-p_2+1} \end{bmatrix}.$$

Let $\hat{\boldsymbol{\vartheta}}_{a,k}$, $\hat{\boldsymbol{\vartheta}}_{b,k}$, $\hat{\boldsymbol{\vartheta}}_{f,k}$, $\hat{\boldsymbol{\vartheta}}_{2,k}$ denote the estimates of $\hat{\boldsymbol{\vartheta}}_a$, $\hat{\boldsymbol{\vartheta}}_b$, $\hat{\boldsymbol{\vartheta}}_f$, $\hat{\boldsymbol{\vartheta}}_2$ at time k, respectively. Finally, we propose the H-MISG algorithm for estimating the parameters of the nonlinear state-space system, as follows:

$$\hat{\boldsymbol{\vartheta}}_{a,k} = \hat{\boldsymbol{\vartheta}}_{a,k-1} + \frac{\hat{\boldsymbol{\Phi}}_a(p_a,k)}{r_{a,k}}\boldsymbol{E}_a(p_a,k), \tag{4.26}$$

$$\boldsymbol{E}_a(p_a,k) = \boldsymbol{Y}_a(p_a,k) - \hat{\boldsymbol{\vartheta}}_{a,k-1}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_a(p_a,k), \tag{4.27}$$

$$\boldsymbol{Y}_a(p_a,k) = [y_{a,k}, y_{a,k-1}, \ldots, y_{a,k-p_a+1}]^{\mathrm{T}}, \tag{4.28}$$

$$r_{a,k} = r_{a,k-1} + \|\hat{\boldsymbol{\Phi}}_a(p_a,k)\|^2,\ r_{a,0} = 1, \tag{4.29}$$

$$\hat{\boldsymbol{\Phi}}_a(p_a,k) = [\hat{\boldsymbol{\phi}}_{a,k}, \hat{\boldsymbol{\phi}}_{a,k-1}, \ldots, \hat{\boldsymbol{\phi}}_{a,k-p_a+1}], \tag{4.30}$$

$$\hat{\boldsymbol{\phi}}_{a,k} = [\hat{x}_{k-1}, \hat{x}_{k-2}, \ldots, \hat{x}_{k-n}]^{\mathrm{T}}, \tag{4.31}$$

$$\hat{\boldsymbol{\vartheta}}_a = [\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_n]^{\mathrm{T}} \in \mathbb{R}^n, \tag{4.32}$$

$$\hat{\boldsymbol{\vartheta}}_{b,k} = \hat{\boldsymbol{\vartheta}}_{b,k-1} + \frac{\hat{\boldsymbol{\Phi}}_b(p_b,k)}{r_{b,k}}\boldsymbol{E}_b(p_b,k), \tag{4.33}$$

$$\boldsymbol{E}_b(p_b,k) = \boldsymbol{Y}_b(p_b,k) - \hat{\boldsymbol{\vartheta}}_{b,k-1}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_b(p_b,k), \tag{4.34}$$

$$\boldsymbol{Y}_b(p_b,k) = [y_{b,k}, y_{b,k-1}, \ldots, y_{b,k-p_b+1}]^{\mathrm{T}}, \tag{4.35}$$

$$r_{b,k} = r_{b,k-1} + \|\hat{\boldsymbol{\Phi}}_b(p_b,k)\|^2,\ r_{b,0} = 1, \tag{4.36}$$

$$\hat{\boldsymbol{\Phi}}_b(p_b,k) = [\hat{\boldsymbol{\phi}}_{b,k}, \hat{\boldsymbol{\phi}}_{b,k-1}, \ldots, \hat{\boldsymbol{\phi}}_{b,k-p_b+1}], \tag{4.37}$$

$$\hat{\boldsymbol{\phi}}_{b,k} = [\hat{\boldsymbol{x}}_{k-1}u_{k-1}, \ldots, \hat{\boldsymbol{x}}_{k-n+1}u_{k-n+1}, \hat{\boldsymbol{x}}_{k-n}u_{k-n}]^{\mathrm{T}}, \tag{4.38}$$

$$\hat{\boldsymbol{\vartheta}}_b = [\hat{\boldsymbol{b}}_1, \hat{\boldsymbol{b}}_2, \ldots, \hat{\boldsymbol{b}}_n]^{\mathrm{T}} \in \mathbb{R}^{n^2}, \tag{4.39}$$

$$\hat{\boldsymbol{\vartheta}}_{f,k} = \hat{\boldsymbol{\vartheta}}_{f,k-1} + \frac{\hat{\boldsymbol{\Phi}}_f(p_f,k)}{r_{f,k}}\boldsymbol{E}_f(p_f,k), \tag{4.40}$$

$$\boldsymbol{E}_f(p_f,k) = \boldsymbol{Y}_f(p_f,k) - \hat{\boldsymbol{\vartheta}}_{f,k-1}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_f(p_f,k), \tag{4.41}$$

$$\boldsymbol{Y}_f(p_f,k) = [y_{f,k}, y_{f,k-1}, \ldots, y_{f,k-p_f+1}]^{\mathrm{T}}, \tag{4.42}$$

$$r_{f,k} = r_{f,k-1} + \|\hat{\boldsymbol{\Phi}}_f(p_f,k)\|^2,\ r_{f,0} = 1, \tag{4.43}$$

$$\hat{\boldsymbol{\Phi}}_f(p_f,k) = [\hat{\boldsymbol{\phi}}_{f,k}, \hat{\boldsymbol{\phi}}_{f,k-1}, \ldots, \hat{\boldsymbol{\phi}}_{f,k-p_f+1}], \tag{4.44}$$

$$\hat{\boldsymbol{\phi}}_{f,k} = [u_{k-1}, u_{k-2}, \ldots, u_{k-n}]^{\mathrm{T}}, \tag{4.45}$$

$$\hat{\boldsymbol{\vartheta}}_f = [\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_n]^{\mathrm{T}} \in \mathbb{R}^n, \tag{4.46}$$

$$\hat{\boldsymbol{\vartheta}}_{2,k} = \hat{\boldsymbol{\vartheta}}_{2,k-1} + \frac{\hat{\boldsymbol{\Phi}}_2(p_2,k)}{r_{2,k}}\boldsymbol{E}_2(p_2,k), \tag{4.47}$$

$$\boldsymbol{E}_2(p_2,k) = \boldsymbol{W}(p_2,k) - \hat{\boldsymbol{\vartheta}}_{2,k-1}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_2(p_2,k), \tag{4.48}$$

$$\boldsymbol{W}(p_2,k) = [w_k, w_{k-1}, \ldots, w_{k-p_2+1}]^{\mathrm{T}}, \tag{4.49}$$

$$r_{2,k} = r_{2,k-1} + \|\hat{\boldsymbol{\Phi}}_2(p_2,k)\|^2,\ r_{2,0} = 1, \tag{4.50}$$

$$\hat{\boldsymbol{\Phi}}_2(p_2,k) = [\hat{\boldsymbol{\phi}}_{2,k}, \hat{\boldsymbol{\phi}}_{2,k-1}, \ldots, \hat{\boldsymbol{\phi}}_{2,k-p_2+1}], \tag{4.51}$$

$$\hat{\boldsymbol{\phi}}_{2,k} = [-\hat{w}_{k-1}, -\hat{w}_{k-2}, \ldots, -\hat{w}_{k-n_c}, v_{k-1}, \ldots, v_{k-n_d}]^{\mathrm{T}}, \tag{4.52}$$

$$\hat{\boldsymbol{\vartheta}}_2 = [\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_{n_c}, \hat{d}_1, \hat{d}_2, \ldots, \hat{d}_{n_d}]^{\mathrm{T}} \in \mathbb{R}^{n_c+n_d}. \tag{4.53}$$

The proposed H-MISG parameter estimation algorithm for nonlinear state space systems in this article can be applied to fields such as chemical process control systems, transportation communication

systems and so on. The computational steps of the H-MISG algorithm in (4.26) to (4.53) for estimating the parameters of the nonlinear state space systems are illustrated as follows.

Algorithm 4.1:

1. To initiate: Preset the recursion length L, preset the innovation length p, set the initial values $r_{a,0}=1$, $r_{b,0}=1$, $r_{f,0}=1$ and $r_{2,0}=1$, $\hat{\boldsymbol{\vartheta}}_{a,0}=\mathbf{1}_n/p_0$, $\hat{\boldsymbol{\vartheta}}_{b,0}=\mathbf{1}_n/p_0$, $\hat{\boldsymbol{\vartheta}}_{f,0}=\mathbf{1}_n/p_0$, $\hat{\boldsymbol{\vartheta}}_{2,0}=\mathbf{1}_n/p_0$, $P_0=10^6$, $\mathbf{1}_n\in\mathbb{R}^n$ with element 1. Set an estimation error ε.
2. Collect the measured data u_k and y_k and form the stacked output vectors $\boldsymbol{Y}_a(p_a,k)$, $\boldsymbol{Y}_b(p_b,k)$, $\boldsymbol{Y}_f(p_f,k)$ and $\boldsymbol{W}(p_2,k)$ using (4.28), (4.35), (4.42) and (4.49), respectively .
3. Compute $\hat{\boldsymbol{\phi}}_{a,k}$, $\hat{\boldsymbol{\phi}}_{b,k}$, $\hat{\boldsymbol{\phi}}_{f,k}$ and $\hat{\boldsymbol{\phi}}_{2,k}$ using (4.31), (4.38), (4.45) and (4.52), respectively.
4. Form the information matrices $\hat{\boldsymbol{\Phi}}_a(p_a,k)$, $\hat{\boldsymbol{\Phi}}_b(p_b,k)$, $\hat{\boldsymbol{\Phi}}_f(p_f,k)$ and $\hat{\boldsymbol{\Phi}}_2(p_2,k)$ using (4.30), (4.37), (4.44) and (4.51), respectively.
5. Compute the innovation vectors $\boldsymbol{E}_a(p_a,k)$, $\boldsymbol{E}_b(p_b,k)$, $\boldsymbol{E}_f(p_f,k)$ and $\boldsymbol{E}_2(p_2,k)$ using (4.27), (4.34), (4.41) and (4.48), respectively.
6. Compute $r_{a,k}$, $r_{b,k}$, $r_{f,k}$ and $r_{2,k}$ using (4.29), (4.36), (4.43) and (4.50), respectively.
7. Update the parameter estimates $\hat{\boldsymbol{\vartheta}}_{a,k}$, $\hat{\boldsymbol{\vartheta}}_{b,k}$, $\hat{\boldsymbol{\vartheta}}_{f,k}$ and $\hat{\boldsymbol{\vartheta}}_{2,k}$ using (4.26), (4.33), (4.40) and (4.47), respectively. If $\|\hat{\boldsymbol{\vartheta}}_{a,k}-\hat{\boldsymbol{\vartheta}}_{a,k-1}\|>\varepsilon$, $\|\hat{\boldsymbol{\vartheta}}_{b,k}-\hat{\boldsymbol{\vartheta}}_{b,k-1}\|>\varepsilon$, $\|\hat{\boldsymbol{\vartheta}}_{f,k}-\hat{\boldsymbol{\vartheta}}_{f,k-1}\|>\varepsilon$ and $\|\hat{\boldsymbol{\vartheta}}_{2,k}-\hat{\boldsymbol{\vartheta}}_{2,k-1}\|>\varepsilon$, then $k:=k+1$ and go to Step 2, else terminate the recursion procedure and obtain the parameter estimates.

Remark 4.1: The proposed H-MISG identification algorithm uses the dynamical batch data with the window width p to calculate the parameter estimates of the nonlinear state space system. Compared with the H-SG identification algorithm, the H-MISG identification algorithm introduces more data to the recursive estimation and can obtain higher parameter estimation accuracy [42].

5. The state estimator for bilinear systems

It is generally appreciated that the Kalman filtering is considered to be one of the most popular state estimation algorithms for linear state space systems. While, the state estimation of nonlinear state space systems is much more complicated. Aiming at this difficulty, some algorithms are proposed, comprising the modified Kalman filter, the particle filter and the extended Kalman filter. In this section, we transform the bilinear state space model into a linear time varying model and design a Kalman state estimator with time variant gains to obtain the system states.

Define $\boldsymbol{A}_{u,k}:=\boldsymbol{A}+\boldsymbol{B}u_k$. By rearranging the bilinear state space model with moving average noise in (2.1)–(2.2) yields:

$$\begin{aligned}\boldsymbol{x}_{k+1}&=\boldsymbol{A}_{u,k}\boldsymbol{x}_k+\boldsymbol{f}u_k,\\ y_k&=\boldsymbol{h}\boldsymbol{x}_k+\frac{D(z)}{C(z)}v_k,\end{aligned}$$

which may be viewed as a "linear time varying state space model". If the parameter matrix $\boldsymbol{A}_{u,k}$ and the vector $\boldsymbol{f}$ are available, and the input output data u_k and y_k are known, then the state estimates are acquired on the basis of the Kalman filtering:

$$\hat{\boldsymbol{x}}_{k+1}=\boldsymbol{A}_{u,k}\hat{\boldsymbol{x}}_k+\boldsymbol{f}u_k+\boldsymbol{L}_k[y_k-\boldsymbol{h}\hat{\boldsymbol{x}}_k-\hat{\boldsymbol{\vartheta}}_2^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{2,k}],\tag{5.54}$$

$$\boldsymbol{L}_k=\boldsymbol{A}_{u,k}\boldsymbol{P}_k\boldsymbol{h}^{\mathrm{T}}[\sigma^2+\boldsymbol{h}\boldsymbol{P}_k\boldsymbol{h}^{\mathrm{T}}]^{-1},\tag{5.55}$$

$$\boldsymbol{P}_{k+1}=\boldsymbol{A}_{u,k}\boldsymbol{P}_k\boldsymbol{A}_{u,k}^{\mathrm{T}}-\boldsymbol{L}_k\boldsymbol{h}\boldsymbol{P}_k\boldsymbol{A}_{u,k}^{\mathrm{T}},\quad \boldsymbol{P}(1)=\boldsymbol{I}_n,\tag{5.56}$$

where $\hat{\boldsymbol{x}}_k:=[\hat{x}_{1,k},\hat{x}_{2,k},\ldots,\hat{x}_{n,k}]^{\mathrm{T}}\in\mathbb{R}^n$ are the state estimates, $\boldsymbol{L}_k\in\mathbb{R}^n$ is the optimal gain vector, $\boldsymbol{P}_{k+1}\in\mathbb{R}^{n\times n}$ is the state estimation error covariance matrix.

The current difficulty is that the parameter matrix $\boldsymbol{A}_{u,k}$ and vector $\boldsymbol{b}$ are unavailable. Consequently, the state estimation in (5.54)–(5.56) cannot be achieved. The solution is that using the parameter

estimates $\hat{a}_{i,k}$, $\hat{\boldsymbol{b}}_{i,k}$ and $\hat{f}_{i,k}$ in $\hat{\boldsymbol{\vartheta}}_{a,k}$, $\hat{\boldsymbol{\vartheta}}_{b,k}$, $\hat{\boldsymbol{\vartheta}}_{f,k}$ and $\hat{\boldsymbol{\vartheta}}_{2,k}$ constructs the estimates $\hat{\boldsymbol{A}}_k$, $\hat{\boldsymbol{B}}_k$ and $\hat{\boldsymbol{f}}_k$ of $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{f}$ as

$$\hat{\boldsymbol{A}}_k = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \hat{a}_{n,k} & \hat{a}_{n-1,k} & \hat{a}_{n-2,k} & \cdots & \hat{a}_{1,k} \end{bmatrix},$$

$$\hat{\boldsymbol{B}}_k = \begin{bmatrix} \hat{\boldsymbol{b}}_{1,k} \\ \hat{\boldsymbol{b}}_{2,k} \\ \vdots \\ \hat{\boldsymbol{b}}_{n,k} \end{bmatrix}, \quad \hat{\boldsymbol{f}}_k = \begin{bmatrix} \hat{f}_{1,k} \\ \hat{f}_{2,k} \\ \vdots \\ \hat{f}_{n,k} \end{bmatrix}.$$

Then, the estimate $\hat{\boldsymbol{A}}_{u,k}$ of $\boldsymbol{A}_u$ can be acquired by $\hat{\boldsymbol{A}}_{u,k} := \hat{\boldsymbol{A}}_k + \hat{\boldsymbol{B}}_k u_k$. Replacing $\boldsymbol{A}_u$ and $\boldsymbol{f}$ in (5.54)–(5.56) with their estimates $\hat{\boldsymbol{A}}_{u,k}$ and $\hat{\boldsymbol{f}}_k$, and replacing the unavailable information vector $\boldsymbol{\phi}_2$, the parameter vector $\boldsymbol{\vartheta}_2$ in (5.54)–(5.56) with their estimates $\hat{\boldsymbol{\phi}}_{2,k}$ and $\hat{\boldsymbol{\vartheta}}_{2,k}$ yield:

$$\hat{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{A}}_{u,k}\hat{\boldsymbol{x}}_k + \hat{\boldsymbol{f}}_k u_k + \boldsymbol{L}_k[y_k - \boldsymbol{h}\hat{\boldsymbol{x}}_k - \hat{\boldsymbol{\vartheta}}_2^{\mathrm{T}}\hat{\boldsymbol{\phi}}_{2,k}], \tag{5.57}$$

$$\boldsymbol{L}_k = \hat{\boldsymbol{A}}_{u,k}\boldsymbol{P}_k\boldsymbol{h}^{\mathrm{T}}[\sigma^2 + \boldsymbol{h}\boldsymbol{P}_k\boldsymbol{h}^{\mathrm{T}}]^{-1}, \tag{5.58}$$

$$\boldsymbol{P}_{k+1} = \hat{\boldsymbol{A}}_{u,k}\boldsymbol{P}_k\hat{\boldsymbol{A}}_{u,k}^{\mathrm{T}} - \boldsymbol{L}_k\boldsymbol{h}\boldsymbol{P}_k\hat{\boldsymbol{A}}_{u,k}^{\mathrm{T}}. \tag{5.59}$$

Equations (5.57)–(5.59) form the Kalman state estimator for bilinear systems to calculate the estimate $\hat{\boldsymbol{x}}_k$ of the state vector $\boldsymbol{x}_k$. The methods proposed in this paper can combine other identification approaches [43, 44, 45, 46, 47, 48] to study the parameter estimation problems of different systems with colored noises [49, 50, 51, 52, 53] and can be applied to other literatures [54, 55, 56] such as signal modeling and communication networked systems.

The system states and parameters can be interactively estimated through the bilinear state observer-based H-MISG identification algorithm.

6. Convergence analysis

In order to analyze the performance of the proposed method from the view of theory. In this section, the convergence of the proposed H-MISG parameter estimation algorithm is analyzed.

Lemma 6.1. Assume that the nonnegative sequences S_k, ζ_k, and ξ_k meet the following inequality

$$S_k \leqslant S_{k-1} + \zeta_k - \xi_k,$$

and $\sum_{k=1}^{\infty}\zeta_k < \infty$, then $\sum_{k=1}^{\infty}\xi_k < \infty$ and $\xi_k \to 0$, S_k is bounded.

The proof is easy and omitted here.

In order to analyze the convergence of the derived H-MISG algorithm, the following assumption is imposed for the nonlinear state-space system in (2.1) and (2.2).

Assumption 6.1. The disturbance v_k is a white noise sequence with zero mean and variance σ^2, that is, $\mathrm{E}[v_k] = 0$, $\mathrm{E}[v_k^2] = \sigma^2$, $\mathrm{E}[v_k v_i] = 0$, $k \neq i$. If $r_{a,k}$, $r_{b,k}$, $r_{f,k}$, $r_{2,k}$ in the H-MISG algorithm in (4.26) to (4.53) satisfies $r_{a,k} \to \infty$, $r_{b,k} \to \infty$, $r_{f,k} \to \infty$, $r_{2,k} \to \infty$ and $\frac{\|\hat{\boldsymbol{\Phi}}_a(p_a,k)\|^2}{r_{a,k}} \to 0$, $\frac{\|\hat{\boldsymbol{\Phi}}_b(p_b,k)\|^2}{r_{b,k}} \to 0$, $\frac{\|\hat{\boldsymbol{\Phi}}_f(p_f,k)\|^2}{r_{f,k}} \to 0$, $\frac{\|\hat{\boldsymbol{\Phi}}_2(p_2,k)\|^2}{r_{2,k}} \to 0$. This shows that the observation data satisfy the persistent excitation conditions. There exists an integer N and the positive constants λ_a, λ_b, λ_f and λ_2 independent of k such that the following strong excitation (SPE) condition holds:

$$\sum_{j_1=0}^{N-1}\frac{\hat{\boldsymbol{\Phi}}_a(p_a, k+j_1)\hat{\boldsymbol{\Phi}}_a^{\mathrm{T}}(p_a, k+j_1)}{r_{a,k+j_1}} \geqslant \lambda_a\boldsymbol{I}, \ a.s.$$

$$\sum_{j_1=0}^{N-1}\frac{\hat{\boldsymbol{\Phi}}_b(p_b,k+j_1)\hat{\boldsymbol{\Phi}}_b^{\mathrm{T}}(p_b,k+j_1)}{r_{b,k+j_1}}\geqslant\lambda_b\boldsymbol{I},\ a.s.$$
$$\sum_{j_1=0}^{N-1}\frac{\hat{\boldsymbol{\Phi}}_f(p_f,k+j_1)\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(p_f,k+j_1)}{r_{f,k+j_1}}\geqslant\lambda_f\boldsymbol{I},\ a.s.$$
$$\sum_{j_2=0}^{N-1}\frac{\hat{\boldsymbol{\Phi}}_2(p_2,k+j_2)\hat{\boldsymbol{\Phi}}_2^{\mathrm{T}}(p_2,k+j_2)}{r_{2,k+j_2}}\geqslant\lambda_2\boldsymbol{I},\ a.s.$$

Theorem 6.1. For the nonlinear state-space system in (2.1) and (2.2) and the derived H-MISG algorithm in (4.26) to (4.53), under the above Assumption, the parameter estimation error given by the H-MISG algorithm converges to zero.

Proof. Refer to the method in Reference [17, 57, 58] and define the parameter estimation vectors $\tilde{\boldsymbol{\vartheta}}_{a,k}:=\hat{\boldsymbol{\vartheta}}_{a,k}-\boldsymbol{\vartheta}_a$, $\tilde{\boldsymbol{\vartheta}}_{b,k}:=\hat{\boldsymbol{\vartheta}}_{b,k}-\boldsymbol{\vartheta}_b$, $\tilde{\boldsymbol{\vartheta}}_{f,k}:=\hat{\boldsymbol{\vartheta}}_{f,k}-\boldsymbol{\vartheta}_f$, $\tilde{\boldsymbol{\vartheta}}_{2,k}:=\hat{\boldsymbol{\vartheta}}_{2,k}-\boldsymbol{\vartheta}_2$. Subtracting $\boldsymbol{\vartheta}_a$, $\boldsymbol{\vartheta}_b$, $\boldsymbol{\vartheta}_f$, $\boldsymbol{\vartheta}_2$ from both sides of Equations (4.26), (4.33), (4.40), (4.47) respectively gives

$$\begin{aligned}\tilde{\boldsymbol{\vartheta}}_{a,k}&=\hat{\boldsymbol{\vartheta}}_{a,k-1}+\frac{\hat{\boldsymbol{\Phi}}_a(p_a,k)}{r_{a,k}}\boldsymbol{E}_a(p_a,k)-\boldsymbol{\vartheta}_a\\&=:\hat{\boldsymbol{\vartheta}}_{a,k-1}+\frac{\hat{\boldsymbol{\Phi}}_a(p_a,k)}{r_{a,k}}[-\tilde{\boldsymbol{Y}}_a(p_a,k)+\Delta(p_a,k)+\boldsymbol{V}(p_a,k)],\end{aligned}\tag{6.60}$$

$$\begin{aligned}\tilde{\boldsymbol{\vartheta}}_{b,k}&=\hat{\boldsymbol{\vartheta}}_{b,k-1}+\frac{\hat{\boldsymbol{\Phi}}_b(p_b,k)}{r_{b,k}}\boldsymbol{E}_b(p_b,k)-\boldsymbol{\vartheta}_b\\&=:\hat{\boldsymbol{\vartheta}}_{b,k-1}+\frac{\hat{\boldsymbol{\Phi}}_b(p_b,k)}{r_{b,k}}[-\tilde{\boldsymbol{Y}}_b(p_b,k)+\Delta(p_b,k)+\boldsymbol{V}(p_b,k)],\end{aligned}\tag{6.61}$$

$$\begin{aligned}\tilde{\boldsymbol{\vartheta}}_{f,k}&=\hat{\boldsymbol{\vartheta}}_{f,k-1}+\frac{\hat{\boldsymbol{\Phi}}_f(p_f,k)}{r_{f,k}}\boldsymbol{E}_f(p_f,k)-\boldsymbol{\vartheta}_f\\&=:\hat{\boldsymbol{\vartheta}}_{f,k-1}+\frac{\hat{\boldsymbol{\Phi}}_f(p_f,k)}{r_{f,k}}[-\tilde{\boldsymbol{Y}}_f(p_f,k)+\Delta(p_f,k)+\boldsymbol{V}(p_f,k)],\end{aligned}\tag{6.62}$$

$$\begin{aligned}\tilde{\boldsymbol{\vartheta}}_{2,k}&=\hat{\boldsymbol{\vartheta}}_{2,k-1}+\frac{\hat{\boldsymbol{\Phi}}_2(p_2,k)}{r_{2,k}}\boldsymbol{E}_2(p_2,k)-\boldsymbol{\vartheta}_2\\&=:\hat{\boldsymbol{\vartheta}}_{2,k-1}+\frac{\hat{\boldsymbol{\Phi}}_2(p_2,k)}{r_{2,k}}[-\tilde{\boldsymbol{W}}(p_2,k)+\Delta(p_2,k)+\boldsymbol{V}_2(p_2,k)].\end{aligned}\tag{6.63}$$

where

$$\begin{aligned}&\tilde{\boldsymbol{Y}}_a(p_a,k):=\hat{\boldsymbol{\Phi}}_a^{\mathrm{T}}(p_a,k)\tilde{\boldsymbol{\vartheta}}_{a,k-1},\quad \Delta(p_a,k):=[\boldsymbol{\Phi}_a(p_a,k)-\hat{\boldsymbol{\Phi}}_a(p_a,k)]^{\mathrm{T}}\boldsymbol{\vartheta}_a,\\&\tilde{\boldsymbol{Y}}_b(p_b,k):=\hat{\boldsymbol{\Phi}}_b^{\mathrm{T}}(p_b,k)\tilde{\boldsymbol{\vartheta}}_{b,k-1},\quad \Delta(p_b,k):=[\boldsymbol{\Phi}_b(p_b,k)-\hat{\boldsymbol{\Phi}}_b(p_b,k)]^{\mathrm{T}}\boldsymbol{\vartheta}_b,\\&\tilde{\boldsymbol{Y}}_f(p_f,k):=\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(p_f,k)\tilde{\boldsymbol{\vartheta}}_{f,k-1},\quad \Delta(p_f,k):=[\boldsymbol{\Phi}_f(p_f,k)-\hat{\boldsymbol{\Phi}}_f(p_f,k)]^{\mathrm{T}}\boldsymbol{\vartheta}_f,\\&\tilde{\boldsymbol{W}}(p_2,k):=\hat{\boldsymbol{\Phi}}_2^{\mathrm{T}}(p_2,k)\tilde{\boldsymbol{\vartheta}}_{2,k-1},\quad \Delta(p_2,k):=[\boldsymbol{\Phi}_2(p_2,k)-\hat{\boldsymbol{\Phi}}_2(p_2,k)]^{\mathrm{T}}\boldsymbol{\vartheta}_2.\end{aligned}$$

Taking the norm of (6.60) to (6.63) yields

$$\begin{aligned}\|\tilde{\boldsymbol{\vartheta}}_{a,k}\|^2&=\|\tilde{\boldsymbol{\vartheta}}_{a,k-1}+\frac{\hat{\boldsymbol{\Phi}}_a(p_a,k)}{r_{a,k}}[-\tilde{\boldsymbol{Y}}_a(p_a,k)+\Delta(p_a,k)+\boldsymbol{V}(p_a,k)]\|^2\\&=\|\tilde{\boldsymbol{\vartheta}}_{a,k-1}\|^2-\frac{r_{a,k}+r_{a,k-1}}{r_{a,k}^2}\|\tilde{\boldsymbol{Y}}_a(p_a,k)\|^2+\frac{2r_{a,k-1}}{r_{a,k}^2}\tilde{\boldsymbol{Y}}_a^{\mathrm{T}}(p_a,k)[\Delta(p_a,k)+\boldsymbol{V}(p_a,k)]\\&\quad+\frac{\|\hat{\boldsymbol{\Phi}}_a(p_a,k)\|^2}{r_{a,k}^2}[\|\Delta(p_a,k)\|^2+\|\boldsymbol{V}(p_a,k)\|^2+2\boldsymbol{V}^{\mathrm{T}}(p_a,k)\Delta(p_a,k)]\\&\leqslant\|\tilde{\boldsymbol{\vartheta}}_{a,k-1}\|^2-\frac{1}{r_{a,k}}\|\tilde{\boldsymbol{Y}}_a(p_a,k)\|^2+\frac{2r_{a,k-1}}{r_{a,k}^2}\tilde{\boldsymbol{Y}}_a^{\mathrm{T}}(p_a,k)[\Delta(p_a,k)+\boldsymbol{V}(p_a,k)]\end{aligned}$$

$$+\frac{\|\hat{\boldsymbol{\Phi}}_a(p_a,k)\|^2}{r_{a,k}^2}[\|\Delta(p_a,k)\|^2+\|\boldsymbol{V}(p_a,k)\|^2+2\boldsymbol{V}^{\mathrm{T}}(p_a,k)\Delta(p_a,k)], \tag{6.64}$$

$$\|\tilde{\boldsymbol{\vartheta}}_{b,k}\|^2 \leqslant \|\tilde{\boldsymbol{\vartheta}}_{b,k-1}\|^2-\frac{1}{r_{b,k}}\|\tilde{\boldsymbol{Y}}_b(p_b,k)\|^2+\frac{2r_{b,k-1}}{r_{b,k}^2}\tilde{\boldsymbol{Y}}_b^{\mathrm{T}}(p_b,k)[\Delta(p_b,k)+\boldsymbol{V}(p_b,k)]$$
$$+\frac{\|\hat{\boldsymbol{\Phi}}_b(p_b,k)\|^2}{r_{b,k}^2}[\|\Delta(p_b,k)\|^2+\|\boldsymbol{V}(p_b,k)\|^2+2\boldsymbol{V}^{\mathrm{T}}(p_b,k)\Delta(p_b,k)],$$

$$\|\tilde{\boldsymbol{\vartheta}}_{f,k}\|^2 \leqslant \|\tilde{\boldsymbol{\vartheta}}_{f,k-1}\|^2-\frac{1}{r_{f,k}}\|\tilde{\boldsymbol{Y}}_f(p_f,k)\|^2+\frac{2r_{f,k-1}}{r_{f,k}^2}\tilde{\boldsymbol{Y}}_f^{\mathrm{T}}(p_f,k)[\Delta(p_f,k)+\boldsymbol{V}(p_f,k)]$$
$$+\frac{\|\hat{\boldsymbol{\Phi}}_f(p_f,k)\|^2}{r_{f,k}^2}[\|\Delta(p_f,k)\|^2+\|\boldsymbol{V}(p_f,k)\|^2+2\boldsymbol{V}^{\mathrm{T}}(p_f,k)\Delta(p_f,k)],$$

$$\|\tilde{\boldsymbol{\vartheta}}_{2,k}\|^2 \leqslant \|\tilde{\boldsymbol{\vartheta}}_{2,k-1}\|^2-\frac{1}{r_{2,k}}\|\tilde{\boldsymbol{W}}(p_2,k)\|^2+\frac{2r_{2,k-1}}{r_{2,k}^2}\tilde{\boldsymbol{W}}^{\mathrm{T}}(p_2,k)[\Delta(p_2,k)+\boldsymbol{V}_2(p_2,k)]$$
$$+\frac{\|\hat{\boldsymbol{\Phi}}_2(p_2,k)\|^2}{r_{2,k}^2}[\|\Delta(p_2,k)\|^2+\|\boldsymbol{V}_2(p_2,k)\|^2+2\boldsymbol{V}_2^{\mathrm{T}}(p_2,k)\Delta(p_2,k)].$$

Assume that v_k is a white noise with zero mean and variance σ^2, $[\boldsymbol{\phi}_{a,k}-\hat{\boldsymbol{\phi}}_{a,k}]^{\mathrm{T}}\boldsymbol{\vartheta}_a$ is bounded with ε_a, which means that $\|\Delta(p_a,k)\|^2 \leqslant p_a\varepsilon_a$, and $\tilde{\boldsymbol{\vartheta}}_{a,k-1}$, $\tilde{\boldsymbol{Y}}_a(p_a,k)$, $\hat{\boldsymbol{\Phi}}_a(p_a,k)$, $r_{a,k}$, and $\Delta(p_a,k)$ are uncorrelated with v_k. Taking the expectation of both sides of (6.64) and using the Assumption give

$$\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{a,k}\|^2] \leqslant \mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{a,k-1}\|^2]-\mathrm{E}\left[\frac{\|\tilde{\boldsymbol{Y}}_a(p_a,k)\|^2}{r_{a,k}}\right]+\mathrm{E}\left[\frac{\|\hat{\boldsymbol{\Phi}}_a(p_a,k)\|^2}{r_{a,k}^2}\right][p_a(\sigma^2+\varepsilon_a)]. \tag{6.65}$$

Similarly,

$$\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{b,k}\|^2] \leqslant \mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{b,k-1}\|^2]-\mathrm{E}\left[\frac{\|\tilde{\boldsymbol{Y}}_b(p_b,k)\|^2}{r_{b,k}}\right]+\mathrm{E}\left[\frac{\|\hat{\boldsymbol{\Phi}}_b(p_b,k)\|^2}{r_{b,k}^2}\right][p_b(\sigma^2+\varepsilon_b)], \tag{6.66}$$

$$\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{f,k}\|^2] \leqslant \mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{f,k-1}\|^2]-\mathrm{E}\left[\frac{\|\tilde{\boldsymbol{Y}}_f(p_f,k)\|^2}{r_{f,k}}\right]+\mathrm{E}\left[\frac{\|\hat{\boldsymbol{\Phi}}_f(p_f,k)\|^2}{r_{f,k}^2}\right][p_f(\sigma^2+\varepsilon_f)], \tag{6.67}$$

$$\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{2,k}\|^2] \leqslant \mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{2,k-1}\|^2]-\mathrm{E}\left[\frac{\|\tilde{\boldsymbol{W}}(p_2,k)\|^2}{r_{2,k}}\right]+\mathrm{E}\left[\frac{\|\hat{\boldsymbol{\Phi}}_2(p_2,k)\|^2}{r_{2,k}^2}\right][p_2(\sigma^2+\varepsilon_2)]. \tag{6.68}$$

Note that the sum of the last term on RHS from $t=1$ to ∞ is finite,

$$\sum_{t=1}^{\infty}\frac{\|\hat{\boldsymbol{\Phi}}_a(p_a,k)\|^2}{r_{a,k}^2}[p_a(\sigma^2+\varepsilon_a)] \leqslant \sum_{t=1}^{\infty}\frac{\|\hat{\boldsymbol{\Phi}}_a(p_a,k)\|^2}{r_{a,k-1}r_{a,k}}[p_a(\sigma^2+\varepsilon_a)]$$
$$=\left[1-\frac{1}{r_a(\infty)}\right][p_a(\sigma^2+\varepsilon_a)]<\infty,$$

$$\sum_{t=1}^{\infty}\frac{\|\hat{\boldsymbol{\Phi}}_b(p_b,k)\|^2}{r_{b,k}^2}[p_b(\sigma^2+\varepsilon_b)] \leqslant \sum_{t=1}^{\infty}\frac{\|\hat{\boldsymbol{\Phi}}_b(p_b,k)\|^2}{r_{b,k-1}r_{b,k}}[p_b(\sigma^2+\varepsilon_b)]$$
$$=\left[1-\frac{1}{r_b(\infty)}\right][p_b(\sigma^2+\varepsilon_b)]<\infty,$$

$$\sum_{t=1}^{\infty}\frac{\|\hat{\boldsymbol{\Phi}}_f(p_f,k)\|^2}{r_{f,k}^2}[p_f(\sigma^2+\varepsilon_f)] \leqslant \sum_{t=1}^{\infty}\frac{\|\hat{\boldsymbol{\Phi}}_f(p_f,k)\|^2}{r_{f,k-1}r_{f,k}}[p_f(\sigma^2+\varepsilon_f)]$$
$$=\left[1-\frac{1}{r_f(\infty)}\right][p_f(\sigma^2+\varepsilon_f)]<\infty,$$

$$\sum_{t=1}^{\infty}\frac{\|\hat{\boldsymbol{\Phi}}_2(p_2,k)\|^2}{r_{2,k}^2}[p_2(\sigma^2+\varepsilon_2)] \leqslant \sum_{t=1}^{\infty}\frac{\|\hat{\boldsymbol{\Phi}}_2(p_2,k)\|^2}{r_{2,k-1}r_{2,k}}[p_2(\sigma^2+\varepsilon_2)]$$

$$= \left[1 - \frac{1}{r_2(\infty)}\right][p_2(\sigma^2 + \varepsilon_2)] < \infty.$$

Applying Lemma 6.1 to (6.65)–(6.68), we can conclude that $\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{a,k}\|^2]$, $\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{b,k}\|^2]$, $\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{f,k}\|^2]$, $\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{2,k}\|^2]$ are bounded, which are denoted by C_a, C_b, C_f, C_2 and $\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{a,k}\|^2] \le C_a$, $\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{b,k}\|^2] \le C_b$, $\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{f,k}\|^2] \le C_f$, $\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{2,k}\|^2] \le C_2$, and

$$\mathrm{E}\left[\sum_{t=1}^{\infty} \frac{\|\tilde{\boldsymbol{Y}}_a(p_a,k)\|^2}{r_{a,k}}\right] < \infty, \tag{6.69}$$

$$\mathrm{E}\left[\sum_{t=1}^{\infty} \frac{\|\tilde{\boldsymbol{Y}}_b(p_b,k)\|^2}{r_{b,k}}\right] < \infty, \tag{6.70}$$

$$\mathrm{E}\left[\sum_{t=1}^{\infty} \frac{\|\tilde{\boldsymbol{Y}}_f(p_f,k)\|^2}{r_{f,k}}\right] < \infty, \tag{6.71}$$

$$\mathrm{E}\left[\sum_{t=1}^{\infty} \frac{\|\tilde{\boldsymbol{W}}(p_2,k)\|^2}{r_{2,k}}\right] < \infty. \tag{6.72}$$

According to the above relation, we can obtain

$$\tilde{\boldsymbol{\vartheta}}_{a,k+j_1} = \tilde{\boldsymbol{\vartheta}}_{a,k} + \sum_{i_1=1}^{j_1} \frac{\hat{\boldsymbol{\Phi}}_a(p_a,k+i_1)}{r_{a,k+i_1}} \boldsymbol{E}_a(p_a,k+i_1),$$

$$\tilde{\boldsymbol{\vartheta}}_{b,k+j_1} = \tilde{\boldsymbol{\vartheta}}_{b,k} + \sum_{i_1=1}^{j_1} \frac{\hat{\boldsymbol{\Phi}}_b(p_b,k+i_1)}{r_{b,k+i_1}} \boldsymbol{E}_b(p_b,k+i_1),$$

$$\tilde{\boldsymbol{\vartheta}}_{f,k+j_1} = \tilde{\boldsymbol{\vartheta}}_{f,k} + \sum_{i_1=1}^{j_1} \frac{\hat{\boldsymbol{\Phi}}_f(p_f,k+i_1)}{r_{f,k+i_1}} \boldsymbol{E}_f(p_f,k+i_1),$$

$$\tilde{\boldsymbol{\vartheta}}_{2,k+j_2} = \tilde{\boldsymbol{\vartheta}}_{2,k} + \sum_{i_2=1}^{j_2} \frac{\hat{\boldsymbol{\Phi}}_2(p_2,k+i_2)}{r_{2,k+i_2}} \boldsymbol{E}_2(p_2,k+i_2).$$

Since $\tilde{\boldsymbol{Y}}_a(p_a,k) = \hat{\boldsymbol{\Phi}}_a^{\mathrm{T}}(p_a,k)\tilde{\boldsymbol{\vartheta}}_{a,k-1}$, $\tilde{\boldsymbol{Y}}_b(p_b,k) = \hat{\boldsymbol{\Phi}}_b^{\mathrm{T}}(p_b,k)\tilde{\boldsymbol{\vartheta}}_{b,k-1}$, $\tilde{\boldsymbol{Y}}_f(p_f,k) = \hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(p_f,k)\tilde{\boldsymbol{\vartheta}}_{f,k-1}$, and $\tilde{\boldsymbol{W}}(p_2,k) = \hat{\boldsymbol{\Phi}}_2^{\mathrm{T}}(p_2,k)\tilde{\boldsymbol{\vartheta}}_{2,k-1}$, and the above equations, we have

$$\tilde{\boldsymbol{\vartheta}}_{a,k+j_1-1} = \tilde{\boldsymbol{\vartheta}}_{a,k} + \sum_{i_1=1}^{j_1-1} \frac{\hat{\boldsymbol{\Phi}}_a(p_a,k+i_1)}{r_{a,k+i_1}} \boldsymbol{E}_a(p_a,k+i_1),$$

$$\tilde{\boldsymbol{\vartheta}}_{b,k+j_1-1} = \tilde{\boldsymbol{\vartheta}}_{b,k} + \sum_{i_1=1}^{j_1-1} \frac{\hat{\boldsymbol{\Phi}}_b(p_b,k+i_1)}{r_{b,k+i_1}} \boldsymbol{E}_b(p_b,k+i_1),$$

$$\tilde{\boldsymbol{\vartheta}}_{f,k+j_1-1} = \tilde{\boldsymbol{\vartheta}}_{f,k} + \sum_{i_1=1}^{j_1-1} \frac{\hat{\boldsymbol{\Phi}}_f(p_f,k+i_1)}{r_{f,k+i_1}} \boldsymbol{E}_f(p_f,k+i_1),$$

$$\tilde{\boldsymbol{\vartheta}}_{2,k+j_2-1} = \tilde{\boldsymbol{\vartheta}}_{2,k} + \sum_{i_2=1}^{j_2-1} \frac{\hat{\boldsymbol{\Phi}}_2(p_2,k+i_2)}{r_{2,k+i_2}} \boldsymbol{E}_2(p_2,k+i_2).$$

Synthesizing the above equations gives

$$\hat{\boldsymbol{\Phi}}_a^{\mathrm{T}}(p_a,k+j_1)\tilde{\boldsymbol{\vartheta}}_{a,k} = \tilde{\boldsymbol{Y}}_a(p_a,k+j_1) - \hat{\boldsymbol{\Phi}}_a^{\mathrm{T}}(p_a,k+j_1) \sum_{i_1=1}^{j_1-1} \frac{\hat{\boldsymbol{\Phi}}_a(p_a,k+i_1)}{r_{a,k+i_1}} \boldsymbol{E}_a(p_a,k+i_1), \tag{6.73}$$

$$\hat{\boldsymbol{\Phi}}_b^{\mathrm{T}}(p_b,k+j_1)\tilde{\boldsymbol{\vartheta}}_{b,k} = \tilde{\boldsymbol{Y}}_b(p_b,k+j_1) - \hat{\boldsymbol{\Phi}}_b^{\mathrm{T}}(p_b,k+j_1) \sum_{i_1=1}^{j_1-1} \frac{\hat{\boldsymbol{\Phi}}_b(p_b,k+i_1)}{r_{b,k+i_1}} \boldsymbol{E}_b(p_b,k+i_1), \tag{6.74}$$

$$\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(p_f,k+j_1)\tilde{\boldsymbol{\vartheta}}_{f,k}=\tilde{\boldsymbol{Y}}_f(p_f,k+j_1)-\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(p_f,k+j_1)\sum_{i_1=1}^{j_1-1}\frac{\hat{\boldsymbol{\Phi}}_f(p_f,k+i_1)}{r_{f,k+i_1}}\boldsymbol{E}_f(p_f,k+i_1), \tag{6.75}$$

$$\hat{\boldsymbol{\Phi}}_2^{\mathrm{T}}(p_2,k+j_2)\tilde{\boldsymbol{\vartheta}}_{2,k}=\tilde{\boldsymbol{Y}}_2(p_2,k+j_2)-\hat{\boldsymbol{\Phi}}_2^{\mathrm{T}}(p_2,k+j_2)\sum_{i_2=1}^{j_2-1}\frac{\hat{\boldsymbol{\Phi}}_2(p_2,k+i_2)}{r_{2,k+i_2}}\boldsymbol{E}_2(p_2,k+i_2). \tag{6.76}$$

Taking the norm of both sides of (6.73)–(6.76) yields

$$\begin{aligned}
&\tilde{\boldsymbol{\vartheta}}_{a,k}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_a(p_a,k+j_1)\hat{\boldsymbol{\Phi}}_a^{\mathrm{T}}(p_a,k+j_1)\tilde{\boldsymbol{\vartheta}}_{a,k}\\
&\leqslant 2\|\tilde{\boldsymbol{Y}}_a(p_a,k+j_1)\|^2+2\|\hat{\boldsymbol{\Phi}}_a(p_a,k+j_1)\|^2\|\sum_{i_1=1}^{j_1-1}\frac{\hat{\boldsymbol{\Phi}}_a(p_a,k+i_1)}{r_{a,k+i_1}}\boldsymbol{E}_a(p_a,k+i_1)\|^2,\\
&\tilde{\boldsymbol{\vartheta}}_{b,k}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_b(p_b,k+j_1)\hat{\boldsymbol{\Phi}}_b^{\mathrm{T}}(p_b,k+j_1)\tilde{\boldsymbol{\vartheta}}_{b,k}\\
&\leqslant 2\|\tilde{\boldsymbol{Y}}_b(p_b,k+j_1)\|^2+2\|\hat{\boldsymbol{\Phi}}_b(p_b,k+j_1)\|^2\|\sum_{i_1=1}^{j_1-1}\frac{\hat{\boldsymbol{\Phi}}_b(p_b,k+i_1)}{r_{b,k+i_1}}\boldsymbol{E}_b(p_b,k+i_1)\|^2,\\
&\tilde{\boldsymbol{\vartheta}}_{f,k}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_f(p_f,k+j_1)\hat{\boldsymbol{\Phi}}_f^{\mathrm{T}}(p_f,k+j_1)\tilde{\boldsymbol{\vartheta}}_{f,k}\\
&\leqslant 2\|\tilde{\boldsymbol{Y}}_f(p_f,k+j_1)\|^2+2\|\hat{\boldsymbol{\Phi}}_f(p_f,k+j_1)\|^2\|\sum_{i_1=1}^{j_1-1}\frac{\hat{\boldsymbol{\Phi}}_f(p_f,k+i_1)}{r_{f,k+i_1}}\boldsymbol{E}_f(p_f,k+i_1)\|^2,\\
&\tilde{\boldsymbol{\vartheta}}_{2,k}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_2(p_2,k+j_2)\hat{\boldsymbol{\Phi}}_2^{\mathrm{T}}(p_2,k+j_2)\tilde{\boldsymbol{\vartheta}}_{2,k}\\
&\leqslant 2\|\tilde{\boldsymbol{W}}(p_2,k+j_2)\|^2+2\|\hat{\boldsymbol{\Phi}}_2(p_2,k+j_2)\|^2\|\sum_{i_2=1}^{j_2-1}\frac{\hat{\boldsymbol{\Phi}}_2(p_2,k+i_2)}{r_{2,k+i_2}}\boldsymbol{E}_2(p_2,k+i_2)\|^2.
\end{aligned}$$

Divide $r_{a,k+j_1}$, $r_{b,k+j_1}$, $r_{f,k+j_1}$, $r_{2,k+j_2}$, both sides of the above equations and sum for $j_1=0$ to $j_1=N-1$, $j_2=0$ to $j_2=N-1$ and use the SPE condition in the Assumption. Then, we have

$$\begin{aligned}
\lambda_a\|\tilde{\boldsymbol{\vartheta}}_{a,k}\|^2&\leqslant\sum_{j_1=1}^{N-1}\frac{\tilde{\boldsymbol{\vartheta}}_{a,k}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_a(p_a,k+j_1)\tilde{\boldsymbol{\vartheta}}_{a,k}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_a(p_a,k+j_1)\tilde{\boldsymbol{\vartheta}}_{a,k}}{r_{a,k+j_1}}\\
&\leqslant\sum_{j_1=1}^{N-1}\left[\frac{2\|\tilde{\boldsymbol{Y}}_a(p_a,k+j_1)\|^2}{r_{a,k+j_1}}+\frac{4\|\hat{\boldsymbol{\Phi}}_a(p_a,k+j_1)\|^2}{r_{a,k+j_1}}\|\tilde{\boldsymbol{\vartheta}}_{a,k+j_1-1}\|^2+\frac{4\|\hat{\boldsymbol{\Phi}}_a(p_a,k+j_1)\|^2}{r_{a,k+j_1}}\|\tilde{\boldsymbol{\vartheta}}_{a,k}\|^2\right],\\
\lambda_b\|\tilde{\boldsymbol{\vartheta}}_{b,k}\|^2&\leqslant\sum_{j_1=1}^{N-1}\left[\frac{2\|\tilde{\boldsymbol{Y}}_b(p_b,k+j_1)\|^2}{r_{b,k+j_1}}+\frac{4\|\hat{\boldsymbol{\Phi}}_b(p_b,k+j_1)\|^2}{r_{b,k+j_1}}\|\tilde{\boldsymbol{\vartheta}}_{b,k+j_1-1}\|^2+\frac{4\|\hat{\boldsymbol{\Phi}}_b(p_b,k+j_1)\|^2}{r_{b,k+j_1}}\|\tilde{\boldsymbol{\vartheta}}_{b,k}\|^2\right],\\
\lambda_f\|\tilde{\boldsymbol{\vartheta}}_{f,k}\|^2&\leqslant\sum_{j_1=1}^{N-1}\left[\frac{2\|\tilde{\boldsymbol{Y}}_f(p_f,k+j_1)\|^2}{r_{f,k+j_1}}+\frac{4\|\hat{\boldsymbol{\Phi}}_f(p_f,k+j_1)\|^2}{r_{f,k+j_1}}\|\tilde{\boldsymbol{\vartheta}}_{f,k+j_1-1}\|^2+\frac{4\|\hat{\boldsymbol{\Phi}}_f(p_f,k+j_1)\|^2}{r_{f,k+j_1}}\|\tilde{\boldsymbol{\vartheta}}_{f,k}\|^2\right],\\
\lambda_2\|\tilde{\boldsymbol{\vartheta}}_{2,k}\|^2&\leqslant\sum_{j_2=1}^{N-1}\left[\frac{2\|\tilde{\boldsymbol{W}}(p_2,k+j_2)\|^2}{r_{2,k+j_2}}+\frac{4\|\hat{\boldsymbol{\Phi}}_2(p_2,k+j_2)\|^2}{r_{2,k+j_2}}\|\tilde{\boldsymbol{\vartheta}}_{2,k+j_2-1}\|^2+\frac{4\|\hat{\boldsymbol{\Phi}}_2(p_2,k+j_2)\|^2}{r_{2,k+j_2}}\|\tilde{\boldsymbol{\vartheta}}_{2,k}\|^2\right].
\end{aligned}$$

Taking the expectation and the limit yields

$$\begin{aligned}
\lim_{t\to\infty}\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{a,k}\|^2]&\leqslant\frac{1}{\lambda_a}\lim_{t\to\infty}\mathrm{E}\Big[\sum_{j_1=1}^{N-1}\frac{2\|\tilde{\boldsymbol{Y}}_a(p_a,k+j_1)\|^2}{r_{a,k+j_1}}+\frac{4\|\hat{\boldsymbol{\Phi}}_1(p_a,k+j_1)\|^2}{r_{a,k+j_1}}\|\tilde{\boldsymbol{\vartheta}}_{a,k+j_1-1}\|^2\\
&\quad+\frac{4\|\hat{\boldsymbol{\Phi}}_a(p_a,k+j_1)\|^2}{r_{a,k+j_1}}\|\tilde{\boldsymbol{\vartheta}}_{a,k}\|^2\Big]\\
&\leqslant\lim_{t\to\infty}\mathrm{E}\left[\frac{2\|\tilde{\boldsymbol{Y}}_a(p_a,k+j_1)\|^2}{r_{a,k+j_1}}+\frac{8C\|\hat{\boldsymbol{\Phi}}_a(p_a,k+j_1)\|^2}{r_{a,k+j_1}}\right],
\end{aligned}$$

$$\lim_{t\to\infty}\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{b,k}\|^2]\leqslant\lim_{t\to\infty}\mathrm{E}\left[\frac{2\|\tilde{\boldsymbol{Y}}_b(p_b,k+j_1)\|^2}{r_{b,k+j_1}}+\frac{8C\|\hat{\boldsymbol{\Phi}}_b(p_b,k+j_1)\|^2}{r_{b,k+j_1}}\right],$$
$$\lim_{t\to\infty}\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{f,k}\|^2]\leqslant\lim_{t\to\infty}\mathrm{E}\left[\frac{2\|\tilde{\boldsymbol{Y}}_f(p_f,k+j_1)\|^2}{r_{f,k+j_1}}+\frac{8C\|\hat{\boldsymbol{\Phi}}_f(p_f,k+j_1)\|^2}{r_{f,k+j_1}}\right],$$
$$\lim_{t\to\infty}\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{2,k}\|^2]\leqslant\lim_{t\to\infty}\mathrm{E}\left[\frac{2\|\tilde{\boldsymbol{W}}(p_2,k+j_2)\|^2}{r_{2,k+j_2}}+\frac{8C\|\hat{\boldsymbol{\Phi}}_2(p_2,k+j_2)\|^2}{r_{2,k+j_2}}\right].$$

According to the Assumption 6.1 and Equations (6.69) to (6.72), we can conclude $\lim_{t\to\infty}\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{a,k}\|^2]=0$, $\lim_{t\to\infty}\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{b,k}\|^2]=0$, $\lim_{t\to\infty}\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{f,k}\|^2]=0$, $\lim_{t\to\infty}\mathrm{E}[\|\tilde{\boldsymbol{\vartheta}}_{2,k}\|^2]=0$. This completes the proof.

The proposed approaches in the paper can combine some mathematical tools and identification methods [59, 60, 61, 62, 63, 64, 65] to study the parameter estimation issues of other linear stochastic systems and nonlinear stochastic systems with different structures and disturbance noises [66, 67, 68, 69, 70, 71] and can be applied to literatures [72, 73, 74, 75, 76, 77, 78] such as paper-making systems, information processing, engineering systems and so on.

7. Examples

Example 7.1: In this section, the simulation example is provided to illustrate the efficiency of the proposed algorithms. Consider the following second-order bilinear state-space model:

$$\begin{aligned}\boldsymbol{x}_{k+1}&=\begin{bmatrix}0&1\\0.07&0.02\end{bmatrix}\boldsymbol{x}_k+\begin{bmatrix}-0.06&0.07\\0.16&-0.07\end{bmatrix}\boldsymbol{x}_k u_k+\begin{bmatrix}-0.70\\1.43\end{bmatrix}u_k,\\ y_k&=[1,\ 0]\boldsymbol{x}_k+0.12w_{k-1}+0.08v_{k-1}+v_k.\end{aligned}$$

The parameter vector to be identified is

$$\begin{aligned}\boldsymbol{\vartheta}&=[a_1,a_2,b_{11},b_{12},b_{21},b_{22},f_1,f_2,c,d]^{\mathrm{T}}\\&=[0.07,0.02,-0.06,0.07,0.16,-0.07,-0.70,1.43,-0.12,0.08]^{\mathrm{T}}.\end{aligned}$$

In simulation, the initial values may be is randomly set [79]. The input $\{u_k\}$ is taken as an independent persistent excitation signal sequence with zero mean and unit variance, and $\{v_k\}$ as a white noise sequence with zero mean and variances $\sigma^2=0.10^2$ and $\sigma^2=0.30^2$, $w_k=0.12w_{k-1}+0.08v_{k-1}+v_k$. The input and output signal vs t are shown in Fig. 1.

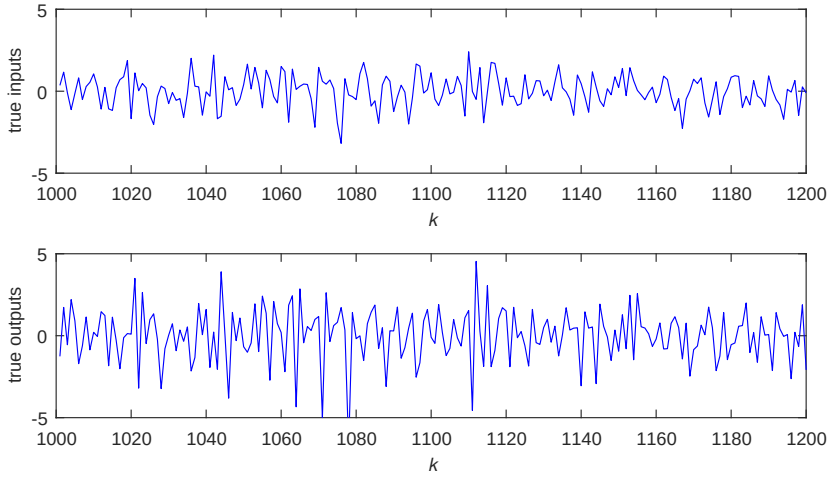

Figure 1: Simulated input–output data

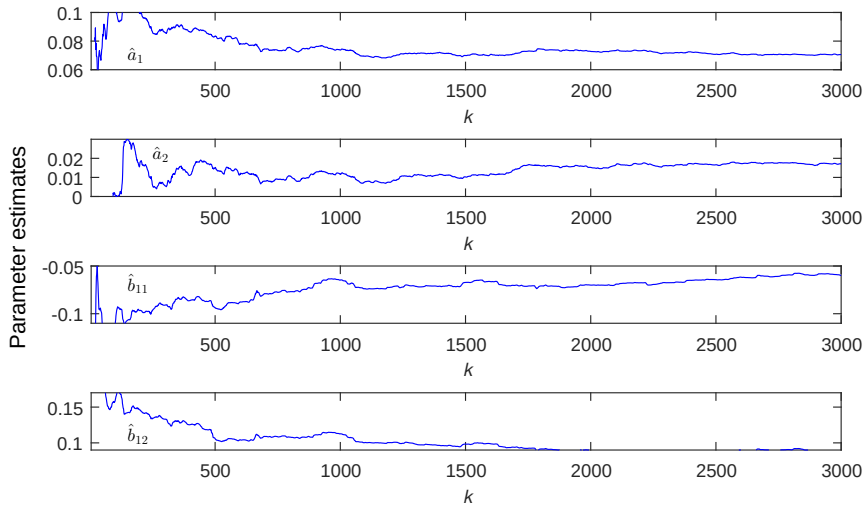

Figure 2: The H-MISG estimates versus k with $\sigma^2 = 0.10^2$

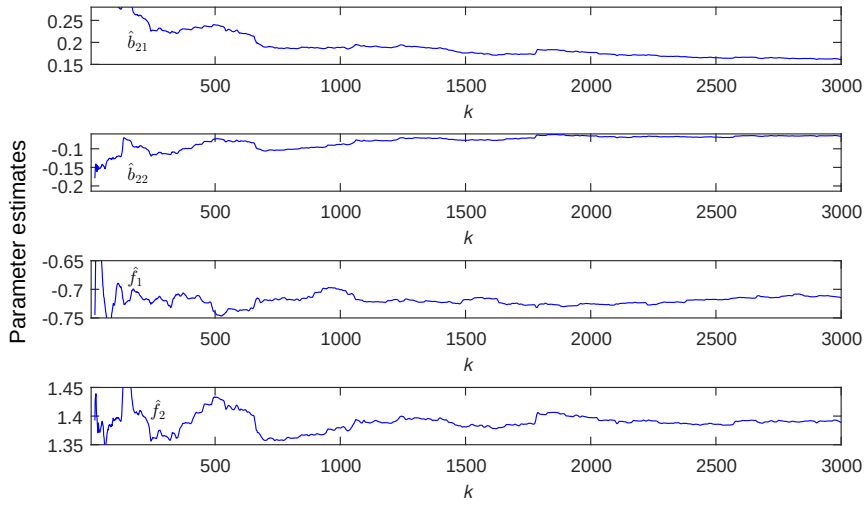

Figure 3: The H-MISG estimates versus k with $\sigma^2 = 0.10^2$

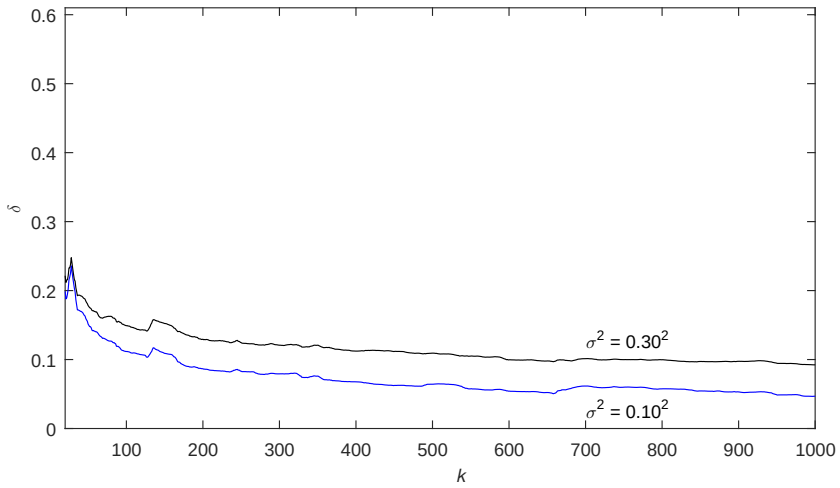

Figure 4: The H-MISG estimation errors δ versus k with $\sigma^2 = 0.10^2$ and $\sigma^2 = 0.30^2$

Table 1: The parameter estimates and errors with $\sigma^2 = 0.10^2$

p	k	100	200	500	1000	2000	3000
1	$a_2 = 0.07000$	-0.03874	-0.02983	-0.01697	-0.01143	-0.00529	-0.00191
	$a_1 = 0.2000$	-0.08393	-0.06923	-0.05593	-0.04931	-0.04183	-0.03822
	$b_{11} = -0.06000$	-0.16508	-0.16704	-0.16936	-0.16788	-0.17115	-0.17196
	$b_{12} = 0.07000$	-0.22008	-0.22229	-0.22384	-0.22531	-0.22954	-0.22887
	$b_{21} = 0.16000$	0.40183	0.40764	0.41847	0.41959	0.42124	0.42159
	$b_{22} = -0.07000$	0.39409	0.41120	0.42619	0.42955	0.42991	0.42956
	$f_1 = -0.70000$	-0.43757	-0.44347	-0.45382	-0.45840	-0.47066	-0.47320
	$f_2 = 1.43000$	0.82878	0.86275	0.90474	0.91906	0.93747	0.94657
	δ (%)	56.45408	55.52851	54.32412	53.80386	53.01530	52.59829
2	$a_2 = 0.07000$	0.07284	0.07716	0.07827	0.07772	0.07920	0.07851
	$a_1 = 0.2000$	-0.03630	-0.02557	-0.01535	-0.01195	-0.00776	-0.00701
	$b_{11} = -0.06000$	-0.17001	-0.16748	-0.17287	-0.17072	-0.17580	-0.17332
	$b_{12} = 0.07000$	-0.15187	-0.15709	-0.15747	-0.15272	-0.15725	-0.15449
	$b_{21} = 0.16000$	0.42306	0.42376	0.41685	0.40146	0.39596	0.38935
	$b_{22} = -0.07000$	0.47303	0.47874	0.46655	0.45267	0.44411	0.43809
	$f_1 = -0.70000$	-0.47864	-0.49345	-0.50881	-0.50848	-0.52650	-0.52706
	$f_2 = 1.43000$	0.94879	0.98670	1.00222	0.99711	1.00676	1.00880
	δ (%)	52.43595	51.16194	49.78256	48.96976	48.03535	47.50741
8	$a_2 = 0.07000$	0.10455	0.09973	0.08342	0.07380	0.07280	0.07052
	$a_1 = 0.2000$	0.00107	0.01782	0.01525	0.01181	0.01524	0.01717
	$b_{11} = -0.06000$	-0.11335	-0.09657	-0.09370	-0.06477	-0.07001	-0.05992
	$b_{12} = 0.07000$	0.15807	0.14580	0.10597	0.11255	0.08974	0.08719
	$b_{21} = 0.16000$	0.28190	0.26207	0.23962	0.18824	0.17705	0.16098
	$b_{22} = -0.07000$	-0.12640	-0.09145	-0.07431	-0.08839	-0.06541	-0.06641
	$f_1 = -0.70000$	-0.72333	-0.71092	-0.74056	-0.70067	-0.72560	-0.71450
	$f_2 = 1.43000$	1.39266	1.41005	1.43234	1.37918	1.39726	1.38895
	δ (%)	11.16175	8.66556	6.42328	4.67386	3.15042	2.92926

Applying the H-SG identification algorithm in (3.10) to (3.21) and the H-MISG identification algorithm in (4.26) to (4.53) to estimate the parameter vector $\boldsymbol{\vartheta}$ of this bilinear system. Figs. 2–3 show the parameter estimates $\hat{a}_{1,k}$, $\hat{a}_{2,k}$, $\hat{b}_{11,k}$, $\hat{b}_{12,k}$, $\hat{b}_{21,k}$, $\hat{b}_{22,k}$, $\hat{f}_{1,k}$, $\hat{f}_{2,k}$, $\hat{c}_k$, $\hat{d}_k$ versus k with $\sigma^2 = 0.10^2$. It can be observed that the parameter estimates reach their true values as k increases.

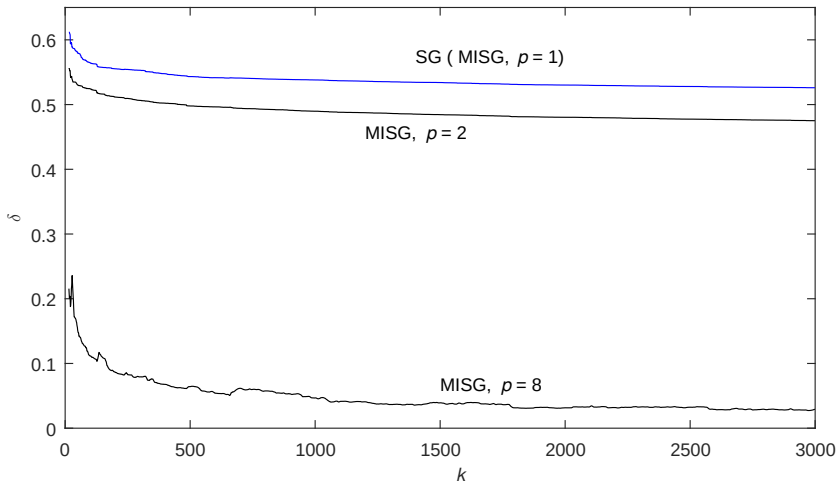

Figure 5: The parameter estimation errors δ versus k with $\sigma^2 = 0.10^2$

Table 2: The parameter estimates and errors with $\sigma^2 = 0.3^2$

p	k	100	200	500	1000	2000	3000
1	$a_2 = 0.07000$	0.06086	0.07631	0.09878	0.10988	0.12121	0.12754
	$a_1 = 0.2000$	0.01213	0.03329	0.05602	0.06741	0.08018	0.08685
	$b_{11} = -0.06000$	-0.15103	-0.15509	-0.15849	-0.15816	-0.16201	-0.16363
	$b_{12} = 0.07000$	-0.22926	-0.23307	-0.23550	-0.23681	-0.24157	-0.24098
	$b_{21} = 0.16000$	0.40575	0.40698	0.41748	0.41865	0.41980	0.42014
	$b_{22} = -0.07000$	0.39671	0.41092	0.42438	0.42727	0.42770	0.42703
	$f_1 = -0.70000$	-0.42036	-0.42878	-0.44117	-0.44628	-0.45943	-0.46267
	$f_2 = 1.43000$	0.84181	0.87031	0.91033	0.92425	0.94208	0.95057
	δ (%)	55.70579	55.00000	54.02931	53.63435	53.01350	52.69224
2	$a_2 = 0.07000$	0.22110	0.23113	0.23865	0.24152	0.24549	0.24651
	$a_1 = 0.2000$	0.09497	0.11063	0.12972	0.13703	0.14402	0.14682
	$b_{11} = -0.06000$	-0.16409	-0.16176	-0.16680	-0.16477	-0.17029	-0.16836
	$b_{12} = 0.07000$	-0.15286	-0.15732	-0.15632	-0.15111	-0.15549	-0.15303
	$b_{21} = 0.16000$	0.42922	0.42550	0.41825	0.40397	0.39869	0.39169
	$b_{22} = -0.07000$	0.46649	0.46800	0.45540	0.44254	0.43466	0.42837
	$f_1 = -0.70000$	-0.47842	-0.49335	-0.50681	-0.50645	-0.52446	-0.52605
	$f_2 = 1.43000$	0.96023	0.98890	1.00317	0.99955	1.01070	1.01211
	δ (%)	52.79990	51.85075	50.79058	50.12199	49.32363	48.85753
8	$a_2 = 0.07000$	0.30275	0.29294	0.28311	0.26852	0.26643	0.26373
	$a_1 = 0.2000$	0.12827	0.14833	0.15975	0.15841	0.16527	0.17024
	$b_{11} = -0.06000$	-0.09519	-0.07462	-0.07708	-0.05004	-0.05716	-0.04766
	$b_{12} = 0.07000$	0.15836	0.15182	0.11504	0.11896	0.09486	0.09225
	$b_{21} = 0.16000$	0.32832	0.29355	0.27141	0.21805	0.20480	0.18532
	$b_{22} = -0.07000$	-0.13995	-0.11417	-0.08772	-0.09785	-0.07227	-0.07463
	$f_1 = -0.70000$	-0.73506	-0.71578	-0.74569	-0.70731	-0.73333	-0.72206
	$f_2 = 1.43000$	1.38266	1.37866	1.41963	1.37098	1.39191	1.38052
	δ (%)	20.80451	19.26749	17.85088	16.32625	15.86152	15.79511

The parameter estimates and errors $\delta = \|\hat{\boldsymbol{\vartheta}}_k - \boldsymbol{\vartheta}\|/\|\boldsymbol{\vartheta}\| \times 100\%$ with different σ are shown in Tabs. 1–2 and Fig. 4 and the parameter estimation errors δ versus k are shown in Fig. 5 with $p = 1$, $p = 2$, $p = 8$, respectively. It can be observed that the parameter estimation errors δ of the H-MISG identification algorithm become smaller with p increasing, and errors δ approach to zero fast if p is large enough as well as the data length tends to infinity.

8. Conclusion

This paper presents a H-MISG identification algorithm for identifying the parameters and states of bilinear systems. Compared with the H-SG identification algorithm, the H-MISG identification algorithm can improve the accuracy of parameter estimation because of the introduction of new information length parameter and the full use of system input and output data. Then, a state estimator is proposed to identify bilinear state-space systems. The combined parameter and state estimation is based on the layered principle. The simulation results show that the proposed H-MISG identification algorithm in this paper based on the state estimator works well and can reduce the influence of colored noise on parameter estimation. The proposed identification algorithm for bilinear stochastic systems in this paper can be extended to MIMO (multi-input and multi-output) bilinear systems and can be applied to other control and schedule areas [80, 81, 82, 83, 84, 85, 86] such as transportation communication systems [87, 88, 89, 90, 91, 92, 93] and process control systems and so on.

References

[1] L. Xu, Separable Newton recursive estimation method through system responses based on dynamically discrete measurements with increasing data length, Int. J. Control Autom. Syst. 20 (2) (2022) 432-443.

[2] L. Xu, Separable multi-innovation Newton iterative modeling algorithm for multi-frequency signals based on the sliding measurement window, Circuits Syst. Signal Process. 41 (2) (2022) 805-830.

[3] F. Ding, T. Chen, Combined parameter and output estimation of dual-rate systems using an auxiliary model, Automatica 40 (10) (2004) 1739-1748.

[4] J. Pan, X. Jiang, X.K. Wan, W. Ding, A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems, Int. J. Control Autom. Syst. 15 (3) (2017) 1189-1197.

[5] F. Ding, T. Chen, Parameter estimation of dual-rate stochastic systems by using an output error method, IEEE Trans. Autom. Control 50 (9) (2005) 1436-1441.

[6] L. Xu, G.L. Song, A recursive parameter estimation algorithm for modeling signals with multi-frequencies, Circuits Syst. Signal Process. 39 (8) (2020) 4198-4224.

[7] Y. Ji, Z. Kang, Model recovery for multi-input signal-output nonlinear systems based on the compressed sensing recovery theory, J. Frankl. Inst. 359(5) (2022) 2317-2339.

[8] C. Zhang, H.B. Liu, Y. Ji, Gradient parameter estimation of a class of nonlinear systems based on the maximum likelihood principle, Int. J. Control Autom. Syst. 20 (5) (2022) 1393-1404.

[9] F. Ding, X.G. Liu, J. Chu, Gradient-based and least-squares-based iterative algorithms for Hammerstein systems using the hierarchical identification principle, IET Control Theory Appl. 7 (2) (2013) 176-184.

[10] J. Chen, M. Gan, Q. Zhu, Y. Mao, Varying infimum gradient descent algorithm for agent-server systems with uncertain communication network, IEEE Trans. Instrum. Meas. 70 (2021) 9510511.

[11] F. Ding, Hierarchical multi-innovation stochastic gradient algorithm for Hammerstein nonlinear system modeling, Appl. Math. Modell. 37 (4) (2013) 1694-1704.

[12] M. Gan, Y. Guan, G.Y. Chen, C.L.P. Chen. Recursive variable projection algorithm for a class of separable nonlinear models. IEEE Trans. Neural Netw. Learn. Syst. 32 (11) (2021) 4971-4982.

[13] M. Gan, G.Y. Chen, L. Chen, C.L.P. Chen, Term selection for a class of separable nonlinear models, IEEE Trans. Neural Netw. Learn. Syst. 31 (2) (2020) 445-451.

[14] H. Xu, B. Champagne, Joint parameter and time-delay estimation for a class of nonlinear time-series models, IEEE Signal Process. Lett. 29 (2022) 947-951.

[15] Z. Kang, Y. Ji, X.M. Liu, Hierarchical recursive least squares algorithms for Hammerstein nonlinear autoregressive output-error systems, Int. J. Adapt. Control Signal Process. 35 (11) (2021) 2276-2295.

[16] X. Zhang, Q.Y. Liu, T. Hayat, Recursive identification of bilinear time-delay systems through the redundant rule, J. Frankl. Inst. 357 (1) (2020) 726-747.

[17] X. Zhang, L. Xu, Recursive parameter estimation methods and convergence analysis for a special class of nonlinear systems, Int. J. Robust Nonlinear Control 30 (4) (2020) 1373-1393.

[18] J. Pan, H. Ma, J. Sheng, Recursive coupled projection algorithms for multivariable output-error-like systems with coloured noises, IET Signal Process. 14 (7) (2020) 455-466.

[19] J.W. Wang, Y. Ji, Two-stage gradient-based iterative algorithms for the fractional-order nonlinear systems by using the hierarchical identification principle, Int. J. Adapt. Control Signal Process. 36 (7) (2022) 1778-1796.

[20] X. Zhang, Hierarchical parameter and state estimation for bilinear systems, Int. J. Syst. Sci. 51 (2) (2020) 275-290.

[21] J.W. Wang, Y. Ji, C. Zhang, Iterative parameter and order identification for fractional-order nonlinear finite impulse response systems using the key term separation, Int. J. Adapt. Control Signal Process. 35 (8) (2021) 1562-1577.

[22] M.H. Li, X.M. Liu, Maximum likelihood hierarchical least squares-based iterative identification for dual-rate stochastic systems, Int. J. Adapt. Control Signal Process. 35 (2) (2021) 240-261.

[23] M.H. Li, X.M. Liu, Iterative identification methods for a class of bilinear systems by using the particle filtering technique, Int. J. Adapt. Control Signal Process. 35 (10) (2021) 2056-2074.

[24] F. Ding, L. Xu, D.D. Meng, Gradient estimation algorithms for the parameter identification of bilinear systems using the auxiliary model, J. Comput. Appl. Math. 369 (2020) Article Number: 112575.

[25] Y. Gu, R. Ding, A least squares identification algorithm for a state space model with multi-state delays, Appl. Math. Lett. 26 (7) (2013) 748-753.

[26] P. Ma, L. Wang, Filtering-based recursive least squares estimation approaches for multivariate equation-error systems by using the multiinnovation theory, Int. J. Adapt. Control Signal Process. 35 (9) (2021) 1898-1915.

[27] J. Chen, B. Huang, M. Gan, C.L.P. Chen, A novel reduced-order algorithm for rational models based on Arnoldi process and Krylov subspace, Automatica 129 (2021) Article Number: 109663.

[28] F. Ding, Coupled-least-squares identification for multivariable systems, IET Control Theory Appl. 7 (1) (2013) 68-79.

[29] F. Ding, Y. Shi, T. Chen, Performance analysis of estimation algorithms of non-stationary ARMA processes, IEEE Trans. Signal Process. 54 (3) (2006) 1041-1053.

[30] J. Chen, Q.M. Zhu, Y.J. Liu, Modified Kalman filtering based multi-step-length gradient iterative algorithm for ARX models with random missing outputs, Automatica 118 (2020). Article Number: 109034.

[31] J. Ding, G. Liu, Hierarchical least squares identification for linear SISO systems with dual-rate sampled-data, IEEE Trans. Autom. Control 56 (11) (2011) 2677-2683.

[32] Y.H. Zhou, X. Zhang, Partially-coupled nonlinear parameter optimization algorithm for a class of multivariate hybrid models, Appl. Math. Computat. 414 (2022) 126663.

[33] Y. Ji, Z. Kang, Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems, Int. J. Robust Nonlinear Control 31 (3) (2021) 971-987.

[34] X. Zhang, Optimal adaptive filtering algorithm by using the fractional-order derivative, IEEE Signal Process. Lett. 29 (2022) 399-403.

[35] Y. Ji, Z. Kang, C. Zhang, Two-stage gradient-based recursive estimation for nonlinear models by using the data filtering, Int. J. Control Autom. Syst. 19 (8) (2021) 2706-2715.

[36] J.L. Ding, W.H. Zhang, Finite-time adaptive control for nonlinear systems with uncertain parameters based on the command filters, Int. J. Adapt. Control Signal Process. 35 (9) (2021) 1754-1767.

[37] L. Xu, Q.M. Zhu, Separable synchronous multi-innovation gradient-based iterative signal modeling from on-line measurements, IEEE Trans. Instrum. Meas. 71 (2022) 6501313.

[38] J. Pan, W. Li, H.P. Zhang, Control algorithms of magnetic suspension systems based on the improved double exponential reaching law of sliding mode control, Int. J. Control Autom. Syst. 16 (6) (2018) 2878-2887.

[39] L. Xu, Q. Zhu, Decomposition strategy-based hierarchical least mean square algorithm for control systems from the impulse responses, Int. J. Syst. Sci. 52 (9) (2021) 1806-1821.

[40] J.X. Xiong, J. Pan, G.Y. Chen, Sliding mode dual-channel disturbance rejection attitude control for a quadrotor, IEEE Trans. Ind. Electron. 69 (10) (2022) 10489-10499.

[41] Y. Gu, J.C. Liu, X.L. Li, State space model identification of multirate processes with time-delay using the expectation maximization, J. Frankl. Inst. 356 (3) (2019) 1623-1639.

[42] L. Xu, F.Y. Chen, T. Hayat, Hierarchical recursive signal modeling for multi-frequency signals based on discrete measured data, Int. J. Adapt. Control Signal Process. 35 (5) (2021) 676-693.

[43] Y. Ji, Z. Kang, X.M. Liu, The data filtering based multiple-stage Levenberg-Marquardt algorithm for Hammerstein nonlinear systems, Int. J. Robust Nonlinear Control 31 (15) (2021) 7007-7025.

[44] F. Ding, Y.J. Liu, B. Bao, Gradient based and least squares based iterative estimation algorithms for multi-input multi-output systems, Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng. 226 (1) (2012) 43-55.

[45] F. Ding, T. Chen, L. Qiu, Bias compensation based recursive least squares identification algorithm for MISO systems, IEEE Trans. Circuits Syst. II Express Briefs 53 (5) (2006) 349-353.

[46] Y. Ji, X.K. Jiang, L.J. Wan, Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems, J. Frankl. Inst. 357 (8) (2020) 5019-5032.

[47] Y.H. Zhou, Modeling nonlinear processes using the radial basis function-based state-dependent autoregressive models, IEEE Signal Process. Lett. 27 (2020) 1600-1604.

[48] Y.J. Liu, Y. Shi, An efficient hierarchical identification method for general dual-rate sampled-data systems, Automatica 50 (3) (2014) 962-970.

[49] Y.M. Fan, X.M. Liu, Two-stage auxiliary model gradient-based iterative algorithm for the input nonlinear controlled autoregressive system with variable-gain nonlinearity, Int. J. Robust Nonlinear Control 30 (14) (2020) 5492-5509.

[50] F. Ding, X.P. Liu, G. Liu, Multiinnovation least squares identification for linear and pseudo-linear regression models, IEEE Trans. Syst. Man Cybern. Part B Cybern. 40 (3) (2010) 767-778.

[51] Y.J. Wang, Novel data filtering based parameter identification for multiple-input multiple-output systems using the auxiliary model, Automatica 71 (2016) 308-313.

[52] X.M. Liu, Y.M. Fan, Maximum likelihood extended gradient-based estimation algorithms for the input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity, Int. J. Robust Nonlinear Control 31 (9) (2021) 4017-4036.

[53] X. Zhang, Adaptive parameter estimation for a general dynamical system with unknown states, Int. J. Robust Nonlinear Control 30 (4) (2020) 1351-1372.

[54] X. Zhang, E.F. Yang, Highly computationally efficient state filter based on the delta operator, Int. J. Adapt. Control Signal Process. 33 (6) (2019) 875-889.

[55] X. Zhang, E.F. Yang, State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors, Int. J. Adapt. Control Signal Process. 33 (7) (2019) 1157-1173.

[56] C.J. Xu, H.C, Xu, C. Liu, H.S. Su, Adaptive bipartite consensus of competitive linear multi-agent systems with asynchronous intermittent communication, Int. J. Robust Nonlinear Control 32 (9) (2022) 5120-5140.

[57] F. Ding, *System Identification - Auxiliary Model Identification Idea and Methods*, Science Press, Beijing (2017).

[58] L. Xu, E. Yang, Auxiliary model multiinnovation stochastic gradient parameter estimation methods for nonlinear sandwich systems, Int. J. Robust Nonlinear Control. 31 (1) (2021) 148-165.
[59, 60, 61, 62]

[59] C.C. Yin, Y.Z. Wen, An extension of Paulsen-Gjessing's risk model with stochastic return on investments, Insur. Math Econ 52 (3) (2013) 469-476.

[60] C.C. Yin, J.S. Zhao, Nonexponential asymptotics for the solutions of renewal equations, with applications, J. Appl. Probab. 43 (3) (2006) 815-824.

[61] C.C. Yin, K.C. Yuen, Optimality of the threshold dividend strategy for the compound Poisson model, Stat. Probabll. Lett. 81 (12) (2011) 1841-1846.

[62] C.C. Yin, K.C. Yuen, Optimal dividend problems for a jump-diffusion model with capital injections and proportional transaction costs, J. Ind. Manag. Optim. 11 (4) (2015) 1247-1262.

[63] H. Wang, H. Fan, J. Pan, A true three-scroll chaotic attractor coined, Discrete. Cont. Dyn-B 27 (5) (2022) 2891-2915.

[64] Y.H. Zhou, Hierarchical estimation approach for RBF-AR models with regression weights based on the increasing data length, IEEE Trans. Circuits Syst. II Express Briefs 68 (12) (2021) 3597-3601.

[65] S.Y. Liu, X. Zhang, L. Xu, Expectation-maximization algorithm for bilinear systems by using the Rauch-Tung-Striebel smoother, Automatica 142 (2022) 110365.

[66] W. Xiong, X. Jia, D. Yang, DP-LinkNet: A convolutional network for historical document image binarization, KSII Trans. Internet Inf. Syst. 15 (5) (2021) 1778-1797.

[67] G. Zhao, T.H. Cao, Y.D. Wang, Optimal sizing of isolated microgrid containing photovoltaic/photothermal/wind/diesel/battery, Int. J. Photoenergy 2021 (2021) 5566597.

[68] X.G. Wang, M. Zhao, Y. Zhou, Design and analysis for multi-disc coreless axial-flux permanent-magnet synchronous machine, IEEE Trans. Appl. Superconduct. 31 (8) (2021) 5203804.

[69] X.G. Wang, Z.W. Wan, L. Tang, Electromagnetic performance analysis of an axial flux hybrid excitation motor for HEV drives, IEEE Trans. Appl. Superconduct. 31 (8) (2021) Article Number: 5205605.

[70] M. Li, G. Xu, Q. Lai, J. Chen, A chaotic strategy-based quadratic opposition-based learning adaptive variable-speed whale optimization algorithm, Math. Comput. Simul. 193 (2022) 71-99.

[71] Y. An, Y.J. Zhang, W.J. Cao, A lightweight and practical anonymous authentication protocol based on bit-self-test PUF. Electronics 11 (5) (2022) 772.

[72] J. Hou, F.W. Chen, P.H. Li, Z.Q. Zhu, Gray-box parsimonious subspace identification of Hammerstein-type systems, IEEE Trans. Ind. Electron. 68 (10) (2021) 9941-9951.

[73] N. Zhao, A. Wu, Y. Pei, Spatial-temporal aggregation graph convolution network for efficient mobile cellular traffic prediction, IEEE Commun. Lett. 26 (3) (2022) 587-591.

[74] Y.F. Chen, C. Zhang, C.Y. Liu, Y.M. Wang, X.K. Wan, Atrial fibrillation detection using feedforward neural network, J. Med. Biol. Eng. 42 (1) (2022) 63-73.

[75] G.H. Wang, Q. Zhai, H. Liu, Cross self-attention network for 3D point cloud, Knowledge-Based Syst. 247 (2022) 108769.

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

[77] L. Xu, The damping iterative parameter identification method for dynamical systems based on the sine signal measurement, Signal Process. 120 (2016) 660-667.

[78] 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 (2016) 135-143.

[79] F. Ding, *System Identification - Performances Analysis for Identification Methods*, Science Press, Beijing (2014).

[80] Y. Cao, Z.X. Zhang, F.L. Cheng, S. Su, Trajectory optimization for high-speed trains via a mixed integer linear programming approach, IEEE Trans. Intell. Transp. Syst. (2023), DOI: 10.1109/TITS.2022.3155628.

[81] Y. Cao, Y.K. Sun, G. Xie, P. Li, A sound-based fault diagnosis method for railway point machines based on two-stage feature selection strategy and ensemble classifier, IEEE Trans. Intell. Transp. Syst. (2022). doi:10.1109/TITS.2021.3109632.

[82] Y. Cao, J.K. Wen, A. Hobiny, Parameter-varying artificial potential field control of virtual coupling system with nonlinear dynamics, Fractals (2022). doi: 10.1142/S0218348X22400990

[83] Y. Cao, J.K. Wen, L.C. Ma, Tracking and collision avoidance of virtual coupling train control system, Alex. Eng. J. 60 (2) (2021) 2115-2125.

[84] Y.K. Sun, Y. Cao, L.C. Ma, A fault diagnosis method for train plug doors via sound signals, IEEE Intel. Transp. Sy. 13 (3) (2021) 107-117.

[85] Y.K. Sun, Y. Cao, G. Xie, T. Wen, Sound based fault diagnosis for RPMs based on multi-scale fractional permutation entropy and two-scale algorithm, IEEE Trans. Veh. Technol. 70 (11) (2021) 11184-11192.

[86] S. Su, X.K. Wang, Y. Cao, J.T. Yin, An energy-efficient train operation approach by integrating the metro timetabling and eco-driving, IEEE Trans. Intell. Transp. Syst. 21 (10) (2020) 4252-4268.

[87] Y. Cao, Z. Wang, F. Liu, Bio-inspired speed curve optimization and sliding mode tracking control for subway trains, IEEE Trans. Veh. Technol. 68 (7) (2019) 6331-6342.

[88] Y. Cao, Y.K. Sun, G. Xie, T. Wen, Fault diagnosis of train plug door based on a hybrid criterion for IMFs selection and fractional wavelet package energy entropy, IEEE Trans. Veh. Technol. 68 (8) (2019) 7544-7551.

[89] Y. Cao, L.C. Ma, S. Xiao, Standard analysis for transfer delay in CTCS-3, Chinese J. Electron. 26 (5) (2017) 1057-1063.

[90] S. Su, J.F. She, K.C. Li, A nonlinear safety equilibrium spacing based model predictive control for virtually coupled train set over gradient terrains, IEEE Trans. Transp. Electrif. 8 (2) (2022) 2810-2824.

[91] S. Su, Q.Y. Zhu, J.Q. Liu, T. Tang, Q.L. Wei, Y. Cao, Eco-driving of trains with a data-driven iterative learning approach, IEEE Trans. Ind. Inf. (2022). doi::10.1109/TII.2022.3195888

[92] S. Su, T. Tang, J. Xun, Design of running grades for energy-efficient train regulation: A case study for Beijing Yizhuang line, IEEE Intell. Transp. Syst. Mag. 13 (2) (2021) 189-200.

[93] J. Wang, C. Ding, M. Wu, Y. Liu, G. Chen, Lightweight multiple scale-patch dehazing network for real-world hazy image, KSII Trans. Internet Inf. Syst. 15 (12) (2022) 4420-4438.