Stochastic gradient identification methods for multivariate pseudo-linear systems using the multi-innovation and the data filtering

Ping Ma[a], Feng Ding[*,a], Quanmin Zhu[b]

[a]*Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, PR China*
[b]*Department of Engineering Design and Mathematics, Bristol BS16 1QY, University of the West of England*

Abstract

This paper researches the parameter identification methods of multivariate pseudo-linear moving average systems. A multivariate generalized stochastic gradient (M-GSG) algorithm is presented as a comparison firstly. In order to improve the parameter estimation accuracy, a multivariate multi-innovation generalized stochastic gradient (M-MI-GSG) algorithm and a filtering based multivariate generalized stochastic gradient (F-M-GSG) algorithm are presented by means of the multi-innovation identification theory and the data filtering technique. The simulation results confirm that the proposed algorithms are more effective than the M-GSG algorithm.

Key words: Parameter estimation, Gradient search, Multi-innovation, Data filtering identification, Multivariate system

---

1. Introduction

For decades, system modeling and identification have extensive applications in industrial control [1, 2, 3]. The mathematical models must be obtained first for control problems from input-output data. Many practical processes are multivariable and thus multivariate systems are causing people's attention [4, 5, 6]. There are many identification methods for identifying multivariate systems. For example, Anderson et al. discussed the structure identification of multivariate auto-regressive (AR) and auto-regressive moving average (ARMA) systems and the spectral density estimation and identifiability of ARMA systems from mixed frequency data [7]. For multivariate systems with colored noise, Wang et al. presented a hierarchical identification algorithm by decomposing a multivariate system into several subsystems [8]; Wang and Ding proposed the recursive parameter estimation algorithms for multivariate Box-Jenkins systems [9].

The multi-innovation identification theory is an important branch of system identification [10, 11,

[☆]This work was supported by the National Natural Science Foundation of China.
[*]Corresponding author.
Email addresses: `pingma12@126.com` (Ping Ma), `fding@jiangnan.edu.cn` (Feng Ding), `quan.zhu@uwe.ac.uk` (Quanmin Zhu)

12], the innovation is the useful information that can improve the accuracy of the parameter estimation or the state estimation [13, 14]. In this aspect, Mao and Ding presented a multi-innovation stochastic gradient algorithm for Hammerstein nonlinear systems [15]; Jin et al. proposed a multi-innovation least squares identification algorithm for multivariable output-error systems with scarce measurements [16]; Wang and Zhu presented a multi-innovation stochastic gradient algorithm for a class of linear-in-parameter systems [17]. Shen and Ding presented a hierarchical multi-innovation extended stochastic gradient algorithm for input nonlinear multivariable output-error moving average (OEMA) systems by the key-term separation principle [18].

The data filtering technique is effective for picking up the useful signals from noisy measurements and has been used in system identification [19, 20, 21]. The main idea is a special filter to filter the input and output data, and then to identify the filtered system model and the filtered noise model with each other for generating corresponding parameter estimates [22, 23]. Ding et al. presented a recursive least squares parameter estimation algorithm for output-error autoregressive moving average system by using the data filtering technique and the auxiliary model identification idea [24]; Wang combine the auxiliary model identification idea with the filtering theory, transform an OEMA system into two identification models and presented a filtering and auxiliary model based recursive least squares identification algorithm [25].

For multivariate pseudo-linear moving average system, this paper gives a multivariate multi-innovation generalized stochastic gradient algorithm and a filtering based multivariate generalized stochastic gradient algorithm. The basic idea is that deriving the mathematical model of multivariate systems firstly, and then using the negative gradient search principle, combing with the multi-innovation identification theory or data filtering technique, obtain corresponding algorithms, to overcome the influence of colored noise.

The rest of this paper is organized as follows. Section 2 introduces the identification problems for multivariate pseudo-linear moving average system. Section 3 gives a generalized stochastic gradient algorithm for comparison. Section 4 derives a multi-innovation stochastic gradient algorithm. Section 5 presents a filtering based stochastic gradient algorithm. Section 6 provides two examples for illustrating the results in this paper. Finally, we offer some concluding remarks in Section 7.

2. System description and identification model

First of all we give some notation in this paper. $\boldsymbol{I}_m$ denotes an identity matrix of size $m \times m$; $\mathbf{1}_n$ stands for an n-dimensional column vector whose elements are 1, that is $\mathbf{1}_n = [1, 1, \cdots, 1]^{\mathrm{T}} \in \mathbb{R}^n$; $\mathbf{1}_{m\times n}$ represents a matrix of size $m \times n$ whose elements are 1; the norm of a matrix $\boldsymbol{X}$ is defined by $\|\boldsymbol{X}\|^2 := \mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$, the superscript T stands for the matrix/vector transpose; the symbol $\otimes$ represents kronecker product, for example, $\boldsymbol{A} := [a_{ij}] \in \mathbb{R}^{m\times n}$, $\boldsymbol{B} := [b_{ij}] \in \mathbb{R}^{p\times q}$, $\boldsymbol{A}\otimes\boldsymbol{B} = [a_{ij}\boldsymbol{B}] \in \mathbb{R}^{(mp)\times(nq)}$, in general, $\boldsymbol{A}\otimes\boldsymbol{B} \neq \boldsymbol{B}\otimes\boldsymbol{A}$; col$[\boldsymbol{X}]$ is defined as a vector consist of all columns of matrix $\boldsymbol{X}$ arranged in order, for example, $\boldsymbol{X} := [\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n] \in \mathbb{R}^{m\times n}$, $\boldsymbol{x}_i \in \mathbb{R}^m$ $(i = 1, 2, \cdots, n)$, col$[\boldsymbol{X}] := [\boldsymbol{x}_1^{\mathrm{T}},\boldsymbol{x}_2^{\mathrm{T}},\cdots,\boldsymbol{x}_n^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{mn}$.

Consider the following multivariate pseudo-linear moving average system,

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}_{\mathrm{s}}(t)\boldsymbol{\theta} + \boldsymbol{C}^{-1}(z)\boldsymbol{v}(t), \tag{1}$$

where $\boldsymbol{y}(t) := [y_1(t), y_2(t), \cdots, y_m(t)]^{\mathrm{T}} \in \mathbb{R}^m$ is the system output vector, $\boldsymbol{\Phi}_{\mathrm{s}}(t) \in \mathbb{R}^{m\times n}$ is the information matrix consist of the input-output data, $\boldsymbol{\theta} \in \mathbb{R}^n$ is the system parameter vector to be identified, $\boldsymbol{v}(t) := [v_1(t),v_2(t),\cdots,v_m(t)]^{\mathrm{T}} \in \mathbb{R}^m$ is a white noise process with zero mean, $\boldsymbol{C}(z) \in \mathbb{R}^{m\times m}$ is a polynomial matrix in the unit backward shift operator $[z^{-1}\boldsymbol{y}(t) = \boldsymbol{y}(t-1)]$,

$$\boldsymbol{C}(z) := \boldsymbol{I}_m + \boldsymbol{C}_1 z^{-1} + \boldsymbol{C}_2 z^{-2} + \cdots + \boldsymbol{C}_{n_c} z^{-n_c}, \quad \boldsymbol{C}_i \in \mathbb{R}^{m\times m}.$$

Without loss of generality, assume that the orders m, n and n_c are known, and $\boldsymbol{y}(t) = \boldsymbol{0}$, $\boldsymbol{\Phi}_{\mathrm{s}}(t) = \boldsymbol{0}$ and $\boldsymbol{v}(t) = \boldsymbol{0}$ for $t \leqslant 0$.

Define the middle vector $\boldsymbol{w}(t)$, the parameter matrix $\boldsymbol{\theta}_{\mathrm{n}}$ and the information vector $\boldsymbol{\psi}(t)$ as

$$\begin{aligned} \boldsymbol{w}(t) &:= \boldsymbol{C}^{-1}(z)\boldsymbol{v}(t) \in \mathbb{R}^m, \\ \boldsymbol{\theta}_{\mathrm{n}} &:= [\boldsymbol{C}_1, \boldsymbol{C}_2, \cdots, \boldsymbol{C}_{n_c}]^{\mathrm{T}} \in \mathbb{R}^{(mn_c)\times m}, \\ \boldsymbol{\psi}(t) &:= [-\boldsymbol{w}^{\mathrm{T}}(t-1), -\boldsymbol{w}^{\mathrm{T}}(t-2), \cdots, -\boldsymbol{w}^{\mathrm{T}}(t-n_c)]^{\mathrm{T}} \in \mathbb{R}^{mn_c}. \end{aligned} \tag{2}$$

From (2), we have

$$\begin{aligned} \boldsymbol{w}(t) &= [\boldsymbol{I}_m - \boldsymbol{C}(z)]\boldsymbol{w}(t) + \boldsymbol{v}(t), \\ &= (-\boldsymbol{C}_1 z^{-1} - \boldsymbol{C}_2 z^{-2} - \cdots - \boldsymbol{C}_{n_c} z^{-n_c})\boldsymbol{w}(t) + \boldsymbol{v}(t) \\ &= -\boldsymbol{C}_1\boldsymbol{w}(t-1) - \boldsymbol{C}_2\boldsymbol{w}(t-2) - \cdots - \boldsymbol{C}_{n_c}\boldsymbol{w}(t-n_c) + \boldsymbol{v}(t) \\ &= \boldsymbol{\theta}_{\mathrm{n}}^{\mathrm{T}}\boldsymbol{\psi}(t) + \boldsymbol{v}(t). \end{aligned} \tag{3}$$

Equation (3) is the noise model. Equation (1) can be rewritten as

$$\begin{aligned} \boldsymbol{y}(t) &= \boldsymbol{\Phi}_{\mathrm{s}}(t)\boldsymbol{\theta} + \boldsymbol{w}(t) \\ &= \boldsymbol{\Phi}_{\mathrm{s}}(t)\boldsymbol{\theta} + \boldsymbol{\theta}_{\mathrm{n}}^{\mathrm{T}}\boldsymbol{\psi}(t) + \boldsymbol{v}(t). \end{aligned} \tag{4}$$

Combining the information matrix $\boldsymbol{\Phi}_{\mathrm{s}}(t)$ and the information vector $\boldsymbol{\psi}(t)$ into an information matrix $\boldsymbol{\Phi}(t)$, and the parameter vector $\boldsymbol{\theta}$ and the parameter matrix $\boldsymbol{\theta}_{\mathrm{n}}$ into a parameter vector $\boldsymbol{\vartheta}$:

$$\begin{aligned} \boldsymbol{\Phi}(t) &:= [\boldsymbol{\Phi}_{\mathrm{s}}(t), \boldsymbol{I}_m \otimes \boldsymbol{\psi}^{\mathrm{T}}(t)] \in \mathbb{R}^{m\times n_0}, \quad n_0 := n + m^2 n_c, \\ \boldsymbol{\vartheta} &:= \begin{bmatrix} \boldsymbol{\theta} \\ \mathrm{col}[\boldsymbol{\theta}_{\mathrm{n}}] \end{bmatrix} \in \mathbb{R}^{n_0}. \end{aligned}$$

Equation (1) can be rewritten as

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}(t)\boldsymbol{\vartheta} + \boldsymbol{v}(t). \tag{5}$$

In this model,the new parameter vector $\boldsymbol{\vartheta}$ contains the parameter vector $\boldsymbol{\theta}$ of the system model and the parameter matrices $\boldsymbol{C}_i'$s of the noise model.

The objective is to derive gradient methods for estimating the parameters $\boldsymbol{\theta}$ and $\boldsymbol{C}_i$ by using the multi-innovation theory or data filtering technique for reducing the computational burden and to evaluate the parameter estimation accuracy of the proposed algorithms by simulations on computers.

Supposing $\hat{\boldsymbol{X}}(t)$ is the estimate of $\boldsymbol{X}$ at time t. That is to say $\hat{\boldsymbol{C}}_i(t)$ are the estimate of $\boldsymbol{C}_i$ at time t, $\hat{\boldsymbol{\vartheta}}(t)$ is the estimate of $\boldsymbol{\vartheta}$ at time t, $\hat{\boldsymbol{\vartheta}}(t) := \begin{bmatrix} \hat{\boldsymbol{\theta}}(t) \\ \mathrm{col}[\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)] \end{bmatrix}$ is the estimate of $\boldsymbol{\vartheta} := \begin{bmatrix} \boldsymbol{\theta} \\ \mathrm{col}[\boldsymbol{\theta}_{\mathrm{n}}] \end{bmatrix}$ at time t, $\hat{\boldsymbol{w}}(t)$ is the estimate of $\boldsymbol{w}(t)$ at time t, and so on.

3. The M-GSG algorithm

In this section, we derive an M-GSG algorithm as the basic algorithm firstly. Define a gradient criterion function

$$J_1(\boldsymbol{\vartheta}) := \frac{1}{2}\|\boldsymbol{y}(t) - \boldsymbol{\Phi}(t)\boldsymbol{\vartheta}\|^2.$$

Assuming that $\mu(t)$ is the step-size, using the negative gradient search and minimizing $J_1(\boldsymbol{\vartheta})$, we have the gradient relationship:

$$\begin{aligned} \hat{\boldsymbol{\vartheta}}(t) &= \hat{\boldsymbol{\vartheta}}(t-1) - \mu(t)\mathrm{grad}[J_1(\hat{\boldsymbol{\vartheta}}(t-1))] \\ &= \hat{\boldsymbol{\vartheta}}(t-1) + \mu(t)\boldsymbol{\Phi}^{\mathrm{T}}(t)[\boldsymbol{y}(t) - \boldsymbol{\Phi}(t)\hat{\boldsymbol{\vartheta}}(t-1)]. \end{aligned}$$

In system identification, $\boldsymbol{e}(t) := \boldsymbol{y}(t) - \boldsymbol{\Phi}(t)\hat{\boldsymbol{\vartheta}}(t-1) \in \mathbb{R}^m$ is called the innovation vector, let the step-size $\mu(t) := 1/r(t), \quad r(t) = r(t-1) + \|\boldsymbol{\Phi}(t)\|^2$. Thus, we have:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\Phi}^{\mathrm{T}}(t)}{r(t)}[\boldsymbol{y}(t) - \boldsymbol{\Phi}(t)\hat{\boldsymbol{\vartheta}}(t-1)], \tag{6}$$

$$r(t) = r(t-1) + \|\boldsymbol{\Phi}(t)\|^2. \tag{7}$$

Here, the difficulty of identification is that $\boldsymbol{\Phi}(t)$ contains the unmeasurable noise terms $\boldsymbol{w}(t-i)$, the estimate $\hat{\boldsymbol{\vartheta}}(t)$ in (6)–(7) is impossible to compute. The solution is to replace the unmeasurable variables $\boldsymbol{w}(t-i)$ with their corresponding estimates $\hat{\boldsymbol{w}}(t-i)$. From (4), we have $\boldsymbol{w}(t) = \boldsymbol{y}(t) - \boldsymbol{\Phi}_{\mathrm{s}}(t)\boldsymbol{\theta}$, replacing the unknown parameter vector $\boldsymbol{\theta}$ with the estimate $\hat{\boldsymbol{\theta}}(t)$, we can compute the estimate of $\boldsymbol{w}(t)$:

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \boldsymbol{\Phi}_{\mathrm{s}}(t)\hat{\boldsymbol{\theta}}(t).$$

Since the information vector $\boldsymbol{\psi}(t)$ contains the unknown term $\boldsymbol{w}(t-i)$, we use the estimate $\hat{\boldsymbol{w}}(t-i)$ to define the estimate of $\boldsymbol{\psi}(t)$:

$$\hat{\boldsymbol{\psi}}(t) := [-\hat{\boldsymbol{w}}^{\mathrm{T}}(t-1), -\hat{\boldsymbol{w}}^{\mathrm{T}}(t-2), \cdots, -\hat{\boldsymbol{w}}^{\mathrm{T}}(t-n_c)]^{\mathrm{T}}, \tag{8}$$

Furthermore, the information matrix $\boldsymbol{\Phi}(t)$ contains the unknown term $\boldsymbol{\psi}(t)$, we use the estimate $\hat{\boldsymbol{\psi}}(t)$ to define the estimate of $\boldsymbol{\Phi}(t)$:

$$\hat{\boldsymbol{\Phi}}(t) := [\boldsymbol{\Phi}_{\mathrm{s}}(t), \boldsymbol{I}_m \otimes \hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)].$$

Afterwards, we can obtain the multivariate generalized stochastic gradient (M-GSG) algorithm for estimating the parameter vector $\boldsymbol{\vartheta}$:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t)}{r(t)}[\boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t-1)], \tag{9}$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\Phi}}(t)\|^2, \tag{10}$$

$$\hat{\boldsymbol{\Phi}}(t) = [\boldsymbol{\Phi}_{\mathrm{s}}(t), \boldsymbol{I}_m \otimes \hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)], \tag{11}$$

$$\hat{\boldsymbol{\psi}}(t) = [-\hat{\boldsymbol{w}}^{\mathrm{T}}(t-1), -\hat{\boldsymbol{w}}^{\mathrm{T}}(t-2), \cdots, -\hat{\boldsymbol{w}}^{\mathrm{T}}(t-n_c)]^{\mathrm{T}}, \tag{12}$$

$$\hat{\boldsymbol{\vartheta}}(t) = \begin{bmatrix} \hat{\boldsymbol{\theta}}(t) \\ \mathrm{col}[\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)] \end{bmatrix}, \tag{13}$$

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \boldsymbol{\Phi}_{\mathrm{s}}(t)\hat{\boldsymbol{\theta}}(t). \tag{14}$$

The procedures of computing the parameter estimation vector $\hat{\boldsymbol{\vartheta}}(t)$ by the M-GSG algorithm in (9)–(14) are listed as follows.

1. Let $t = 1$, set the initial values $\hat{\boldsymbol{\vartheta}}(0) = \mathbf{1}_{n_0}/p_0$, $r(0) = 1$, $\hat{\boldsymbol{w}}(t-i) = 0$, $i = 0, 1, \cdots, n_c$, $p_0 = 10^6$.
2. Collect the observation data $\boldsymbol{\Phi}_{\mathrm{s}}(t)$ and $\boldsymbol{y}(t)$, construct the information vector $\hat{\boldsymbol{\psi}}(t)$ using (12) and construct the information matrix $\hat{\boldsymbol{\Phi}}(t)$ using (11).
3. Compute $r(t)$ using (10).
4. Update the parameter estimate vector $\hat{\boldsymbol{\vartheta}}(t)$ using (9).
5. Read $\hat{\boldsymbol{\theta}}(t)$ form $\hat{\boldsymbol{\vartheta}}(t)$ in (13) and compute $\hat{\boldsymbol{w}}(t)$ using (14).
6. Increase t by 1, go to Step 2.

4. The M-MI-GSG algorithm

In order to improve the parameter estimation accuracy of the M-GSG algorithm, we can introduce an innovation length to expand the innovation vector to a large innovation vector. This section discusses an M-MI-GSG algorithm by using the multi-innovation identification theory.

Based on the M-GSG algorithm (9)–(14), define the stacked information matrix $\boldsymbol{\Gamma}(p,t)$ and the stacked output vector $\boldsymbol{Y}(p,t)$ as

$$\boldsymbol{\Gamma}(p,t) := [\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t), \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t-1), \cdots, \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t-p+1)] \in \mathbb{R}^{n_0 \times (mp)},$$

$$\boldsymbol{Y}(p,t) := \begin{bmatrix} \boldsymbol{y}(t) \\ \boldsymbol{y}(t-1) \\ \vdots \\ \boldsymbol{y}(t-p+1) \end{bmatrix} \in \mathbb{R}^{mp}.$$

$\boldsymbol{e}(t) := \boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t-1) \in \mathbb{R}^m$ in (9) is the innovation vector. Define $p \geqslant 1$ as the innovation length and expand the innovation vector $\boldsymbol{e}(t)$ into a large innovation vector:

$$\boldsymbol{E}(p,t) := \begin{bmatrix} \boldsymbol{e}(t) \\ \boldsymbol{e}(t-1) \\ \vdots \\ \boldsymbol{e}(t-p+1) \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t-1) \\ \boldsymbol{y}(t-1) - \hat{\boldsymbol{\Phi}}(t-1)\hat{\boldsymbol{\vartheta}}(t-2) \\ \vdots \\ \boldsymbol{y}(t-p+1) - \hat{\boldsymbol{\Phi}}(t-p+1)\hat{\boldsymbol{\vartheta}}(t-p) \end{bmatrix} \in \mathbb{R}^{mp}.$$

Normally, we reach an agreement that the estimate $\hat{\boldsymbol{\vartheta}}(t-1)$ at time $(t-1)$ is closer to the true value $\boldsymbol{\vartheta}$ than the estimate $\hat{\boldsymbol{\vartheta}}(t-i)$ at time $(t-i)$ $(i \geqslant 2)$. Therefore, replacing the term $\hat{\boldsymbol{\vartheta}}(t-i)$ $(i \geqslant 2)$ in $\boldsymbol{E}(p,t)$ with $\hat{\boldsymbol{\vartheta}}(t-1)$, then the large innovation vector can be modified into

$$\begin{aligned} \boldsymbol{E}(p,t) &:= \begin{bmatrix} \boldsymbol{y}(t) - \hat{\boldsymbol{\Phi}}(t)\hat{\boldsymbol{\vartheta}}(t-1) \\ \boldsymbol{y}(t-1) - \hat{\boldsymbol{\Phi}}(t-1)\hat{\boldsymbol{\vartheta}}(t-1) \\ \vdots \\ \boldsymbol{y}(t-p+1) - \hat{\boldsymbol{\Phi}}(t-p+1)\hat{\boldsymbol{\vartheta}}(t-1) \end{bmatrix} \\ &= \boldsymbol{Y}(p,t) - \boldsymbol{\Gamma}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1). \end{aligned}$$

We can obtain the following multivariate multi-innovation generalized stochastic gradient (M-MI-GSG) algorithm with an innovation length p:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \frac{\boldsymbol{\Gamma}(p,t)}{r(t)}\boldsymbol{E}(p,t), \tag{15}$$

$$\boldsymbol{E}(p,t) = \boldsymbol{Y}(p,t) - \boldsymbol{\Gamma}^{\mathrm{T}}(p,t)\hat{\boldsymbol{\vartheta}}(t-1), \tag{16}$$

$$r(t) = r(t-1) + \|\hat{\boldsymbol{\Phi}}(t)\|^2, \tag{17}$$

$$\boldsymbol{Y}(p,t) = [\boldsymbol{y}^{\mathrm{T}}(t), \boldsymbol{y}^{\mathrm{T}}(t-1), \cdots, \boldsymbol{y}^{\mathrm{T}}(t-p+1)]^{\mathrm{T}}, \tag{18}$$

$$\boldsymbol{\Gamma}(p,t) = [\hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t), \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t-1), \cdots, \hat{\boldsymbol{\Phi}}^{\mathrm{T}}(t-p+1)], \tag{19}$$

$$\hat{\boldsymbol{\Phi}}(t) = [\boldsymbol{\Phi}_{\mathrm{s}}(t), \boldsymbol{I}_m \otimes \hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)], \tag{20}$$

$$\hat{\boldsymbol{\psi}}(t) = [-\hat{\boldsymbol{w}}^{\mathrm{T}}(t-1), -\hat{\boldsymbol{w}}^{\mathrm{T}}(t-2), \cdots, -\hat{\boldsymbol{w}}^{\mathrm{T}}(t-n_c)]^{\mathrm{T}}, \tag{21}$$

$$\hat{\boldsymbol{\vartheta}}(t) = \begin{bmatrix} \hat{\boldsymbol{\theta}}(t) \\ \mathrm{col}[\hat{\boldsymbol{\theta}}_{\mathrm{n}}(t)] \end{bmatrix}, \tag{22}$$

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \boldsymbol{\Phi}_{\mathrm{s}}(t)\hat{\boldsymbol{\theta}}(t). \tag{23}$$

The procedures of computing the parameter estimation vector $\hat{\boldsymbol{\vartheta}}(t)$ by the M-MI-GSG algorithm in (15)–(23) are listed as follows.

1. Let $t=1$, choose an innovation length p, set the initial values $\hat{\boldsymbol{\vartheta}}(0)=\mathbf{1}_{n_0}/p_0$, $r(0)=1$, $\hat{\boldsymbol{w}}(t-i)=0$, $i=0,\,1,\,\cdots,\,n_c$, $p_0=10^6$.
2. Collect the observation data $\boldsymbol{\Phi}_{\rm s}(t)$ and $\boldsymbol{y}(t)$, construct the information vector $\hat{\boldsymbol{\psi}}(t)$ using (21) and construct the information matrix $\hat{\boldsymbol{\Phi}}(t)$ using (20).
3. Construct the stacked output vector $\boldsymbol{Y}(p,t)$ and the stacked information matrix $\boldsymbol{\Gamma}(p,t)$ using (18)–(19).
4. Compute the innovation vector $\boldsymbol{E}(p,t)$ using (16) and compute $r(t)$ using (17).
5. Update the parameter estimate vector $\hat{\boldsymbol{\vartheta}}(t)$ using (15).
6. Read $\hat{\boldsymbol{\theta}}(t)$ form $\hat{\boldsymbol{\vartheta}}(t)$ in (22) and compute $\hat{\boldsymbol{w}}(t)$ using (23).
7. Increase t by 1, go to Step 2.

When the innovation length $p=1$, we can obtain the M-GSG algorithm from Equations (9)–(14). Compared with the M-GSG algorithm, the M-MI-GSG algorithm not only use the current data and innovation but also use the past data and innovation when it recursively compute the parameter estimation at each step. Using the data information of the system repeatedly makes the accuracy of the parameter estimation improved.

5. The F-M-GSG algorithm

In this section, we use the data filtering technique to improve the parameter estimation accuracy of the M-GSG algorithm. The identification thought based on the data filtering is an effective method to deal with the colored noise, the main idea is a special filter to filter the input and output data, and then to identify the filtered system model and the filtered noise model with each other for generating corresponding parameter estimates. The choice of this special filter depends on the construction of the colored noise.

For the multivariate pseudo-linear moving average system in (1). Choose the matrix polynomial $\boldsymbol{C}(z)$ as the filter, multipling both sides of (1) by the filter $\boldsymbol{C}(z)$ gives

$$\boldsymbol{C}(z)\boldsymbol{y}(t)=\boldsymbol{C}(z)\boldsymbol{\Phi}_{\rm s}(t)\boldsymbol{\theta}+\boldsymbol{v}(t). \tag{24}$$

Define the filtered output vector $\boldsymbol{y}_{\rm f}(t)$ and the filtered information matrix $\boldsymbol{\Phi}_{\rm f}(t)$ as

$$\boldsymbol{y}_{\rm f}(t):=\boldsymbol{C}(z)\boldsymbol{y}(t)\in\mathbb{R}^m,$$
$$\boldsymbol{\Phi}_{\rm f}(t):=\boldsymbol{C}(z)\boldsymbol{\Phi}_{\rm s}(t)\in\mathbb{R}^{m\times n}.$$

Equation (24) can be written as

$$\boldsymbol{y}_{\rm f}(t)=\boldsymbol{\Phi}_{\rm f}(t)\boldsymbol{\theta}+\boldsymbol{v}(t), \tag{25}$$

This is the filtered model. Define and minimize the gradient criterion function

$$J_2(\boldsymbol{\theta}) := \frac{1}{2}\|\boldsymbol{y}_{\rm f}(t) - \boldsymbol{\Phi}_{\rm f}(t)\boldsymbol{\theta}\|^2.$$

Furthermore, we have the following gradient recursive relation,

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\boldsymbol{\Phi}_{\rm f}^{\rm T}(t)}{r_1(t)}[\boldsymbol{y}_{\rm f}(t) - \boldsymbol{\Phi}_{\rm f}(t)\hat{\boldsymbol{\theta}}(t-1)], \quad \hat{\boldsymbol{\theta}}(0) = \mathbf{1}_n/p_0, \tag{26}$$

$$r_1(t) = r_1(t-1) + \|\boldsymbol{\Phi}_{\rm f}(t)\|^2, \quad r_1(0) = 1. \tag{27}$$

However, the parameter matrices $\boldsymbol{C}_i$ are unknown, so are $\boldsymbol{y}_{\rm f}(t)$ and $\boldsymbol{\Phi}_{\rm f}(t)$, the estimate $\hat{\boldsymbol{\theta}}(t)$ in (26)–(27) is impossible to compute. If we define

$$\boldsymbol{\varphi}_y(t) := \begin{bmatrix} \boldsymbol{y}(t-1) \\ \boldsymbol{y}(t-2) \\ \vdots \\ \boldsymbol{y}(t-n_c) \end{bmatrix} \in \mathbb{R}^{mn_c},$$

$$\boldsymbol{\varphi}_{\rm s}(t) := \begin{bmatrix} \boldsymbol{\Phi}_{\rm s}(t-1) \\ \boldsymbol{\Phi}_{\rm s}(t-2) \\ \vdots \\ \boldsymbol{\Phi}_{\rm s}(t-n_c) \end{bmatrix} \in \mathbb{R}^{(mn_c)\times n}.$$

Then the filtered output vector $\boldsymbol{y}_{\rm f}(t)$ and the filtered information matrix $\boldsymbol{\Phi}_{\rm f}(t)$ can be expressed as

$$\begin{aligned} \boldsymbol{y}_{\rm f}(t) &= \boldsymbol{C}(z)\boldsymbol{y}(t) \\ &= (\boldsymbol{I}_m + \boldsymbol{C}_1 z^{-1} + \boldsymbol{C}_2 z^{-2} + \cdots + \boldsymbol{C}_{n_c} z^{-n_c})\boldsymbol{y}(t) \\ &= \boldsymbol{y}(t) + \boldsymbol{C}_1\boldsymbol{y}(t-1) + \boldsymbol{C}_2\boldsymbol{y}(t-2) + \cdots + \boldsymbol{C}_{n_c}\boldsymbol{y}(t-n_c) \\ &= \boldsymbol{y}(t) + \boldsymbol{\theta}_{\rm n}^{\rm T}\boldsymbol{\varphi}_y(t), \end{aligned} \tag{28}$$

$$\begin{aligned} \boldsymbol{\Phi}_{\rm f}(t) &= \boldsymbol{C}(z)\boldsymbol{\Phi}_{\rm s}(t) \\ &= (\boldsymbol{I}_m + \boldsymbol{C}_1 z^{-1} + \boldsymbol{C}_2 z^{-2} + \cdots + \boldsymbol{C}_{n_c} z^{-n_c})\boldsymbol{\Phi}_{\rm s}(t) \\ &= \boldsymbol{\Phi}_{\rm s}(t) + \boldsymbol{C}_1\boldsymbol{\Phi}_{\rm s}(t-1) + \boldsymbol{C}_2\boldsymbol{\Phi}_{\rm s}(t-2) + \cdots + \boldsymbol{C}_{n_c}\boldsymbol{\Phi}_{\rm s}(t-n_c) \\ &= \boldsymbol{\Phi}_{\rm s}(t) + \boldsymbol{\theta}_{\rm n}^{\rm T}\boldsymbol{\varphi}_{\rm s}(t). \end{aligned} \tag{29}$$

According to (28) and (29), replacing the unknown noise model parameter matrix $\boldsymbol{\theta}_{\rm n}$ with the estimate $\hat{\boldsymbol{\theta}}_{\rm n}(t)$, the filtered output vector estimate $\hat{\boldsymbol{y}}_{\rm f}(t)$ and the filtered information matrix estimate $\hat{\boldsymbol{\Phi}}_{\rm f}(t)$ can be computed by

$$\hat{\boldsymbol{y}}_{\rm f}(t) = \boldsymbol{y}(t) + \hat{\boldsymbol{\theta}}_{\rm n}^{\rm T}(t)\boldsymbol{\varphi}_y(t), \tag{30}$$

$$\hat{\boldsymbol{\Phi}}_{\rm f}(t) = \boldsymbol{\Phi}_{\rm s}(t) + \hat{\boldsymbol{\theta}}_{\rm n}^{\rm T}(t)\boldsymbol{\varphi}_{\rm s}(t). \tag{31}$$

From the noise model (3),we have

$$\boldsymbol{w}(t) = \boldsymbol{\theta}_{\rm n}^{\rm T}\boldsymbol{\psi}(t) + \boldsymbol{v}(t).$$

Defining and minimizing the gradient criterion function

$$J_3(\boldsymbol{\theta}_{\rm n}) := \frac{1}{2}\|\boldsymbol{w}(t) - \boldsymbol{\theta}_{\rm n}^{\rm T}\boldsymbol{\psi}(t)\|^2,$$

we can get the gradient recursive relation of parameter matrix estimate $\hat{\boldsymbol{\theta}}_n(t)$

$$\hat{\boldsymbol{\theta}}_{\rm n}(t) = \hat{\boldsymbol{\theta}}_{\rm n}(t-1) + \frac{\boldsymbol{\psi}(t)}{r_2(t)}[\boldsymbol{w}(t) - \hat{\boldsymbol{\theta}}_{\rm n}^{\rm T}(t-1)\boldsymbol{\psi}(t)]^{\rm T}, \quad \hat{\boldsymbol{\theta}}_{\rm n}(0) = \mathbf{1}_{(mn_c)\times m}/p_0, \tag{32}$$

$$r_2(t) = r_2(t-1) + \|\boldsymbol{\psi}(t)\|^2, \quad r_2(0) = 1. \tag{33}$$

However, $\boldsymbol{w}(t-i)$ in the filtered noise information vector $\boldsymbol{\psi}(t)$ are unknown, Equations (32)–(33) cannot be used to compute the parameter matrix estimate $\hat{\boldsymbol{\theta}}_n(t)$. The solution is the same as before, use the estimate $\hat{\boldsymbol{w}}(t-i)$ of $\boldsymbol{w}(t-i)$ to define

$$\hat{\boldsymbol{\psi}}(t) := [-\hat{\boldsymbol{w}}^{\rm T}(t-1), -\hat{\boldsymbol{w}}^{\rm T}(t-2), \cdots, -\hat{\boldsymbol{w}}^{\rm T}(t-n_c)]^{\rm T}.$$

In order to identify the filtered system model and the filtered noise model with each other, replacing the parameter estimate $\hat{\boldsymbol{\theta}}(t)$ with $\hat{\boldsymbol{\theta}}(t-1)$

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \boldsymbol{\Phi}_{\rm s}(t)\hat{\boldsymbol{\theta}}(t-1). \tag{34}$$

Therefore, we can summarize the filtering based multivariate generalized stochastic gradient (F-M-GSG) algorithm as:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \frac{\hat{\boldsymbol{\Phi}}_{\rm f}^{\rm T}(t)}{r_1(t)}[\hat{\boldsymbol{y}}_{\rm f}(t) - \hat{\boldsymbol{\Phi}}_{\rm f}(t)\hat{\boldsymbol{\theta}}(t-1)], \tag{35}$$

$$r_1(t) = r_1(t-1) + \|\hat{\boldsymbol{\Phi}}_{\rm f}(t)\|^2, \tag{36}$$

$$\hat{\boldsymbol{y}}_{\rm f}(t) = \boldsymbol{y}(t) + \hat{\boldsymbol{\theta}}_{\rm n}^{\rm T}(t)\boldsymbol{\varphi}_y(t), \tag{37}$$

$$\hat{\boldsymbol{\Phi}}_{\rm f}(t) = \boldsymbol{\Phi}_{\rm s}(t) + \hat{\boldsymbol{\theta}}_{\rm n}^{\rm T}(t)\boldsymbol{\varphi}_{\rm s}(t), \tag{38}$$

$$\boldsymbol{\varphi}_y(t) = [\boldsymbol{y}^{\rm T}(t-1), \boldsymbol{y}^{\rm T}(t-2), \cdots, \boldsymbol{y}^{\rm T}(t-n_c)]^{\rm T}, \tag{39}$$

$$\boldsymbol{\varphi}_{\rm s}(t) = [\boldsymbol{\Phi}_{\rm s}^{\rm T}(t-1), \boldsymbol{\Phi}_{\rm s}^{\rm T}(t-2), \cdots, \boldsymbol{\Phi}_{\rm s}^{\rm T}(t-n_c)]^{\rm T}, \tag{40}$$

$$\hat{\boldsymbol{\theta}}_{\rm n}(t) = \hat{\boldsymbol{\theta}}_{\rm n}(t-1) + \frac{\hat{\boldsymbol{\psi}}(t)}{r_2(t)}[\hat{\boldsymbol{w}}(t) - \hat{\boldsymbol{\theta}}_{\rm n}^{\rm T}(t-1)\hat{\boldsymbol{\psi}}(t)]^{\rm T}, \tag{41}$$

$$r_2(t) = r_2(t-1) + \|\hat{\boldsymbol{\psi}}(t)\|^2, \tag{42}$$

$$\hat{\boldsymbol{\psi}}(t) = [-\hat{\boldsymbol{w}}^{\rm T}(t-1), -\hat{\boldsymbol{w}}^{\rm T}(t-2), \cdots, -\hat{\boldsymbol{w}}^{\rm T}(t-n_c)]^{\rm T}, \tag{43}$$

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \boldsymbol{\Phi}_{\rm s}(t)\hat{\boldsymbol{\theta}}(t-1). \tag{44}$$

The computation procedures of the F-M-GSG algorithm in (35)–(44) are listed as follows.

1. Let $t = 1$, set the initial values $\hat{\boldsymbol{\theta}}(0) = \mathbf{1}_n/p_0$, $\hat{\boldsymbol{\theta}}_n(0) = \mathbf{1}_{(mn_c)\times m}/p_0$, $r_1(0) = 1$, $r_2(0) = 1$, $\hat{\boldsymbol{w}}(t-i) = 0$, $i = 0, 1, \cdots, n_c$, $p_0 = 10^6$.

2. Collect the observation data $\mathbf{\Phi}_{\rm s}(t)$ and $\boldsymbol{y}(t)$, compute $\hat{\boldsymbol{w}}(t)$ using (44) and construct the information vector $\hat{\boldsymbol{\psi}}(t)$ using (43).
3. Compute $r_2(t)$ using (42) and update the noise parameter estimate matrix $\hat{\boldsymbol{\theta}}_{\rm n}(t)$ using (41).
4. Construct $\boldsymbol{\varphi}_y(t)$ and $\boldsymbol{\varphi}_{\rm s}(t)$ using (39)–(40) and compute $\hat{\boldsymbol{y}}_{\rm f}(t)$ and $\hat{\mathbf{\Phi}}_{\rm f}(t)$ using (37)–(38).
5. Compute $r_1(t)$ using (36) and update the parameter estimate vector $\hat{\boldsymbol{\theta}}(t)$ using (35).
6. Increase t by 1, go to Step 2.

6. Examples

Example 1. Consider the following multivariate pseudo-linear moving average system,

$$\begin{aligned}
&\boldsymbol{y}(t)=\mathbf{\Phi}_{\rm s}(t)\boldsymbol{\theta}+\boldsymbol{C}^{-1}(z)\boldsymbol{v}(t),\\
&\boldsymbol{C}(z)=\boldsymbol{I}_2+\left[\begin{array}{cc} c_{11} & c_{12}\\ c_{21} & c_{22}\end{array}\right]z^{-1}=\left[\begin{array}{cc} 1 & 0\\ 0 & 1\end{array}\right]+\left[\begin{array}{cc} -0.62 & 0.31\\ -0.86 & -0.25\end{array}\right]z^{-1},\\
&\quad\boldsymbol{\theta}=[\theta_1,\theta_2,\theta_3,\theta_4]^{\rm T}=[2.25,3.27,1.61,2.53]^{\rm T}.\\
&\quad\boldsymbol{\vartheta}=[2.25,3.27,1.61,2.53,-0.62,0.31,-0.86,-0.25]^{\rm T}.
\end{aligned}$$

In simulation, $\boldsymbol{y}(t)=\begin{bmatrix} y_1(t)\\ y_2(t)\end{bmatrix}\in\mathbb{R}^2$ is the output vector, $\{\mathbf{\Phi}_{\rm s}(t)\}$ is a 2×4 matrix sequence, $\boldsymbol{v}(t)=\begin{bmatrix} v_1(t)\\ v_2(t)\end{bmatrix}\in\mathbb{R}^2$ is a white noise vector with zero mean, σ_1^2 and σ_2^2 are the variance of $v_1(t)$ and $v_2(t)$. Taking the noise variances $\sigma_1^2=0.80^2$ and $\sigma_2^2=0.70^2$, using the M-GSG algorithm (i.e. the M-MI-GSG algorithm with $p=1$) and the M-MI-GSG algorithm with $p=2$ and $p=4$ to estimate the parameters of this example system, we obtain the parameter estimates and their errors $\delta:=\|\hat{\boldsymbol{\vartheta}}(t)-\boldsymbol{\vartheta}\|/\|\boldsymbol{\vartheta}\|$ shown in Table 1. The parameter estimation errors versus t are shown in Figure 1. The parameter estimates $\hat{\theta}_1(t)$, $\hat{\theta}_2(t)$, $\hat{\theta}_3(t)$, $\hat{\theta}_4(t)$ and $\hat{c}_{11}(t)$, $\hat{c}_{12}(t)$, $\hat{c}_{21}(t)$, $\hat{c}_{22}(t)$ versus t with $p=4$ are shown in Figures 2–3.

Example 2. Consider the following another multivariate pseudo-linear moving average system,

$$\begin{aligned}
&\boldsymbol{y}(t)=\mathbf{\Phi}_{\rm s}(t)\boldsymbol{\theta}+\boldsymbol{C}^{-1}(z)\boldsymbol{v}(t),\\
&\boldsymbol{C}(z)=\boldsymbol{I}_2+\left[\begin{array}{cc} c_{11} & c_{12}\\ c_{21} & c_{22}\end{array}\right]z^{-1}=\left[\begin{array}{cc} 1 & 0\\ 0 & 1\end{array}\right]+\left[\begin{array}{cc} -0.63 & 0.32\\ -0.88 & -0.43\end{array}\right]z^{-1},\\
&\quad\boldsymbol{\theta}=[\theta_1,\theta_2,\theta_3,\theta_4]^{\rm T}=[2.33,3.46,1.62,2.45]^{\rm T}.\\
&\quad\boldsymbol{\vartheta}=[2.33,3.46,1.62,2.45,-0.63,0.32,-0.88,-0.43]^{\rm T}.
\end{aligned}$$

The simulation conditions are similar to those in Example 1. Taking the noise variances $\sigma_1^2=1.80^2$ and $\sigma_2^2=1.60^2$, applying the M-GSG algorithm and the F-M-GSG algorithm to estimate the parameters of this example system, the parameter estimates and their errors are shown in Table 2. The parameter estimation errors versus t are shown in Figure 4. The parameter estimates $\hat{\theta}_1(t)$, $\hat{\theta}_2(t)$,

Table 1: The M-MI-GSG estimates and errors ($\sigma_1^2 = 0.80^2$, $\sigma_2^2 = 0.70^2$)

p	t	θ_1	θ_2	θ_3	θ_4	c_{11}	c_{12}	c_{21}	c_{22}	δ (%)
1	100	1.77664	3.41774	1.47256	2.53234	0.01877	-0.16936	0.25595	-0.55773	29.35171
	200	1.90594	3.36654	1.47910	2.51537	-0.02887	-0.09990	0.13809	-0.51001	25.75203
	500	2.04871	3.31407	1.53064	2.51175	-0.13207	-0.00046	-0.08876	-0.48735	19.93708
	1000	2.12216	3.32319	1.55175	2.52029	-0.20001	0.05353	-0.21950	-0.45671	16.60565
	2000	2.18849	3.30431	1.59315	2.51747	-0.26627	0.11338	-0.33916	-0.43824	13.52102
	3000	2.20036	3.29903	1.59841	2.52158	-0.30285	0.13883	-0.39845	-0.43156	12.07245
2	100	1.98167	3.31418	1.60859	2.76648	-0.53509	0.17174	-0.56441	-0.32913	9.79672
	200	2.06616	3.27617	1.60759	2.61526	-0.51918	0.18171	-0.60748	-0.27547	7.12501
	500	2.17932	3.29302	1.60291	2.54617	-0.52133	0.24322	-0.70431	-0.29540	4.21885
	1000	2.22328	3.30681	1.59035	2.54105	-0.53404	0.26226	-0.74583	-0.27088	3.14310
	2000	2.26231	3.29163	1.62401	2.52387	-0.54868	0.29110	-0.77819	-0.28317	2.32660
	3000	2.25981	3.28374	1.62380	2.53298	-0.56081	0.29595	-0.79277	-0.29207	2.00594
4	100	2.21712	3.16691	1.64507	2.60887	-0.51391	0.21934	-0.66129	-0.42176	6.42996
	200	2.10873	3.20570	1.63544	2.50367	-0.51317	0.20340	-0.72539	-0.28446	5.09270
	500	2.22994	3.29594	1.59818	2.51843	-0.54017	0.26763	-0.79732	-0.28448	2.37104
	1000	2.24826	3.30395	1.58219	2.54047	-0.55536	0.28815	-0.81371	-0.23706	1.85983
	2000	2.27111	3.29091	1.63294	2.51451	-0.57429	0.31066	-0.82765	-0.27079	1.41584
	3000	2.26479	3.27685	1.62799	2.53502	-0.58748	0.31028	-0.83207	-0.28538	1.19297
True values		2.25000	3.27000	1.61000	2.53000	-0.62000	0.31000	-0.86000	-0.25000	

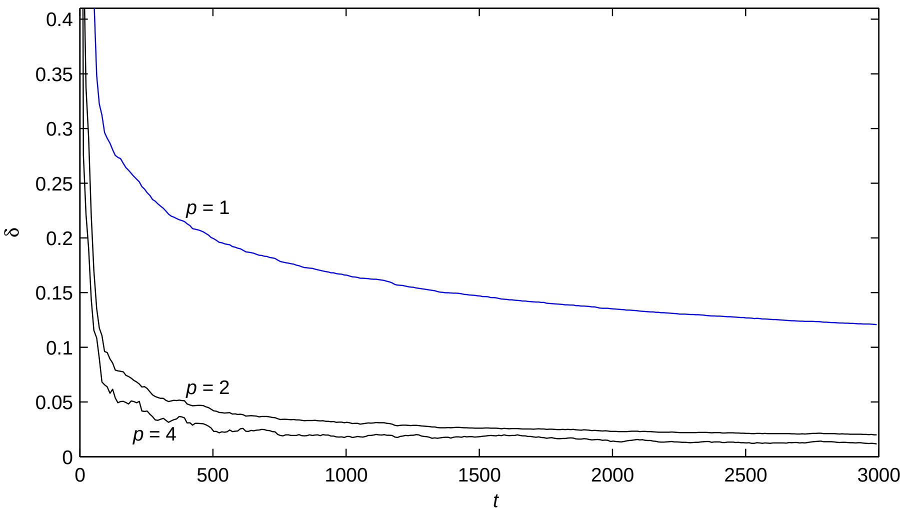

Figure 1: The M-MI-GSG estimation errors versus t ($\sigma_1^2 = 0.80^2$, $\sigma_2^2 = 0.70^2$)

$\hat{\theta}_3(t)$, $\hat{\theta}_4(t)$ and $\hat{c}_{11}(t)$, $\hat{c}_{12}(t)$, $\hat{c}_{21}(t)$, $\hat{c}_{22}(t)$ versus t are shown in Figures 5–6. When $\sigma_1^2 = 0.80^2$ and $\sigma_2^2 = 0.60^2$, the simulation results are shown in Table 3 and Figures 7–9.

From Tables 1–3 and Figures 1–9, we can draw the following conclusions.

1. The parameter estimation errors of the M-MI-GSG and the F-M-GSG algorithms become smaller with the data length t increases.
2. Under the same noise variances, the M-MI-GSG and the F-M-GSG algorithms have faster convergence rates and higher accurate parameter estimates than the M-GSG algorithm.
3. Introducing the innovation length p can effectively improve the parameter estimation accuracy of the M-MI-GSG algorithm, as the innovation length p increases, the parameter estimates are

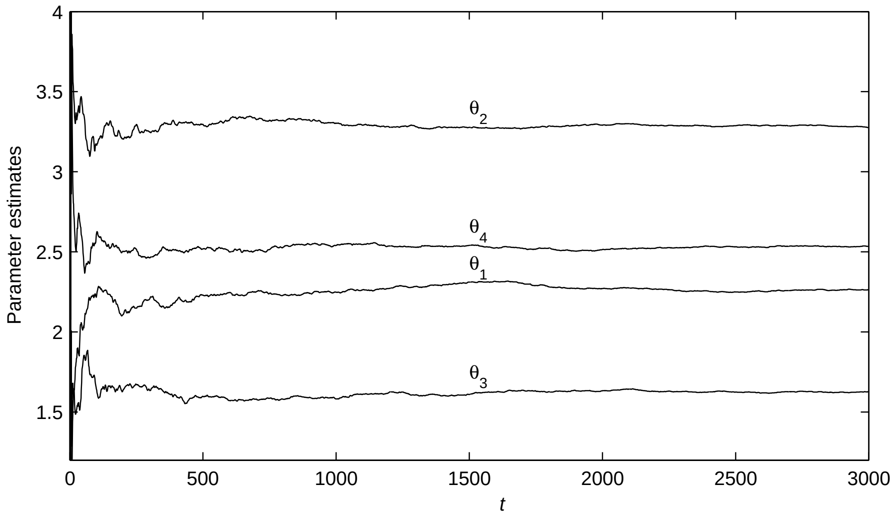

Figure 2: The M-MI-GSG parameter estimates $\hat{\theta}_1(t)$, $\hat{\theta}_2(t)$, $\hat{\theta}_3(t)$, $\hat{\theta}_4(t)$ versus t with $p = 4$ ($\sigma_1^2 = 0.80^2$, $\sigma_2^2 = 0.70^2$)

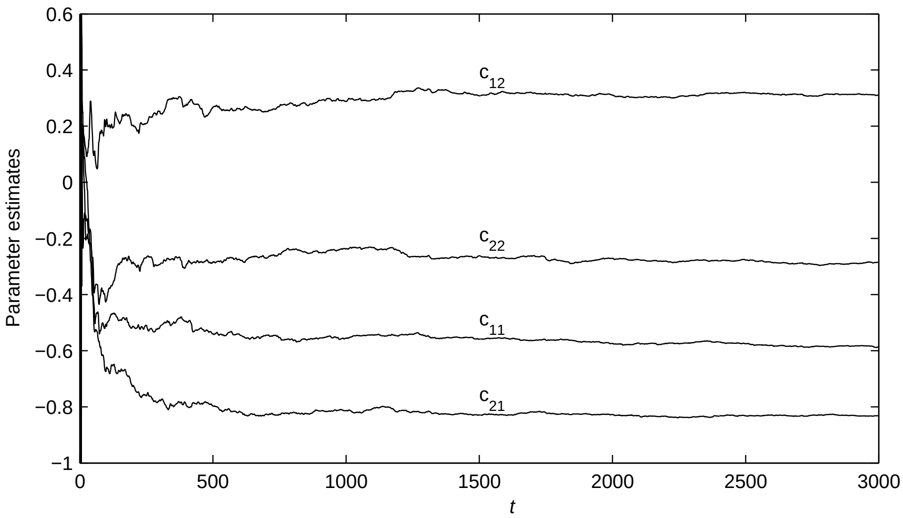

Figure 3: The M-MI-GSG parameter estimates $\hat{c}_{11}(t)$, $\hat{c}_{12}(t)$, $\hat{c}_{21}(t)$, $\hat{c}_{22}(t)$ versus t with $p = 4$ ($\sigma_1^2 = 0.80^2$, $\sigma_2^2 = 0.70^2$)

Table 2: The estimates and errors ($\sigma_1^2 = 1.80^2$, $\sigma_2^2 = 1.60^2$)

Algorithms	t	θ_1	θ_2	θ_3	θ_4	c_{11}	c_{12}	c_{21}	c_{22}	δ (%)
M-GSG	100	1.36962	3.54923	0.99212	2.98082	-0.39978	0.38957	-0.54754	-0.46207	25.41724
	200	1.56766	3.60879	1.14174	2.86660	-0.46991	0.33664	-0.59683	-0.38940	20.12600
	500	1.78048	3.56345	1.31961	2.76905	-0.56875	0.24258	-0.70544	-0.42934	14.08025
	1000	1.91831	3.53356	1.39184	2.68761	-0.57034	0.27428	-0.80406	-0.45107	10.36475
	2000	2.01772	3.52404	1.47041	2.60750	-0.57248	0.29495	-0.84126	-0.44443	7.49614
	3000	2.07341	3.50928	1.49417	2.58116	-0.58830	0.30477	-0.85331	-0.44600	6.15842
F-M-GSG	100	2.27167	3.37648	1.44798	2.48495	-0.44643	0.36492	-0.83050	-0.54665	5.81558
	200	2.30122	3.50583	1.49200	2.41249	-0.54057	0.34221	-0.77083	-0.44302	3.87619
	500	2.31571	3.49197	1.56782	2.46458	-0.65835	0.27329	-0.83294	-0.41943	1.85649
	1000	2.32599	3.47423	1.59081	2.45255	-0.62890	0.30290	-0.88724	-0.44111	0.75006
	2000	2.32988	3.46934	1.62144	2.42810	-0.61312	0.31242	-0.90044	-0.43879	0.71536
	3000	2.33518	3.46415	1.61649	2.43686	-0.62136	0.31642	-0.89416	-0.44391	0.50846
True values		2.33000	3.46000	1.62000	2.45000	-0.63000	0.32000	-0.88000	-0.43000	

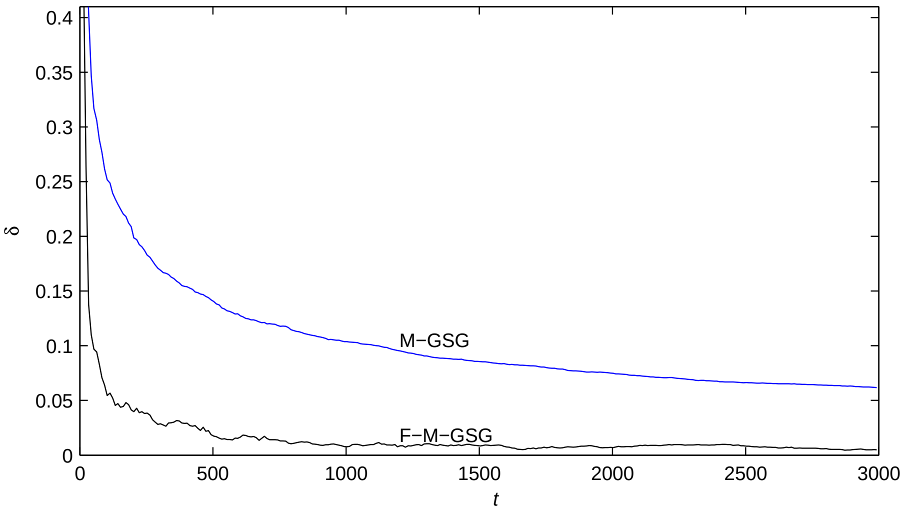

Figure 4: The estimation error δ versus t ($\sigma_1^2 = 1.80^2$, $\sigma_2^2 = 1.60^2$)

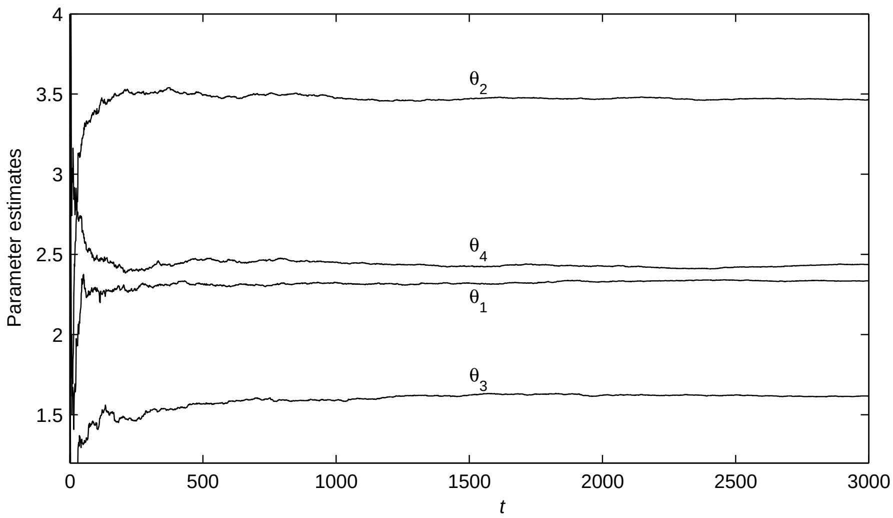

Figure 5: The F-M-GSG parameter estimates $\hat{\theta}_1(t)$, $\hat{\theta}_2(t)$, $\hat{\theta}_3(t)$, $\hat{\theta}_4(t)$ versus t ($\sigma_1^2 = 1.80^2$, $\sigma_2^2 = 1.60^2$)

Table 3: The estimates and errors ($\sigma_1^2 = 0.80^2$, $\sigma_2^2 = 0.60^2$)

Algorithms	t	θ_1	θ_2	θ_3	θ_4	c_{11}	c_{12}	c_{21}	c_{22}	δ (%)
M-GSG	100	2.36059	3.36174	1.52543	2.37678	-0.22746	0.29337	-0.64495	-0.46858	9.42917
	200	2.32838	3.43731	1.57221	2.42129	-0.30840	0.30513	-0.66997	-0.46562	7.45238
	500	2.31144	3.45197	1.61052	2.44073	-0.40548	0.25269	-0.71432	-0.45921	5.52263
	1000	2.31267	3.45215	1.61623	2.44438	-0.45699	0.26450	-0.76076	-0.46659	4.22132
	2000	2.31301	3.45843	1.62410	2.43621	-0.48922	0.26931	-0.79150	-0.46634	3.41418
	3000	2.32003	3.45404	1.62420	2.44100	-0.50737	0.27503	-0.80387	-0.46564	2.97713
F-M-GSG	100	2.37457	3.40748	1.56329	2.40797	-0.32292	0.36867	-0.77594	-0.48398	6.61117
	200	2.34361	3.46705	1.57772	2.40914	-0.43228	0.34339	-0.76522	-0.46120	4.57294
	500	2.33364	3.46939	1.60658	2.44398	-0.55317	0.26140	-0.80315	-0.43878	2.38529
	1000	2.33170	3.46413	1.61215	2.44530	-0.58169	0.28975	-0.84671	-0.45045	1.33237
	2000	2.33118	3.46367	1.62213	2.43850	-0.59055	0.29722	-0.86782	-0.44937	1.00027
	3000	2.33265	3.46142	1.61993	2.44315	-0.59981	0.30315	-0.87212	-0.44941	0.78411
True values		2.33000	3.46000	1.62000	2.45000	-0.63000	0.32000	-0.88000	-0.43000	

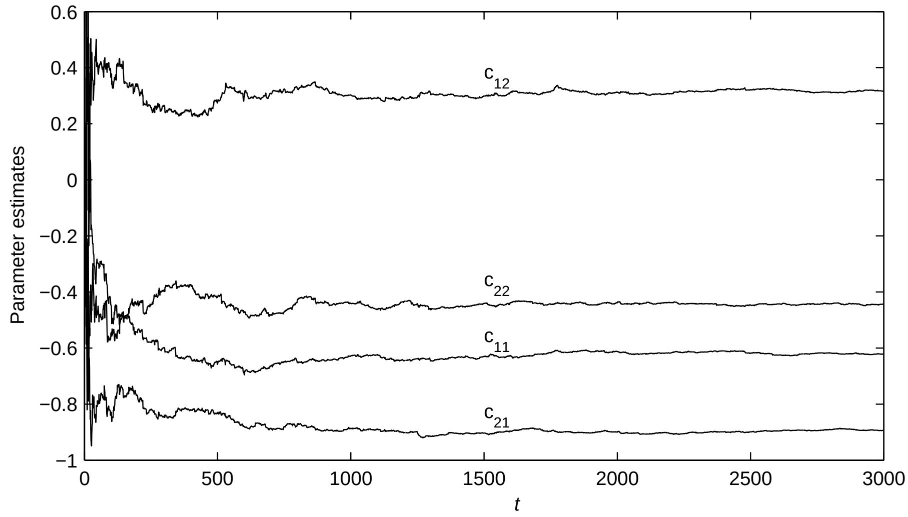

Figure 6: The F-M-GSG parameter estimates $\hat{c}_{11}(t)$, $\hat{c}_{12}(t)$, $\hat{c}_{21}(t)$, $\hat{c}_{22}(t)$ versus t ($\sigma_1^2 = 1.80^2$, $\sigma_2^2 = 1.60^2$)

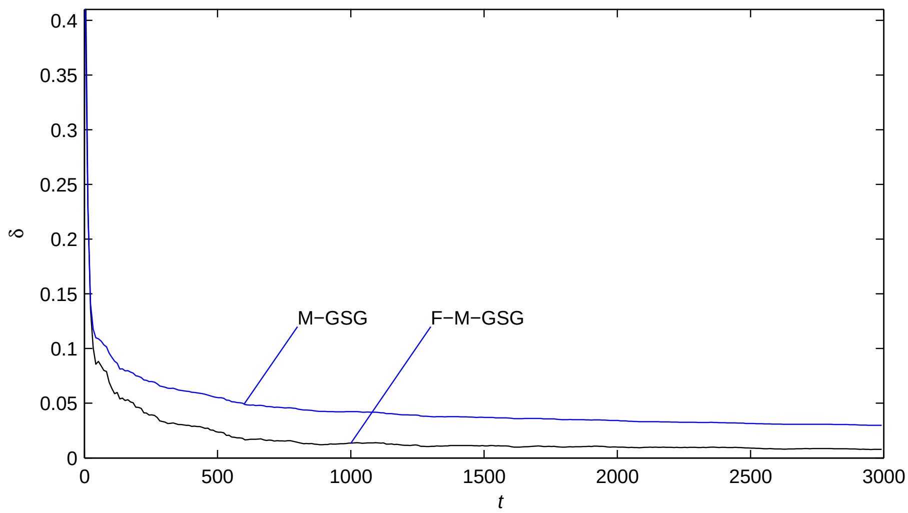

Figure 7: The estimation errors δ versus t ($\sigma_1^2 = 0.80^2$, $\sigma_2^2 = 0.60^2$)

getting more stationary.

7. Conclusions

This paper derives an M-MI-GSG algorithm and an F-M-GSG algorithm for identifying the multivariate pseudo-linear moving average system based on the multi-innovation identification theory and data filtering technique. The simulation results indicate that the proposed algorithms provide more accurate parameter estimation than the M-GSG algorithm. The identification idea of the proposed methods can be extended to study the parameter estimation problems of other nonlinear and dual-rate/multirate systems with colored noises.

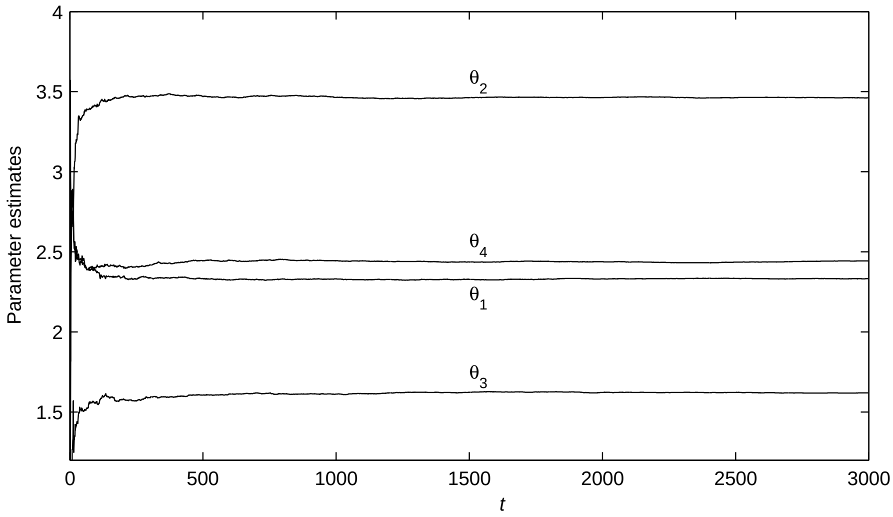

Figure 8: The F-M-GSG parameter estimates $\hat{\theta}_1(t)$, $\hat{\theta}_2(t)$, $\hat{\theta}_3(t)$, $\hat{\theta}_4(t)$ versus t ($\sigma_1^2 = 0.80^2$, $\sigma_2^2 = 0.60^2$)

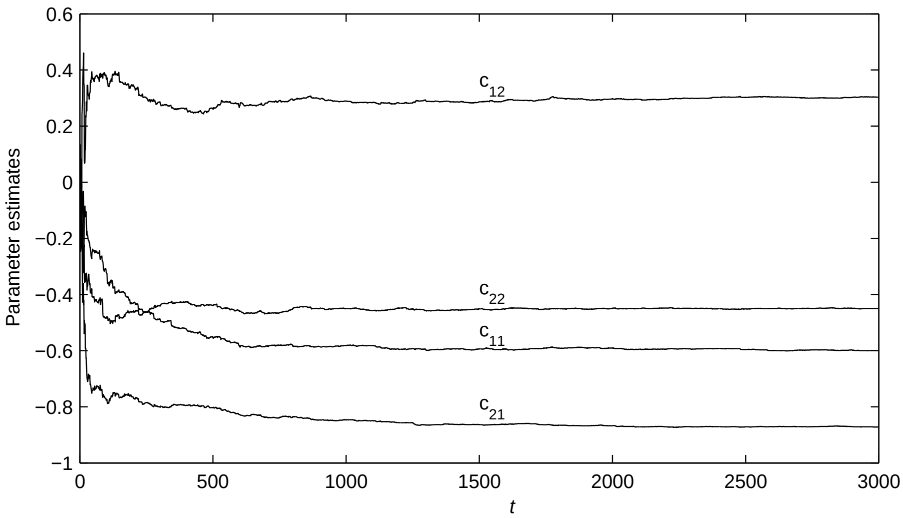

Figure 9: The F-M-GSG parameter estimates $\hat{c}_{11}(t)$, $\hat{c}_{12}(t)$, $\hat{c}_{21}(t)$, $\hat{c}_{22}(t)$ versus t ($\sigma_1^2 = 0.80^2$, $\sigma_2^2 = 0.60^2$)

References

[1] Ahmad, S.M., Linear and nonlinear system identification techniques for modelling of a remotely operated underwater vehicle, International Journal of Modelling, Identification and Control 24 (1) (2015) 75-87.

[2] Q.M. Zhu, Y.J. Wang, D.Y. Zhao, S.Y. Li, Billings, S.A., Review of rational (total) nonlinear dynamic system modelling, identification, and control, International Journal of Systems Science 46 (12) (2015) 2122-2133.

[3] Guta, M., Yamamoto, N., System identification for passive linear quantum systems, IEEE Transactions on Automatic Control 61 (4) (2016) 921-936.

[4] Nassirharand, A., Teh, S.H., Describing function-based identification of nonlinear transfer functions for nonlinear systems from experimental/simulation data, International Journal of Modelling, Identification and Control 25 (2) (2016) 93-101.

[5] Weijtjens, W., De, S.G., Devriendt, C., Guillaume, P., Operational modal parameter estimation of MIMO systems using transmissibility functions, Automatica 50 (2) (2014) 559-564.

[6] Yoshimura, T., Adaptive fuzzy backstepping control for MIMO uncertain discrete-time nonlinear systems using a set of noisy measurements, International Journal of Modelling, Identification and Control 23 (4) (2015) 336-345.

[7] Anderson, B.D.O., Deisler, M., Felsenstein, E., Koelbl, L., The structure of multivariate AR and ARMA systems: Regular and singular systems; the single and the mixed frequency case, Journal of Econometrics, 192 (2) (2016) 366-373.

[8] D.Q. Wang, R. Ding, X.Z. Dong, Iterative parameter estimation for a class of multivariable systems based on the hierarchical identification principle and the gradient search, Circuits Systems and Signal Process 31 (6) (2012) 2167-2177.

[9] X.H. Wang, F. Ding, Performance analysis of the recursive parameter estimation algorithms for multivariable Box-Jenkins systems, Journal of the Franklin Institute 351 (10) (2014) 4749-4764.

[10] Y.W. Mao, F. Ding, Data filtering-based multi-innovation stochastic gradient algorithm for nonlinear output error autoregressive systems, Circuits Systems and Signal Processing 35 (2) (2016) 651-667.

[11] H.Y. Hu, Y.S. Xiao, R. Ding, Multi-innovation stochastic gradient identification algorithm for Hammerstein controlled autoregressive autoregressive systems based on the key term separation principle and on the model decomposition, Journal of Applied Mathematics, 2013, Article ID 596141, 7 pages. http://dx.doi.org/10.1155/2013/596141

[12] J. Chen, L.X. Lv, R.F. Ding, Multi-innovation stochastic gradient algorithms for dual-rate sampled systems with preload nonlinearity, Applied Mathematics Letters 26 (1) (2013) 124-129.

[13] Z.Y. Wang, Y. Wang, Z.C. Ji, Filtering based multi-innovation extended stochastic gradient algorithm for Hammerstein nonlinear system modeling, Applied Mathematical Modelling 39 (18) (2015) 5724-5732.

[14] Y.B. Hu, B.L. Liu, Q. Zhou, A multi-innovation generalized extended stochastic gradient algorithm for output nonlinear autoregressive moving average systems, Applied Mathematics and Computation 247 (2014) 218-224.

[15] Y.W. Mao, F. Ding, A novel data filtering based multi-innovation stochastic gradient algorithm for Hammerstein nonlinear systems, Digital Signal Processing 46 (2015) 215-225.

[16] Q.B. Jin, Z. Wang, X.P. Liu, Auxiliary model-based interval-varying multi-innovation least squares identification for multivariable OE-like systems with scarce measurements, Journal of Process Control 35 (2015) 154-168.

[17] C. Wang, L. Zhu, Parameter identification of a class of nonlinear systems based on the multi-innovation identification theory, Journal of the Franklin Institute 352 (10) (2015) 4624-4637.

[18] Q.Y. Shen, F. Ding, Hierarchical multi-innovation extended stochastic gradient algorithms for input nonlinear multivariable OEMA systems by the key-term separation principle, Nonlinear Dynamics 85 (1) (2016) 499-507.

[19] Basin, M., P. Shi, Calderon, D., Joint state filtering and parameter estimation for linear stochastic time-delay systems, Signal Processing 91 (4) (2011) 782-792.

[20] Novara, C., Ruiz, F., Milanese, M., Direct Filtering: A new approach to optimal filter design for nonlinear systems, IEEE Transactions on Automatic Control, 58 (1) (2013) 86-99.

[21] Basin, M., Loukianov, A., Hernandez, M., Mean-square filtering for uncertain linear stochastic systems, Signal Processing 90 (6) (2010) 1916-1923.

[22] F.F. Wang, F. Ding, Gradient-based iterative identification methods for multivariate pseudo-linear moving average systems using the data filtering, Nonlinear Dynamics 84 (4) (2016) 2003-2015.

[23] S. Zhao, Yuriy S. Shmaliy, F. Liu, Fast Kalman-like optimal unbiased FIR filtering with applications, IEEE Transactions on Signal Processing 64 (9) (2016) 2284-2297.

[24] F. Ding, Y.J. Wang, J. Ding, Recursive least squares parameter identification algorithms for systems with colored noise using the filtering technique and the auxiliary model, Digital Signal Processing 37 (2015) 100-108.

[25] D.Q. Wang, Least squares-based recursive and iterative estimation for output error moving average systems using data filtering, IET Control Theory and Applications 5 (14) (2011) 1648-1657.