

The data filtering based generalized stochastic gradient parameter estimation algorithms for multivariate output-error autoregressive systems using the auxiliary model

Qinyao Liu¹ · Feng Ding^{1,2}

Received: 26 April 2017 / Revised: 21 September 2017 / Accepted: 24 October 2017 /
Published online: 31 October 2017
© Springer Science+Business Media, LLC 2017

Abstract Parameter estimation has wide applications in one-dimensional and multidimensional signal processing and filtering. This paper focuses on the parameter estimation problem of multivariate output-error autoregressive systems. Based on the data filtering technique and the auxiliary model identification idea, we derive a filtering based auxiliary model generalized stochastic gradient algorithm. The key is to choose an appropriate filter to filter the input-output data and to study a novel method to get the system model parameters and noise model parameters respectively. By employing the multi-innovation identification theory, a filtering based auxiliary model multi-innovation generalized stochastic gradient algorithm is proposed. Compared with the auxiliary model generalized stochastic gradient algorithm, the proposed algorithms can generate more accurate parameter estimates. Finally, an illustrative example is provided to verify the effectiveness of the proposed algorithms.

Keywords Parameter estimation · Filtering technique · Multi-innovation identification · Multivariate system · Auxiliary model

1 Introduction

With the development of control theory and the demand of engineering practices, the multivariable systems widely exist in all kinds of control processes (Cheng and Ugrinovskii 2016) and signal processing (Li and Zhang 2015). Compared with single variable systems, multivariable systems have more complex structures (Saikrishna et al. 2017) and uncertain

✉ Feng Ding
fding@jiangnan.edu.cn

Qinyao Liu
qyliu12@126.com

¹ Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China

² Nonlinear Analysis and Applied Mathematics (NAAM) Research Group, Department of Mathematics, King Abdulaziz University, Jidda 21589, Saudi Arabia

disturbances (Xing et al. 2016; Jafari et al. 2014). As a consequence, the identification problems for multivariable systems have attracted a lot of attention (Mercère and Bako 2011; Mu and Chen 2013; Cham et al. 2017). Katayama and Ase (2016) considered the linear approximation and identification of multi-input multi-output (MIMO) Wiener–Hammerstein systems and applied the separable least squares algorithm to estimate the parameters. In order to improve the estimation accuracy, Wang and Ding (2016) proposed an auxiliary model based recursive least squares parameter estimation algorithm for MIMO systems by filtering input-output data. Cheng et al. (2017) proposed the subspace identification methods for two-dimensional causal, recursive, and separable-in-denominator systems, which are applicable to both open-loop and closed-loop data. Recently, a gradient based parameter estimation method has been proposed for multivariate pseudo-linear systems using the multi-innovation and the data filtering (Ma and Ding 2017). In this paper, we use the multidimensional input and output signals to study the parameter identification problems for the multivariate system.

The data filtering technique is used to eliminate noises and outliers to reduce the noise-to-signal ratio in signal processing (Dhabal and Venkateswaran 2017; Ding et al. 2017; Tseng and Lee 2017; Zhao et al. 2017b), and has been employed to deal with the parameter estimation problem of systems which contains colored noises (Afshari et al. 2017). Differing from eliminating noises, the data filtering method in system identification only changes the structure of the system and does not change the relationship between the inputs and outputs (Pan et al. 2016; Ding et al. 2017a). Pan et al. (2017) used the filtering technique and the multi-innovation identification theory to identify the multivariable system with moving average noise, and proposed the filtering based multi-innovation extended stochastic gradient algorithm for improving the parameter estimation accuracy.

Many identification methods has been studied for linear systems and nonlinear systems (Li et al. 2017a, b, c), such as the least squares (Wan et al. 2016), the Newton iteration (Xu 2016, 2015; Xu et al. 2015) and the stochastic methods (Xu and Ding 2017b). Compared with the recursive least square algorithm (Wang et al. 2016; Zhang and Mao 2017), the stochastic gradient algorithm requires less computational cost and has been applied in parameter estimation (Liang et al. 2014; Levanony and Berman 2004). However, the stochastic gradient algorithm has a slow convergence rate and can not reach a satisfactory estimation accuracy (Li et al. 2014). To solve this problem, we use the multi-innovation identification theory to improve the performance of the stochastic gradient algorithm (Xu and Ding 2017c; Xu et al. 2017). Ding et al. (2017) presented a filtering based multi-innovation gradient algorithm for linear state space systems with time-delay by adopting the data filtering technique.

This paper studies the parameter estimation methods for multivariate output-error systems whose disturbance is an autoregressive noise (Aslam 2016) using the data filtering and the auxiliary model (Ding et al. 2017b; Jin et al. 2015). The main idea is to use a filter to filter the input-output data, then the system can be transformed into two identification models: a multivariate output-error model with white noise and an autoregressive noise model. The difficulty is that the two models have unmeasurable variables. We employ the auxiliary model identification idea to establish the auxiliary models and replace the unknown variables with the outputs of the auxiliary models. The main contributions of this paper are in the following aspects.

- A filtering based auxiliary model generalized stochastic gradient algorithm is derived for multivariate output-error autoregressive systems by using the data filter technique and the auxiliary model. Compared with the auxiliary model based generalized stochastic gradient algorithm, the proposed algorithm can generate more accurate estimates.

- A filtering based auxiliary model multi-innovation generalized stochastic gradient algorithm is proposed in order to improve the performance of the filtering based auxiliary model generalized stochastic gradient algorithm and it can get more accurate parameter estimates than the auxiliary model multi-innovation generalized stochastic gradient algorithm for the same innovation length.

The rest of this paper is organized as follows. In Sect. 2, we give some definitions and the identification model for multivariate output-error autoregressive systems. Section 3 gives the auxiliary model based stochastic gradient algorithm and the auxiliary model based multi-innovation stochastic gradient algorithm for comparisons. Section 4 derives the filtering identification model and presents the filtering based auxiliary model generalized stochastic gradient algorithm and the filtering based auxiliary model multi-innovation generalized stochastic gradient algorithm. An illustrative example is shown to verify the effectiveness of the proposed algorithms in Sect. 5. Finally, we offer some concluding remarks in Sect. 6.

2 The system description

Some notation is introduced. “ $A =: X$ ” or “ $X := A$ ” stands for “ A is defined as X ”; the superscript T stands for the vector/matrix transpose; the symbol I_n denotes an identity matrix of appropriate size ($n \times n$); $\hat{\boldsymbol{\theta}}(t)$ denotes the estimate of $\boldsymbol{\theta}$ at time t ; $\mathbf{1}_n$ stands for an n -dimensional column vector whose elements are 1; the norm of a matrix (or a column vector) X is defined by $\|X\|^2 := \text{tr}[XX^T]$.

Consider the following multivariate output-error autoregressive (M-OEAR) system in Fig. 1,

$$\mathbf{y}(t) = \frac{\boldsymbol{\Phi}_s(t)}{A(z)}\boldsymbol{\theta} + \frac{1}{C(z)}\mathbf{v}(t), \tag{1}$$

where $\mathbf{y}(t) := [y_1(t), y_2(t), \dots, y_m(t)]^T \in \mathbb{R}^m$ is the output vector of the system, $\boldsymbol{\Phi}_s(t) \in \mathbb{R}^{m \times n}$ is the information matrix consisting of the input signals $\mathbf{u}(t) = [u_1(t), u_2(t), \dots, u_r(t)] \in \mathbb{R}^r$ and output signals $\mathbf{y}(t)$, $\boldsymbol{\theta} \in \mathbb{R}^n$ is the parameter vector to be identified and $\mathbf{v}(t) := [v_1(t), v_2(t), \dots, v_m(t)]^T \in \mathbb{R}^m$ is a white noise vector with zero mean, $A(z)$ and $C(z)$ are the polynomials in the unit backward shift operator z^{-1} [$z^{-1}y(t) = y(t - 1)$], and defined as

$$\begin{aligned} A(z) &:= 1 + a_1z^{-1} + a_2z^{-2} + \dots + a_{n_a}z^{-n_a}, \quad a_i \in \mathbb{R}, \\ C(z) &:= 1 + c_1z^{-1} + c_2z^{-2} + \dots + c_{n_c}z^{-n_c}, \quad c_i \in \mathbb{R}. \end{aligned}$$

Assume that the orders m, n, n_a and n_c are known and $\mathbf{y}(t) = \mathbf{0}$, $\boldsymbol{\Phi}_s(t) = \mathbf{0}$ and $\mathbf{v}(t) = \mathbf{0}$ for $t \leq 0$.

Define the parameter vectors:

$$\begin{aligned} \mathbf{a} &:= [a_1, a_2, \dots, a_{n_a}]^T \in \mathbb{R}^{n_a}, \\ \mathbf{c} &:= [c_1, c_2, \dots, c_{n_c}]^T \in \mathbb{R}^{n_c}, \\ \boldsymbol{\theta} &:= [\boldsymbol{\theta}^T, \mathbf{a}^T, \mathbf{c}^T]^T \in \mathbb{R}^{n+n_a+n_c} \end{aligned}$$

and the information matrices:

$$\begin{aligned} \boldsymbol{\phi}_a(t) &:= [-\mathbf{x}(t - 1), -\mathbf{x}(t - 2), \dots, -\mathbf{x}(t - n_a)] \in \mathbb{R}^{m \times n_a}, \\ \boldsymbol{\phi}_c(t) &:= [-\mathbf{w}(t - 1), -\mathbf{w}(t - 2), \dots, -\mathbf{w}(t - n_c)] \in \mathbb{R}^{m \times n_c}, \\ \boldsymbol{\Phi}(t) &:= [\boldsymbol{\Phi}_s(t), \boldsymbol{\phi}_a(t), \boldsymbol{\phi}_c(t)] \in \mathbb{R}^{m \times (n+n_a+n_c)}. \end{aligned}$$

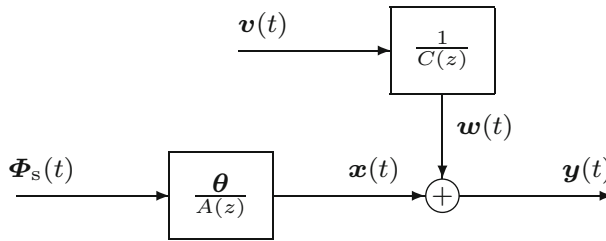


Fig. 1 A multivariate output-error autoregressive system

Define the intermediate variables:

$$\begin{aligned}
 \mathbf{x}(t) &:= \frac{\Phi_s(t)}{A(z)} \boldsymbol{\theta} \\
 &= [1 - A(z)]\mathbf{x}(t) + \Phi_s(t)\boldsymbol{\theta} \\
 &= - \sum_{j=1}^{n_a} a_j \mathbf{x}(t - j) + \Phi_s(t)\boldsymbol{\theta} \\
 &= \Phi_s(t)\boldsymbol{\theta} + \boldsymbol{\phi}_a(t)\mathbf{a},
 \end{aligned} \tag{2}$$

$$\begin{aligned}
 \mathbf{w}(t) &:= \frac{1}{C(z)} \mathbf{v}(t) \\
 &= [1 - C(z)]\mathbf{w}(t) + \mathbf{v}(t) \\
 &= - \sum_{j=1}^{n_c} c_j \mathbf{w}(t - j) + \mathbf{v}(t) \\
 &= \boldsymbol{\phi}_c(t)\mathbf{c} + \mathbf{v}(t).
 \end{aligned} \tag{3}$$

Equation (3) is the noise identification model.

Substituting (2) and (3) into (1) gives

$$\begin{aligned}
 \mathbf{y}(t) &= \mathbf{x}(t) + \mathbf{w}(t) \\
 &= \Phi_s(t)\boldsymbol{\theta} + \boldsymbol{\phi}_a(t)\mathbf{a} + \mathbf{w}(t)
 \end{aligned} \tag{4}$$

$$\begin{aligned}
 &= \Phi_s(t)\boldsymbol{\theta} + \boldsymbol{\phi}_a(t)\mathbf{a} + \boldsymbol{\phi}_c(t)\mathbf{c} + \mathbf{v}(t) \\
 &= \boldsymbol{\Phi}(t)\boldsymbol{\vartheta} + \mathbf{v}(t).
 \end{aligned} \tag{5}$$

Equation (5) is the identification model for the M-OEAR system and the parameter vector $\boldsymbol{\vartheta}$ contains all the parameters to be estimated. The object of this paper is to derive a new algorithm for the M-OEAR system by using the data filtering technique and the multi-innovation theory, which can generate more accurate estimates.

This paper studies the parameter estimation problem for the multivariate system by using the multidimensional input and output signals. The information matrix $\Phi_s(t)$ is composed of the m -dimensional output signals and the r -dimensional input signals. Therefore, the problem under consideration belongs to the multidimensional signal processing and estimation.

3 The auxiliary model based generalized stochastic gradient identification algorithm

In this section, we give the generalized stochastic gradient (AM-GSG) identification algorithm based on the auxiliary model identification idea. In order to improve the convergence rate and parameter estimation accuracy, an auxiliary model based multi-innovation stochastic gradient (AM-MI-GSG) algorithm is derived.

3.1 The AM-GSG algorithm

The stochastic gradient algorithm has been used to identify the multivariable system, and its convergence has been analyzed under weak conditions (Ding et al. 2008). On the basis of the work in Ding et al. (2008), we derive the AM-GSG algorithm for M-OEAR systems.

According to the identification model (5), we can define a gradient criterion function

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

Using the negative gradient search and minimizing the criterion function $J_1(\vartheta)$ give

$$\begin{aligned} \hat{\vartheta}(t) &= \hat{\vartheta}(t-1) - \frac{1}{r(t)} \text{grad}[J_1(\hat{\vartheta}(t-1))] \\ &= \hat{\vartheta}(t-1) + \frac{\Phi^T(t)}{r(t)} [y(t) - \Phi(t)\hat{\vartheta}(t-1)], \end{aligned} \tag{6}$$

$$r(t) = r(t-1) + \|\Phi(t)\|^2, \quad r(0) = 1. \tag{7}$$

Here, some problems arise. The information matrix $\Phi(t)$ contains the unknown terms $\{x(t-i), i = 1, 2, \dots, n_a\}$ and $\{w(t-i), i = 1, 2, \dots, n_c\}$, thus the estimate $\hat{\vartheta}(t)$ in (6–7) is impossible to compute. An effective method to solve this problem is to employ the auxiliary model identification idea. Establish the appropriate auxiliary models, use their outputs $x_a(t-i)$ and $\hat{w}(t-i)$ to replace the unknown variables $x(t-i)$ and $w(t-i)$. The estimates $\hat{\phi}_a(t)$ and $\hat{\phi}_c(t)$ of $\phi_a(t)$ and $\phi_c(t)$ can be formed by $x_a(t-i)$ and $\hat{w}(t-i)$. Then we can get the estimate $\hat{\Phi}(t)$ by using $\Phi_s(t)$, $\hat{\phi}_a(t)$ and $\hat{\phi}_c(t)$. Define

$$\begin{aligned} \hat{\phi}_a(t) &:= [-x_a(t-1), -x_a(t-2), \dots, -x_a(t-n_a)] \in \mathbb{R}^{m \times n_a}, \\ \hat{\phi}_c(t) &:= [-\hat{w}(t-1), -\hat{w}(t-2), \dots, -\hat{w}(t-n_c)] \in \mathbb{R}^{m \times n_c}, \\ \hat{\Phi}(t) &:= [\Phi_s(t), \hat{\phi}_a(t), \hat{\phi}_c(t)] \in \mathbb{R}^{m \times (n+n_a+n_c)}. \end{aligned}$$

According to (2), replacing the $\phi_a(t)$, θ and a with their estimates $\hat{\phi}_a(t)$, $\hat{\theta}(t)$ and $\hat{a}(t)$, the output $x_a(t)$ of the auxiliary model can be computed by

$$x_a(t) := \Phi_s(t)\hat{\theta}(t) + \hat{\phi}_a(t)\hat{a}(t).$$

Similarly, from (4), $\hat{w}(t)$ can be computed through

$$\begin{aligned} \hat{w}(t) &:= y(t) - \Phi_s(t)\hat{\theta}(t) - \hat{\phi}_a(t)\hat{a}(t) \\ &= y(t) - x_a(t). \end{aligned}$$

For convenience, define the innovation vector

$$e(t) := y(t) - \hat{\Phi}(t)\hat{\vartheta}(t-1) \in \mathbb{R}^m.$$

Replacing $\Phi(t)$ in (6–7) with its estimate $\hat{\Phi}(t)$, we can obtain the auxiliary model based generalized stochastic gradient (AM-GSG) algorithm:

$$\hat{\vartheta}(t) = \hat{\vartheta}(t - 1) + \frac{\hat{\Phi}^T(t)}{r(t)} e(t), \tag{8}$$

$$e(t) = \mathbf{y}(t) - \hat{\Phi}(t)\hat{\vartheta}(t - 1), \tag{9}$$

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

$$\hat{\Phi}(t) = [\Phi_s(t), \hat{\phi}_a(t), \hat{\phi}_c(t)], \tag{11}$$

$$\hat{\phi}_a(t) = [-\mathbf{x}_a(t - 1), -\mathbf{x}_a(t - 2), \dots, -\mathbf{x}_a(t - n_a)], \tag{12}$$

$$\hat{\phi}_c(t) = [-\hat{\mathbf{w}}(t - 1), -\hat{\mathbf{w}}(t - 2), \dots, -\hat{\mathbf{w}}(t - n_c)], \tag{13}$$

$$\mathbf{x}_a(t) = \Phi_s(t)\hat{\boldsymbol{\theta}}(t) + \hat{\phi}_a(t)\hat{\mathbf{a}}(t), \tag{14}$$

$$\hat{\mathbf{w}}(t) = \mathbf{y}(t) - \mathbf{x}_a(t), \tag{15}$$

$$\hat{\vartheta}(t) = [\hat{\boldsymbol{\theta}}^T(t), \hat{\mathbf{a}}^T(t), \hat{\mathbf{c}}^T(t)]^T. \tag{16}$$

The procedure for computing the parameter estimation vector $\hat{\vartheta}(t)$ in the AM-GSG algorithm (8–16) is as follows.

1. Set the initial values: let $t = 1$, $\hat{\vartheta}(0) = \mathbf{1}_{n+n_a+n_c}/p_0$, $r(0) = 1$, $\mathbf{x}_a(t - i) = \mathbf{1}_m/p_0$, $\hat{\mathbf{w}}(t - i) = \mathbf{1}_m/p_0$, $i = 1, 2, \dots, \max[n_a, n_c]$, $p_0 = 10^6$ and set a small positive number ε .
2. Collect the observation data $\mathbf{y}(t)$ and $\Phi_s(t)$, and construct the information matrices $\hat{\phi}_a(t)$, $\hat{\phi}_c(t)$ and $\hat{\Phi}(t)$ using (12–13) and (11).
3. Compute the innovation vector $e(t)$ and the step-size $r(t)$ according to (9) and (10).
4. Update the parameter estimation vector $\hat{\vartheta}(t)$ using (8).
5. Compute $\mathbf{x}_a(t)$ and $\hat{\mathbf{w}}(t)$ using (14–15).
6. Compare $\hat{\vartheta}(t)$ with $\hat{\vartheta}(t - 1)$: if $\|\hat{\vartheta}(t) - \hat{\vartheta}(t - 1)\| < \varepsilon$, terminate recursive calculation procedure and obtain $\hat{\vartheta}(t)$; otherwise, increase t by 1 and go to Step 2.

Remark 1 : In order to improve the transient performance and parameter estimation accuracy of the AM-GSG algorithm, we can introduce a forgetting factor (FF) λ in (10):

$$r(t) = \lambda r(t - 1) + \|\hat{\Phi}(t)\|^2, \quad 0 \leq \lambda < 1, \quad r(0) = 1. \tag{17}$$

Equations (8–9), (11–16) and (17) form the auxiliary model based forgetting factor gradient stochastic gradient (AM-FF-GSG) algorithm for the M-OEAR system. When $\lambda=1$, the AM-FF-GSG algorithm reduces to the AM-GSG algorithm (8–16).

Remark 2 In order to improve the stable state performance, a convergence index ε can be introduced in (8),

$$\hat{\vartheta}(t) = \hat{\vartheta}(t - 1) + \frac{\hat{\Phi}^T(t)}{r^\varepsilon(t)} e(t), \quad \frac{1}{2} < \varepsilon \leq 1. \tag{18}$$

Then Eqs. (9–16) and (18) form the Modified AM-GSG (M-AM-GSG) algorithm. When $\varepsilon = 1$, the M-AM-GSG algorithm reduces to the AM-GSG algorithm in (8–16).

3.2 The AM-MI-GSG algorithm

In order to improve the convergence rate and parameter estimation accuracy of the AM-GSG algorithm, we expand the dimension of the innovation vector $e(t)$ by employing the

multi-innovation identification theory and derive the auxiliary model based multi-innovation generalized stochastic gradient algorithm for the M-OEAR system.

Let p represents the innovation length, consider p data from $j = t - p + 1$ to $j = t$ and define a new multi-innovation vector:

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

It is usually considered that the estimate $\hat{\vartheta}(t-1)$ is more closer to the true value than $\hat{\vartheta}(t-i)$ and $E(p, t)$ is modified as

$$E(p, t) := \begin{bmatrix} y(t) - \hat{\Phi}(t)\hat{\vartheta}(t-1) \\ y(t-1) - \hat{\Phi}(t-1)\hat{\vartheta}(t-1) \\ \vdots \\ y(t-p+1) - \hat{\Phi}(t-p+1)\hat{\vartheta}(t-1) \end{bmatrix}.$$

Define the stacked output vector $Y(p, t)$ and the stacked information matrix $\hat{\Gamma}(p, t)$ as

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

$$\hat{\Gamma}(p, t) := \begin{bmatrix} \hat{\Phi}(t) \\ \hat{\Phi}(t-1) \\ \vdots \\ \hat{\Phi}(t-p+1) \end{bmatrix} \in \mathbb{R}^{(mp) \times (n+n_a+n_c)}.$$

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

$$E(p, t) = Y(p, t) - \hat{\Gamma}(p, t)\hat{\vartheta}(t-1) \in \mathbb{R}^{mp}.$$

Thus, we have

$$\hat{\vartheta}(t) = \hat{\vartheta}(t-1) + \frac{\hat{\Gamma}^T(p, t)}{r(t)} E(p, t), \tag{19}$$

$$E(p, t) = Y(p, t) - \hat{\Gamma}(p, t)\hat{\vartheta}(t-1), \tag{20}$$

$$r(t) = r(t-1) + \|\hat{\Phi}(t)\|^2, \tag{21}$$

$$Y(p, t) = [y^T(t), y^T(t-1), \dots, y^T(t-p+1)]^T, \tag{22}$$

$$\hat{\Gamma}(p, t) = [\hat{\Phi}^T(t), \hat{\Phi}^T(t-1), \dots, \hat{\Phi}^T(t-p+1)]^T. \tag{23}$$

Equations (19–23) and (11–16) consist the AM-MI-GSG algorithm. When the innovation length $p = 1$, the AM-MI-GSG algorithm reduces to the AM-GSG algorithm in (8–16). That is to say, the AM-GSG is a special case of the AM-MI-GSG algorithm. Similarly, a forgetting factor λ can be introduced in (21),

$$r(t) = \lambda r(t-1) + \|\hat{\Phi}(t)\|^2, \quad r(0) = 1, \quad 0 < \lambda \leq 1. \tag{24}$$

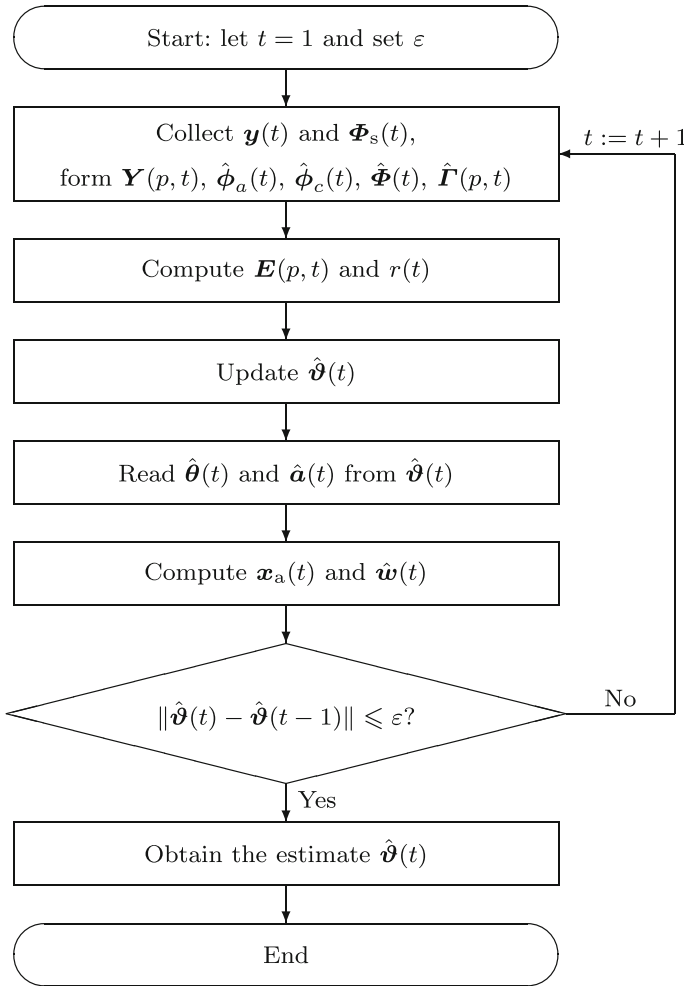


Fig. 2 The flowchart of computing the AM-MI-GSG parameter estimates $\hat{\vartheta}(t)$

Equations (19–20), (24), (22–23) and (11–16) form the auxiliary model based forgetting factor multi-innovation stochastic generalized gradient (AM-FF-MI-GSG) algorithm for the M-OEAR system.

To initialize the AM-MI-GSG algorithm, we should set the innovation length p and some initial values, e.g., $\hat{\vartheta}(0) = \mathbf{1}_{n+n_a+n_c}/p_0$, $\mathbf{x}_a(t-i) = \mathbf{1}_m/p_0$, $\hat{\mathbf{w}}(t-i) = \mathbf{1}_m/p_0$, $i = 1, 2, \dots, \max[n_a, n_c]$ and $p_0 = 10^6$. The flowchart of computing $\hat{\vartheta}(t)$ in the AM-MI-GSG algorithm is shown in Fig. 2.

4 The filtering based AM-GSG identification algorithm

In this section, a linear filter $L(z)$ is introduced to deal with the colored noises. We derive two identification models by filtering the input and output data including a system model

and a noise model, and identify each subsystems respectively. A filtering base AM-GSG (F-AM-GSG) and a filtering base AM-MI-GSG (F-AM-MI-GSG) identification algorithms are proposed in order to improve the convergence rate and parameter estimation accuracy.

4.1 The F-AM-GSG algorithm

For the M-OEAR model in (1), choose the polynomial $C(z)$ as the filter, that is to say, $L(z) = C(z)$. Multiplying the both sides of (1) by $C(z)$ gives

$$C(z)y(t) = C(z) \frac{\Phi_s(t)}{A(z)} \theta + v(t). \tag{25}$$

Define the filtered output vector $y_f(t)$ and the filtered information matrix $\Phi_f(t)$ as

$$\begin{aligned} y_f(t) &:= L(z)y(t) \\ &= C(z)y(t) \in \mathbb{R}^m, \end{aligned} \tag{26}$$

$$\begin{aligned} \phi_f(t) &:= L(z)\Phi_s(t) \\ &= C(z)\Phi_s(t) \in \mathbb{R}^{m \times n}, \end{aligned} \tag{27}$$

which can be expressed as the following recursive forms:

$$\begin{aligned} y_f(t) &= C(z)y(t) \\ &= y(t) + [y(t-1), y(t-2), \dots, y(t-n_c)]c, \end{aligned} \tag{28}$$

$$\begin{aligned} \phi_f(t) &= C(z)\Phi_s(t) \\ &= \Phi_s(t) + [\Phi_s(t-1), \Phi_s(t-2), \dots, \Phi_s(t-n_c)]c. \end{aligned} \tag{29}$$

From (25), we have

$$y_f = \frac{\phi_f(t)}{A(z)} \theta + v(t). \tag{30}$$

Define an inner variable

$$\begin{aligned} x_f(t) &:= \frac{\phi_f(t)}{A(z)} \theta \\ &= [1 - A(z)]x_f(t) + \phi_f(t)\theta \\ &= - \sum_{j=1}^{n_a} a_j x_f(t-j) + \phi_f(t)\theta \\ &= \Phi_f(t)\theta_f \in \mathbb{R}^m, \end{aligned} \tag{31}$$

where

$$\begin{aligned} \Phi_f(t) &:= [\phi_f(t), -x_f(t-1), -x_f(t-2), \dots, -x_f(t-n_a)] \in \mathbb{R}^{m \times (n+n_a)}, \\ \theta_f &:= \begin{bmatrix} \theta \\ a \end{bmatrix} \in \mathbb{R}^{n+n_a}. \end{aligned}$$

Then Eq. (30) can be rewritten as

$$y_f(t) = \Phi_f(t)\theta_f + v(t). \tag{32}$$

For the filter identification model in (32) and the noise identification model in (3), defining and minimizing the two gradient criterion functions

$$J_2(\theta_f) := \frac{1}{2} \|y_f(t) - \Phi_f(t)\theta_f\|^2,$$

$$J_3(c) := \frac{1}{2} \|w(t) - \phi_c(t)c\|^2,$$

result in the following gradient recursive relations:

$$\hat{\theta}_f(t) = \hat{\theta}_f(t - 1) + \frac{\Phi_f^T(t)}{r_1(t)} [y_f(t) - \Phi_f(t)\hat{\theta}_f(t - 1)], \tag{33}$$

$$r_1(t) = r_1(t - 1) + \|\Phi_f(t)\|^2, \quad r_1(0) = 1. \tag{34}$$

$$\hat{c}(t) = \hat{c}(t - 1) + \frac{\phi_c^T(t)}{r_2(t)} [w(t) - \phi_c(t)\hat{c}(t - 1)]$$

$$= \hat{c}(t - 1) + \frac{\phi_c^T(t)}{r_2(t)} [y(t) - \Phi_s(t)\theta - \phi_a(t)a - \phi_c(t)\hat{c}(t - 1)], \tag{35}$$

$$r_2(t) = r_2(t - 1) + \|\phi_c(t)\|^2, \quad r_2(0) = 1. \tag{36}$$

As we can see, Eqs. (33–36) cannot generate the estimates of $\hat{\theta}_f(t)$ and $\hat{c}(t)$, because the filter $C(z)$ is unknown, then the filtered matrix Φ_f is unknown, in addition, $x(t - i)$ and $w(t - i)$ are unmeasurable, the information matrices $\phi_a(t)$ and $\phi_c(t)$ contain those unknown terms. In order to solve those difficulties, we employ the auxiliary model method to replace the unknown variables with the output of the auxiliary model.

Use the outputs of the auxiliary models $x_a(t - i)$ and $\hat{w}(t - i)$ to construct the estimates

$$\hat{\phi}_a(t) = [-x_a(t - 1), -x_a(t - 2), \dots, -x_a(t - n_a)] \in \mathbb{R}^{m \times n_a},$$

$$\hat{\phi}_c(t) = [-\hat{w}(t - 1), -\hat{w}(t - 2), \dots, -\hat{w}(t - n_c)] \in \mathbb{R}^{m \times n_c}.$$

Similarly, we use the estimate $x_{fa}(t - i)$ of $x_f(t - i)$ and the estimate $\hat{\phi}_f(t)$ of $\Phi_f(t)$ to define

$$\hat{\Phi}_f(t) := [\hat{\phi}_f(t), -x_{fa}(t - 1), -x_{fa}(t - 2), \dots, -x_{fa}(t - n_a)] \in \mathbb{R}^{m \times (n+n_a)}.$$

From (2), (4) and (31), we can compute the outputs $x_a(t)$, $\hat{w}(t)$ and $x_{fa}(t)$ of the auxiliary models by

$$x_a(t) = \Phi_s(t)\hat{\theta}(t) + \hat{\phi}_a(t)\hat{a}(t),$$

$$\hat{w}(t) = y(t) - x_a(t),$$

$$x_{fa}(t) = \hat{\Phi}_f(t)\hat{\theta}_f(t).$$

Use the parameter estimates of the noise model

$$\hat{c}(t) := [\hat{c}_1(t), \hat{c}_2(t), \dots, \hat{c}_{n_c}(t)]^T \in \mathbb{R}^{n_c},$$

to construct the estimate of $C(z)$:

$$\hat{C}(t, z) := 1 + \hat{c}_1(t)z^{-1} + \hat{c}_2(t)z^{-2} + \dots + \hat{c}_{n_c}(t)z^{-n_c}.$$

Replacing the $C(z)$ in (26) and (27) with $\hat{C}(t, z)$, the estimates of the filtered output vector $\mathbf{y}_f(t)$ and the filtered information matrix $\Phi_f(t)$ can be computed by

$$\begin{aligned} \hat{\mathbf{y}}_f(t) &= \hat{C}(t, z)\mathbf{y}(t) \\ &= \mathbf{y}(t) + [\mathbf{y}(t - 1), \mathbf{y}(t - 2), \dots, \mathbf{y}(t - n_c)]\hat{\mathbf{c}}(t), \\ \hat{\Phi}_f(t) &= \hat{C}(t, z)\Phi_s(t) \\ &= \Phi_s(t) + [\Phi_s(t - 1), \Phi_s(t - 2), \dots, \Phi_s(t - n_c)]\hat{\mathbf{c}}(t). \end{aligned}$$

Replace the unknown information matrix $\Phi_f(t)$ in (33–34) with its estimate $\hat{\Phi}_f(t)$, $\phi_c(t)$ in (35–36) with $\hat{\phi}_c(t)$, $\phi_a(t)$ in (36) with $\hat{\phi}_a(t)$, and the filtered output $\mathbf{y}_f(t)$ in (33) with $\hat{\mathbf{y}}_f(t)$, the parameter vector θ and \mathbf{a} in (36) with $\hat{\theta}(t - 1)$ and $\hat{\mathbf{a}}(t - 1)$ respectively. Furthermore, define the innovation vectors:

$$\begin{aligned} \mathbf{e}_1(t) &:= \hat{\mathbf{y}}_f(t) - \hat{\Phi}_f(t)\hat{\theta}_f(t - 1) \in \mathbb{R}^m, \\ \mathbf{e}_2(t) &:= \mathbf{y}(t) - \Phi_s(t)\hat{\theta}(t - 1) - \hat{\phi}_a(t)\hat{\mathbf{a}}(t - 1) - \hat{\phi}_c(t)\hat{\mathbf{c}}(t - 1) \in \mathbb{R}^m. \end{aligned}$$

Then, we can derive a filtering based auxiliary model generalized stochastic gradient (F-AM-GSG) algorithm to estimate the parameter vectors $\hat{\theta}_f(t)$ and $\hat{\mathbf{c}}(t)$ for the M-OEAR system:

$$\hat{\theta}_f(t) = \hat{\theta}_f(t - 1) + \frac{\hat{\Phi}_f^T(t)}{r_1(t)}\mathbf{e}_1(t), \tag{37}$$

$$\mathbf{e}_1(t) = \hat{\mathbf{y}}_f(t) - \hat{\Phi}_f(t)\hat{\theta}_f(t - 1), \tag{38}$$

$$r_1(t) = r_1(t - 1) + \|\hat{\Phi}_f(t)\|^2, \tag{39}$$

$$\hat{\mathbf{c}}(t) = \hat{\mathbf{c}}(t - 1) + \frac{\hat{\phi}_c^T(t)}{r_2(t)}\mathbf{e}_2(t), \tag{40}$$

$$\mathbf{e}_2(t) = \mathbf{y}(t) - \Phi_s(t)\hat{\theta}(t - 1) - \hat{\phi}_a(t)\hat{\mathbf{a}}(t - 1) - \hat{\phi}_c(t)\hat{\mathbf{c}}(t - 1), \tag{41}$$

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

$$\hat{\Phi}_f(t) = [\hat{\Phi}_f(t), -\mathbf{x}_{fa}(t - 1), -\mathbf{x}_{fa}(t - 2), \dots, -\mathbf{x}_{fa}(t - n_a)], \tag{43}$$

$$\hat{\mathbf{y}}_f(t) = \mathbf{y}(t) + [\mathbf{y}(t - 1), \mathbf{y}(t - 2), \dots, \mathbf{y}(t - n_c)]\hat{\mathbf{c}}(t), \tag{44}$$

$$\hat{\Phi}_f(t) = \Phi_s(t) + [\Phi_s(t - 1), \Phi_s(t - 2), \dots, \Phi_s(t - n_c)]\hat{\mathbf{c}}(t), \tag{45}$$

$$\hat{\phi}_a(t) = [-\mathbf{x}_a(t - 1), -\mathbf{x}_a(t - 2), \dots, -\mathbf{x}_a(t - n_a)], \tag{46}$$

$$\hat{\phi}_c(t) = [-\hat{\mathbf{w}}(t - 1), -\hat{\mathbf{w}}(t - 2), \dots, -\hat{\mathbf{w}}(t - n_c)], \tag{47}$$

$$\mathbf{x}_{fa}(t) = \hat{\Phi}_f(t)\hat{\theta}_f(t), \tag{48}$$

$$\mathbf{x}_a(t) = \Phi_s(t)\hat{\theta}(t) + \hat{\phi}_a(t)\hat{\mathbf{a}}(t), \tag{49}$$

$$\hat{\mathbf{w}}(t) = \mathbf{y}(t) - \mathbf{x}_a(t), \tag{50}$$

$$\hat{\theta}_f(t) = [\hat{\theta}^T(t), \hat{\mathbf{a}}^T(t)]^T. \tag{51}$$

The steps involved in the F-AM-GSG algorithm are listed as follows.

1. Set the initial values: let $t = 1$, $\hat{\theta}_f(0) = \mathbf{1}_{n+n_a}/p_0$, $\hat{\mathbf{c}}(0) = \mathbf{1}_{n_c}/p_0$, $r_1(0) = 1$, $r_2(0) = 1$, $\mathbf{x}_{fa}(t - i) = \mathbf{1}_m/p_0$, $\mathbf{x}_a(t - i) = \mathbf{1}_m/p_0$, $\hat{\mathbf{w}}(t - i) = \mathbf{1}_m/p_0$, $i = 1, 2, \dots, \max[n_a, n_c]$, $p_0 = 10^6$, and set a small positive number ε .
2. Collect the observation data $\mathbf{y}(t)$ and $\Phi_s(t)$, construct the information matrices $\hat{\phi}_a(t)$ and $\hat{\phi}_c(t)$ by (46) and (47).
3. Compute $\mathbf{e}_2(t)$ by (41), $r_2(t)$ by (42).
4. Update the parameter estimate $\hat{\mathbf{c}}(t)$ by (40).

5. Compute the filtered output vector $\hat{y}_f(t)$ by (44) and the filtered information matrix $\hat{\Phi}_f(t)$ by (45), and form $\hat{\Phi}_f(t)$ by (43).
6. Compute $e_1(t)$ by (38) and $r_1(t)$ by (39).
7. Update the parameter estimate $\hat{\theta}_f(t)$ by (37).
8. Compute $x_{fa}(t)$ by (48). Read $\hat{\theta}(t)$ and $\hat{a}(t)$ from $\hat{\theta}_f(t)$ and compute $x_a(t)$ and $\hat{w}(t)$ by (49–50).
9. Compare $\hat{\theta}_f(t)$ with $\hat{\theta}_f(t-1)$ and compare $\hat{c}(t)$ with $\hat{c}(t-1)$: if $\|\hat{\theta}_f(t) - \hat{\theta}_f(t-1)\| < \varepsilon$ and $\|\hat{c}(t) - \hat{c}(t-1)\| < \varepsilon$, terminate recursive calculation procedure and obtain $\hat{\theta}(t)$; otherwise, increase t by 1 and go to Step 2.

4.2 The F-AM-MI-GSG algorithm

The F-AM-GSG algorithm can identify the parameter vectors $\hat{\theta}_f(t)$ and $\hat{c}(t)$, but has slow convergence speed. To improve the convergence rate and parameter estimation accuracy of the F-AM-GSG algorithm, according to the multi-innovation identification theory, we expand the innovation vectors $e_1(t) = \hat{y}_f(t) - \hat{\Phi}_f(t)\hat{\theta}_f(t-1) \in \mathbb{R}^m$ in (38) and $e_2(t) = y(t) - \Phi_s(t)\hat{\theta}(t-1) - \hat{\phi}_a(t)\hat{a}(t-1) - \hat{\phi}_c(t)\hat{c}(t-1) \in \mathbb{R}^m$ in (41) into large innovation vectors (p represents the innovation length):

$$E_1(p, t) := \begin{bmatrix} \hat{y}_f(t) - \hat{\Phi}_f(t)\hat{\theta}_f(t-1) \\ \hat{y}_f(t-1) - \hat{\Phi}_f(t-1)\hat{\theta}_f(t-1) \\ \vdots \\ \hat{y}_f(t-p+1) - \hat{\Phi}_f(t-p+1)\hat{\theta}_f(t-1) \end{bmatrix} \in \mathbb{R}^{mp},$$

$$E_2(p, t) := \begin{bmatrix} y(t) - \Phi_s(t)\hat{\theta}(t-1) - \hat{\phi}_a(t)\hat{a}(t-1) - \hat{\phi}_c(t)\hat{c}(t-1) \\ y(t-1) - \Phi_s(t-1)\hat{\theta}(t-1) - \hat{\phi}_a(t-1)\hat{a}(t-1) - \hat{\phi}_c(t-1)\hat{c}(t-1) \\ \vdots \\ y(t-p+1) - \Phi_s(t-p+1)\hat{\theta}(t-1) - \hat{\phi}_a(t-p+1)\hat{a}(t-1) - \hat{\phi}_c(t-p+1)\hat{c}(t-1) \end{bmatrix} \in \mathbb{R}^{mp}.$$

Define the stacked filtered output vector $\hat{Y}_f(p, t)$, the stacked filtered information matrix $\hat{\Gamma}_f(p, t)$, the stacked output vector $Y(p, t)$, the stacked information matrices $\Gamma_s(p, t)$, $\hat{\Omega}_a(p, t)$ and $\hat{\Omega}_c(p, t)$ as

$$\begin{aligned} \hat{Y}_f(p, t) &:= [\hat{y}_f^T(t), \hat{y}_f^T(t-1), \dots, \hat{y}_f^T(t-p+1)]^T \in \mathbb{R}^{mp}, \\ \hat{\Gamma}_f(p, t) &:= [\hat{\Phi}_f^T(t), \hat{\Phi}_f^T(t-1), \dots, \hat{\Phi}_f^T(t-p+1)]^T \in \mathbb{R}^{(mp) \times (n+n_a)}, \\ Y(p, t) &:= [y^T(t), y^T(t-1), \dots, y^T(t-p+1)]^T \in \mathbb{R}^{mp}, \\ \Gamma_s(p, t) &:= [\Phi_s^T(t), \Phi_s^T(t-1), \dots, \Phi_s^T(t-p+1)]^T \in \mathbb{R}^{(mp) \times n}, \\ \hat{\Omega}_a(p, t) &:= [\hat{\phi}_a^T(t), \hat{\phi}_a^T(t-1), \dots, \hat{\phi}_a^T(t-p+1)]^T \in \mathbb{R}^{(mp) \times n_a}, \\ \hat{\Omega}_c(p, t) &:= [\hat{\phi}_c^T(t), \hat{\phi}_c^T(t-1), \dots, \hat{\phi}_c^T(t-p+1)]^T \in \mathbb{R}^{(mp) \times n_c}. \end{aligned}$$

Then $E_1(p, t)$ and $E_2(p, t)$ can be equivalently expressed as

$$\begin{aligned} E_1(p, t) &:= \hat{Y}_f(p, t) - \hat{\Gamma}_f(p, t)\hat{\theta}_f(t-1) \in \mathbb{R}^{mp}, \\ E_2(p, t) &:= Y(p, t) - \Gamma_s(p, t)\hat{\theta}(t-1) - \hat{\Omega}_a(p, t)\hat{a}(t-1) - \hat{\Omega}_c(p, t)\hat{c}(t-1) \in \mathbb{R}^{mp}. \end{aligned}$$

According to the F-AM-GSG algorithm in (37–51), we can obtain the following equations:

$$\hat{\theta}_f(t) = \hat{\theta}_f(t - 1) + \frac{\hat{\Gamma}_f^T(p, t)}{r_1(t)} E_1(p, t), \tag{52}$$

$$E_1(p, t) = \hat{Y}_f(p, t) - \hat{\Gamma}_f(p, t)\hat{\theta}_f(t - 1), \tag{53}$$

$$r_1(t) = r_1(t - 1) + \|\hat{\Phi}_f(t)\|^2, \tag{54}$$

$$\hat{c}(t) = \hat{c}(t - 1) + \frac{\hat{\Omega}_c^T(p, t)}{r_2(t)} E_2(p, t), \tag{55}$$

$$E_2(p, t) = Y(p, t) - \Gamma_s(p, t)\hat{\theta}(t - 1) - \hat{\Omega}_a(p, t)\hat{a}(t - 1) - \hat{\Omega}_c(p, t)\hat{c}(t - 1), \tag{56}$$

$$r_2(t) = r_2(t - 1) + \|\hat{\Phi}_c(t)\|^2, \tag{57}$$

$$\hat{Y}_f(p, t) = [\hat{y}_f^T(t), \hat{y}_f^T(t - 1), \dots, \hat{y}_f^T(t - p + 1)]^T, \tag{58}$$

$$\hat{\Gamma}_f(p, t) = [\hat{\Phi}_f^T(t), \hat{\Phi}_f^T(t - 1), \dots, \hat{\Phi}_f^T(t - p + 1)]^T, \tag{59}$$

$$Y(p, t) = [y^T(t), y^T(t - 1), \dots, y^T(t - p + 1)]^T, \tag{60}$$

$$\Gamma_s(p, t) = [\Gamma_s^T(t), \Gamma_s^T(t - 1), \dots, \Gamma_s^T(t - p + 1)]^T, \tag{61}$$

$$\hat{\Omega}_a(p, t) = [\hat{\Phi}_a^T(t), \hat{\Phi}_a^T(t - 1), \dots, \hat{\Phi}_a^T(t - p + 1)]^T, \tag{62}$$

$$\hat{\Omega}_c(p, t) = [\hat{\Phi}_c^T(t), \hat{\Phi}_c^T(t - 1), \dots, \hat{\Phi}_c^T(t - p + 1)]^T. \tag{63}$$

The equations above and (43–51) form the data filtering based auxiliary model multi-innovation generalized stochastic gradient (F-AM-MI-GSG) algorithm. When $p = 1$, the F-AM-MI-GSG algorithm reduce to the F-AM-GSG algorithm in (37–51). The proposed method in this paper can be extended to study the parameter estimation of transfer functions (Xu 2014; Xu and Ding 2017a), time-varying systems (Ding et al. 2016) and signal models (Xu 2017; Wang et al. 2018), and applied to other fields (Feng et al. 2016; Li et al. 2017d; Ji and Ding 2017; Cao and Zhu 2017; Chu et al. 2017; Zhao et al. 2017a).

5 Example

Consider the following multivariate output-error autoregressive system:

$$y(t) = \frac{\Phi_s(t)}{A(z)}\theta + \frac{1}{C(z)}v(t),$$

$$\Phi_s(t) = \begin{bmatrix} -y_1(t - 1), & y_1(t - 2) \sin(y_2(t - 2)), & y_2(t - 1), & y_2(t - 2)u_1(t - 2), \\ -y_1(t - 1), & y_1(t - 2) \sin(t/\pi), & y_2(t - 1), & y_1(t - 2)u_2(t - 2), \\ u_1(t - 1), & u_1(t - 2)u_2(t - 2), & u_2(t - 1) \cos(t) \\ u_1^2(t - 1), & \sin(u_2(t - 2)), & u_1(t - 1) + u_2(t - 2) \end{bmatrix} \in \mathbb{R}^{2 \times 7},$$

$$A(z) = 1 + a_1z^{-1} + a_2z^{-2}$$

$$= 1 + 0.12z^{-1} + 0.3z^{-2},$$

$$C(z) = 1 + c_1z^{-1} + c_2z^{-2}$$

$$= 1 - 0.11z^{-1} + 0.8z^{-2},$$

$$\theta = [0.70, 0.41, 0.47, -0.52, 0.10, -0.30, 0.40]^T,$$

$$a = [a_1, a_2] = [0.12, 0.3]^T,$$

$$c = [c_1, c_2] = [-0.11, 0.8]^T,$$

Table 1 The AM-MI-GSG parameter estimates and errors ($\sigma^2 = 0.50^2$)

p	t	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	a_1	a_2	c_1	c_2	δ (%)
1	100	0.34978	0.12918	0.20189	-0.24104	0.10994	-0.09709	-0.00062	0.02632	0.04545	0.08232	0.11916	72.13706
	200	0.37198	0.15726	0.21482	-0.25214	0.10379	-0.09849	0.01949	0.03118	0.05689	0.08025	0.17568	67.49535
	500	0.41027	0.18784	0.23976	-0.30250	0.10128	-0.09721	0.04610	0.03697	0.07662	0.07998	0.21569	62.06408
	1000	0.42793	0.19894	0.25435	-0.32449	0.10153	-0.09775	0.05732	0.03816	0.08712	0.07875	0.23785	59.44171
	2000	0.44525	0.20925	0.26502	-0.34865	0.10103	-0.09900	0.06790	0.04097	0.09819	0.07811	0.25900	56.94354
2	3000	0.45316	0.21469	0.26951	-0.35911	0.10081	-0.09943	0.07405	0.04231	0.10338	0.07760	0.26859	55.78055
	100	0.47937	0.19108	0.29883	-0.32963	0.10276	-0.111024	0.07210	0.04569	0.09746	0.09641	0.19146	59.43663
	200	0.50693	0.23555	0.31769	-0.34270	0.09341	-0.111948	0.10151	0.05937	0.12294	0.08412	0.27300	52.39111
	500	0.55290	0.27701	0.34468	-0.40309	0.08998	-0.12189	0.13937	0.07151	0.16000	0.07885	0.31778	46.01479
	1000	0.57009	0.29070	0.36260	-0.42587	0.09115	-0.12568	0.15468	0.07312	0.17701	0.07379	0.34312	43.08157
6	2000	0.58549	0.30113	0.37341	-0.44683	0.09085	-0.13033	0.16828	0.07744	0.19252	0.06912	0.36641	40.52256
	3000	0.59145	0.30680	0.37732	-0.45635	0.09105	-0.13262	0.17587	0.07949	0.19923	0.06623	0.37708	39.33515
	100	0.64289	0.29261	0.41703	-0.45316	0.08052	-0.16505	0.21771	0.10105	0.21485	0.07941	0.36815	37.59413
	200	0.65868	0.35858	0.43897	-0.45185	0.06874	-0.19519	0.25138	0.13414	0.25294	0.03910	0.48010	28.09700
	500	0.67945	0.39698	0.45640	-0.50334	0.07060	-0.20542	0.29294	0.14025	0.28149	0.02311	0.52267	23.31238
12	1000	0.68086	0.40335	0.47371	-0.51594	0.07856	-0.21587	0.30820	0.13595	0.28749	0.01063	0.55141	20.73968
	2000	0.68148	0.40179	0.47773	-0.51671	0.08028	-0.22564	0.31954	0.13649	0.28778	0.00069	0.58009	18.47261
	3000	0.68019	0.40260	0.47689	-0.52122	0.08235	-0.23025	0.32538	0.13767	0.28871	-0.00552	0.59268	17.40080
	100	0.67702	0.34765	0.46363	-0.49807	0.08714	-0.21926	0.28167	0.12816	0.27335	0.03975	0.50488	25.05067
	200	0.67495	0.40825	0.48147	-0.47989	0.07305	-0.25357	0.30966	0.17104	0.29908	-0.01750	0.62421	16.00446
3000	500	0.68434	0.43232	0.48576	-0.52285	0.07459	-0.25651	0.34922	0.16275	0.30251	-0.03259	0.65884	12.50473
	1000	0.68601	0.42825	0.49471	-0.53064	0.08871	-0.26724	0.36182	0.15482	0.29974	-0.04453	0.68524	10.21010
	2000	0.68473	0.41719	0.49305	-0.52220	0.08780	-0.27488	0.36978	0.15086	0.29702	-0.05119	0.71081	8.29086
	3000	0.68332	0.41465	0.48820	-0.52598	0.09005	-0.27721	0.37359	0.15183	0.29774	-0.05748	0.72019	7.48340
	True values	0.70000	0.41000	0.47000	-0.52000	0.10000	-0.30000	0.40000	0.12000	0.30000	-0.11000	0.80000	

Table 2 The F-AM-MI-GSG parameter estimates and errors ($\sigma^2 = 0.50^2$)

p	t	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	a_1	a_2	c_1	c_2	δ (%)
1	100	0.29885	0.08587	0.16789	-0.24924	0.06655	-0.07513	-0.02014	0.01704	0.03706	0.114567	0.44414	65.54516
	200	0.32244	0.10869	0.16836	-0.25843	0.06942	-0.07518	0.00412	0.02351	0.03936	0.111874	0.56391	60.25032
	500	0.36034	0.13810	0.18342	-0.31553	0.07066	-0.07686	0.03583	0.03035	0.04784	0.111147	0.59101	55.23704
	1000	0.37672	0.15022	0.19196	-0.33726	0.07459	-0.07896	0.05236	0.03253	0.05226	0.10253	0.61176	52.75038
	2000	0.39358	0.16162	0.19649	-0.35795	0.07652	-0.08267	0.06544	0.03625	0.05700	0.09992	0.63315	50.59019
	3000	0.40130	0.16763	0.20008	-0.36575	0.07825	-0.08379	0.07398	0.03739	0.05957	0.09832	0.63842	49.57277
2	100	0.43512	0.15654	0.23404	-0.34350	0.05303	-0.07988	0.04480	0.04535	0.07518	0.03057	0.56273	49.47825
	200	0.46498	0.18803	0.23580	-0.34909	0.06185	-0.09150	0.07943	0.05570	0.08449	0.01616	0.70163	43.63691
	500	0.50183	0.22633	0.26175	-0.40408	0.06473	-0.10188	0.12151	0.05819	0.11301	0.03006	0.69064	38.33105
	1000	0.51676	0.24132	0.27455	-0.42213	0.07131	-0.10916	0.14226	0.05827	0.12492	0.02440	0.70572	35.62568
	2000	0.53089	0.25364	0.28309	-0.43672	0.07426	-0.11810	0.15778	0.06010	0.13783	0.01998	0.72116	33.31103
	3000	0.53734	0.26046	0.28843	-0.44224	0.07678	-0.12156	0.16745	0.06044	0.14386	0.01536	0.72097	32.17219
6	100	0.59096	0.29950	0.35947	-0.47323	0.07045	-0.15298	0.18572	0.07524	0.19637	-0.09101	0.74435	23.84063
	200	0.62292	0.33510	0.36647	-0.44800	0.07992	-0.18509	0.22670	0.09007	0.22389	-0.10465	0.81319	18.97508
	500	0.65189	0.36801	0.39760	-0.49365	0.07975	-0.19987	0.27386	0.09331	0.26797	-0.07560	0.75843	13.76003
	1000	0.66366	0.37746	0.41010	-0.50821	0.08954	-0.21220	0.29643	0.09422	0.27646	-0.08395	0.78985	11.09294
	2000	0.67257	0.38073	0.41602	-0.51083	0.08982	-0.22510	0.31032	0.09742	0.28516	-0.10371	0.79630	9.40169
	3000	0.67606	0.38461	0.41957	-0.51192	0.09136	-0.22987	0.31787	0.09844	0.28814	-0.11036	0.79005	8.66687
12	100	0.63819	0.35562	0.43949	-0.51341	0.08484	-0.21814	0.26616	0.07679	0.26942	-0.06686	0.77756	13.20438
	200	0.67719	0.39012	0.43604	-0.47442	0.09032	-0.24896	0.29877	0.10520	0.28874	-0.16258	0.80187	9.65927
	500	0.69992	0.41170	0.45446	-0.52490	0.08465	-0.25236	0.34245	0.11417	0.31275	-0.10166	0.77787	5.62008
	1000	0.70441	0.41433	0.46283	-0.53260	0.09833	-0.26289	0.36259	0.11251	0.30771	-0.09442	0.81640	4.10800
	2000	0.70547	0.40769	0.46199	-0.52417	0.09349	-0.27374	0.37029	0.11485	0.30728	-0.11755	0.80204	2.94015
	3000	0.70493	0.40960	0.46210	-0.52210	0.09380	-0.27658	0.37358	0.11539	0.30681	-0.11905	0.79641	2.66569
True values		0.70000	0.41000	0.47000	-0.52000	0.10000	-0.30000	0.40000	0.12000	0.30000	-0.11000	0.80000	

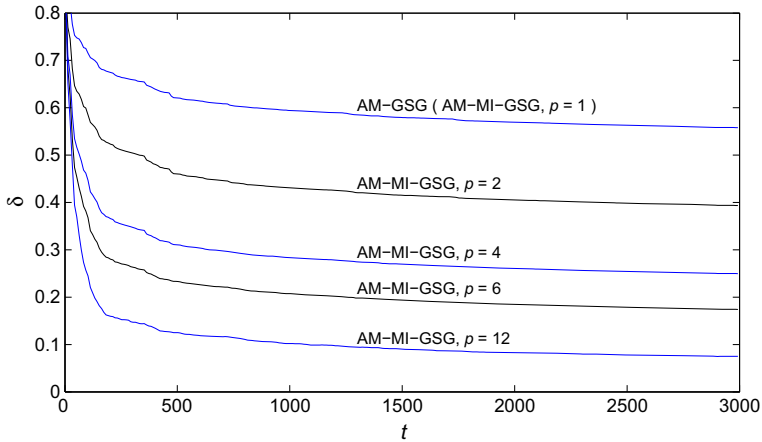


Fig. 3 The AM-MI-GSG parameter estimation errors δ versus t

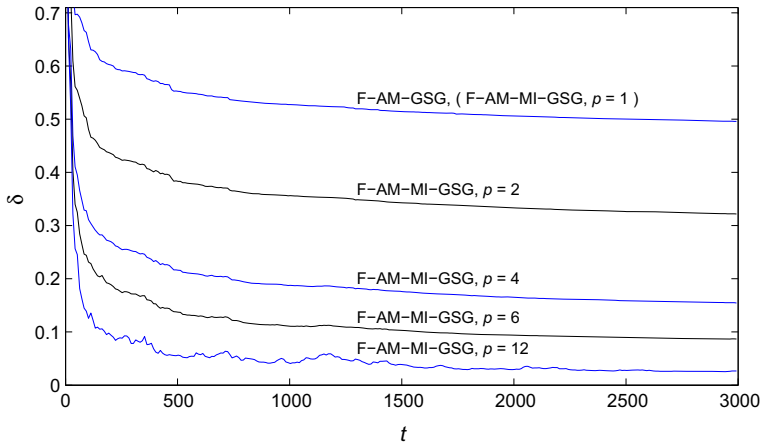


Fig. 4 The F-AM-MI-GSG parameter estimation errors δ versus t

$$\begin{aligned} \vartheta &= [\theta^T, a^T, c^T]^T, \\ \theta_f &= [\theta^T, a^T]^T. \end{aligned}$$

In simulation, the inputs $\{u_1(t)\}$ and $\{u_2(t)\}$ are taken as two independent persistent excitation signal sequences with zero mean and unit variances, $\{v_1(t)\}$ and $\{v_2(t)\}$ are taken as two white noise sequences with zero mean and variances σ_1^2 for $v_1(t)$ and σ_2^2 for $v_2(t)$. Taking $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 0.50^2$ and based on the above model, we generate the system's output signals $y(t) = [y_1(t), y_2(t)]^T$. By using $u_1(t), u_2(t), y_1(t)$ and $y_2(t)$ and applying the AM-MI-GSG algorithm and the F-AM-MISG algorithm to estimate the parameters of this system, the parameter estimates and errors are shown in Tables 1 and 2 with $p = 1, 2, 4, 6$ and 12. The estimation errors $\delta := \|\hat{\vartheta}(t) - \vartheta\|/\|\vartheta\|$ versus t are shown in Figs. 3 and 4.

From Tables 1 and 2 and Figs. 3 and 4, we can draw the following conclusions.

1. The parameter estimation errors of the AM-MI-GSG and the F-AM-MI-GSG algorithms become smaller with the data length t increasing – see the estimation errors of the last columns in Tables 1 and 2.
2. A larger innovation length p leads to smaller parameter estimation errors both for the AM-MI-GSG algorithm and the F-AM-MI-GSG algorithm—see from Figs. 3 and 4. The estimation errors can almost converge to zero if the innovation length p is large enough and the data length goes to infinity.
3. Under the same innovation length p , the F-AM-MI-GSG algorithm can give more accurate parameter estimates than the AM-MI-GSG algorithm.

6 Conclusions

In this paper, we employ the data filtering technique to propose an F-AM-GSG algorithm for M-OEAR systems and derive an F-AM-MI-GSG algorithm by adopting the multi-innovation identification theory, taking account into an autoregressive noise. This work can be extended to the case with an autoregressive moving average noise. Compared with the AM-MI-GSG algorithm, the F-AM-MI-GSG algorithm has smaller estimation errors under the same innovation length. The proposed filtering based identification method can be applied to study the parameter estimation of other multivariate systems with different structures and disturbance noise, e.g., nonlinear multidimensional multivariate systems with colored noise. These are worth further studying in multidimensional multivariate systems in the future.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 61273194) and the Fundamental Research Funds for the Central Universities (JUSRP51733B).

References

- Afshari, H. H., Gadsden, S. A., & Habibi, S. (2017). Gaussian filters for parameter and state estimation: A general review of theory and recent trends. *Signal Processing*, 135, 218–238.
- Aslam, M. S. (2016). Maximum likelihood least squares identification method for active noise control systems with autoregressive moving average noise. *Automatica*, 69, 1–11.
- Cao, X., & Zhu, D. Q. (2017). Multi-AUV task assignment and path planning with ocean current based on biological inspired self-organizing map and velocity synthesis algorithm. *Intelligent Automation and Soft Computing*, 23(1), 31–39.
- Cham, C. L., Tan, A. H., & Tan, W. H. (2017). Identification of a multivariable nonlinear and time-varying mist reactor system. *Control Engineering Practice*, 63, 13–23.
- Cheng, J. X., Fang, M. Q., & Wang, Y. Q. (2017). Subspace identification for closed-loop 2-D separable-in-denominator systems. *Multidimensional Systems and Signal Processing*, 28(4), 1499–1521.
- Cheng, Y., & Ugrinovskii, V. (2016). Event-triggered leader-following tracking control for multivariable multi-agent systems. *Automatica*, 70, 204–210.
- Chu, Z. Z., Zhu, D. Q., & Yang, S. X. (2017). Adaptive sliding mode control for depth trajectory tracking of remotely operated vehicle with thruster nonlinearity. *Journal of Navigation*, 70(1), 149–164.
- Dhabal, S., & Venkateswaran, P. (2017). A novel accelerated artificial bee colony algorithm for optimal design of two dimensional FIR filter. *Multidimensional Systems and Signal Processing*, 28(2), 471–493.
- Ding, D. R., Wang, Z. D., Ho, D. W. C., et al. (2017). Distributed recursive filtering for stochastic systems under uniform quantizations and deception attacks through sensor networks. *Automatica*, 78, 231–240.
- Ding, F., Wang, F. F., Xu, L., et al. (2017a). Decomposition based least squares iterative identification algorithm for multivariate pseudo-linear ARMA systems using the data filtering. *Journal of the Franklin Institute*, 354(3), 1321–1339.
- Ding, F., Wang, F. F., Xu, L., et al. (2017b). Parameter estimation for pseudo-linear systems using the auxiliary model and the decomposition technique. *IET Control Theory and Applications*, 11(3), 390–400.

- Ding, F., Wang, X. H., Mao, L., et al. (2017). Joint state and multi-innovation parameter estimation for time-delay linear systems and its convergence based on the Kalman filtering. *Digital Signal Processing*, *62*, 211–223.
- Ding, F., Xu, L., & Zhu, Q. M. (2016). Performance analysis of the generalised projection identification for time-varying systems. *IET Control Theory and Applications*, *10*(18), 2506–2514.
- Ding, F., Yang, H. Z., & Liu, F. (2008). Performance analysis of stochastic gradient algorithms under weak conditions. *Science in China Series F-Information Sciences*, *51*(9), 1269–1280.
- Feng, L., Wu, M. H., Li, Q. X., et al. (2016). Array factor forming for image reconstruction of one-dimensional nonuniform aperture synthesis radiometers. *IEEE Geoscience and Remote Sensing Letters*, *13*(2), 237–241.
- Jafari, M., Salimifard, M., & Dehghani, M. (2014). Identification of multivariable nonlinear systems in the presence of colored noises using iterative hierarchical least squares algorithm. *ISA Transactions*, *53*(4), 1243–1252.
- Ji, Y., & Ding, F. (2017). Multiperiodicity and exponential attractivity of neural networks with mixed delays. *Circuits, Systems and Signal Processing*, *36*(6), 2558–2573.
- Jin, Q. B., Wang, Z., & Liu, X. P. (2015). Auxiliary model-based interval-varying multi-innovation least squares identification for multivariable OE-like systems with scarce measurements. *Journal of Process Control*, *35*, 154–168.
- Katayama, T., & Ase, H. (2016). Linear approximation and identification of MIMO Wiener-Hammerstein systems. *Automatica*, *71*, 118–124.
- Levanony, D., & Berman, N. (2004). Recursive nonlinear system identification by a stochastic gradient algorithm: Stability, performance, and model nonlinearity considerations. *IEEE Transactions on Signal Processing*, *52*(9), 2540–2550.
- Liang, J. L., Wang, Z. D., & Liu, X. H. (2014). Robust state estimation for two-dimensional stochastic time-delay systems with missing measurements and sensor saturation. *Multidimensional Systems and Signal Processing*, *25*(1), 157–177.
- Li, J. F., & Zhang, X. F. (2015). Sparse representation-based joint angle and Doppler frequency estimation for MIMO radar. *Multidimensional Systems and Signal Processing*, *26*(1), 179–192.
- Li, J. P., Hua, C. C., Tang, Y. J., et al. (2014). A time-varying forgetting factor stochastic gradient combined with Kalman filter algorithm for parameter identification of dynamic systems. *Nonlinear Dynamics*, *78*(3), 1943–1952.
- Li, M. H., Liu, X. M., & Ding, F. (2017a). Least-squares-based iterative and gradient-based iterative estimation algorithms for bilinear systems. *Nonlinear Dynamics*, *89*(1), 197–211.
- Li, M. H., Liu, X. M., & Ding, F. (2017b). The maximum likelihood least squares based iterative estimation algorithm for bilinear systems with autoregressive moving average noise. *Journal of the Franklin Institute*, *354*(12), 4861–4881.
- Li, M. H., Liu, X. M., & Ding, F. (2017c). The gradient based iterative estimation algorithms for bilinear systems with autoregressive noise. *Circuits, Systems and Signal Processing*, *36*(11), 4541–4568.
- Li, X. F., Chu, Y. D., Leung, A. Y. T., & Zhang, H. (2017d). Synchronization of uncertain chaotic systems via complete-adaptive-impulsive controls. *Chaos Solitons & Fractals*, *100*, 24–30.
- Ma, P., & Ding, F. (2017). New gradient based identification methods for multivariate pseudo-linear systems using the multi-innovation and the data filtering. *Journal of the Franklin Institute*, *354*(3), 1568–1583.
- Mercère, G., & Bako, L. (2011). Parameterization and identification of multivariable state-space systems: A canonical approach. *Automatica*, *47*(8), 1547–1555.
- Mu, B. Q., & Chen, H. F. (2013). Recursive identification of MIMO Wiener systems. *IEEE Transactions on Automatic Control*, *58*(3), 802–808.
- Pan, J., Jiang, X., Wan, X. K., et al. (2017). A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems. *International Journal of Control, Automation and Systems*, *15*(3), 1189–1197.
- Pan, J., Yang, X. H., Cai, H. F., et al. (2016). Image noise smoothing using a modified Kalman filter. *Neuro-computing*, *173*, 1625–1629.
- Saikrishna, P. S., Pasumarthi, R., & Bhatt, N. P. (2017). Identification and multivariable gain-scheduling control for cloud computing systems. *IEEE Transactions on Control Systems Technology*, *25*(3), 792–807.
- Tseng, C. C., & Lee, S. L. (2017). Closed-form designs of digital fractional order Butterworth filters using discrete transforms. *Signal Processing*, *137*, 80–97.
- Wan, X. K., Li, Y., Xia, C., et al. (2016). A T-wave alternans assessment method based on least squares curve fitting technique. *Measurement*, *86*, 93–100.
- Wang, Y. J., & Ding, F. (2016). Novel data filtering based parameter identification for multiple-input multiple-output systems using the auxiliary model. *Automatica*, *71*, 308–313.

- Wang, Y.J., Ding, F., & Xu, L. (2018) Some new results of designing an IIR filter with colored noise for signal processing. *Digital Signal Processing*, 72, 44–58.
- Wang, Z., Jin, Q. B., & Liu, X. P. (2016). Recursive least squares identification of hybrid Box-Jenkins model structure in open-loop and closed-loop. *Journal of the Franklin Institute*, 353(2), 265–278.
- Xing, H. L., Li, D. H., Li, J., et al. (2016). Linear extended state observer based sliding mode disturbance decoupling control for nonlinear multivariable systems with uncertainty. *International Journal of Control, Automation and Systems*, 14(4), 967–976.
- Xu, L. (2014). A proportional differential control method for a time-delay system using the Taylor expansion approximation. *Applied Mathematics and Computation*, 236, 391–399.
- Xu, L. (2015). Application of the Newton iteration algorithm to the parameter estimation for dynamical systems. *Journal of Computational and Applied Mathematics*, 288, 33–43.
- Xu, L. (2016). The damping iterative parameter identification method for dynamical systems based on the sine signal measurement. *Signal Processing*, 120, 660–667.
- Xu L. (2017). The parameter estimation algorithms based on the dynamical response measurement data. *Advances in Mechanical Engineering*, 9. <https://dx.doi.org/10.1177/1687814017730003>.
- Xu, L., Chen, L., & Xiong, W. L. (2015). Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. *Nonlinear Dynamics*, 79(3), 2155–2163.
- Xu, L., & Ding, F. (2017a). Parameter estimation for control systems based on impulse responses. *International Journal of Control, Automation and Systems*, 15(6). <https://dx.doi.org/10.1007/s12555-016-0224-2>.
- Xu, L., & Ding, F. (2017b). Recursive least squares and multi-innovation stochastic gradient parameter estimation methods for signal modeling. *Circuits, Systems and Signal Processing*, 36(4), 1735–1753.
- Xu, L., & Ding, F. (2017c). Parameter estimation algorithms for dynamical response signals based on the multi-innovation theory and the hierarchical principle. *IET Signal Processing*, 11(2), 228–237.
- Xu, L., Ding, F., Gu, Y., et al. (2017). A multi-innovation state and parameter estimation algorithm for a state space system with d-step state-delay. *Signal Processing*, 140, 97–103.
- Zhang, B., & Mao, Z. Z. (2010). Bias compensation principle based recursive least squares identification method for Hammerstein nonlinear systems. *Journal of the Franklin Institute*, 354(3), 1340–1355.
- Zhao, N., Chen, Y., Liu, R., Wu, M. H., & Xiong, W. (2017a). Monitoring strategy for relay incentive mechanism in cooperative communication networks. *Computers & Electrical Engineering*, 60, 14–29.
- Zhao, N., Wu, M. H., & Chen, J. J. (2017b). Android-based mobile educational platform for speech signal processing. *International Journal of Electrical Engineering Education*, 54(1), 3–16.



Qinyao Liu was born in Chengdu, Sichuan Province, China. She received her B.Sc. degree from Jiangnan University, Wuxi, China in 2016, and now is a Ph.D. student in the School of Internet of Things Engineering, Jiangnan University, Wuxi, China. Her interests include system modeling, system identification and process control.



Feng Ding was born in Guangshui, Hubei Province, China. He received the B.Sc. degree from the Hubei University of Technology (Wuhan, China) in 1984, and the M.Sc. and Ph.D. degrees in automatic control both from the Department of Automation, Tsinghua University, Beijing, in 1991 and 1994, respectively. He has been a Professor in the School of Internet of Things Engineering, Jiangnan University, Wuxi, China since 2004. He is a Colleges and Universities Blue Project Middle-Aged Academic Leader (Jiangsu, China). His current research interests include model identification and adaptive control. He authored four books on System Identification.