Journal of Computational and Applied Mathematics 427 (2023) 115104

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Separable synthesis gradient estimation methods and convergence analysis for multivariable systems☆

Ling Xu [a,d], Feng Ding [b,c,d,*]

[a] School of Microelectronics and Control Engineering, Changzhou University, Changzhou 213159, PR China
[b] College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, PR China
[c] College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, PR China
[d] School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, PR China

ARTICLE INFO

Article history:
Received 1 July 2022
Received in revised form 27 October 2022

Keywords:
Multivariable system
Large-scale system
Parameter estimation
Multi-innovation
Stochastic gradient

ABSTRACT

This paper studies the parameter identification problem for a large-scale multivariable systems. In terms of the identification obstacle causing by huge amounts of parameters of large-scale systems, a separable gradient (synthesis) identification algorithm is developed in accordance with the hierarchical computation principle. For the large-scale multivariable equation-error systems, the whole parameters are detached into several sub-parameter matrices based on the scales of the coefficient matrices of the inputs and outputs. On the basis of the detached parameter matrices, multiple parameter estimation sub-algorithms are presented for estimating the parameters of each sub-matrix through using the gradient search and multi-innovation theory from real-time measurements. Concerning the problem that the sub-algorithms are not effective because of the unknown parameters existing in the recursive computation, the previous estimates of the unknown parameters and the interactive estimation are introduced into the sub-algorithms to eliminate the associated items that make the sub-algorithms impossible to implement. In order to analyze the convergence of the proposed algorithms theoretically, we prove the convergence by using martingale convergence theory and stochastic principle. Finally, the performance tests of the proposed identification approaches for large-scale multivariable systems are carried out on several numerical examples and the simulation results demonstrate the effectiveness of the proposed methods.

1. Introduction

With the development of the network technology, the system scale is getting more complicated and larger. Recently, the study regarding on large-scale systems is used widely in many applications such as biological systems, communication systems, smart power grids and traffic systems [1–4]. Due to the huge amounts of parameters and complex structure of the large-scale systems, the system identification for large-scale systems faces many challenges from the perspectives of computational complexity [5–9]. Generally, the large-scale systems contain multiple input variables and output variables, which are described as multi-input multi-output (MIMO) systems [10–12]. For the study of the identification on the MIMO systems, the main researches are concentrated on enhancing the parameter estimation accuracy and reducing

☆ This work was supported by the National Natural Science Foundation of China (No. 61873111) and the 111 Project (B23008).
* Corresponding author at: College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, PR China.
E-mail addresses: lingxu0848@163.com (L. Xu), fding@jiangnan.edu.cn (F. Ding).

https://doi.org/10.1016/j.cam.2023.115104
0377-0427/

the computational complexity. The hierarchical identification method is effective for reducing computation amount by separating large-scale systems into smaller systems. Many hierarchical identification algorithms based on the model separation or key term separation are presented [13,14]. This hierarchical technique is also used in other research areas such as the signal modeling [15], the image processing and solving high-dimension matrix equations [16,17]. Zhou et al. adopted the hierarchical technique to train the parameters for RBF-AR models with regression weights [18]. Ji et al. proposed a three-stage stochastic gradient identification method for dealing with the parameter estimation of the nonlinear systems [19,20]. Wang et al. used the separable method to identify the tensor systems for reducing the complexity of the algorithms [21–23]. In view of the effectiveness of the separation technique in solving identification problems, we extend this method to the multivariable system identification in this study.

Because the large-scale systems are mostly made up of networks and various complex interferences existing in the actual networks, it is necessary to identify the large-scale systems through on-line means to track the changes of the system information in real time. Moreover, the dynamical characteristics of the control systems usually change with the operating conditions. So it is important to develop the on-line identification algorithms to track the changes of system dynamics, in which the system models are refreshed when the real-time measurements are sampled. For instance, the on-line identification is proposed to identify the critical parameters of the winding resistances and the load torque in induction machines [24]. Generally, the iterative estimation is used in off-line identification [25–28] and the recursive estimation is used in on-line identification [29–31]. The discrete-time Frequency-Locked-Loop non-linear filter is applied to estimate the characteristic parameters of a time-varying sinusoidal signal simultaneously [32]. In view of on-line identification methods, the recursive algorithms can fulfill the estimation computation in accordance with time sequence or time changes via dynamical data, which are used widely in on-line identification for dynamical systems [33]. In practical applications, many identification problems utilize the recursive criterions to derive the identification algorithms such as the prediction error algorithms, the subspace identification methods and the instrumental variable methods [34]. The recursive least squares method is very effective for solving linear problems and it can get highly performance [35]. However, the least squares methods are no longer suitable for dealing with nonlinear problems. Therefore, this paper studies the identification methods which can be used for the quadratic optimization and the nonlinear optimization.

The convergence is important for identification methods, which can ensure the feasibility of the algorithm application. Lewis et al. analyzed the convergence and provided the solvers for the positone, nonexistence and semipositone problems by means of finite difference methods [36]. As the theoretical tools for the convergence analysis, the Cramér–Rao bounds are used widely in mathematical areas and engineering areas [37]. Schmitt et al. presented a recursive state estimation method for nonlinear stochastic systems and analyzed the mean-square error lower bound in accordance with the Cramér–Rao lower bounds theory [38]. However, the Cramér–Rao lower bounds only provide the estimation error lower bound theoretically. In practical industrial processes, we need to judge and restrict the maximal values of the control plants the errors. In terms of the identification problem, the estimation error upper bounds are more meaningful than the estimation error lower bounds for the reason that the estimation upper bounds can assessment the estimation accuracy, i.e., the maximal estimation errors. The difference between the Cramér–Rao lower bound and the martingale convergence is that the Cramér–Rao lower bounds give the lower bound of the estimation error theoretically while the martingale convergence provides the upper bound of the estimation error. Therefore, this paper studies and presents the convergence theory of the identification methods for the large-scale systems by means of the martingale convergence theorem [39,40]. The main contributions of this paper are outlined as follows.

- The separable synthesis gradient parameter identification methods are developed for the large-scale equation-error systems based on the system separation and hierarchical computation by means of dynamical observations.
- A separable synthesis stochastic gradient parameter estimation method is presented through the negative gradient search. In order to simplify the determination of the step-size, an adaptive step-size method is proposed in accordance with the real-time dynamic information, which can avoid complicated one-dimensional search.
- For the purpose of enhancing the accuracy of the identified model parameters, a separable synthesis multi-innovation stochastic gradient identification by introducing dynamical window data by merging the historical data with the real-time data, which can make full use of the prediction ability of the historical data and real-time modal information of the system to get highly accurate model parameters.
- From the perspective of the engineering application, the parameter estimation upper bounds are analyzed based on the Martingale convergence theorem, which can give a more meaningful convergence analysis proof for engineering applications.

The other sections of this paper are organized as follows. The basic descriptions of the identification problem for the large-scale multivariable systems are given in Section 2. The details of the separable synthesis stochastic gradient (SS-SG) algorithm for the large-scale multivariable systems and its convergence proof are presented in Section 3. The details of the separable synthesis multi-innovation stochastic gradient (SS-MISG) algorithm for the large-scale multivariable systems and its convergence proof are presented in Section 4. Moreover, the implementation of the SS-SG algorithm and the SS-MISG algorithm and their simulation results are provided to show the effectiveness in Section 5. Finally, the conclusive remarks are given in Section 6.

2. Problem description

The system tackled in this paper regards the large-scale systems. Consider the large-scale multivariable system described by a multivariable equation-error model [41]:

$$\boldsymbol{\Theta}(z)\boldsymbol{y}(t)=\boldsymbol{\Omega}(z)\boldsymbol{u}(t)+\boldsymbol{v}(t), \tag{1}$$

where $\boldsymbol{u}(t)$ is the input measurement vector, $\boldsymbol{y}(t)$ is the output measurement vector, $\boldsymbol{v}(t)\in\mathbb{R}^m$ is the noise vector, $\boldsymbol{\Theta}(z)$ and $\boldsymbol{\Omega}(z)$ are the unknown polynomial matrices to be identified. The detailed definitions are as follows:

$$\begin{aligned}
&\boldsymbol{y}(t):=[y_1(t),y_2(t),\ldots,y_m(t)]^{\mathrm{T}}\in\mathbb{R}^m,\\
&\boldsymbol{u}(t):=[u_1(t),u_2(t),\ldots,u_r(t)]^{\mathrm{T}}\in\mathbb{R}^r,\\
&\boldsymbol{\Theta}(z):=\boldsymbol{I}_m+\boldsymbol{\Theta}_1z^{-1}+\boldsymbol{\Theta}_2z^{-2}+\cdots+\boldsymbol{\Theta}_{n_a}z^{-n_a}\in\mathbb{R}^{m\times m},\\
&\boldsymbol{\Omega}(z):=\boldsymbol{\Omega}_1z^{-1}+\boldsymbol{\Omega}_2z^{-2}+\cdots+\boldsymbol{\Omega}_{n_b}z^{-n_b}\in\mathbb{R}^{m\times r},
\end{aligned}$$

where z^{-1} is a shift operator and $z^{-1}y(t)=y(t-1)$.

Since the system scale is large, the parameter matrices contain the sum of multiple sub-matrices, which results in difficulty for identification. In general, in order to identify the system parameters in (1), all of the system parameter matrices $\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2,\cdots,\boldsymbol{\Theta}_{n_a},\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2,\cdots,\boldsymbol{\Omega}_{n_b}$ are combined and defined as a large parameter matrix

$$\boldsymbol{\theta}:=[\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2,\ldots,\boldsymbol{\Theta}_{n_a},\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2,\ldots,\boldsymbol{\Omega}_{n_b}]\in\mathbb{R}^{m\times(mn_a+rn_b)}.$$

Here $\boldsymbol{\theta}$ is a matrix which consists of the system unknown parameters. The goal of this paper is to develop the identification methods for estimating the unknown system parameter matrix $\boldsymbol{\theta}$ based on the collected observations $\boldsymbol{y}(t)$ and $\boldsymbol{u}(t)$ through constructing objective functions and optimizing them.

From $\boldsymbol{\theta}$, we can see that the dimension of $\boldsymbol{\theta}$ is very large, which will lead to very complicated computation. In system identification, if the system parameters have high dimensions, the identification algorithms will have numerous computational amount. For the on-line identification, the complicated on-line identification computation will make the systems unstable. Moreover, for the large-scale multivariable systems, numerous input and output variables make the identification processes very intricate. Only the identification methods with small amount of calculation can meet the needs of the rapidity and stability of on-line estimation. This problem inspires us to develop a high-efficient modeling method with low computational burden to identify the multivariable system.

Rewriting (1) yields

$$\begin{aligned}
\boldsymbol{y}(t)&=[\boldsymbol{I}_m-\boldsymbol{\Theta}(z)]\boldsymbol{y}(t)+\boldsymbol{\Omega}(z)\boldsymbol{u}(t)+\boldsymbol{v}(t)\\
&=-\boldsymbol{\Theta}_1\boldsymbol{y}(t-1)-\boldsymbol{\Theta}_2\boldsymbol{y}(t-2)-\cdots-\boldsymbol{\Theta}_{n_a}\boldsymbol{y}(t-n_a)\\
&\quad+\boldsymbol{\Omega}_1\boldsymbol{u}(t-1)+\boldsymbol{\Omega}_2\boldsymbol{u}(t-2)+\cdots+\boldsymbol{\Omega}_{n_b}\boldsymbol{u}(t-n_b)+\boldsymbol{v}(t).
\end{aligned} \tag{2}$$

The identification models are important and basic for deriving the parameter identification algorithms. Many identification methods are derived based on the identification models of the systems [42–46] and these methods can be used to estimate the parameters of other linear systems and nonlinear systems [47–51] and can be applied to other fields [52–56] such as chemical process control systems. In order to obtain the identification models to derive highly efficient parameter estimation algorithms, separating system (2) into multiple virtual sub-systems gives

$$\begin{aligned}
&\boldsymbol{y}_{a,i}(t)=-\boldsymbol{\Theta}_i\boldsymbol{y}(t-i)+\boldsymbol{v}(t),\quad i=1,2,\ldots,n_a,\\
&\boldsymbol{y}_{b,j}(t)=\boldsymbol{\Omega}_j\boldsymbol{u}(t-j)+\boldsymbol{v}(t),\quad j=1,2,\ldots,n_b,
\end{aligned}$$

where $\boldsymbol{y}(t-i)\in\mathbb{R}^m$ and $\boldsymbol{u}(t-j)\in\mathbb{R}^r$.

After system separation, the original system is divided into many sub-models. The virtual outputs of the separated sub-models $\boldsymbol{y}_{a,i}(t)$ and $\boldsymbol{y}_{b,j}(t)$ are given by

$$\boldsymbol{y}_{a,i}(t)=\boldsymbol{y}(t)+\sum_{q=1,q\neq i}^{n_a}\boldsymbol{\Theta}_q\boldsymbol{y}(t-q)-\sum_{j=1}^{n_b}\boldsymbol{\Omega}_j\boldsymbol{u}(t-j), \tag{3}$$

$$\boldsymbol{y}_{b,j}(t)=\boldsymbol{y}(t)-\sum_{l=1,l\neq j}^{n_b}\boldsymbol{\Omega}_l\boldsymbol{u}(t-l)+\sum_{i=1}^{n_a}\boldsymbol{\Theta}_i\boldsymbol{y}(t-i). \tag{4}$$

From the separable sub-models in (3)–(4), we can see that the original system parameters can be divided into n_a+n_b sub-matrices. Meanwhile, we can form the objective functions regarding the separable parameter matrices. Consequently, define the objective functions by using the observation of the current sampling instant with respect to the parameter

matrices $\boldsymbol{\Theta}_i$ and $\boldsymbol{\Omega}_j$ as

$$J(\boldsymbol{\Theta}_i) := \frac{1}{2}\|\boldsymbol{y}_{a,i}(t) + \boldsymbol{\Theta}_i\boldsymbol{y}(t-i)\|^2,$$
$$J(\boldsymbol{\Omega}_j) := \frac{1}{2}\|\boldsymbol{y}_{b,j}(t) - \boldsymbol{\Omega}_j\boldsymbol{u}(t-j)\|^2.$$

In order to derive the identification method to obtain the parameter estimates $\hat{\boldsymbol{\Theta}}_i(t)$ and $\hat{\boldsymbol{\Omega}}_j(t)$, minimizing the objective functions $J(\boldsymbol{\Theta}_i)$ and $J(\boldsymbol{\Omega}_j)$ by means of the negative search gives

$$\text{grad}[J(\boldsymbol{\Theta}_i)] := \frac{\partial J(\boldsymbol{\Theta}_i)}{\partial \boldsymbol{\Theta}_i} = \boldsymbol{y}(t-i)[\boldsymbol{y}_{a,i}(t) + \boldsymbol{\Theta}_i\boldsymbol{y}(t-i)]^{\mathrm{T}} \in \mathbb{R}^{m\times m},$$
$$\text{grad}[J(\boldsymbol{\Omega}_j)] := \frac{\partial J(\boldsymbol{\Omega}_j)}{\partial \boldsymbol{\Omega}_j} = -\boldsymbol{u}(t-j)[\boldsymbol{y}_{b,j}(t) - \boldsymbol{\Omega}_j\boldsymbol{u}(t-j)]^{\mathrm{T}} \in \mathbb{R}^{r\times m}.$$

Based on the aforementioned gradient matrices grad$[J(\boldsymbol{\Theta}_i)]$ and grad$[J(\boldsymbol{\Omega}_j)]$ and the dynamical observations, we can derive some on-line estimation algorithms according to different data usage schemes. In system identification, the iterative estimation is employed to identify system parameters through a way of off-line for the reason that the observations are the data with finite datum length and the data are unchangeable with time increasing. Another way is the recursive estimation which employs the dynamical data of one instant or batch. Since the recursive estimation utilizes dynamical data, it is widely used in on-line estimation. This paper tries to develop the recursive estimation algorithms which can be applied to solve the identification problem for large-scale multivariable systems aiming to enhance the identification efficiency.

3. Separable synthesis stochastic gradient method

Le $\hat{\boldsymbol{\Theta}}_i(t)$ and $\hat{\boldsymbol{\Omega}}_j(t)$ be the estimates of the parameter matrices $\boldsymbol{\Theta}_i$ and $\boldsymbol{\Omega}_j$ at time t. In accordance with the stochastic gradient principle, the following relations are obtained:

$$\begin{aligned}\hat{\boldsymbol{\Theta}}_i(t) &= \hat{\boldsymbol{\Theta}}_i(t-1) - \frac{1}{r_i(t)}\text{grad}[J(\hat{\boldsymbol{\Theta}}_i(t-1))]\\ &= \hat{\boldsymbol{\Theta}}_i(t-1) - \frac{\boldsymbol{y}(t-i)}{r_i(t)}[\boldsymbol{y}_{a,i}(t) + \hat{\boldsymbol{\Theta}}_i(t-1)\boldsymbol{y}(t-i)]^{\mathrm{T}}, \qquad (5)\\ \hat{\boldsymbol{\Omega}}_j(t) &= \hat{\boldsymbol{\Omega}}_j(t-1) - \frac{1}{r_j(t)}\text{grad}[J(\hat{\boldsymbol{\Omega}}_j(t-1))]\\ &= \hat{\boldsymbol{\Omega}}_j(t-1) + \frac{\boldsymbol{u}(t-j)}{r_j(t)}[\boldsymbol{y}_{b,j}(t) - \hat{\boldsymbol{\Omega}}_j(t-1)\boldsymbol{u}(t-j)]^{\mathrm{T}}. \qquad (6)\end{aligned}$$

Substituting (3) and (4) into (5) and (6) yields

$$\hat{\boldsymbol{\Theta}}_i(t) = \hat{\boldsymbol{\Theta}}_i(t-1) - \frac{\boldsymbol{y}(t-i)}{r_i(t)}\left[\boldsymbol{y}(t) + \sum_{q=1,q\neq i}^{n_a}\boldsymbol{\Theta}_q\boldsymbol{y}(t-q) - \sum_{j=1}^{n_b}\boldsymbol{\Omega}_j\boldsymbol{u}(t-j) + \hat{\boldsymbol{\Theta}}_i(t-1)\boldsymbol{y}(t-i)\right]^{\mathrm{T}}, \qquad (7)$$

$$\hat{\boldsymbol{\Omega}}_j(t) = \hat{\boldsymbol{\Omega}}_j(t-1) + \frac{\boldsymbol{u}(t-j)}{r_j(t)}\left[\boldsymbol{y}(t) - \sum_{l=1,l\neq j}^{n_b}\boldsymbol{\Omega}_l\boldsymbol{u}(t-l) + \sum_{i=1}^{n_a}\boldsymbol{\Theta}_i\boldsymbol{y}(t-i) - \hat{\boldsymbol{\Omega}}_j(t-1)\boldsymbol{u}(t-j)\right]^{\mathrm{T}}. \qquad (8)$$

For simplification, the step-sizes $r_i(t)$ and $r_j(t)$ are set as the same one, i.e.,

$$r_i(t) = r_j(t) = r(t) = r(t-1) + \sum_{q=1}^{n_a}\|\boldsymbol{y}(t-q)\|^2 + \sum_{l=1}^{n_b}\|\boldsymbol{u}(t-l)\|^2, \quad r(0) = 1.$$

Replacing $\boldsymbol{\Theta}_i$ and $\boldsymbol{\Omega}_j$ in (7) and (8) with their estimates $\hat{\boldsymbol{\Theta}}_i(t-1)$ and $\hat{\boldsymbol{\Omega}}_j(t-1)$ at time $t-1$ yields the separable synthesis stochastic gradient (SS-SG) algorithm for the large-scale multivariable systems:

$$\begin{aligned}\hat{\boldsymbol{\Theta}}_i(t) &= \hat{\boldsymbol{\Theta}}_i(t-1) - \frac{\boldsymbol{y}(t-i)}{r(t)}\left[\boldsymbol{y}(t) + \sum_{q=1,q\neq i}^{n_a}\hat{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-q) - \sum_{j=1}^{n_b}\hat{\boldsymbol{\Omega}}_j(t-1)\boldsymbol{u}(t-j) + \hat{\boldsymbol{\Theta}}_i(t-1)\boldsymbol{y}(t-i)\right]^{\mathrm{T}}\\ &= \hat{\boldsymbol{\Theta}}_i(t-1) - \frac{\boldsymbol{y}(t-i)}{r(t)}\left[\boldsymbol{y}(t) + \sum_{q=1}^{n_a}\hat{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-q) - \sum_{l=1}^{n_b}\hat{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-l)\right]^{\mathrm{T}}, \qquad (9)\end{aligned}$$

$$\begin{aligned}\hat{\boldsymbol{\Omega}}_j(t) &= \hat{\boldsymbol{\Omega}}_j(t-1)+\frac{\boldsymbol{u}(t-j)}{r(t)}\left[\boldsymbol{y}(t)-\sum_{l=1,l\neq j}^{n_b}\hat{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-l)+\sum_{i=1}^{n_a}\hat{\boldsymbol{\Theta}}_i(t-1)\boldsymbol{y}(t-i)-\hat{\boldsymbol{\Omega}}_j(t-1)\boldsymbol{u}(t-j)\right]^{\mathrm{T}}\\ &= \hat{\boldsymbol{\Omega}}_j(t-1)+\frac{\boldsymbol{u}(t-j)}{r(t)}\left[\boldsymbol{y}(t)+\sum_{q=1}^{n_a}\hat{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-q)-\sum_{l=1}^{n_b}\hat{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-l)\right]^{\mathrm{T}}, \qquad (10)\end{aligned}$$

$$r(t)=r(t-1)+\sum_{q=1}^{n_a}\|\boldsymbol{y}(t-q)\|^2+\sum_{l=1}^{n_b}\|\boldsymbol{u}(t-l)\|^2,\quad r(0)=1,\ i=1,2,\ldots,n_a,\ j=1,2,\ldots,n_b. \qquad (11)$$

Remark 1. Because the system to be identified is a large-scale system, there are numerous parameters to be estimated. We can combine all of the system parameters into one parameter matrix and define the cost function regarding this defined parameter matrix. The separable methods show their advantages for dealing with the models which are composed of linear and nonlinear relation to reduce the parameter dimensions and the complexity of algorithms. Especially for the least squares and nonlinear least squares algorithms or the Newton algorithms, the separable methods are very effective for reducing the computational complexity. Considering the feature of the high dimension of the system parameter matrices and for the purpose of enhancing the algorithm performance, this paper separates the system parameter matrix $\boldsymbol{\theta}$ into the parameter sub-matrices $\boldsymbol{\Theta}_i$ ($i=1,2,\ldots,n_a$) and $\boldsymbol{\Omega}_j$ ($j=1,2,\ldots,n_b$). Then n_a+n_b identification sub-models are constructed and n_a+n_b identification sub-algorithms are derived on the basis of the separated parameter matrices.

Remark 2. It is noted that the SS-SG algorithm in (9)–(11) is proposed by separating the original complicated identification model into many simple identification sub-models based on parameter decomposition. In order to deal with the associated parameters among the sub-algorithms, the unknown associated terms are replaced by using their estimates at the previous moment and interactive estimation technique to ensure the accomplishment of the combined sub-algorithms. The proposed parameter identification algorithm for multivariable stochastic systems in this paper can be extended to other multi-input multi-output bilinear and nonlinear systems [57–61] and can be applied to other control and schedule areas [62–66] such as the information processing and transportation communication systems [67–72] and so on.

In order to evaluate the convergence performance of the proposed SS-SG algorithm, some assumptions are set firstly. Suppose that $\{\boldsymbol{v}(t),\mathscr{D}_t\}$ is a sequence of martingale vectors defined in probability space, where $\{\mathscr{D}_t\}$ is the σ sequence generated by $\boldsymbol{v}(t)$, i.e.,

$$\mathscr{D}_t=\sigma(\boldsymbol{v}(t),\boldsymbol{v}(t-1),\boldsymbol{v}(t-2),\ldots).$$

The noise vector sequence $\boldsymbol{v}(t)$ satisfies the following assumptions:

$$\begin{aligned}&\text{(A1)}\quad \mathrm{E}[\boldsymbol{v}(t)|\mathscr{D}_t]=\boldsymbol{0},\quad \text{a.s.},\\ &\text{(A2)}\quad \mathrm{E}[\|\boldsymbol{v}(t)\|^2|\mathscr{D}_t]=\sigma_v^2(t)\leqslant\sigma_v^2<\infty,\quad \text{a.s.},\\ &\text{(A3)}\quad \limsup_{t\to\infty}\frac{1}{t}\sum_{i=1}^{t}\|\boldsymbol{v}(i)\|^2\leqslant\sigma_v^2<\infty,\quad \text{a.s.}\end{aligned} \qquad (12)$$

Lemma 1. *Suppose that $\{X(t)\}$, $\{\zeta(t)\}$ and $\{\xi(t)\}$ are the nonnegative variables and satisfy the following inequality*

$$X(t)\leqslant X(t-1)+\zeta(t)-\xi(t), \qquad (13)$$

and $\sum_{t=1}^{\infty}\zeta(t)<\infty$. Then when $t\to\infty$, we have $\sum_{t=1}^{\infty}\xi(t)<\infty$ and $X(t)\to X_0$, i.e., $X(t)$ converges to a finite constant.

Lemma 2. *Martingale convergence theorem: Lemma D.5.3 in [73]*

If T_t, α_t, β_t are non-negative random variables, measurable with respect to a non-decreasing sequence of σ algebra $\mathscr{D}_{t-1}$, and satisfy

$$\mathrm{E}[T_t|\mathscr{D}_{t-1}]\leqslant T_{t-1}+\alpha_t-\beta_t,\ \text{a.s.},$$

then when $\sum_{t=1}^{\infty}\alpha_t<\infty$, a.s., we have $\sum_{t=1}^{\infty}\beta_t<\infty$, a.s. and $T_t\to T$, a.s. (a.s.: almost surely) a finite nonnegative random variable.

Lemma 3. *For the large-scale multivariable system in (1) and the SS-SG algorithm in (9)–(11), the following inequality holds:*

$$\sum_{t=1}^{\infty}\left[\sum_{i=1}^{n_a}\frac{\|\boldsymbol{y}(t-i)\|^2}{r^2(t)}+\sum_{j=1}^{n_b}\frac{\|\boldsymbol{u}(t-j)\|^2}{r^2(t)}\right]<\infty,\quad \text{a.s.}$$

Proof. According to the definition of $r(t)$, we have

$$\begin{aligned}\sum_{t=1}^{\infty}\left[\sum_{i=1}^{n_a}\frac{\|\boldsymbol{y}(t-i)\|^2}{r^2(t)}+\sum_{j=1}^{n_b}\frac{\|\boldsymbol{u}(t-j)\|^2}{r^2(t)}\right] &\leqslant \sum_{t=1}^{\infty}\left[\sum_{i=1}^{n_a}\frac{\|\boldsymbol{y}(t-i)\|^2}{r(t-1)r(t)}+\sum_{j=1}^{n_b}\frac{\|\boldsymbol{u}(t-j)\|^2}{r(t-1)r(t)}\right]\\ &=\sum_{t=1}^{\infty}\frac{\sum_{i=1}^{n_a}\|\boldsymbol{y}(t-i)\|^2+\sum_{j=1}^{n_b}\|\boldsymbol{u}(t-j)\|^2}{r(t-1)r(t)}\\ &=\sum_{t=1}^{\infty}\left[\frac{1}{r(t-1)}-\frac{1}{r(t)}\right]\\ &=\frac{1}{r(0)}-\frac{1}{r(\infty)}<\infty. \quad \square\end{aligned} \tag{14}$$

Lemma 4. *[39] Suppose that vector sequences $\boldsymbol{x}(t)\in\mathbb{R}^{m\times n}$ and $\boldsymbol{\varphi}(t)\in\mathbb{R}^n$ satisfy the following equations:*

$$\boldsymbol{\varphi}^T(t)\boldsymbol{x}(t)=\boldsymbol{0}, \quad \text{if} \quad t\to 0, \tag{15}$$

$$\lim_{t\to\infty}[\boldsymbol{x}(t)-\boldsymbol{x}(t-k)]=\boldsymbol{0}, \quad \text{a.s. } 0<k<\infty. \tag{16}$$

If there exist two positive constants α, β and an integer $N\geqslant mn$ to make the following strong persistent excitation condition holds,

$$\text{(SPE)} \quad \alpha\boldsymbol{I}\leqslant\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^T(t+i)\leqslant\beta\boldsymbol{I}, \quad t\geqslant 0. \tag{17}$$

Then $\lim_{t\to\infty}\boldsymbol{x}(t)=\boldsymbol{0}$.

Theorem 1. *For the large-scale multivariable system in (1) and the SS-SG algorithm in (9)–(11), the parameter estimation error is consistently bounded, i.e.,*

$$\sum_{i=1}^{n_a}\|\hat{\boldsymbol{\Theta}}_i(t)-\boldsymbol{\Theta}_i\|^2+\sum_{j=1}^{n_b}\|\hat{\boldsymbol{\Omega}}_j(t)-\boldsymbol{\Omega}_j\|^2\to W_0<\infty,$$

where

$$\mathrm{E}[W_0]\leqslant\sum_{i=1}^{n_a}\|\hat{\boldsymbol{\Theta}}_i(0)-\boldsymbol{\Theta}_i\|^2+\sum_{j=1}^{n_b}\|\hat{\boldsymbol{\Omega}}_j(0)-\boldsymbol{\Omega}_j\|^2+\frac{\sigma_v^2}{r(0)}.$$

Proof. Define n_a+n_b parameter estimation error matrices:

$$\tilde{\boldsymbol{\Theta}}_i(t):=\hat{\boldsymbol{\Theta}}_i(t)-\boldsymbol{\Theta}_i\in\mathbb{R}^{m\times m}, \quad i=1,2,\ldots,n_a, \tag{18}$$

$$\tilde{\boldsymbol{\Omega}}_j(t):=\hat{\boldsymbol{\Omega}}_j(t)-\boldsymbol{\Omega}_j\in\mathbb{R}^{m\times r}, \quad j=1,2,\ldots,n_b. \tag{19}$$

Substituting (2) and (9) into (18) yields

$$\begin{aligned}\tilde{\boldsymbol{\Theta}}_i(t)&=\hat{\boldsymbol{\Theta}}_i(t-1)-\frac{\boldsymbol{y}(t-i)}{r(t)}\left[\boldsymbol{y}(t)+\sum_{q=1}^{n_a}\hat{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-q)-\sum_{l=1}^{n_b}\hat{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-l)\right]^{\mathrm{T}}-\boldsymbol{\Theta}_i\\ &=\tilde{\boldsymbol{\Theta}}_i(t-1)-\frac{\boldsymbol{y}(t-i)}{r(t)}[\boldsymbol{\Gamma}(t)-\boldsymbol{\Xi}(t)+\boldsymbol{v}(t)]^{\mathrm{T}}.\end{aligned} \tag{20}$$

Substituting (2) and (10) into (19) yields

$$\tilde{\boldsymbol{\Omega}}_j(t)=\tilde{\boldsymbol{\Omega}}_j(t-1)+\frac{\boldsymbol{u}(t-j)}{r(t)}[\boldsymbol{\Gamma}(t)-\boldsymbol{\Xi}(t)+\boldsymbol{v}(t)]^{\mathrm{T}}, \tag{21}$$

where

$$\boldsymbol{\Gamma}(t):=\sum_{q=1}^{n_a}\tilde{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-q)\in\mathbb{R}^m, \tag{22}$$

$$\boldsymbol{\Xi}(t) := \sum_{l=1}^{n_b} \tilde{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-l) \in \mathbb{R}^m. \tag{23}$$

Define a random Lyapunov function as

$$W(t) := \sum_{i=1}^{n_a} \|\tilde{\boldsymbol{\Theta}}_i(t)\|^2 + \sum_{j=1}^{n_b} \|\tilde{\boldsymbol{\Omega}}_j(t)\|^2.$$

In accordance with (20) and (21), we have

$$\begin{aligned}
W(t) &= \sum_{i=1}^{n_a} \left\| \tilde{\boldsymbol{\Theta}}_i(t-1) - \frac{\boldsymbol{y}(t-i)}{r(t)}[\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)]^{\mathrm{T}} \right\|^2 + \sum_{j=1}^{n_b} \left\| \tilde{\boldsymbol{\Omega}}_j(t-1) + \frac{\boldsymbol{u}(t-j)}{r(t)}[\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)]^{\mathrm{T}} \right\|^2 \\
&= \sum_{i=1}^{n_a} \left\{ \|\tilde{\boldsymbol{\Theta}}_i(t-1)\|^2 - \frac{2[\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)]^{\mathrm{T}} \tilde{\boldsymbol{\Theta}}_i(t-1)\boldsymbol{y}(t-i)}{r(t)} + \frac{\|\boldsymbol{y}(t-i)\|^2}{r^2(t)} \|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)\|^2 \right\} \\
&\quad + \sum_{j=1}^{n_b} \left\{ \|\tilde{\boldsymbol{\Omega}}_j(t-1)\|^2 + \frac{2[\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)]^{\mathrm{T}} \tilde{\boldsymbol{\Omega}}_j(t-1)\boldsymbol{u}(t-j)}{r(t)} + \frac{\|\boldsymbol{u}(t-j)\|^2}{r^2(t)} \|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)\|^2 \right\} \\
&= \sum_{i=1}^{n_a} \left\{ \|\tilde{\boldsymbol{\Theta}}_i(t-1)\|^2 - \frac{2}{r(t)} \boldsymbol{\Gamma}_i^{\mathrm{T}}(t)[\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)] + \frac{\|\boldsymbol{y}(t-i)\|^2}{r^2(t)} \|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)\|^2 \right\} \\
&\quad + \sum_{j=1}^{n_b} \left\{ \|\tilde{\boldsymbol{\Omega}}_j(t-1)\|^2 + \frac{2}{r(t)} \boldsymbol{\Xi}_j^{\mathrm{T}}(t)[\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)] + \frac{\|\boldsymbol{u}(t-j)\|^2}{r^2(t)} \|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)\|^2 \right\} \\
&= W(t-1) - \frac{2}{r(t)} \left[\sum_{i=1}^{n_a} \boldsymbol{\Gamma}_i^{\mathrm{T}}(t) - \sum_{j=1}^{n_b} \boldsymbol{\Xi}_j^{\mathrm{T}}(t) \right] [\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)] \\
&\quad + \left[\sum_{i=1}^{n_a} \frac{\|\boldsymbol{y}(t-i)\|^2}{r^2(t)} + \sum_{j=1}^{n_b} \frac{\|\boldsymbol{u}(t-j)\|^2}{r^2(t)} \right] \|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)\|^2 \\
&= W(t-1) - \frac{2}{r(t)} [\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)]^{\mathrm{T}} [\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t) + \boldsymbol{v}(t)] \\
&\quad + \left[\sum_{i=1}^{n_a} \frac{\|\boldsymbol{y}(t-i)\|^2}{r^2(t)} + \sum_{j=1}^{n_b} \frac{\|\boldsymbol{u}(t-j)\|^2}{r^2(t)} \right] \{\|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)\|^2 + 2[\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)]^{\mathrm{T}} \boldsymbol{v}(t) + \|\boldsymbol{v}(t)\|^2\} \\
&= W(t-1) - \frac{2}{r(t)} \|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)\|^2 - \frac{2}{r(t)} [\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)]^{\mathrm{T}} \boldsymbol{v}(t) \\
&\quad + \left[\sum_{i=1}^{n_a} \frac{\|\boldsymbol{y}(t-i)\|^2}{r^2(t)} + \sum_{j=1}^{n_b} \frac{\|\boldsymbol{u}(t-j)\|^2}{r^2(t)} \right] \{\|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)\|^2 + 2[\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)]^{\mathrm{T}} \boldsymbol{v}(t) + \|\boldsymbol{v}(t)\|^2\} \\
&= W(t-1) - \left\{ \frac{2}{r(t)} - \frac{1}{r^2(t)} \left[\sum_{i=1}^{n_a} \|\boldsymbol{y}(t-i)\|^2 + \sum_{j=1}^{n_b} \|\boldsymbol{u}(t-j)\|^2 \right] \right\} \|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)\|^2 \\
&\quad - \left\{ \frac{2}{r(t)} - \frac{2}{r^2(t)} \left[\sum_{i=1}^{n_a} \|\boldsymbol{y}(t-i)\|^2 + \sum_{j=1}^{n_b} \|\boldsymbol{u}(t-j)\|^2 \right] \right\} [\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)]^{\mathrm{T}} \boldsymbol{v}(t) \\
&\quad + \left[\sum_{i=1}^{n_a} \frac{\|\boldsymbol{y}(t-i)\|^2}{r^2(t)} + \sum_{j=1}^{n_b} \frac{\|\boldsymbol{u}(t-j)\|^2}{r^2(t)} \right] \|\boldsymbol{v}(t)\|^2 \\
&\leqslant W(t-1) - \frac{1}{r(t)} \|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)\|^2 - \frac{2r(t) - 2\left[\sum_{i=1}^{n_a} \|\boldsymbol{y}(t-i)\|^2 + \sum_{j=1}^{n_b} \|\boldsymbol{u}(t-j)\|^2\right]}{r^2(t)} [\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)]^{\mathrm{T}} \boldsymbol{v}(t) \\
&\quad + \left[\sum_{i=1}^{n_a} \frac{\|\boldsymbol{y}(t-i)\|^2}{r^2(t)} + \sum_{j=1}^{n_b} \frac{\|\boldsymbol{u}(t-j)\|^2}{r^2(t)} \right] \|\boldsymbol{v}(t)\|^2,
\end{aligned} \tag{24}$$

where $\boldsymbol{\Gamma}_i(t) := \tilde{\boldsymbol{\Theta}}_i(t-1)\boldsymbol{y}(t-i) \in \mathbb{R}^m$ and $\boldsymbol{\Xi}(t)_j := \tilde{\boldsymbol{\Omega}}_j(t-1)\boldsymbol{u}(t-j) \in \mathbb{R}^m$. Because $\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)$, $r(t)$, $\boldsymbol{y}(t-i)$ and $\boldsymbol{u}(t-j)$ are not correlated with noise $\boldsymbol{v}(t)$, and are measurable with respect to $\mathscr{D}_t$, taking conditional expectation to both sides of (24) and using Assumptions (A1)–(A3) yield

$$\mathrm{E}[W(t)|\mathscr{D}_t] \leqslant W(t-1) - \frac{1}{r(t)}\|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)\|^2 + \frac{\sum_{i=1}^{n_a}\|\boldsymbol{y}(t-i)\|^2 + \sum_{j=1}^{n_b}\|\boldsymbol{u}(t-j)\|^2}{r^2(t)}\sigma_v^2, \tag{25}$$

$$\leqslant W(t-1) + \frac{\sum_{i=1}^{n_a}\|\boldsymbol{y}(t-i)\|^2 + \sum_{j=1}^{n_b}\|\boldsymbol{u}(t-j)\|^2}{r^2(t)}\sigma_v^2. \tag{26}$$

In accordance with Lemma 3, we can conclude that

$$\sum_{t=1}^{\infty}\frac{1}{r^2(t)}\sum_{i=1}^{n_a}\|\boldsymbol{y}(t-i)\|^2 + \sum_{j=1}^{n_b}\|\boldsymbol{u}(t-j)\|^2 < \infty,$$

which means that the last term of (25) is limited. Applying Lemma 2 to (25), we can know that $W(t)$ converges to a finite random variable, say W_0, and

$$\sum_{t=1}^{\infty}\frac{1}{r(t)}\|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)\|^2 < \infty, \quad \text{a.s.}$$

Taking mathematical expectation to both sides of (26) and applying Lemma 3, we have

$$\begin{aligned}
\mathrm{E}[W(t)] &\leqslant \mathrm{E}[W(t-1)] + \mathrm{E}\left[\frac{\sum_{i=1}^{n_a}\|\boldsymbol{y}(t-i)\|^2 + \sum_{j=1}^{n_b}\|\boldsymbol{u}(t-j)\|^2}{r^2(t)}\sigma_v^2\right] \\
&\leqslant \mathrm{E}[W(0)] + \mathrm{E}\left[\sum_{k=1}^{t}\frac{\sum_{i=1}^{n_a}\|\boldsymbol{y}(k-i)\|^2 + \sum_{j=1}^{n_b}\|\boldsymbol{u}(k-j)\|^2}{r^2(k)}\sigma_v^2\right] \\
&= \mathrm{E}[W(0)] + \mathrm{E}\left[\sum_{k=1}^{t}\frac{r(k)-r(k-1)}{r^2(k)}\sigma_v^2\right] \\
&\leqslant \mathrm{E}[W(0)] + \mathrm{E}\left[\sum_{k=1}^{t}\frac{r(k)-r(k-1)}{r(k)r(k-1)}\sigma_v^2\right] \\
&\leqslant \mathrm{E}[W(0)] + \frac{\sigma_v^2}{r(0)}.
\end{aligned} \tag{27}$$

This means

$$\mathrm{E}[W(t)] \leqslant \sum_{i=1}^{n_a}\|\hat{\boldsymbol{\Theta}}_i(0) - \boldsymbol{\Theta}_i\|^2 + \sum_{j=1}^{n_b}\|\hat{\boldsymbol{\Omega}}_j(0) - \boldsymbol{\Omega}_j\|^2 + \frac{\sigma_v^2}{r(0)}. \quad \square$$

Theorem 2. *For the large-scale multivariable system in (1) and the SS-SG algorithm in (9)–(11), if Assumptions (A1)–(A3) hold and the information vector $\boldsymbol{\varphi}(t)$ composed by $\boldsymbol{y}(t-i)$ and $\boldsymbol{u}(t-j)$*

$$\boldsymbol{\varphi}(t) := \begin{bmatrix} \boldsymbol{y}(t-i) \\ \boldsymbol{u}(t-j) \end{bmatrix} \in \mathbb{R}^{m\times n_a + r\times n_b}, \quad i = 1, 2, \ldots, n_a, \quad j = 1, 2, \ldots, n_b$$

is sufficiently rich, then the parameter estimation errors consistently converge to zero, i.e.,

$$\lim_{t\to\infty}\sum_{i=1}^{n_a}\|\hat{\boldsymbol{\Theta}}_i(t) - \boldsymbol{\Theta}_i\|^2 + \sum_{j=1}^{n_b}\|\hat{\boldsymbol{\Omega}}_j(t) - \boldsymbol{\Omega}_j\|^2 = 0, \quad \text{a.s.},$$

This means

$$\lim_{t\to\infty}\hat{\boldsymbol{\Theta}}_i(t) = \boldsymbol{\Theta}_i, \quad i = 1, 2, \ldots, n_a, \quad \text{a.s.}, \quad \lim_{t\to\infty}\hat{\boldsymbol{\Omega}}_j(t) = \boldsymbol{\Omega}_j, \quad j = 1, 2, \ldots, n_b, \quad \text{a.s.}$$

Proof. Because $\boldsymbol{\varphi}(t)$ is sufficiently rich, it can be obtained that

$$r(t) = O(t), \quad \text{when} \quad t \to \infty. \tag{28}$$

From (31), we have $\boldsymbol{\Gamma}(t) = \boldsymbol{\Xi}(t)$, i.e.,

$$\sum_{i=1}^{n_a}\tilde{\boldsymbol{\Theta}}_i(t-1)\boldsymbol{y}(t-i) = \sum_{j=1}^{n_b}\tilde{\boldsymbol{\Omega}}_j(t-1)\boldsymbol{u}(t-j), \quad t \to \infty.$$

Define

$$\boldsymbol{x}(t) := [\tilde{\boldsymbol{\Theta}}_i(t-1), -\tilde{\boldsymbol{\Omega}}_j(t-1)]^{\mathrm{T}} \in \mathbb{R}^{(mn_a+rn_b)\times m}, \quad i = 1, 2, \ldots, a_n, \quad j = 1, 2, \ldots, n_b.$$

Then we have

$$\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{x}(t) = 0. \tag{29}$$

In accordance with (A3), (31), (28), (20) and (21), we have

$$\lim_{t\to\infty}[\boldsymbol{x}(t) - \boldsymbol{x}(t-k)] = \boldsymbol{0}, \quad 0 < k < \infty. \tag{30}$$

Based on Lemma 4, (49) and (50), we have

$$\lim_{t\to\infty}\boldsymbol{x}(t) = \boldsymbol{0},$$

i.e.,

$$\lim_{t\to\infty}\hat{\boldsymbol{\Theta}}_i(t) = \boldsymbol{\Theta}_i, \quad i = 1, 2, \ldots, n_a, \quad \text{a.s.}, \quad \lim_{t\to\infty}\hat{\boldsymbol{\Omega}}_j(t) = \boldsymbol{\Omega}_j, \quad j = 1, 2, \ldots, n_b, \quad \text{a.s.}$$

or

$$\lim_{t\to\infty}\sum_{i=1}^{n_a}\|\hat{\boldsymbol{\Theta}}_i(t) - \boldsymbol{\Theta}_i\|^2 + \sum_{j=1}^{n_b}\|\hat{\boldsymbol{\Omega}}_j(t) - \boldsymbol{\Omega}_j\|^2 = 0, \quad \text{a.s.}$$

Assume that $r(t) = O(t) \to \infty$ and $\sum_{i=1}^{n_a}\|\boldsymbol{y}(t-i)\|^2 + \sum_{j=1}^{n_b}\|\boldsymbol{u}(t-j)\|^2 < \infty$. Then the convergence rates of $W(t)-W(t-1)$ and $\|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)\|^2$ are $O(1/t)$, respectively, i.e.,

$$\|\boldsymbol{\Gamma}(t) - \boldsymbol{\Xi}(t)\|^2 \to 0, \quad \text{a.s.} \quad t \to \infty. \quad \square \tag{31}$$

Remark 3. Compared with the least squares algorithms, the identification methods based on the gradient search have less computation but their convergence speed is slower. In order to enhance the tracing ability of the proposed SS-SG algorithm, a forgetting factor $0 < \eta \leqslant 1$ is introduced in $r(t)$ and the forgetting factor SS-SG (FF-SS-SG) method is obtained, in which $r(t)$ is computed as follows:

$$r(t) = \eta r(t-1) + \sum_{q=1}^{n_a}\|\boldsymbol{y}(t-q)\|^2 + \sum_{l=1}^{n_b}\|\boldsymbol{u}(t-l)\|^2, \quad r(0) = 1, \quad i = 1, 2, \ldots, n_a, \quad j = 1, 2, \ldots, n_b. \tag{32}$$

As a result, Eqs. (9)–(11) and (32) constitute the FF-SS-SG algorithm.

4. Separable synthesis multi-innovation stochastic gradient method

In order to enhance parameter estimation accuracy, the moving window data with the length p are introduced to join in the computational process. Because the sliding window data accumulate dynamically, it can obtain more system information and build a highly accurate model. Assume that the current sampling instant is t and the batch data with datum length p are composed of the current data and the previous data at p sampling instants. As a result, the dynamical data with window width p are represented by $\{y(t), y(t-1), \ldots, y(t-p+1)\}$. Based on the stacked data, the stacked information matrix components are defined as

$$\begin{aligned}&\boldsymbol{\Upsilon}(p, t-i) := [\boldsymbol{y}(t-i), \boldsymbol{y}(t-1-i), \ldots, \boldsymbol{y}(t-p+1-i)] \in \mathbb{R}^{m\times p},\\ &\boldsymbol{\Lambda}(p, t-j) := [\boldsymbol{u}(t-j), \boldsymbol{u}(t-1-j)\cdots, \boldsymbol{u}(t-p+1-j)] \in \mathbb{R}^{r\times p}.\end{aligned} \tag{33}$$

The multi-innovation matrix is defined as

$$\boldsymbol{E}(p, t) := \begin{bmatrix} \left[\boldsymbol{y}(t) + \sum_{q=1}^{n_a}\hat{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-q) - \sum_{l=1}^{n_b}\hat{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-l)\right]^{\mathrm{T}} \\ \left[\boldsymbol{y}(t-1) + \sum_{q=1}^{n_a}\hat{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-1-q) - \sum_{l=1}^{n_b}\hat{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-1-l)\right]^{\mathrm{T}} \\ \vdots \\ \left[\boldsymbol{y}(t-p+1) + \sum_{q=1}^{n_a}\hat{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-p+1-q) - \sum_{l=1}^{n_b}\hat{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-p+1-l)\right]^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{m\times p}.$$

Based on the stochastic gradient algorithm in (9)–(11) and by expanding innovation vector or matrix into multi-innovation matrix, the separable synthesis multi-innovation stochastic gradient (SS-MISG) algorithm is given by

$$\hat{\boldsymbol{\Theta}}_i(t) = \hat{\boldsymbol{\Theta}}_i(t-1) - \frac{\boldsymbol{\Upsilon}(p, t-i)}{r(t)}\boldsymbol{E}^{\mathrm{T}}(p, t), \quad i = 1, 2, \ldots, n_a, \tag{34}$$

$$\hat{\boldsymbol{\Omega}}_j(t) = \hat{\boldsymbol{\Omega}}_j(t-1) + \frac{\boldsymbol{\Lambda}(p,t-j)}{r(t)}\boldsymbol{E}^{\mathrm{T}}(p,t),\ j=1,2,\ldots,n_b, \tag{35}$$

$$\boldsymbol{E}(p,t) = \begin{bmatrix} \left[\boldsymbol{y}(t) + \sum_{q=1}^{n_a}\hat{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-q) - \sum_{l=1}^{n_b}\hat{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-l)\right]^{\mathrm{T}} \\ \left[\boldsymbol{y}(t-1) + \sum_{q=1}^{n_a}\hat{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-1-q) - \sum_{l=1}^{n_b}\hat{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-1-l)\right]^{\mathrm{T}} \\ \vdots \\ \left[\boldsymbol{y}(t-p+1) + \sum_{q=1}^{n_a}\hat{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-p+1-q) - \sum_{l=1}^{n_b}\hat{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-p+1-l)\right]^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \tag{36}$$

$$r(t) = r(t-1) + \sum_{i=1}^{n_a}\|\boldsymbol{\Upsilon}(p,t-i)\|^2 + \sum_{j=1}^{n_b}\|\boldsymbol{\Lambda}(p,t-j)\|^2, \quad r(0)=1. \tag{37}$$

The proposed algorithms in this paper can combine other mathematical tools [74–80] to study parameter identification of different systems [81–87] and can be applied to other fields [88–92] such as information processing.

Remark 4. The SS-MISG algorithm is presented by using the observations into sliding window observations, which can capture more dynamical information of real-time systems. Therefore, the proposed SS-MISG method can obtain higher estimation accuracy than the SS-SG algorithm theoretically.

For the purposed of evaluate the convergence of the proposed SS-MISG method in theory, we give the convergence analysis proof in terms of the proposed SS-MISG algorithm in (34)–(37).

Lemma 5. *For the large-scale multivariable system in (1) and the SS-MISG algorithm in (34)–(37), the following inequality holds*

$$\sum_{t=1}^{\infty}\left[\sum_{i=1}^{n_a}\frac{\|\boldsymbol{\Upsilon}(p,t-i)\|^2}{r^2(t)} + \sum_{j=1}^{n_b}\frac{\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r^2(t)}\right] < \infty, \quad \text{a.s.}$$

Proof. According to the definition of $r(t)$ in (37), we have

$$\begin{aligned}\sum_{t=1}^{\infty}\left[\sum_{i=1}^{n_a}\frac{\|\boldsymbol{\Upsilon}(p,t-i)\|^2}{r^2(t)} + \sum_{j=1}^{n_b}\frac{\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r^2(t)}\right] &\leqslant \sum_{t=1}^{\infty}\left[\sum_{i=1}^{n_a}\frac{\|\boldsymbol{\Upsilon}(p,t-i)\|^2}{r(t-1)r(t)} + \sum_{j=1}^{n_b}\frac{\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r(t-1)r(t)}\right] \\ &= \sum_{t=1}^{\infty}\frac{\sum_{i=1}^{n_a}\|\boldsymbol{\Upsilon}(p,t-i)\|^2 + \sum_{j=1}^{n_b}\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r(t-1)r(t)} \\ &= \sum_{t=1}^{\infty}\left[\frac{1}{r(t-1)} - \frac{1}{r(t)}\right] \\ &= \frac{1}{r(0)} - \frac{1}{r(\infty)} \\ &< \infty. \quad \text{a.s.} \quad \square\end{aligned}$$

Theorem 3. *For the large-scale multivariable system in (1) and the SS-MISG algorithm in (34)–(37), the parameter estimation error is consistently bounded, i.e.,*

$$\sum_{i=1}^{n_a}\|\hat{\boldsymbol{\Theta}}_i(t) - \boldsymbol{\Theta}_i\|^2 + \sum_{j=1}^{n_b}\|\hat{\boldsymbol{\Omega}}_j(t) - \boldsymbol{\Omega}_j\|^2 \to W(0) < \infty.$$

Proof. Define stacked matrices

$$\boldsymbol{\Gamma}(p,t) \coloneqq [\boldsymbol{\Gamma}(t), \boldsymbol{\Gamma}(t-1), \ldots, \boldsymbol{\Gamma}(t-p+1)] \in \mathbb{R}^{m\times p}, \tag{38}$$

$$\boldsymbol{\Xi}(p,t) \coloneqq [\boldsymbol{\Xi}(t), \boldsymbol{\Xi}(t-1), \ldots, \boldsymbol{\Xi}(t-p+1)] \in \mathbb{R}^{m\times p}. \tag{39}$$

where $\boldsymbol{\Gamma}(t) = \sum_{q=1}^{n_a}\tilde{\boldsymbol{\Theta}}_q(t-1)\boldsymbol{y}(t-q) \in \mathbb{R}^m$ and $\boldsymbol{\Xi}(t) = \sum_{l=1}^{n_b}\tilde{\boldsymbol{\Omega}}_l(t-1)\boldsymbol{u}(t-l) \in \mathbb{R}^m$. Define stacked noise vector:

$$\boldsymbol{V}(p,t) \coloneqq [v(t), v(t-1), \ldots, v(t-p+1)]^{\mathrm{T}} \in \mathbb{R}^p.$$

Then, the parameter estimation error matrices $\tilde{\boldsymbol{\Theta}}_i(t)$ and $\tilde{\boldsymbol{\Omega}}_j(t)$ are as follows:

$$\begin{aligned}\tilde{\boldsymbol{\Theta}}_i(t) &:= \hat{\boldsymbol{\Theta}}_i(t)-\boldsymbol{\Theta}_i\\ &= \hat{\boldsymbol{\Theta}}_i(t-1)-\frac{\boldsymbol{\Upsilon}(p,t-i)}{r(t)}\boldsymbol{E}^{\mathrm{T}}(p,t)-\boldsymbol{\Theta}_i\\ &= \tilde{\boldsymbol{\Theta}}_i(t-1)-\frac{\boldsymbol{\Upsilon}(p,t-i)}{r(t)}[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)]^{\mathrm{T}}. \end{aligned}\tag{40}$$

$$\begin{aligned}\tilde{\boldsymbol{\Omega}}_j(t) &:= \hat{\boldsymbol{\Omega}}_j(t)-\boldsymbol{\Omega}_j\\ &= \hat{\boldsymbol{\Omega}}_j(t-1)+\frac{\boldsymbol{\Lambda}(p,t-j)}{r(t)}\boldsymbol{E}^{\mathrm{T}}(p,t)-\boldsymbol{\Omega}_j\\ &= \tilde{\boldsymbol{\Omega}}_j(t-1)+\frac{\boldsymbol{\Lambda}(p,t-j)}{r(t)}[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)]^{\mathrm{T}}. \end{aligned}\tag{41}$$

Define a random Lyapunov function as

$$W(t) := \sum_{i=1}^{n_a}\|\tilde{\boldsymbol{\Theta}}_i(t)\|^2+\sum_{j=1}^{n_b}\|\tilde{\boldsymbol{\Omega}}_j(t)\|^2.$$

Substituting (40) and (41) into the above equation gives

$$\begin{aligned}W(t) &= \sum_{i=1}^{n_a}\|\tilde{\boldsymbol{\Theta}}_i(t-1)-\frac{\boldsymbol{\Upsilon}(p,t-i)}{r(t)}[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)]^{\mathrm{T}}\|^2\\ &\quad+\sum_{j=1}^{n_b}\|\tilde{\boldsymbol{\Omega}}_j(t-1)+\frac{\boldsymbol{\Lambda}(p,t-j)}{r(t)}[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)]^{\mathrm{T}}\|^2 \qquad (42)\\ &= \sum_{i=1}^{n_a}\Bigg\{\|\tilde{\boldsymbol{\Theta}}_i(t-1)\|^2-\frac{2[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)]^{\mathrm{T}}\tilde{\boldsymbol{\Theta}}_i(t-1)\boldsymbol{\Upsilon}(p,t-i)}{r(t)}\\ &\quad+\frac{\|\boldsymbol{\Upsilon}(p,t-i)\|^2}{r^2(t)}\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)\|^2\Bigg\}\\ &\quad+\sum_{j=1}^{n_b}\Bigg\{\|\tilde{\boldsymbol{\Omega}}_j(t-1)\|^2+\frac{2[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)]^{\mathrm{T}}\tilde{\boldsymbol{\Omega}}_j(t-1)\boldsymbol{\Lambda}(p,t-j)}{r(t)}\\ &\quad+\frac{\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r^2(t)}\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)\|^2\Bigg\}\\ &= \sum_{i=1}^{n_a}\Bigg\{\|\tilde{\boldsymbol{\Theta}}_i(t-1)\|^2-\frac{2}{r(t)}\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)]\\ &\quad+\frac{\|\boldsymbol{\Upsilon}(p,t-i)\|^2}{r^2(t)}\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)\|^2\Bigg\}\\ &\quad+\sum_{j=1}^{n_b}\Bigg\{\|\tilde{\boldsymbol{\Omega}}_j(t-1)\|^2+\frac{2}{r(t)}\boldsymbol{\Xi}_j^{\mathrm{T}}(p,t)[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)]\\ &\quad+\frac{\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r^2(t)}\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)\|^2\Bigg\}\\ &= W(t-1)-\frac{2}{r(t)}\left[\sum_{i=1}^{n_a}\boldsymbol{\Gamma}_i^{\mathrm{T}}(p,t)-\sum_{j=1}^{n_b}\boldsymbol{\Xi}_j^{\mathrm{T}}(p,t)\right][\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)]\\ &\quad+\left[\sum_{i=1}^{n_a}\frac{\|\boldsymbol{\Upsilon}(p,t-i)\|^2}{r^2(t)}+\sum_{j=1}^{n_b}\frac{\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r^2(t)}\right]\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)\|^2\\ &= W(t-1)-\frac{2}{r(t)}[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)]^{\mathrm{T}}[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)+\boldsymbol{V}(p,t)]\end{aligned}$$

$$
\begin{aligned}
&+\left[\sum_{i=1}^{n_a}\frac{\|\boldsymbol{\Upsilon}(p,t-i)\|^2}{r^2(t)}+\sum_{j=1}^{n_b}\frac{\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r^2(t)}\right]\{\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)\|^2\\
&+2[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)]^{\mathrm{T}}\boldsymbol{V}(p,t)+\|\boldsymbol{V}(p,t)\|^2\}\\
=\ & W(t-1)-\left\{\frac{2}{r(t)}-\frac{1}{r^2(t)}\left[\sum_{i=1}^{n_a}\|\boldsymbol{\Upsilon}(p,t-i)\|^2+\sum_{j=1}^{n_b}\|\boldsymbol{\Lambda}(p,t-j)\|^2\right]\right\}\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)\|^2\\
&-\left\{\frac{2}{r(t)}-\frac{2}{r^2(t)}\left[\sum_{i=1}^{n_a}\|\boldsymbol{\Upsilon}(p,t-i)\|^2+\sum_{j=1}^{n_b}\|\boldsymbol{\Lambda}(p,t-j)\|^2\right]\right\}[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)]^{\mathrm{T}}\boldsymbol{V}(p,t)\\
&+\left[\sum_{i=1}^{n_a}\frac{\|\boldsymbol{\Upsilon}(p,t-i)\|^2}{r^2(t)}+\sum_{j=1}^{n_b}\frac{\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r^2(t)}\right]\|\boldsymbol{V}(p,t)\|^2\\
\leqslant\ & W(t-1)-\frac{2r(t)-2\left[\sum_{i=1}^{n_a}\|\boldsymbol{\Upsilon}(p,t-i)\|^2+\sum_{j=1}^{n_b}\|\boldsymbol{\Lambda}(p,t-j)\|^2\right]}{r^2(t)}[\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)]^{\mathrm{T}}\boldsymbol{V}(p,t)\\
&-\frac{1}{r(t)}\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)\|^2+\left[\sum_{i=1}^{n_a}\frac{\|\boldsymbol{\Upsilon}(p,t-i)\|^2}{r^2(t)}+\sum_{j=1}^{n_b}\frac{\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r^2(t)}\right]\|\boldsymbol{V}(p,t)\|^2, \qquad (43)
\end{aligned}
$$

where $\boldsymbol{\Gamma}_i(p,t):=\tilde{\boldsymbol{\Theta}}_i(t-1)\boldsymbol{\Upsilon}(p,t-i)\in\mathbb{R}^{m\times p}$ and $\boldsymbol{\Xi}_j(p,t):=\tilde{\boldsymbol{\Omega}}_j(t-1)\boldsymbol{\Lambda}(p,t-j)\in\mathbb{R}^{m\times p}$.

Since $\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)$, $r(t)$, $\boldsymbol{\Upsilon}(p,t-i)$ and $\boldsymbol{\Lambda}(p,t-j)$ are not corrective with noise $\boldsymbol{V}(p,t)$, taking condition expectation both sides of (43) and using Assumptions (A1)–(A3) yield

$$
\begin{aligned}
\mathrm{E}[W(t)|\mathscr{D}_t]&\leqslant W(t-1)-\frac{1}{r(t)}\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)\|^2+\frac{\sum_{i=1}^{n_a}\|\boldsymbol{\Upsilon}(p,t-i)\|^2+\sum_{j=1}^{n_b}\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r^2(t)}p\sigma_v^2, \qquad (44)\\
&\leqslant W(t-1)+\frac{\sum_{i=1}^{n_a}\|\boldsymbol{\Upsilon}(p,t-i)\|^2+\sum_{j=1}^{n_b}\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r^2(t)}p\sigma_v^2. \qquad (45)
\end{aligned}
$$

In accordance with Lemma 5, we can conclude that

$$
\sum_{t=1}^{\infty}\frac{1}{r^2(t)}\sum_{i=1}^{n_a}\|\boldsymbol{\Upsilon}(p,t-i)\|^2+\sum_{j=1}^{n_b}\|\boldsymbol{\Lambda}(p,t-j)\|^2<\infty,
$$

which means that the last term of (44) is limited. Applying Lemma 2 to (44), we can know that $W(t)$ converges to a finite random variable $W(0)$ and

$$
\sum_{t=1}^{\infty}\frac{1}{r(t)}\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)\|<\infty,\quad \text{a.s.}
$$

Taking mathematical expectation both sides of (45) and applying Lemma 5 yield

$$
\begin{aligned}
\mathrm{E}[W(t)]&\leqslant \mathrm{E}[W(t-1)]+\mathrm{E}\left[\frac{\sum_{i=1}^{n_a}\|\boldsymbol{\Upsilon}(p,t-i)\|^2+\sum_{j=1}^{n_b}\|\boldsymbol{\Lambda}(p,t-j)\|^2}{r^2(t)}p\sigma_v^2\right]\\
&\leqslant \mathrm{E}[W(0)]+\mathrm{E}\left[\sum_{k=1}^{t}\frac{\sum_{i=1}^{n_a}\|\boldsymbol{\Upsilon}(p,k-i)\|^2+\sum_{j=1}^{n_b}\|\boldsymbol{\Lambda}(p,k-j)\|^2}{r^2(k)}p\sigma_v^2\right]\\
&= \mathrm{E}[W(0)]+\mathrm{E}\left[\sum_{k=1}^{t}\frac{r(k)-r(k-1)}{r^2(k)}p\sigma_v^2\right]\\
&\leqslant \mathrm{E}[W(0)]+\mathrm{E}\left[\sum_{k=1}^{t}\frac{r(k)-r(k-1)}{r(k)r(k-1)}p\sigma_v^2\right]\\
&\leqslant \mathrm{E}[W(0)]+\frac{p\sigma_v^2}{r(0)}. \qquad (46)
\end{aligned}
$$

Assume that $r(t)=O(t)\to\infty$ and $\sum_{i=1}^{n_a}\|\boldsymbol{\Upsilon}(p,t-i)\|^2+\sum_{j=1}^{n_b}\|\boldsymbol{\Lambda}(p,t-j)\|^2<\infty$. From (44) and (45), we can obtain the convergence rates of $W(t)-W(t-1)$ and $\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)\|^2$ are $O(1/t^2)$ and $O(1/t)$, respectively. As a result, we

have

$$\|\boldsymbol{\Gamma}(p,t)-\boldsymbol{\Xi}(p,t)\|^2\to 0,\quad \text{a.s.}\quad t\to\infty. \tag{47}$$

Theorem 4. *For the large-scale multivariable system in (1) and the SS-MISG algorithm in (34)–(37), if Assumptions* (A1)-(A3) *exist and the information matrix* $\boldsymbol{\Psi}(t)$ *composed by* $\boldsymbol{\Upsilon}(p,t-i)$ *and* $\boldsymbol{\Lambda}(p,t-j)$

$$\boldsymbol{\Psi}(p,t):=\left[\begin{array}{c}\boldsymbol{\Upsilon}(p,t-i)\\ \boldsymbol{\Lambda}(p,t-j)\end{array}\right]\in\mathbb{R}^{(m\times n_a+r\times n_b)\times p},\quad i=1,2,\ldots,n_a,\quad j=1,2,\ldots,n_b$$

is sufficiently rich, i.e.,

$$\alpha\boldsymbol{I}\leqslant\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\Psi}(p,t+k)\boldsymbol{\Psi}^T(p,t+k)\leqslant\beta\boldsymbol{I}.$$

Then the parameter estimation errors consistently converge to zero, i.e.,

$$\lim_{t\to\infty}\sum_{i=1}^{n_a}\|\hat{\boldsymbol{\Theta}}_i(t)-\boldsymbol{\Theta}_i\|^2+\sum_{j=1}^{n_b}\|\hat{\boldsymbol{\Omega}}_j(t)-\boldsymbol{\Omega}_j\|^2=0,\quad \text{a.s., or}$$

$$\lim_{t\to\infty}\hat{\boldsymbol{\Theta}}_i(t)=\boldsymbol{\Theta}_i,\ i=1,2,\ldots,n_a,\quad \text{a.s.},\quad \lim_{t\to\infty}\hat{\boldsymbol{\Omega}}_j(t)=\boldsymbol{\Omega}_j,\ j=1,2,\ldots,n_b,\quad \text{a.s.}$$

Proof. Since $\boldsymbol{\Psi}(p,t)$ is sufficiently rich and persistently excited, from the definition of $r(t)$ in (37) we can get

$$r(t)=O(t),\quad \text{when}\quad t\to\infty. \tag{48}$$

From (47), we have $\boldsymbol{\Gamma}(p,t)=\boldsymbol{\Xi}(p,t)$ when $t\to\infty$, i.e.,

$$\sum_{i=1}^{n_a}\tilde{\boldsymbol{\Theta}}_i(t-1)\boldsymbol{\Upsilon}(p,t-i)=\sum_{j=1}^{n_b}\tilde{\boldsymbol{\Omega}}_j(t-1)\boldsymbol{\Lambda}(p,t-j),\quad t\to\infty.$$

Define

$$\boldsymbol{x}(t):=[\tilde{\boldsymbol{\Theta}}_i(t-1),-\tilde{\boldsymbol{\Omega}}_j(t-1)]^{\mathrm{T}}\in\mathbb{R}^{(mn_a+rn_b)\times m},\quad i=1,2,\ldots,a_n,\quad j=1,2,\ldots,n_b.$$

Then we have

$$\boldsymbol{\Psi}^{\mathrm{T}}(p,t)\boldsymbol{x}(t)=0. \tag{49}$$

In accordance with (A3), (47), (48), (40) and (41), we have

$$\lim_{t\to\infty}[\boldsymbol{x}(t)-\boldsymbol{x}(t-k)]=\boldsymbol{0},\quad 0<k<\infty. \tag{50}$$

Based on Lemma 4, (49) and (50), we have

$$\lim_{t\to\infty}\boldsymbol{x}(t)=\boldsymbol{0},$$

i.e.,

$$\lim_{t\to\infty}\hat{\boldsymbol{\Theta}}_i(t)=\boldsymbol{\Theta}_i,\quad i=1,2,\ldots,n_a,\quad \text{a.s.},\quad \lim_{t\to\infty}\hat{\boldsymbol{\Omega}}_j(t)=\boldsymbol{\Omega}_j,\quad j=1,2,\ldots,n_b,\quad \text{a.s.}$$

or

$$\lim_{t\to\infty}\sum_{i=1}^{n_a}\|\hat{\boldsymbol{\Theta}}_i(t)-\boldsymbol{\Theta}_i\|^2+\sum_{j=1}^{n_b}\|\hat{\boldsymbol{\Omega}}_j(t)-\boldsymbol{\Omega}_j\|^2=0,\quad \text{a.s.}\quad \square$$

Remark 5. The proposed SS-MISG algorithm is presented by introducing more observations into the recursive computation. For the reason that more system information is employed to the identification process, the SS-MISG algorithm has better performance than the SS-SG algorithm. In fact, when the innovation length is $p=1$, the SS-MISG algorithm becomes the SS-SG algorithm.

Remark 6. The convergence analysis of the proposed SS-SG algorithm and the SS-MISG algorithm for the large-scale systems is based on the martingale convergence theorem in stochastic processes. The martingale convergence theorem provides us an effective tool for the convergence analysis in system identification and the convergence of many identification algorithms is proved through the martingale convergence theorem [39,40].

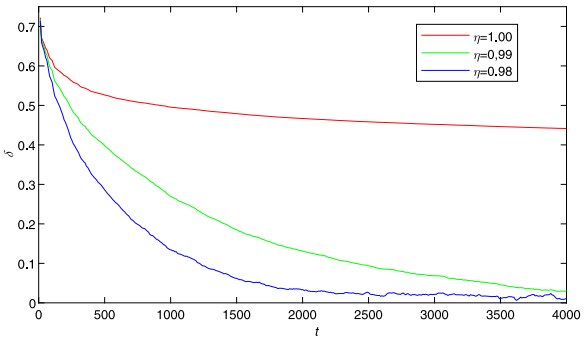

Fig. 1. The FF-SS-SG parameter estimation errors versus t with different forgetting factors.

Remark 7. The SS-SG algorithm and the SS-MISG algorithm are proposed based on the negative gradient search which are similar to gradient decent. Generally, the existing gradient decent algorithms need to determine the step-size during the recursion or iteration processes. But the step-size of the proposed algorithms in this paper can be determined automatically according to the information vector. Therefore, the proposed algorithms are more efficient than other existing gradient algorithms. The proposed SS-SG and SS-MISG algorithms in this paper can joint other identification approaches [93–97] to develop new gradient and multi-innovation gradient aaproaches for linear systems, bilinear systems and nonlinear systems [98–102].

5. Simulation study

Example 1. In this example, we test the performance of the proposed FF-SS-SG algorithm by introducing different forgetting factor for the first-order large-scale multivariable system.

Consider a multivariable system with the following mathematical model:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} + \begin{bmatrix} 0.81 & 0.30 \\ -0.28 & -0.68 \end{bmatrix} \begin{bmatrix} y_1(t-1) \\ y_2(t-1) \end{bmatrix} = \begin{bmatrix} 1.31 & -0.45 \\ -0.48 & 1.38 \end{bmatrix} \begin{bmatrix} u_1(t-1) \\ u_2(t-1) \end{bmatrix} + \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix},$$

where the true parameter values are as follows:

$$\boldsymbol{\Theta}(z) = \boldsymbol{I}_2 + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} z^{-1}, \quad \boldsymbol{\Omega}(z) = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} z^{-1}.$$

In the simulation, the simulation conditions are set as follows: (1) The input sequence $\{u_1(t), u_2(t)\}$ adopts the persistent excitation vector signal sequence with zero mean and unit variance. (2) The noise $\{v_1(t), v_2(t)\}$ uses white noise vector sequence with zero mean and variance $\sigma^2 = 0.30^2$ and $\sigma^2 = 0.50^2$. (3) The recursion is 4000. (4) The forgetting factor is set as $\eta = 1.00$, $\eta = 0.99$ and $\eta = 0.98$. (5) The parameter estimation error is defined as

$$\delta := \sqrt{\frac{\sum_{i=1}^{n_a} \|\hat{\boldsymbol{\Theta}}_i(t) - \boldsymbol{\Theta}_i\|^2 + \sum_{j=1}^{n_b} \|\hat{\boldsymbol{\Omega}}_j(t) - \boldsymbol{\Omega}_j\|^2}{\sum_{i=1}^{n_a} \|\boldsymbol{\Theta}_i\|^2 + \sum_{j=1}^{n_b} \|\boldsymbol{\Omega}_j\|^2}} \times 100\%.$$

By employing the proposed FF-SS-SG algorithm, the parameter estimates and their estimation errors are illustrated in Table 1 and the parameter estimation errors versus t are shown in Fig. 1. The parameter estimates obtained by the proposed FF-SS-SG algorithm vary with time t when the forgetting factors are $\eta = 1.00$ and $\eta = 0.98$ in Figs. 2–3.

Remark 8. From the simulation results of Example 1, we can conclude that the FF-SS-SG algorithm can obtain higher estimation accuracy when the forgetting factor $\eta = 0.98$ and $\eta = 0.99$ than $\eta = 1.00$. It is noted that when $\eta = 1.00$, the FF-SS-SG algorithm is the SS-SG algorithm. The simulation experiments show that the forgetting factor can improve the performance of the proposed SS-SG algorithm. Figs. 2–3 illustrate that the parameter estimates obtained by the FF-SS-SG algorithm when $\eta = 0.98$ is more accurate than those when $\eta = 1.00$.

Example 2. In this example, the proposed SS-MISG algorithm is applied to a first-order multivariable system compare the performance of different innovation length p.

Consider a first-order two-input two-output system:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} + \begin{bmatrix} 0.26 & 0.46 \\ -0.24 & -0.28 \end{bmatrix} \begin{bmatrix} y_1(t-1) \\ y_2(t-1) \end{bmatrix} = \begin{bmatrix} 0.76 & -0.18 \\ -0.25 & 0.98 \end{bmatrix} \begin{bmatrix} u_1(t-1) \\ u_2(t-1) \end{bmatrix} + \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix},$$

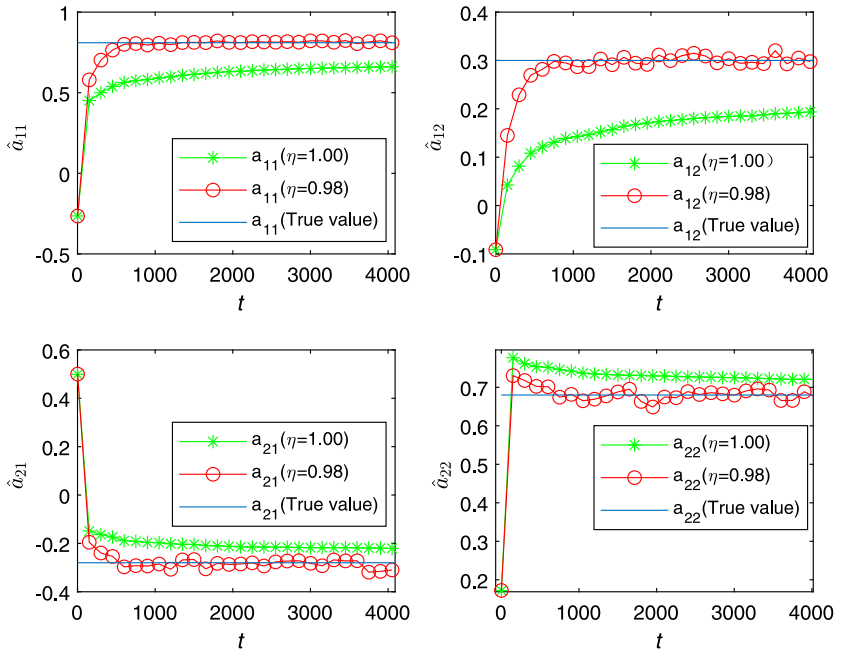

Fig. 2. The FF-SS-SG estimates of the parameters $\boldsymbol{\Theta}$ versus t with different forgetting factor.

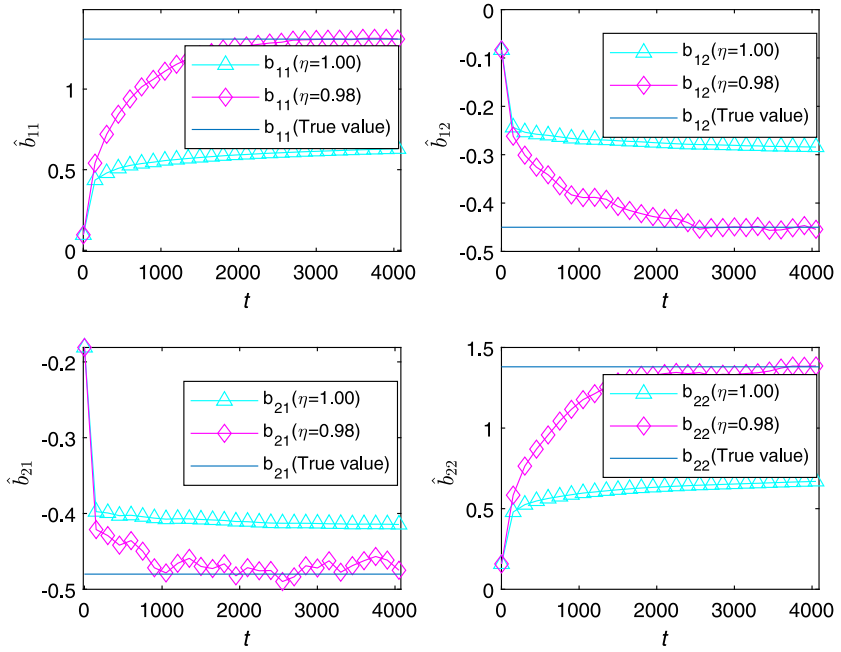

Fig. 3. The FF-SS-SG estimates of the parameters $\boldsymbol{\Omega}$ versus t with different forgetting factor.

where the true parameter values are as follows:

$$\boldsymbol{\Theta}(z)=\boldsymbol{I}_2+\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}z^{-1}, \quad \boldsymbol{\Omega}(z)=\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}z^{-1}.$$

In the simulation experiment, the simulation conditions are set as follows: (1) The noise variances of $v_1(t)$ and $v_2(t)$ are $\sigma^2=0.30^2$ and $\sigma^2=0.50^2$; (2) The innovation length is set as $p=1$, $p=3$, $p=5$, $p=10$ and $p=20$. (3) The input sequences $u_1(t)$ and $u_2(t)$ adopt the persistent excitation vector signal with zero mean and unit variance.

Table 1
The FF-SS-SG estimates and their estimation errors.

η	t	a_{11}	a_{12}	b_{11}	b_{21}	a_{21}	a_{22}	b_{21}	b_{22}	$\delta\%$
1.00	100	0.40559	0.00756	0.40308	−0.23243	−0.12199	0.78711	−0.38556	0.45509	61.49836
	200	0.46854	0.05427	0.45334	−0.24545	−0.14739	0.76797	−0.39703	0.49613	57.50804
	500	0.54603	0.11640	0.51331	−0.25807	−0.18144	0.75428	−0.40168	0.55371	52.62705
	1000	0.58974	0.14219	0.55371	−0.26828	−0.19642	0.73857	−0.40741	0.59412	49.56040
	2000	0.63085	0.17127	0.59200	−0.27539	−0.21269	0.72952	−0.41064	0.63335	46.68620
	3000	0.65061	0.18428	0.61382	−0.28065	−0.21762	0.72510	−0.41322	0.65295	45.20192
	4000	0.66096	0.19311	0.62898	−0.28413	−0.22029	0.72118	−0.41449	0.66855	44.14380
0.99	100	0.44589	0.04132	0.43014	−0.23401	−0.12949	0.77065	−0.38747	0.48542	59.01645
	200	0.54612	0.11677	0.52369	−0.25695	−0.16811	0.73861	−0.40687	0.56325	52.05615
	500	0.70179	0.23232	0.69717	−0.29305	−0.25158	0.72709	−0.41848	0.73082	39.77211
	1000	0.78466	0.27986	0.88613	−0.34302	−0.26491	0.67495	−0.44892	0.93904	26.94844
	2000	0.81601	0.29366	1.10435	−0.39303	−0.28467	0.66912	−0.46465	1.16580	13.11648
	3000	0.81891	0.30327	1.21912	−0.42645	−0.27840	0.68394	−0.46880	1.25212	6.89665
	4000	0.81027	0.29705	1.26972	−0.44117	−0.29717	0.67928	−0.46692	1.32740	3.04231
0.98	100	0.48961	0.07703	0.46328	−0.23515	−0.13670	0.75186	−0.39008	0.52220	56.17280
	200	0.61833	0.17678	0.60839	−0.27064	−0.18760	0.71095	−0.41849	0.64437	46.01542
	500	0.77619	0.27885	0.87476	−0.32825	−0.29283	0.71783	−0.43531	0.90478	28.54954
	1000	0.80710	0.29469	1.09439	−0.39107	−0.26865	0.67220	−0.47543	1.16488	13.44130
	2000	0.81821	0.29502	1.26293	−0.43135	−0.28357	0.65399	−0.48051	1.32897	3.33760
	3000	0.82220	0.30363	1.30745	−0.44934	−0.28153	0.67969	−0.47247	1.33119	2.21089
	4000	0.80802	0.29493	1.30801	−0.45228	−0.30122	0.67509	−0.47056	1.37631	1.07407
True values		0.81000	0.30000	1.31000	−0.45000	−0.28000	0.68000	−0.48000	1.38000	

Table 2
The SS-MISG estimates and their estimation errors.

p	t	a_{11}	a_{12}	b_{11}	b_{21}	a_{21}	a_{22}	b_{21}	b_{22}	$\delta\%$
1	100	0.16819	0.44919	0.55254	0.00205	−0.06871	0.27680	−0.02301	0.64968	36.62419
	500	0.19252	0.45323	0.61317	−0.05184	−0.11563	0.28111	−0.08863	0.75738	25.54540
	1000	0.20125	0.45196	0.63304	−0.06999	−0.13273	0.28271	−0.11121	0.79157	21.87703
	4000	0.21583	0.45223	0.66763	−0.10183	−0.15994	0.28628	−0.14864	0.84301	15.95572
5	100	0.17067	0.43784	0.68672	−0.10765	−0.12382	0.30749	−0.17798	0.88184	15.33284
	500	0.20821	0.45140	0.71695	−0.14015	−0.17066	0.29009	−0.20817	0.93898	8.42557
	1000	0.21889	0.45078	0.72638	−0.14881	−0.18596	0.28699	−0.21956	0.94968	6.51497
	4000	0.23624	0.45363	0.74020	−0.16411	−0.20877	0.28455	−0.23238	0.96072	3.78088
10	100	0.20333	0.44277	0.70917	−0.15605	−0.19780	0.29843	−0.23601	0.95442	6.85958
	500	0.22918	0.45730	0.73206	−0.16565	−0.21428	0.27961	−0.24009	0.97878	3.63435
	1000	0.23630	0.45547	0.73906	−0.16803	−0.22128	0.27791	−0.24535	0.97906	2.74305
	4000	0.24842	0.45742	0.74849	−0.17569	−0.23156	0.27760	−0.24724	0.97684	1.37363
20	100	0.24267	0.44928	0.71593	−0.17844	−0.23365	0.27586	−0.23714	0.97732	3.55613
	500	0.24969	0.46344	0.73764	−0.17478	−0.23434	0.27002	−0.24141	0.98911	2.13545
	1000	0.25223	0.46005	0.74405	−0.17443	−0.23676	0.27090	−0.24749	0.98643	1.54127
	4000	0.25780	0.46049	0.75172	−0.17939	−0.24070	0.27372	−0.24859	0.97996	0.75243
True values		0.26000	0.46000	0.76000	−0.18000	−0.24000	0.28000	−0.25000	0.98000	

Employing the proposed SS-MISG algorithm to estimate the two-input two-output system of Example 2, the parameter estimates and their estimation errors obtained under different innovation lengths are illustrated in Table 2. The parameter estimation errors versus t under different innovation p are shown in Fig. 4.

Remark 9. From the simulation results of Example 2, we can see that the innovation length is an important factor which can influence the parameter estimation accuracy. When the innovation length is large, more observed information of systems is used to the identification process. Fig. 4 and Table 2 show that the estimation accuracy improves with the increase of the innovation length p, which indicates that increasing the innovation length can enhance the parameter accuracy and the SS-MISG algorithm is effective for identifying the multivariable systems.

Example 3. In order to further test the performance of the proposed SS-MISG algorithm for identifying the large-scale multivariable systems, a second-order two-input two-output system is provided to show the parameter estimates and their estimation errors.

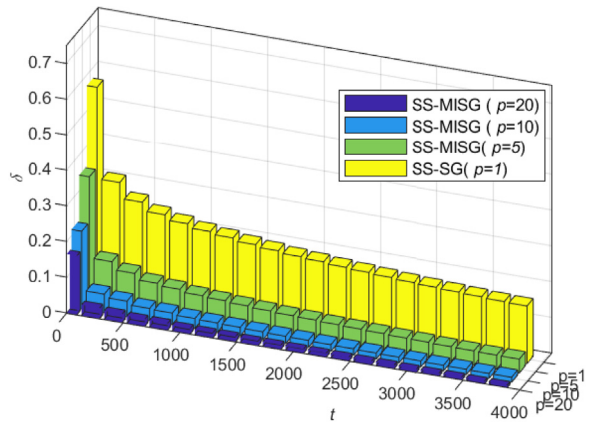

Fig. 4. The SS-MISG parameter estimation errors versus t with different p of Example 2.

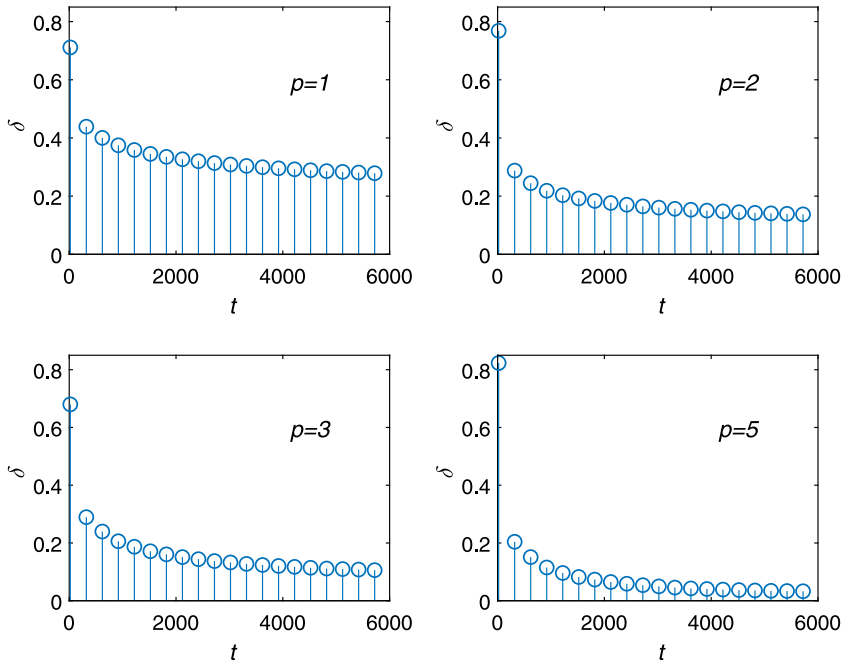

Fig. 5. The SS-MISG parameter estimation errors versus t with different p of Example 3.

Consider a second-order two-input two-output multivariable system:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} + \begin{bmatrix} 0.94 & 0.23 \\ -0.21 & 0.35 \end{bmatrix} \begin{bmatrix} y_1(t-1) \\ y_2(t-1) \end{bmatrix} + \begin{bmatrix} 0.70 & 0.38 \\ -0.65 & 0.28 \end{bmatrix} \begin{bmatrix} y_1(t-2) \\ y_2(t-2) \end{bmatrix}$$
$$= \begin{bmatrix} 1.21 & -0.53 \\ -0.42 & 1.38 \end{bmatrix} \begin{bmatrix} u_1(t-1) \\ u_2(t-1) \end{bmatrix} + \begin{bmatrix} 1.13 & 0.22 \\ -0.29 & 1.82 \end{bmatrix} \begin{bmatrix} u_1(t-2) \\ u_2(t-2) \end{bmatrix} + \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix},$$

where the true parameter values are as follows:

$$\boldsymbol{\Theta}_1 = \begin{bmatrix} a1_{11} & a1_{12} \\ a1_{21} & a1_{22} \end{bmatrix} = \begin{bmatrix} 0.94 & 0.23 \\ -0.21 & 0.35 \end{bmatrix}, \quad \boldsymbol{\Theta}_2 = \begin{bmatrix} a2_{11} & a2_{12} \\ a2_{21} & a2_{22} \end{bmatrix} = \begin{bmatrix} 0.70 & 0.38 \\ -0.65 & 0.28 \end{bmatrix},$$
$$\boldsymbol{\Omega}_1 = \begin{bmatrix} b1_{11} & b1_{12} \\ b1_{21} & b1_{22} \end{bmatrix} = \begin{bmatrix} 1.21 & -0.53 \\ -0.42 & 1.38 \end{bmatrix}, \quad \boldsymbol{\Omega}_2 = \begin{bmatrix} b2_{11} & b2_{12} \\ b2_{21} & b2_{22} \end{bmatrix} = \begin{bmatrix} 1.13 & 0.22 \\ -0.29 & 1.82 \end{bmatrix}.$$

The simulation conditions are set as follows: (1) The input sequence $\{u_1(t), u_2(t)\}$ adopts the persistent excitation vector signal with zero mean and unit variance, which is generated by "(rand(length1,2)-0.5)*sqrt(12)" through MATLAB

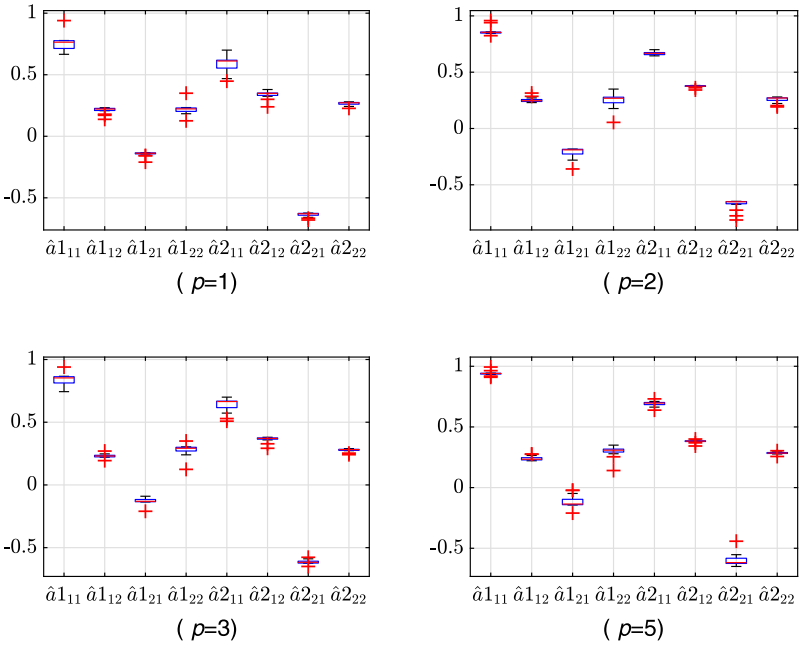

Fig. 6. The SS-MISG parameter estimation distribution of Θ_1 and Θ_2.

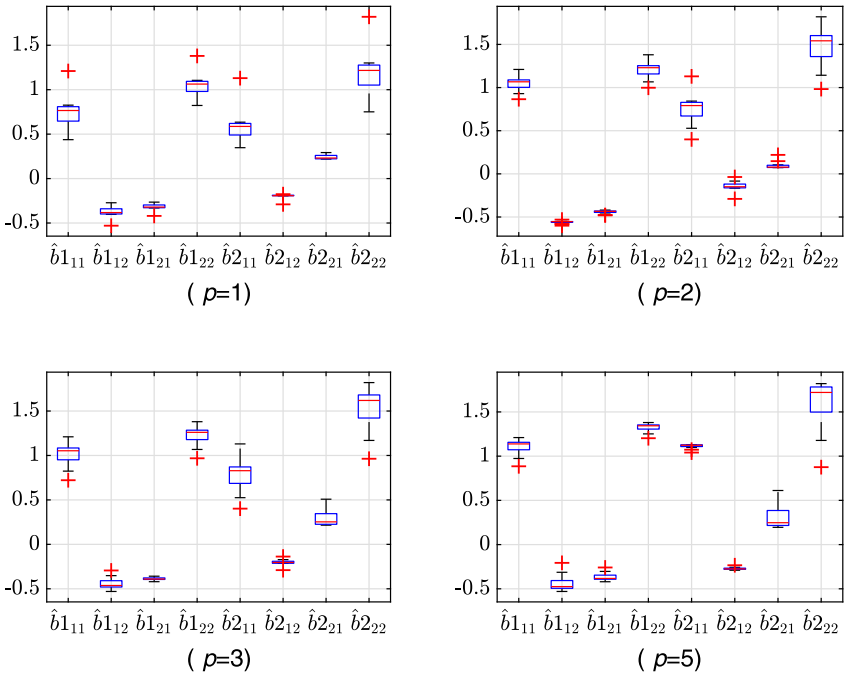

Fig. 7. The SS-MISG parameter estimation distribution of Ω_1 and Ω_2.

R2018; (2) The noise variance of $\{v_1(t), v_2(t)\}$ is $\{\sigma^2 = 0.80^2, \sigma^2 = 0.10^2\}$; (3) The innovation length is set as $p = 1$, $p = 2$, $p = 3$ and $p = 5$.

By employing the proposed SS-MISG algorithm to identify the system parameters in this example, the parameter estimation errors versus t are illustrated in Fig. 5. The parameter estimates are shown in Tables 3–4. The parameter estimation distributions after recursion 6000 are illustrated in Figs. 6–7.

In order to test the performance of the identified system model of Example 3, we apply an M sequence pseudo random signal with the amplitude $[-1, 1]$ to the true system and the estimated system.

Table 3
The SS-MISG parameter estimates of Θ_1 and Θ_2.

p	t	$a1_{11}$	$a1_{12}$	$a1_{21}$	$a1_{22}$	$a2_{11}$	$a2_{12}$	$a2_{21}$	$a2_{22}$
1	100	0.66582	0.17055	−0.15701	0.12561	0.44793	0.23930	−0.66582	0.25523
	500	0.69463	0.22937	−0.15790	0.18512	0.54828	0.30035	−0.63282	0.24260
	2000	0.75656	0.21946	−0.13717	0.21885	0.60967	0.35113	−0.62584	0.26257
	4000	0.77278	0.22634	−0.13525	0.22828	0.61584	0.34994	−0.62956	0.27740
	6000	0.77782	0.21882	−0.13426	0.23341	0.62125	0.35137	−0.62972	0.27339
2	100	0.95838	0.31480	−0.35928	0.05446	0.69080	0.34190	−0.81259	0.19170
	500	0.84359	0.28580	−0.25452	0.21006	0.64752	0.36121	−0.67339	0.22186
	2000	0.84810	0.24507	−0.18968	0.26133	0.68486	0.38216	−0.65262	0.26208
	4000	0.85391	0.24637	−0.18432	0.27562	0.66822	0.37670	−0.65023	0.27450
	6000	0.85473	0.23386	−0.18329	0.28151	0.66704	0.37187	−0.64828	0.27234
3	100	0.79464	0.22299	−0.12997	0.12411	0.50851	0.29223	−0.57688	0.25359
	500	0.79933	0.27011	−0.11955	0.24004	0.60871	0.32800	−0.58912	0.27558
	2000	0.84577	0.23173	−0.12484	0.28963	0.68275	0.37599	−0.61521	0.27533
	4000	0.86007	0.23595	−0.13112	0.29712	0.66451	0.37316	−0.62319	0.28748
	6000	0.86235	0.22391	−0.13590	0.30327	0.66613	0.36924	−0.62434	0.28406
5	100	0.99372	0.26490	−0.02264	0.14072	0.66620	0.34282	−0.44192	0.30390
	500	0.96386	0.27676	−0.09048	0.25398	0.68114	0.36883	−0.56946	0.29453
	2000	0.93580	0.23392	−0.12820	0.30577	0.73113	0.38229	−0.61525	0.27726
	4000	0.94056	0.23281	−0.13615	0.31293	0.69619	0.38514	−0.62520	0.28778
	6000	0.93603	0.22123	−0.14342	0.31972	0.69678	0.37942	−0.62665	0.28576
True values		0.94000	0.23000	−0.21000	0.35000	0.70000	0.38000	−0.65000	0.28000

Table 4
The SS-MISG parameter estimates of Ω_1 and Ω_2.

p	t	$b1_{11}$	$b1_{12}$	$b1_{21}$	$b1_{22}$	$b2_{11}$	$b2_{12}$	$b2_{21}$	$b2_{22}$
1	100	0.43776	−0.27206	−0.26506	0.82250	0.34748	−0.17543	0.29257	0.75087
	500	0.61594	−0.32893	−0.29019	0.95954	0.46466	−0.19345	0.26881	1.00685
	2000	0.74123	−0.37407	−0.31583	1.04316	0.56563	−0.18746	0.23703	1.18074
	4000	0.79675	−0.39386	−0.32624	1.08509	0.60916	−0.19310	0.22460	1.25944
	6000	0.82599	−0.40299	−0.33051	1.10516	0.63434	−0.19246	0.21935	1.30067
2	100	0.86501	−0.60049	−0.48104	0.99797	0.39942	−0.03640	0.14738	0.98258
	500	0.98569	−0.56165	−0.44967	1.13801	0.63544	−0.11072	0.09754	1.30546
	2000	1.05358	−0.55654	−0.43617	1.21366	0.76712	−0.13726	0.07850	1.50441
	4000	1.08407	−0.55200	−0.43162	1.24695	0.81641	−0.15899	0.07394	1.58505
	6000	1.09689	−0.55023	−0.42769	1.26221	0.84297	−0.16554	0.07337	1.62583
3	100	0.72157	−0.29378	−0.35823	0.96830	0.40192	−0.13739	0.50811	0.96321
	500	0.92341	−0.39197	−0.36995	1.15642	0.64447	−0.19011	0.37305	1.36331
	2000	1.03347	−0.45470	−0.38753	1.24232	0.80029	−0.19308	0.27195	1.57936
	4000	1.07754	−0.47404	−0.39368	1.27818	0.85731	−0.21073	0.23381	1.66379
	6000	1.09503	−0.48372	−0.39593	1.29238	0.88682	−0.21413	0.21592	1.70427
5	100	0.88532	−0.20657	−0.25880	1.20271	1.04058	−0.23355	0.61162	0.87555
	500	1.05395	−0.38129	−0.33330	1.29807	1.09794	−0.27116	0.42761	1.42876
	2000	1.12449	−0.46667	−0.37696	1.33388	1.12043	−0.26477	0.27514	1.67903
	4000	1.15502	−0.48611	−0.38914	1.35099	1.12602	−0.27890	0.22035	1.76559
	6000	1.16175	−0.49891	−0.39444	1.35567	1.12914	−0.27991	0.19527	1.80465
True values		1.21000	−0.53000	−0.42000	1.38000	1.13000	−0.29000	0.22000	1.82000

Applying the input sequence to the estimated system which is obtained by the parameter estimates when $t = 6000$ and $p = 5$. Adding the noise to the inputs with different combination by different noise variance and taking the outputs of the estimated system and the true system from the first 100 s, the system outputs $\hat{y}_1(t)$, $y_1(t)$, $\hat{y}_2(t)$ and $y_2(t)$ are illustrated in Figs. 8–9.

Remark 10. From the simulation results of Example 3, we can see that the proposed SS-MISG algorithm is effective for identifying the multivariable system with large amounts of parameters. Tables 3–4 show that the estimation accuracy when $p = 5$ is higher than those when $p = 1$, $p = 2$ and $p = 3$, which means that the multi-innovation can improve the estimation accuracy effectively. Fig. 5 shows that the parameter estimation accuracy becomes smaller gradually with the increasing of t, which means that the proposed SS-MISG algorithm has good convergence from the view of experiment. After the system response test between the true system and the identified system, we can conclude that the estimated system can track the system dynamical characteristic of the true system from Figs. 8–9.

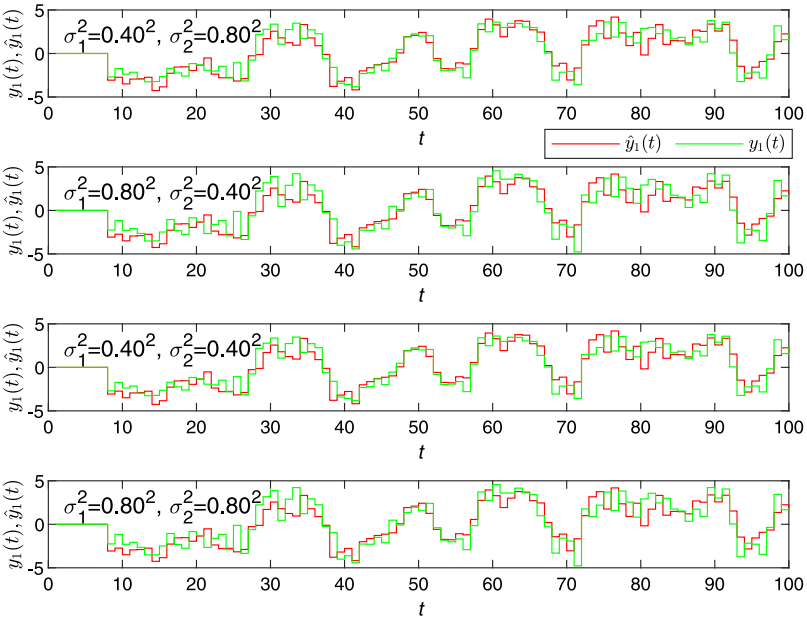

Fig. 8. The system output $y_1(t)$ and model output $\hat{y}_1(t)$.

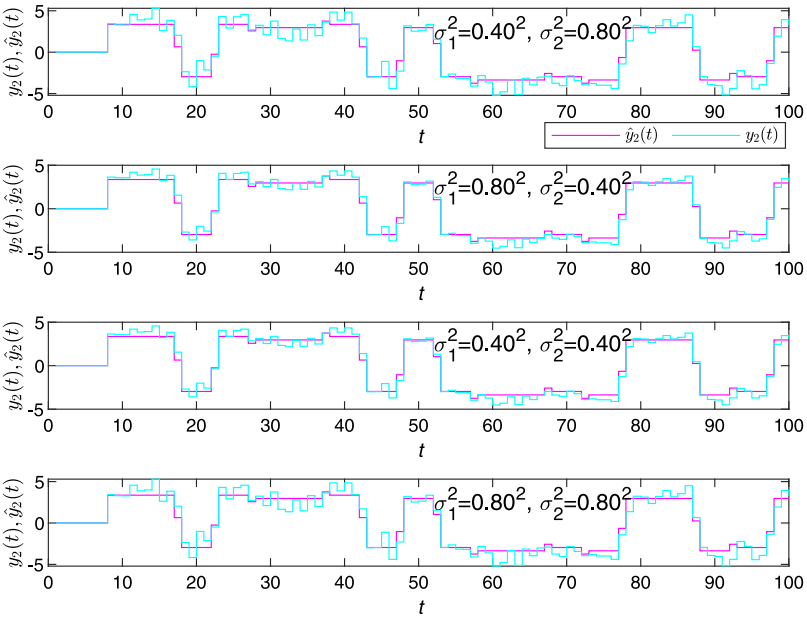

Fig. 9. The system output $y_2(t)$ and model output $\hat{y}_2(t)$.

Remark 11. The separation principle is the main innovation of the proposed SS-SG algorithm, the FF-SS-SG algorithm and the SS-MISG algorithm. For the identification problem of large-scale systems, the separable synthesis technique is inspired by the hierarchical control for multi-level systems in industrial processed, which can reduce the complexity resulting from the numerous parameters, long instant transmission and complicated structure. Expanding the hierarchical control into the field of system identification can lower the calculated amount of the identification algorithm obviously.

6. Conclusions

This paper has proposed a novel SS-SG identification algorithm and an SS-MISG identification algorithm for the large-scale systems with multiple inputs and outputs by combining the model separation and the multi-innovation principle.

Moreover, the convergence analysis of these proposed identification algorithm is provided by the Martingale convergence theorem. Because the proposed algorithms are developed by means of the dynamical observations, these algorithms can by used in on-line identification for all kinds of large-scale systems such as the communication systems, complicated industrial processes and intelligent power systems. The separable gradient identification approaches proposed in this paper can integrate some relevant identification approaches [103–108] to study new parameter identification of dynamical stochastic systems and can be applied to other literatures [109–114] such as paper-making and control engineering systems and information processing systems [115–121] and so on.

Data availability

No data was used for the research described in the article.

References

[1] Y. Cao, Z. Wang, F. Liu, P. Li, G. Xie, Bio-inspired speed curve optimization and sliding mode tracking control for subway trains, IEEE Trans. Veh. Technol. 68 (7) (2019) 6331–6342.
[2] 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.
[3] 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.
[4] Y. Cao, L.C. Ma, S. Xiao, et al., Standard analysis for transfer delay in CTCS-3, Chin. J. Electron. 26 (5) (2017) 1057–1063.
[5] J.Y. You, C.P. Yu, J. Sun, J. Chen, Generalized maximum entropy based identification of graphical ARMA models, Automatica 141 (2022) 110319.
[6] C.P. Yu, Y. Li, H. Fang, J. Chen, System identification approach for inverse optimal control of finite-horizon linear quadratic regulators, Automatica 129 (2021) 109636.
[7] M.H. Li, X.M. Liu, Maximum likelihood hierarchical least squares-based iterative identification for dual-rate stochastic systems, Internat. J. Adapt. Control Signal Process. 35 (2) (2021) 240–261.
[8] Z.W. Shi, H.D. Yang, M. Dai, The data-filtering based bias compensation recursive least squares identification for multi-input single-output systems with colored noises, J. Franklin Inst. 360 (2023) http://dx.doi.org/10.1016/j.jfranklin.2023.01.040.
[9] M.H. Li, X.M. Liu, The filtering-based maximum likelihood iterative estimation algorithms for a special class of nonlinear systems with autoregressive moving average noise using the hierarchical identification principle, Internat. J. Adapt. Control Signal Process. 33 (7) (2019) 1189–1211.
[10] 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.
[11] H. Ma, J. Pan, W. Ding, Partially-coupled least squares based iterative parameter estimation for multi-variable output-error-like autoregressive moving average systems, IET Control Theory Appl. 13 (18) (2019) 3040–3051.
[12] J.X. Ma, W.L. Xiong, J. Chen, et al., Hierarchical identification for multivariate Hammerstein systems by using the modified Kalman filter, IET Control Theory Appl. 11 (6) (2017) 857–869.
[13] 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, Internat. J. Adapt. Control Signal Process. 35 (8) (2021) 1562–1577.
[14] J.W. Wang, Y. Ji, X. Zhang, et al., Two-stage gradient-based iterative algorithms for the fractional-order nonlinear systems by using the hierarchical identification principle, Internat. J. Adapt. Control Signal Process. 36 (7) (2022) 1778–1796.
[15] L. Xu, F.Y. Chen, T. Hayat, Hierarchical recursive signal modeling for multi-frequency signals based on discrete measured data, Internat. J. Adapt. Control Signal Process. 35 (5) (2021) 676–693.
[16] Y. Shen, T.J. Ypma, Solving separable nonlinear least squares problems using the QR factorization, J. Comput. Appl. Math. 345 (2019) 48–58.
[17] L. Xu, F. Ding, Q.M. Zhu, Separable synchronous multi-innovation gradient based iterative signal modeling from on-line measurements, IEEE Trans. Instrum. Meas. 71 (2022) 6501313.
[18] Y.H. Zhou, Partially-coupled nonlinear parameter optimization algorithm for a class of multivariate hybrid models, Appl. Math. Comput. 414 (2022) 126663.
[19] Y. Ji, Z. Kang, Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems, Internat. J. Robust Nonlinear Control 31 (3) (2021) 971–987.
[20] Y. Ji, Z. Kang, X.M. Liu, The data filtering based multiple-stage levenberg–marquardt algorithm for Hammerstein nonlinear systems, Internat. J. Robust Nonlinear Control 31 (15) (2021) 7007–7025.
[21] Y.J. Wang, S.H. Tang, M.Q. Deng, Modeling nonlinear systems using the tensor network B-spline and the multi-innovation identification theory, Internat. J. Robust Nonlinear Control 32 (13) (2022) 7304–7318.
[22] Y.J. Wang, S.H. Tang, X.B. Gu, Parameter estimation for nonlinear Volterra systems by using the multi-innovation identification theory and tensor decomposition, J. Franklin Inst. 359 (2) (2022) 1782–1802.
[23] Y.J. Wang, L. Yang, An efficient recursive identification algorithm for multilinear systems based on tensor decomposition, Internat. J. Robust Nonlinear Control 31 (16) (2021) 7920–7936.
[24] M. Verrelli, A. Savoia, M. Mengoni, et al., On-line identification of winding resistances and load torque in induction machines, IEEE Trans. Control Syst. Technol. 22 (4) (2014) 1629–1637.
[25] F. Ding, L. Lv, J. Pan, et al., Two-stage gradient-based iterative estimation methods for controlled autoregressivevsystems using the measurement data, Int. J. Control Autom. Syst. 18 (4) (2020) 886–896.
[26] M.H. Li, X.M. Liu, Iterative identification methods for a class of bilinear systems by using the particle filtering technique, Internat. J. Adapt. Control Signal Process. 35 (10) (2021) 2056–2074.
[27] 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.
[28] M.H. Li, X.M. Liu, Particle filtering-based iterative identification methods for a class of nonlinear systems with interval-varying measurements, Int. J. Control Autom. Syst. 20 (7) (2022) 2239–2248.
[29] Y.M. Fan, X.M. Liu, Auxiliary model-based multi-innovation recursive identification algorithms for an input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity, Internat. J. Adapt. Control Signal Process. 36 (3) (2022) 690–707.
[30] 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, Internat. J. Robust Nonlinear Control 31 (9) (2021) 4017–4036.

[31] F. Ding, F.F. Wang, L. Xu, M.H. Wu, Decomposition based least squares iterative identification algorithm for multivariate pseudo-linear ARMA systems using the data filtering, J. Franklin Inst. 354 (3) (2017) 1321–1339.
[32] F. Tedesco, A. Casavola, G. Fedele, Unbiased estimation of sinusoidal signal parameters via discrete-time frequency-locked-loop filters, IEEE Trans. Automat. Control 62 (3) (2017) 1484–1490.
[33] L. Li, H. Zhang, X. Ren, J. Zhang, A novel recursive learning identification scheme for Box-Jenkins model based on error data, Appl. Math. Model. 90 (2021) 200–216.
[34] S.C. Chan, J.Q. Lin, X. Sun, et al., A new variable forgetting factor-based bias-compensation algorithm for recursive identification of time-varying multi-input single-output systems with measurement noise, IEEE Trans. Instrum. Meas. 69 (7) (2020) 4555–4568.
[35] Z. Kang, Y. Ji, X.M. Liu, Hierarchical recursive least squares algorithms for Hammerstein nonlinear autoregressive output-error systems, Internat. J. Adapt. Control Signal Process. 35 (11) (2021) 2276–2295.
[36] T. Lewis, Q. Morris, Y. Zhang, Convergence, stability analysis and solvers for approximating sublinear positone and semipositone boundary value problems using finite difference methods, J. Comput. Appl. Math. 404 (2022) 113880.
[37] Z. Song, H. Hofmann, X. Lin, X. Han, J. Hou, Parameter identification of lithium-ion battery pack for different applications based on Cramer–Rao bound analysis and experimental study, Appl. Energy 231 (2018) 1307–1318.
[38] L. Schmitt, W. Fichter, Cramér-rao lower bound for state-constrained nonlinear filtering, IEEE Signal Process. Lett. 24 (12) (2017) 1882–1885.
[39] F. Ding, System Identification Performances Analysis for Identification Methods, Science Press, Beijing, 2014.
[40] L. Xu, F. Ding, E.F. Yang, Auxiliary model multiinnovation stochastic gradient parameter estimation methods for nonlinear sandwich systems, Internat. J. Robust Nonlinear Control 31 (1) (2021) 148–165.
[41] F. Ding, H.Z. Yang, F. Liu, Performance analysis of stochastic gradient algorithms under weak conditions, Sci. China Ser. F Inf. Sci. 51 (9) (2008) 1269–1280.
[42] 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.
[43] Y. Ji, C. Zhang, Z. Kang, T. Yu, Parameter estimation for block-oriented nonlinear systems using the key term separation, Internat. J. Robust Nonlinear Control 30 (9) (2020) 3727–3752.
[44] J.M. Li, F. Ding, T. Hayat, A novel nonlinear optimization method for fitting a noisy Gaussian activation function, Internat. J. Adapt. Control Signal Process. 36 (3) (2022) 690–707.
[45] Y. Ji, X.K. Jiang, L.J. Wan, Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems, J. Franklin Inst. 357 (8) (2020) 5019–5032.
[46] H.B. Liu, J.W. Wang, Y. Ji, Maximum likelihood recursive generalized extended least squares estimation methods for a bilinear-parameter systems with ARMA noise based on the over-parameterization model, Int. J. Control Autom. Syst. 20 (8) (2022) 2606–2615.
[47] J. Pan, Y.Q. Liu, J. Shu, Gradient-based parameter estimation for an exponential nonlinear autoregressive time-series model by using the multi-innovation, Int. J. Control Autom. Syst. 21 (1) (2023) 140–150.
[48] Y.J. Wang, F. Ding, M.H. Wu, Recursive parameter estimation algorithm for multivariate output-error systems, J. Franklin Inst. 355 (12) (2018) 5163–5181.
[49] X. Meng, Y. Ji, J. Wang, Iterative parameter estimation for photovoltaic cell models by using the hierarchical principle, Int. J. Control Autom. Syst. 20 (8) (2022) 2583–2593.
[50] J.X. Ma, F. Ding, Filtering-based multistage recursive identification algorithm for an input nonlinear output-error autoregressive system by using the key term separation technique, Circuits Syst. Signal Process. 36 (2) (2017) 577–599.
[51] Y. Ji, A.N. Jiang, Filtering-based accelerated estimation approach for generalized time-varying systems with disturbances and colored noises, IEEE Trans. Circuits Syst. II: Express Briefs 70 (1) (2023) 206–210.
[52] P. Ma, F. Ding, New gradient based identification methods for multivariate pseudo-linear systems using the multi-innovation and the data filtering, J. Franklin Inst. 354 (3) (2017) 1568–1583.
[53] J. Pan, S.D. Liu, J. Shu, X.K. Wan, Hierarchical recursive least squares estimation algorithm for secondorder Volterra nonlinear systems, Int. J. Control Autom. Syst. 20 (12) (2022) 3940–3950.
[54] 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.
[55] J.X. Xiong, J. Pan, G. Chen, et al., Sliding mode dual-channel disturbance rejection attitude control for a quadrotor, IEEE Trans. Ind. Electron. 69 (10) (2022) 10489–10499.
[56] J. Pan, Q. Chen, J. Xiong, G. Chen, A novel quadruple-boost nine-level switched capacitor inverter, J. Electr. Eng. Technol. 18 (1) (2023) 467–480.
[57] F. Ding, H. Ma, J. Pan, E.F. Yang, Hierarchical gradient- and least squares-based iterative algorithms for input nonlinear output-error systems using the key term separation, J. Franklin Inst. B 358 (9) (2021) 5113–5135.
[58] Y. Gu, Q.M. Zhu, H. Nouri, Identification and U-control of a state-space system with time-delay, Internat. J. Adapt. Control Signal Process. 36 (1) (2022) 138–154.
[59] J.L. Ding, W.H. Zhang, Finite-time adaptive control for nonlinear systems with uncertain parameters based on the command filters, Internat. J. Adapt. Control Signal Process. 35 (9) (2021) 1754–1767.
[60] P. Ma, L. Wang, Filtering-based recursive least squares estimation approaches for multivariate equation-error systems by using the multiinnovation theory, Internat. J. Adapt. Control Signal Process. 35 (9) (2021) 1898–1915.
[61] 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) 109034.
[62] Y. Cao, Y. An, S. Su, G. Xie, Y. Sun, A statistical study of railway safety in China and Japan 1990–2020, Accid. Anal. Prev. 175 (2022) 106764.
[63] Y. Cao, Y.R. Yang, L.C. Ma, J.K. Wen, Research on virtual coupled train control method based on GPC & VAPF, Chin. J. Electron. 31 (5) (2022) 897–905.
[64] Y.K. Sun, Y. Cao, P. Li, Contactless fault diagnosis for railway point machines based on multi-scale fractional wavelet packet energy entropy and synchronous optimization strategy, IEEE Trans. Veh. Technol. 71 (6) (2022) 5906–5914.
[65] Y. Cao, Y.S. Ji, Y.K. Sun, S. Su, The fault diagnosis of a switch machine based on deep random forest fusion, IEEE Intell. Transp. Syst. Mag. 15 (1) 20230, 437-452.
[66] X. Wang, S. Su, Y. Cao, X.L. Wang, Robust control for dynamic train regulation in fully automatic operation system under uncertain wireless transmissions, IEEE Trans. Intell. Transp. Syst. (2023) http://dx.doi.org/10.1109/TITS.2022.3170950.
[67] 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. 23 (10) (2022) 17666–17676.
[68] 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. 23 (8) (2022) 12074–12083.
[69] Y. Cao, J.K. Wen, A. Hobiny, P. Li, T. Wen, Parameter-varying artificial potential field control of virtual coupling system with nonlinear dynamics, Fractals 30 (2) (2022) 2240099.
[70] 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.

[71] Y.K. Sun, Y. Cao, L.C. Ma, A fault diagnosis method for train plug doors via sound signals, IEEE Intell. Transp. Syst. Mag. 13 (3) (2021) 107–117.
[72] 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.
[73] G.C. Goodwin, K.S. Sin, Adaptive Filtering Prediction and Control, Prentice Hall, Englewood Cliffs, New Jersey, 1984.
[74] F.Z. Geng, X.Y. Wu, A novel kernel functions algorithm for solving impulsive boundary value problems, Appl. Math. Lett. 134 (2022) 108318.
[75] X.Y. Li, B.Y. Wu, A kernel regression approach for identification of first order differential equations based on functional data, Appl. Math. Lett. 127 (2022) 107832.
[76] H. Wang, H. Fan, J. Pan, A true three-scroll chaotic attractor coined, Discr. Contin. Dyn. Syst. Ser. B 27 (5) (2022) 2891–2915.
[77] C.C. Yin, Y.Z. Wen, An extension of Paulsen-Gjessing's risk model with stochastic return on investments, Insurance Math. Econom. 52 (3) (2013) 469–476.
[78] C.C. Yin, J.S. Zhao, Nonexponential asymptotics for the solutions of renewal equations, with applications, J. Appl. Probab. 43 (3) (2006) 815–824.
[79] C.C. Yin, K.C. Yuen, Optimality of the threshold dividend strategy for the compound Poisson model, Statist. Probab. Lett. 81 (12) (2011) 1841–1846.
[80] 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.
[81] M. Li, G. Xu, Q. Lai, J. Chen, A chaotic strategy-based quadratic opposition-based learning adaptive variable-speed whale optimization algorithm, Math. Comput. Simulation 193 (2022) 71–99.
[82] M.D. Li, G.H. Xu, L. Zeng, Q. Lai, Hybrid whale optimization algorithm based on symbiosis strategy for global optimization, Appl. Intell. 202 (2023).
[83] C. Xu, H. Xu, Z.H. Guan, Y. Ge, Observer-based dynamic event-triggered semi-global bipartite consensus of linear multi-agent systems with input saturation, IEEE Trans. Cybern. (2023) http://dx.doi.org/10.1109/TCYB.2022.3164048.
[84] C. Xu, Y. Qin, H. Su, Observer-based dynamic event-triggered bipartite consensus of discrete-time multi-agent systems, IEEE Trans. Circuits Syst. II: Express Briefs (2023) http://dx.doi.org/10.1109/TCSII.2022.3217918.
[85] 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.
[86] J. Hou, H. Su, C.P. Yu, F.W. Chen, P.H. Li, Bias-correction errors-in-variables Hammerstein model identification, IEEE Trans. Ind. Electron. (2023) http://dx.doi.org/10.1109/TIE.2022.3199931.
[87] J. Hou, H. Su, C.P. Yu, F.W. Chen, P.H. Li, H.F. Xie, T.F. Li, Consistent subspace identification of errors-in-variables Hammerstein systems, IEEE Trans. Syst. Man Cybern. Syst. (2023) http://dx.doi.org/10.1109/TSMC.2022.3213809.
[88] Y.F. Chen, C. Zhang, C.Y. Liu, Y.M. Wang, X.K. Wan, Atrial fibrillation detection using feedforward neural network, J. Med. Biolog. Eng. 42 (1) (2022) 63–73.
[89] Y. Wang, G. Yang, Arrhythmia classification algorithm based on multi-head self-attention mechanism, Biomed. Signal Process. Control 79 (2023) 104206.
[90] Y. Li, G.C. Yang, Z.D. Su, et al., Human activity recognition based on multienvironment sensor data, Inf. Fusion 91 (2023) 47–63.
[91] J. Lin, Y. Li, G.C. Yang, FPGAN: Face de-identification method with generative adversarial networks for social robots, Neural Netw. 133 (2021) 132–147.
[92] G.C. Yang, Z.J. Chen, Y. Li, Z.D. Su, Rapid relocation method for mobile robot based on improved ORB-SLAM2 algorithm, Remote Sens. 11 (2) (2019) 149.
[93] L. Xu, F. Ding, Q. Zhu, Decomposition strategy-based hierarchical least mean square algorithm for control systems from the impulse responses, Internat. J. Systems Sci. 52 (9) (2021) 1806–1821.
[94] X. Zhang, Hierarchical parameter and state estimation for bilinear systems, Internat. J. Systems Sci. 51 (2) (2020) 275–290.
[95] X. Zhang, State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors, Internat. J. Adapt. Control Signal Process. 33 (7) (2019) 1157–1173.
[96] X. Zhang, Optimal adaptive filtering algorithm by using the fractional-order derivative, IEEE Signal Process. Lett. 29 (2022) 399–403.
[97] Y.H. Zhou, Modeling nonlinear processes using the radial basis function-based state-dependent autoregressive models, IEEE Signal Process. Lett. 27 (2020) 1600–1604.
[98] 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.
[99] L. Xu, F. Ding, L.J. Wan, J. Sheng, Separable multi-innovation stochastic gradient estimation algorithm for the nonlinear dynamic responses of systems, Internat. J. Adapt. Control Signal Process. 34 (7) (2020) 937–954.
[100] L. Xu, F. Ding, X. Lu, L.J. Wan, J. Sheng, Hierarchical multi-innovation generalised extended stochastic gradient methods for multivariable equation-error autoregressive moving average systems, IET Control Theory Appl. 14 (10) (2020) 1276–1286.
[101] F. Ding, X.P. Liu, Auxiliary model based stochastic gradient algorithm for multivariable output error systems, Acta Automat. Sinica 36 (7) (2010) 993–998.
[102] F. Ding, X. Zhang, L. Xu, The innovation algorithms for multivariable state-space models, Internat. J. Adapt. Control Signal Process. 33 (11) (2019) 1601–1618.
[103] H. Ma, A novel multi-innovation gradient support vector machine regression method, ISA Trans. 130 (2022) 343–359.
[104] T. Cui, Moving data window-based partially-coupled estimation approach for modeling a dynamical system involving unmeasurable states, ISA Trans. 128 (2022) 437–452.
[105] Y.M. Fan, Two-stage auxiliary model gradient-based iterative algorithm for the input nonlinear controlled autoregressive system with variable-gain nonlinearity, Internat. J. Robust Nonlinear Control 30 (14) (2020) 5492–5509.
[106] X.H. Wang, Modified particle filtering-based robust estimation for a networked control system corrupted by impulsive noise, Internat. J. Robust Nonlinear Control 32 (2) (2022) 830–850.
[107] J. Chen, Varying infimum gradient descent algorithm for agent-server systems with uncertain communication network, IEEE Trans. Instrum. Meas. 70 (2021) 9510511.
[108] L.J. Wan, Decomposition- and gradient-based iterative identification algorithms for multivariable systems using the multi-innovation theory, Circuits Systems Signal Process. 38 (7) (2019) 2971–2991.
[109] S. Su, J.F. She, K.C. Li, X. Wang, Y. Zhou, 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.
[110] S. Su, T. Tang, J. Xun, F. Cao, Y.H. Wang, 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.
[111] S.Y. Liu, Hierarchical principle-based iterative parameter estimation algorithm for dual-frequency signals, Circuits Syst. Signal Process. 38 (7) (2019) 3251–3268.
[112] S. Su, X.K. Wang, T. Tang, G. Wang, Y. Cao, Energy-efficient operation by cooperative control among trains: A multi-agent reinforcement learning approach, Control Eng. Pract. 116 (2021) 104901.

[113] 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. Inform. (2023) http://dx.doi.org/10.1109/TII.2022.3195888.
[114] X. Jin, Z. Wang, J. Kong, Y. Bai, T. Su, H. Ma, P. Chakrabarti, Deep spatio-temporal graph network with self-optimization for air quality prediction, Entropy 25 (2023) 247, http://dx.doi.org/10.3390/e25020247.
[115] X. Jin, Z. Wang, W. Gong, J. Kong, Y. Bai, T. Su, H. Ma, P. Chakrabarti, Variational bayesian network with information interpretability filtering for air quality forecasting, Mathematics 11 (4) (2023) 837, http://dx.doi.org/10.3390/math11040837.
[116] J. Pan, X. Jiang, X. Wan, W. Ding, A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems, Int. J. Control Autom. Syst. 15 (3) (2017).
[117] N. Zhao, A. Wu, Y. Pei, Y. Liang, D. Niyato, Patial-temporal aggregation graph convolution network for efficient mobile cellular traffic prediction, IEEE Commun. Lett 26 (3) (2022) 587–591.
[118] J. Pan, W. Li, H. 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.
[119] F. Ding, Least squares parameter estimation and multi-innovation least squares methods for linear fitting problems from noisy data, J. Comput. Appl. Math. (2023) 115107, http://dx.doi.org/10.1016/j.cam.2023.115107.
[120] X. Zhang, Highly computationally efficient state filter based on the delta operator, Internat. J. Adapt. Control Signal Process. 33 (6) (2019) 875–889.
[121] J.M. Li, Synchronous optimization schemes for dynamic systems through the kernel-based nonlinear observer canonical form, IEEE Trans. Instrum. Meas. 71 (2022) 3210952.